

Adaptive Coded Modulation With Receive Antenna Diversity and Imperfect Channel Knowledge at Receiver and Transmitter

Duc V. Duong, *Student Member, IEEE*, Geir E. Øien, *Member, IEEE*, and Kjell J. Hole, *Member, IEEE*

Abstract—Unifying and generalizing the works of Cai and Giannakis and Øien *et al.*, the performance of an adaptive trellis-coded modulation system where receive antenna diversity is implemented by means of maximum ratio combining is analyzed and optimized. As in the work of Cai and Giannakis, the analysis is done in the presence of both estimation and prediction errors. Rayleigh fading on each subchannel is considered, with the estimation and prediction being performed independently on each subchannel. The system optimization process is done in such a way that the throughput is maximized under a bit-error-rate (BER) constraint. The numerical example employs a Jakes-fading spectrum and shows how the power should be distributed between pilot and data symbols and how often pilot symbols should be transmitted for maximum throughput under an instantaneous (with respect to the predicted channel) BER constraint.

Index Terms—Adaptive coded modulation (ACM), antenna diversity, estimation and prediction error, maximum ratio combining (MRC), optimal pilot spacing, optimal power allocation, pilot-symbol-assisted modulation (PSAM).

I. INTRODUCTION

ADAPTIVE coded modulation (ACM) is an efficient tool to counteract fading in wireless channel communications and to improve spectral efficiency (SE) [3]. In addition, it is well known that receive diversity stabilizes the channel quality. The scheme presented in this paper is a generalization of two recent papers—one by Cai and Giannakis [1], in which an adaptive modulation system based on uncoded M -quadrature-amplitude modulation (M -QAM) constellations with single transmit and receive antenna is investigated, and one by Øien *et al.* [2], where an ACM system with maximum ratio combining (MRC) reception was investigated assuming perfect receiver channel state information (CSI). We extend and unify the idea of these papers by analyzing the case when, as in [2], coding is included and when MRC is implemented on the receiver side. The

system parameters are optimized, as in [1], for maximal SE under an instantaneous bit-error-rate (BER) constraint.

As in [1] and [2], the estimator and the predictor in our adaptive scheme are based on received known pilot symbols that are sent from the transmitter along with data information. This technique is known as pilot-symbol-assisted modulation (PSAM) [4]. The predicted CSI is fed back to the transmitter via a zero-error non-zero-delay feedback channel. Based on this CSI, the transmitter dynamically adapts the rate and the power to a mode which maximizes the average spectral efficiency (ASE) while maintaining the instantaneous BER below some predefined target value BER_0 . The instantaneous BER here is derived with respect to the predicted channel-signal-to-noise ratio (CSNR). As in [2] and [5], four-dimensional (4-D)—thus, multidimensional—trellis codes designed for additive white Gaussian noise (AWGN) channels are used as component codes.

In [6]–[8], multiple-input multiple-output (MIMO) systems (with single-input multiple-output as a special case) are analyzed where perfect but delayed CSI (thus, outdated) was assumed. No pilot symbol is transmitted in their system, and the BER constraint is subject to instantaneous BER and average BER, respectively. Note also that in [6], two-dimensional (2-D) trellis codes are utilized.

Throughout this paper, we will use the following notations: Superscripts $(\cdot)^T$, $(\cdot)^*$, and $(\cdot)^H$ stand for transpose, complex conjugate, and Hermitian transpose, respectively. $[z]$ means the integer part of z , and $E[z]$ stands for the expectation of z .

The remainder of this paper is outlined as follows. Our system is described in Section II. Channel estimation and prediction is dealt with in Section III. BER performance and ASE performance analyses can be found in Sections IV and V, respectively. A numerical example is given in Section VI, before the conclusions are drawn in Section VII.

II. SYSTEM MODEL

The system we are considering is illustrated in Fig. 1, where we have n_R independent identically distributed (i.i.d.) receive antennas. The adaptive encoder always chooses to transmit symbols from the constellation of size M_n , which is best suited to the predicted channel state, out of a set of constellations of sizes $\{M_n\}_{n=1}^N$, corresponding to the set of SEs $\{R_n\}_{n=1}^N$. The SEs are ordered such that $R_n < R_{n+1}$, and, hence, $M_n < M_{n+1} \forall n$. The choice of code is based on the CSI fed back from the receiver; code n is chosen if the predicted

Manuscript received November 15, 2004; revised May 18, 2005, August 19, 2005, and October 11, 2005. This work was supported by the Research Council of Norway under Bandwidth-Efficient and Adaptive Transmission Schemes for Wireless Multimedia Communications (BEATS) project (URL: <http://www.iet.ntnu.no/projects/beats/>). The review of this paper was coordinated by Dr. M. Stojanovic.

D. V. Duong and G. E. Øien are with the Department of Electronics and Telecommunications, Norwegian University of Science and Technology, N-7491 Trondheim, Norway (e-mail: duong@iet.ntnu.no; oien@iet.ntnu.no).

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

Digital Object Identifier 10.1109/TVT.2005.863348

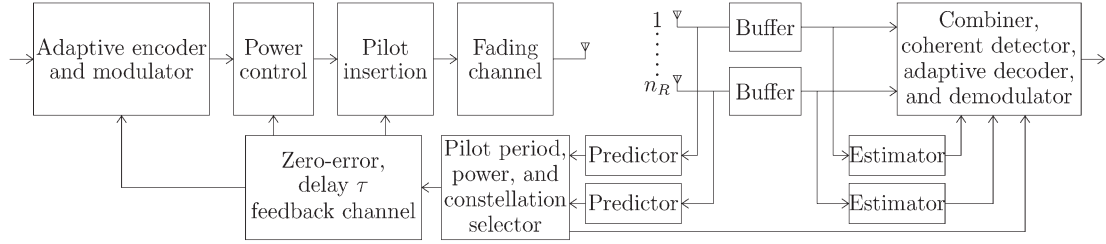


Fig. 1. Adaptive PSAM system model combined with multireception of receiving signal.

CSNR falls between the switching thresholds $\hat{\gamma}_n$ and $\hat{\gamma}_{n+1}$. By letting $\hat{\gamma}_0 = 0$ and $\hat{\gamma}_{N+1} = \infty$, we have $\hat{\gamma}_n < \hat{\gamma}_{n+1}$ for all $n \in \{0, 1, \dots, N\}$.

The power allocation between data and pilot symbols will be chosen in an optimal way. The criterion for choosing the code and the transmit power distribution is to fulfill a target BER performance while maximizing the throughput in terms of ASE. We include MRC on the receiver side with the assumption that the branches are mutually uncorrelated. This assumption can be fulfilled if the antennas are spaced at least half the wavelength from each other [9, p. 330]. However, because of the difference of the scattering environment around the base station and the remote station, about quarter wavelength spacing of the antennas is sufficient to provide enough spatial decorrelation between the antennas at the mobile terminal [10]. On the other hand, separation of 10–20 wavelengths is required at the base station [10]. It is noted that the impact of spatial correlation between different receive antennas on our system is considered in [11] and [12]. In addition, we have extended the system further to operate on a MIMO channel in [13]. The issue about the needed accuracy of PSAM-based channel prediction for MIMO channels using transmit beamforming is considered in [14].

Similar to [1] and [15], the transmitted data stream is divided into frames—each of length L —with a pilot symbol at the beginning of each frame and is followed by $L - 1$ information symbols. Both the estimator and the predictor are linear and are made optimal in a maximum *a posteriori* (MAP) sense [16]. The difference between the two is that the estimator is noncausal and uses information from both past and future symbols, whereas the predictor is strictly causal and predicts relatively far into the future. Because of this, the order of the estimator can usually be lower than the order of the predictor.

The received, noisy, and faded pilot symbols of the j th branch are written in a complex baseband as

$$y_{\text{pl};j}(k; l) = \sqrt{\mathcal{E}_{\text{pl}}} h_j(k; l) s(k; l) + n_j(k; l) \quad l = 0; \quad j = 1, \dots, n_R \quad (1)$$

and the received data symbols as

$$y_{d;j}(k; l) = \sqrt{\mathcal{E}_d} h_j(k; l) s(k; l) + n_j(k; l) \quad l = 1, \dots, L - 1; \quad j = 1, \dots, n_R. \quad (2)$$

The index k counts the frame, and l is the symbol index in that frame. \mathcal{E}_{pl} and \mathcal{E}_d is the power per pilot and per data symbol, respectively, to be optimized later. $s(k; 0)$ is the pilot symbol and $\{s(k; l)\}_{l=1}^{L-1}$ are data symbols in the k th frame. For

simplicity, we assume that $E[|s(k; l)|^2] = |s(k; 0)|^2 = 1$. Furthermore, $n_j(\cdot, \cdot)$ denotes zero-mean complex-valued AWGN with variance $N_0/2$ per dimension and dimensions being uncorrelated. The fading envelope of the j th branch $h_j(k; l)$ is assumed to be a stationary complex Gaussian random process with zero mean and variance $\sigma_{h_j}^2 = 1$. Each subchannel is assumed to be slowly varying so that it remains constant over many channel symbols, and we perform the estimation and prediction independently on each subchannel. Note that the slowly fading conditions are in the sense that the channel does not reveal its ergodic properties over one data packet. That is, it is assumed that the data packet length is significantly smaller than the coherence time of the channel, implying low terminal velocity in the system.

The total average power for both pilot and data symbols is denoted by \mathcal{E} . The average power per data symbol is $\bar{\mathcal{E}}_d = \alpha \mathcal{L} \mathcal{E} / (L - 1)$, and the average power per pilot symbol is $\mathcal{E}_{\text{pl}} = (1 - \alpha) \mathcal{L} \mathcal{E}$. The constant α determines how power is distributed between pilot and data symbols. Equal data and pilot power is obtained when $\alpha = 1 - 1/L$.

III. CHANNEL ESTIMATION AND PREDICTION

The analysis in this section is done assuming that the channel is wide-sense stationary and that the correlation function of the channel is known.

A. Estimation

The optimally estimated channel in the MAP sense for the complex Gaussian case is a linear combination of the noisy observations [16, pp. 741–742]. Following the approach in [1], we can find the estimated channel $h_{e;j}(k; l)$, with the corresponding minimum mean square error (MMSE) of the estimated j th branch as

$$\sigma_{e;j}^2(l) = 1 - \sum_{\kappa=1}^{K_e} \frac{|\mathbf{u}_{\kappa}^H \mathbf{r}_e|^2 (1 - \alpha) L \bar{\gamma}_j}{(1 - \alpha) L \bar{\gamma}_j \lambda_{\kappa} + 1}. \quad (3)$$

In (3), K_e is the estimator order, $\bar{\gamma}_j = \mathcal{E}/N_0$ is the expected CSNR on any receive branch, $\{\mathbf{u}_{\kappa}\}$ denotes the eigenvectors of the covariance matrix of the channel gain $\mathbf{R}_e = E[\mathbf{h}\mathbf{h}^H]$, where $\mathbf{h} = [h_j(k - \lfloor K_e/2 \rfloor; 0), \dots, h_j(k + \lfloor (K_e - 1)/2 \rfloor; 0)]^T$, $\{\lambda_{\kappa}\}$ is the corresponding eigenvalues, and \mathbf{r}_e is the covariance vector defined by $\mathbf{r}_e = E[\mathbf{h}h_j^*(k; l)]$.

Due to the orthogonality principle [17, p. 264], $h_{e;j}(k; l)$ and the estimation error $\epsilon_{e;j}(k; l) = h_j(k; l) - h_{e;j}(k; l)$ are

uncorrelated. Both are zero-mean Gaussian random variables [18, Ch. 7.2].

B. Prediction

For notational simplicity, we restrict the system feedback delay¹ to be $\tau = DLT_s$, where D is a positive integer and T_s is the duration of a channel symbol. The predictor uses K_p pilot symbols from the past to predict one sample in the set $\{h_j(k;l)\}_{l=1}^{L-1}$ of the k th frame. Similar to the estimation case, we can find the predicted channel $h_{p;j}(k;l)$, which is Gaussian distributed with zero mean and variance $\sigma_{h_{p;j}}^2(l) = \sigma_{h_j}^2 - \sigma_{p;j}^2(l) = 1 - \sigma_{p;j}^2(l)$. The corresponding MMSE of the prediction error is

$$\sigma_{p;j}^2(l) = 1 - \sum_{\kappa=1}^{K_p} \frac{|\mathbf{u}_{\kappa}^H \mathbf{r}_p|^2 (1-\alpha)L\bar{\gamma}_j}{(1-\alpha)L\bar{\gamma}_j\lambda_{\kappa} + 1}. \quad (4)$$

Note that $\{\mathbf{u}_{\kappa}\}$ now denote the eigenvectors of the covariance matrix of the channel gain $\mathbf{R}_p = E[\mathbf{h}\mathbf{h}^H]$, where $\mathbf{h} = [h_j(k-D;0), \dots, h_j(k-D-K_p+1;0)]^T$, $\{\lambda_{\kappa}\}$ are the corresponding eigenvalues, and $\mathbf{r}_p = E[\mathbf{h}h_j^*(k;l)]$ is the covariance vector.

IV. BER PERFORMANCE ANALYSIS

A. BER Analysis in the Presence of Estimation Errors

To reduce the complexity of the receiver, we use the suboptimal symbol-by-symbol maximum likelihood (ML) detection on each subchannel as $z(k;l) = y_{d;j}(k;l)/(\sqrt{\mathcal{E}_d}h_{e;j}(k;l))$. As a result of that detection rule, the CSNR of a single branch at the time of detection can be found in [1]. The total CSNR after MRC is the sum of the individual branch CSNRs given as

$$\gamma(k;l) = \frac{\sum_{j=1}^{n_R} \mathcal{E}_d |h_{e;j}(k;l)|^2}{N_0 + g\mathcal{E}_d\sigma_{e;j}^2(l)} = \frac{\mathcal{E}_d \sum_{j=1}^{n_R} |h_{e;j}(k;l)|^2}{N_0 + g\mathcal{E}_d\sigma_e^2(l)} \quad (5)$$

where the last equality is obtained by assuming that the variance of the estimation error is the same for all branches. This is reasonable, because the branches are assumed uncorrelated, and the same estimation algorithm is used on all the branches. The constant $g = 1$ for 4-QAM and $g = 1.3$ for M -QAM when $M > 4$ [1].

ACM systems using 2-D trellis codes were first introduced by Goldsmith and Chua [19]. However, to the best of our knowledge, extending ACM to 4-D trellis codes was first done by Hole *et al.* [20]. In this paper, the 4-D trellis codes in [20] are utilized. Tight approximations for 4-D trellis-coded modulation (TCM) BER performance on AWGN channels can be found in [2], [5], and [20]. However, in order to obtain a closed-form and mathematically tractable solution when MRC is considered, we use a somewhat looser BER approximation expression, which,

¹The feedback delay here includes the time it takes to perform prediction, the actual transmission delay on the feedback channel, and the processing time needed by the transmitter to activate the code to be transmitted.

TABLE I
CODE-DEPENDENT CONSTANTS $\{a_n(\ell)\}_{\ell=1}^3$ AND $\{b_n(\ell)\}_{\ell=1}^3$
FOR THE EXAMPLE 4-D TRELLIS CODES

n	$a_n(1)$	$b_n(1)$	$a_n(2)$	$b_n(2)$	$a_n(3)$	$b_n(3)$
1	233.8034	12.4335	-280.8712	11.4405	51.3394	8.6131
2	210.6415	8.4208	-242.0657	7.9916	34.3732	6.0432
3	246.0565	7.7677	-334.6900	8.1130	89.4924	9.1087
4	99.7887	7.7426	-160.6040	8.3843	61.5091	9.4667
5	78.0083	7.0135	-100.8414	7.4793	23.4319	9.0714
6	86.2181	7.3704	-96.0270	7.6780	10.3583	10.3191
7	87.6912	7.0471	-94.5020	7.2852	7.3344	10.1898
8	89.3099	7.2848	-95.6889	7.4987	6.8972	10.4428

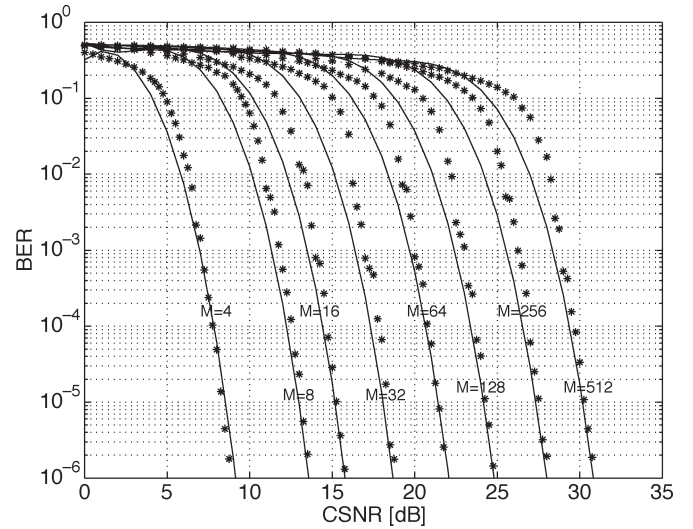


Fig. 2. BER performance of TCM codes on AWGN channels for different M -QAM constellations. Solid lines denote the approximations, whereas stars represent the simulated values.

similar to [21], is given by

$$\text{BER}(M_n|\gamma) = \sum_{\ell=1}^{\mathcal{L}} a_n(\ell) \exp\left(-\frac{b_n(\ell)}{M_n}\gamma\right). \quad (6)$$

Again, M_n is the number of points in the symbol constellation, and $a_n(\ell)$ and $b_n(\ell)$ are constellation-dependent constants (given in Table I) that can be found by first simulating the codes' BER performance and then using curve fitting with the least squares method. \mathcal{L} is the number of exponential functions that approximate the simulated BER. The approximation is illustrated in Fig. 2 for different constellation sizes. At first sight, the approximations seem to be quite coarse. However, it is shown later that the ACM performance results of this approximation are very close to those obtained when the tighter approximations in [2] and [5] are used. This is seen by comparing the results for the single-antenna system in [5] and what we achieve here for $n_R = 1$.

Inserting (5) into (6), the BER becomes

$$\text{BER}(M_n|\{h_{e;j}\}) = \sum_{\ell=1}^{\mathcal{L}} a_n(\ell) \prod_{j=1}^{n_R} \exp\left(-A_n \mathcal{E}_d |h_{e;j}(k;l)|^2\right) \quad (7)$$

where $A_n = b_n(\ell)/(M_n(\mathcal{E}/\bar{\gamma}_j + g\mathcal{E}_d\sigma_e^2(l)))$. This equation will be useful when deriving the BER in the presence of both prediction and estimation errors in the next subsection.

B. BER Analysis in the Presence of Both Estimation and Prediction Errors

We may express the estimated channel as

$$h_{e;j}(k;l) = \underbrace{h_{p;j}(k;l) + \epsilon_{p;j}(k;l)}_{h_j(k;l)} - \epsilon_{e;j}(k;l). \quad (8)$$

For a single-antenna system, $p(|h_{e;j}| | h_{p;j})$ is Ricean distributed [1]. Now, we consider the MRC technique with the branches assumed independent and identical. Thus, $p(\{|h_e|\} | \{h_p\}) = \prod_{j=1}^{n_R} p(|h_{e;j}| | h_{p;j})$. Each branch is still Ricean distributed, with the Rice factor $K = |(1 - \rho)h_{p;j}(k;l)|^2 / \tilde{\sigma}_{h_{e;j}}^2$, where $\rho = E[\epsilon_{e;j}(k;l)h_{p;j}^*(k;l)] / \sigma_{h_{p;j}}^2(l)$ is the correlation between the predicted channel and the estimation error and $\tilde{\sigma}_{h_{e;j}}^2 = \text{Var}(h_{e;j}(k;l) | h_{p;j}(k;l)) = E[|\epsilon_{p;j}(k;l) - \epsilon_{e;j}(k;l) + \rho h_{p;j}(k;l)|^2]$. It is shown in [1] that ρ typically takes on very small values and that the deviations in the BER performance obtained by using the exact value of ρ instead of setting $\rho = 0$ is less than 0.2 dB. We therefore choose to set $\rho = 0$ in our analysis. Moreover, in contrast to [1], $\epsilon_{e;j}(k;l)$ is correlated with $\epsilon_{p;j}(k;l)$. Thus, as a result, $\tilde{\sigma}_{h_{e;j}}^2 = \sigma_{p;j}^2(l) - \sigma_{e;j}^2(l)$. Hence, the BER conditioned on the combined predicted channels is given by

$$\begin{aligned} & \text{BER}(M_n | \{h_{p;j}\}) \\ &= \underbrace{\int_0^\infty \cdots \int_0^\infty}_{n_R\text{-fold}} \text{BER}(M_n | \{|h_{e;j}\}) \\ & \quad \times p(\{|h_{e;j}\} | \{h_{p;j}\}) d|h_{e;1}| \cdots d|h_{e;n_R}| \\ &= \sum_{\ell=1}^{\mathcal{L}} a_n(\ell) \underbrace{\int_0^\infty \cdots \int_0^\infty}_{n_R\text{-fold}} \prod_{j=1}^{n_R} \exp(-A_n \mathcal{E}_d |h_{e;j}|^2) \\ & \quad \times p(|h_{e;j}| | h_{p;j}) d|h_{e;1}| \cdots d|h_{e;n_R}|. \end{aligned} \quad (9)$$

With the help of [22, eq. (6.633-4)], by assuming $\tilde{\sigma}_{h_{e;j}}^2 = \tilde{\sigma}_{h_e}^2 \forall j$ and letting $d_n = 1/(A_n \mathcal{E}_d \tilde{\sigma}_{h_e}^2 + 1)$, (9) becomes

$$\text{BER}(M_n | \{h_{p;j}\}) = \sum_{\ell=1}^{\mathcal{L}} a_n(\ell) d_n^{n_R} \exp\left(-d_n A_n \mathcal{E}_d \sum_{j=1}^{n_R} |h_{p;j}|^2\right). \quad (10)$$

The predicted CSNR on each branch is defined as in [2] by $\hat{\gamma}_j = \bar{\mathcal{E}}_d |h_{p;j}(k;l)|^2 / N_0$, and the combined predicted CSNR

using the MRC scheme is obtained as [23, eq. (26)]

$$\hat{\gamma} = \frac{\bar{\mathcal{E}}_d}{N_0} \sum_{j=1}^{n_R} |h_{p;j}(k;l)|^2. \quad (11)$$

Solving this with respect to $\sum |h_{p;j}|^2$ and inserting the result into (10) gives

$$\text{BER}(M_n | \hat{\gamma}) = \sum_{\ell=1}^{\mathcal{L}} a_n(\ell) d_n^{n_R} \exp\left(-\frac{\hat{\gamma} d_n A_n \mathcal{E}_d}{\bar{\gamma}_j \bar{\mathcal{E}}_d}\right). \quad (12)$$

The combined predicted CSNR $\hat{\gamma}$ is fed back to the transmitter via the return channel and is used for deciding which code to use. If $\hat{\gamma}_n < \hat{\gamma} < \hat{\gamma}_{n+1}$, code n (or constellation of size M_n) is activated to transmit. To find the optimal switching thresholds $\{\hat{\gamma}_n\}_{n=1}^N$ in a maximal ASE sense, subject to BER and power constraints, we set (12) equal to BER_0 and solve for $\hat{\gamma}_n$. In contrast to [1], here, we have to use a numerical approach to obtain the solutions; this will be explained in Section V. The optimal switching thresholds here will be dependent on the average expected subchannel CSNR, which is in contrast to [2], where constant switching thresholds optimized for perfect CSI were used.

From (11), we can find the average predicted CSNR as

$$\bar{\gamma} = \frac{\bar{\mathcal{E}}_d}{N_0} \sum_{j=1}^{n_R} E[|h_{p;j}(k;l)|^2] = r \bar{\gamma}_j n_R \quad (13)$$

where $r = \bar{\mathcal{E}}_d(1 - \sigma_p^2)/\mathcal{E}$. The second equality is obtained by assuming that $\sigma_{p;j}^2 = \sigma_p^2 \forall j$, because the branches are assumed uncorrelated, and the same prediction algorithm is used on all the branches. The overall predicted CSNR with MRC of n_R branches will follow a Gamma distribution [2], i.e., $\hat{\gamma} \sim \mathcal{G}(n_R, r\bar{\gamma}_j) = (\hat{\gamma}^{n_R-1} / (\Gamma(n_R)(r\bar{\gamma}_j)^{n_R})) \exp(-\hat{\gamma}/(r\bar{\gamma}_j))$.

C. Overall Average BER Performance Analysis

The average BER for the n th constellation is found by averaging (12) over the Gamma probability density function of $\hat{\gamma}$ as

$$\begin{aligned} \text{BER}(M_n) &= \int_{\hat{\gamma}_n}^{\hat{\gamma}_{n+1}} \text{BER}(M_n | \hat{\gamma}) p(\hat{\gamma}) d\hat{\gamma} \\ &= \sum_{\ell=1}^{\mathcal{L}} a_n(\ell) \left(\frac{d_n \bar{\mathcal{E}}_d}{r d_n A_n \mathcal{E}_d + \bar{\mathcal{E}}_d}\right)^{n_R} \\ & \quad \times \left\{ \bar{\Gamma}\left(n_R, \hat{\gamma}_n \frac{r d_n A_n \mathcal{E}_d + \bar{\mathcal{E}}_d}{r \bar{\gamma}_j \bar{\mathcal{E}}_d}\right) \right. \\ & \quad \left. - \bar{\Gamma}\left(n_R, \hat{\gamma}_{n+1} \frac{r d_n A_n \mathcal{E}_d + \bar{\mathcal{E}}_d}{r \bar{\gamma}_j \bar{\mathcal{E}}_d}\right) \right\} \end{aligned} \quad (14)$$

where $\bar{\Gamma}(a, z) = (\int_z^\infty t^{a-1} e^{-t} dt) / \Gamma(a)$ is the normalized incomplete gamma function [24, eq. (11.3)].

The most frequently used estimate of the average BER (over all codes) is the ratio between the average number of bits in

error and the number of bits transmitted in total [1], [25], [26] expressed as

$$\text{BER} = \frac{\sum_{n=1}^N \text{BER}(M_n)R_n}{\sum_{n=1}^N P_n R_n} \quad (15)$$

where P_n is the probability that $\hat{\gamma} \in [\hat{\gamma}_n, \hat{\gamma}_{n+1})$; $P_n = \int_{\hat{\gamma}_n}^{\hat{\gamma}_{n+1}} p(\hat{\gamma})d\hat{\gamma} = \bar{\Gamma}(n_R, \hat{\gamma}_n/r\bar{\gamma}_j) - \bar{\Gamma}(n_R, \hat{\gamma}_{n+1}/r\bar{\gamma}_j)$.

V. OPTIMIZATION OF ASE

The goal of this section is to design an adaptive system where both α and L are optimized in such a way that the ASE is maximized while $\text{BER}(M_n|\hat{\gamma}) \leq \text{BER}_0$.

It is obvious that the variance of the prediction error is largest when predicting the last symbol in a frame, i.e., $l = L - 1$. On the other hand, the variance of the estimation error is almost the same for all l if the order of the estimator is $K_e \geq 20$ [1]. Thus, we use the variance of the estimation error $\sigma_{e;j}^2(L - 1)$ and the conservative choice of the prediction error variance $\sigma_{p;j}^2(L - 1)$ when finding the optimal switching thresholds $\{\hat{\gamma}_n\}$, as well as in the further optimization process.

A. ASE Performance Analysis

The system is experiencing an outage when the predicted CSNR falls below $\hat{\gamma}_1$, because, by definition, there is no code in our code set that guarantees the BER performance. In that case, the system does not send anything but the pilots—in order to perform the channel estimation and prediction—whereas the data must be buffered at the transmitter. Furthermore, because no transmission is allowed when $\hat{\gamma} < \hat{\gamma}_1$, we do not use any data power during that time. Therefore, the actual transmitted power per data symbol can be set to

$$\mathcal{E}_d = \frac{\bar{\mathcal{E}}_d}{\int_{\hat{\gamma}_1}^{\infty} p(\hat{\gamma})d\hat{\gamma}} = \frac{\bar{\mathcal{E}}_d}{\bar{\Gamma}\left(n_R, \frac{\hat{\gamma}_1}{r\bar{\gamma}_j}\right)}. \quad (16)$$

The SE of the n th constellation used by the $2G$ -dimensional² trellis code is $R_n = (1 - 1/L)(\log_2(M_n) - 1/G)$. Hence, the ASE is given by

$$\begin{aligned} \text{ASE} &= \sum_{n=1}^N R_n P_n \\ &= \frac{L-1}{L} \sum_{n=1}^N \left(\log_2(M_n) - \frac{1}{G} \right) \\ &\quad \times \left\{ \bar{\Gamma}\left(n_R, \frac{\hat{\gamma}_n}{r\bar{\gamma}_j}\right) - \bar{\Gamma}\left(n_R, \frac{\hat{\gamma}_{n+1}}{r\bar{\gamma}_j}\right) \right\} \end{aligned} \quad (17)$$

where the term $(L - 1)/L$ accounts for the fact that every L th symbol is a pilot in which no information is transmitted.

When using Nyquist sampling, L must be less than $L_{\max} = \lfloor 1/(2f_d T_s) \rfloor$ [4], where f_d is the maximum Doppler shift. Thus,

²As in [2] and [20], G is equal to 2 for our example codes in Section VI.

for $L \in [2, \dots, L_{\max}]$, we have the following optimization problem:

$$\begin{aligned} &\max_{\alpha} \quad \text{ASE} \\ &\text{subject to} \quad 0 < \alpha < 1. \end{aligned} \quad (18)$$

In order to solve $\text{BER}(M_n|\hat{\gamma}) = \text{BER}_0$ with respect to $\hat{\gamma}$, we need the value of \mathcal{E}_d , which, again, is a function of $\hat{\gamma}_1$ [cf. (16)]. The problem of finding $\{\hat{\gamma}_n\}$ is reduced to finding the $\hat{\gamma}_1$, because $\{\hat{\gamma}_n\}_{n=2}^N$ is easily found when \mathcal{E}_d is given. Thus, for the first constellation M_1 and given an α , we let \mathcal{E}_d vary as a function of the whole range of $\hat{\gamma}$ (note that it is not $\hat{\gamma}_1$, i.e., $\mathcal{E}_d(\hat{\gamma})$). The solution of $\text{BER}(M_1|\hat{\gamma}) = \text{BER}_0$ for that $\mathcal{E}_d(\hat{\gamma})$ yields the $\hat{\gamma}_1$, which is optimal if α is optimal. Once $\hat{\gamma}_1$ is known, \mathcal{E}_d is explicitly given by (16), and it can be used to find other thresholds. In that way, $\text{ASE}(\alpha)$ is reduced from a 2-D function $S(\alpha, \mathcal{E}_d)$ in [1, eq. (23)] to a one-dimensional (1-D) function. Thus, the optimization of the 1-D function of (18) is easily done by picking an $\alpha \in (0, 1)$ by numerical search, which maximizes the ASE.³

After solving (18) for all the possible L values, the maximum ASE is found by searching over all L in order to find the α and the L values that simultaneously maximize ASE.

VI. NUMERICAL EXAMPLE

At this point, we consider an ACM system that has a set of $N = 8$ QAM signal constellations of sizes $\{M_n\} = \{4, 8, 16, 32, 64, 128, 256, 512\}$ to switch between. Those constellations are used to represent eight 4-D trellis codes. The carrier frequency is 2 GHz, and the length of a channel symbol is $5 \mu\text{s}$, which corresponds to a channel bandwidth of 200 kHz using Nyquist sampling. The mobile velocity is $v = 30 \text{ m/s}$, and the system delay considered is $\tau = 1 \text{ ms}$. With those parameters, the Doppler frequency is $f_d = 200 \text{ Hz}$ and the normalized delay $\tau f_d = 0.2$. We require the system to tolerate a $\text{BER}_0 = 10^{-5}$, and, as in [1], we choose the order of the estimator and the predictor to be $K_e = 20$ and $K_p = 250$, respectively. That choice of K_p leads to a suboptimal but satisfactory predictor [27]. Furthermore, we assume that the expected subchannel CSNR is the same for all the branches, i.e., $\bar{\gamma} = \bar{\gamma}_1 = \bar{\gamma}_2 = \dots = \bar{\gamma}_{n_R}$.

Figs. 3 and 4 show the optimum pilot symbol period⁴ and the optimum fraction of power allocation to the pilot symbols, respectively, for different numbers of branches to combine. As can be seen in Fig. 3, when the power distribution between pilot and data symbols is optimized, fewer pilot symbols are needed. The spacing between two pilot symbols is also larger when there are more antennas available to combine. The power allocated to pilot symbols is lower for higher diversity orders for both the equal power and optimal power allocation cases (cf. Fig. 4). This is due to the array gain and to the fact that MRC is optimal in the sense that it maximizes the output CSNR,

³For this purpose, we have used the function `fminbnd` in MATLAB.

⁴Note that because $\tau = DLT_s = 1 \text{ ms}$, and both D and L must be positive integers, the value of L must follow $L = 200/D$. Thus, L has to be in the set $L \in \{2, 4, 5, 8, 10, 20, 25, 40, 50, 100, 200\}$.

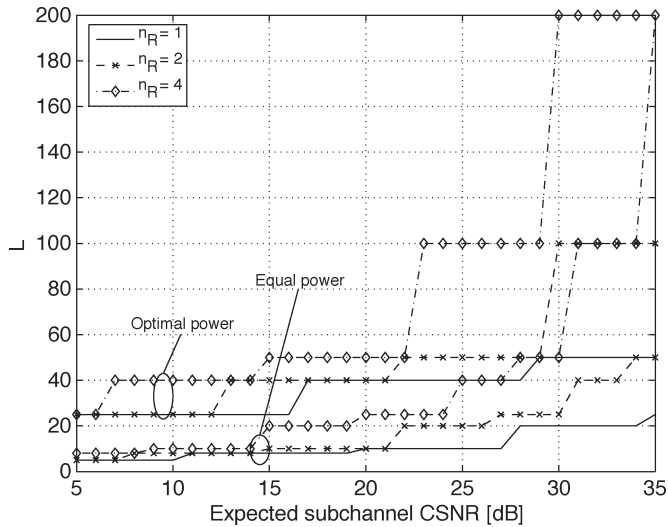


Fig. 3. Optimum pilot period L when the power is equally and optimally allocated between pilot and data symbols. The number of antennas to combine is $n_R = 1, 2,$ and 4 .

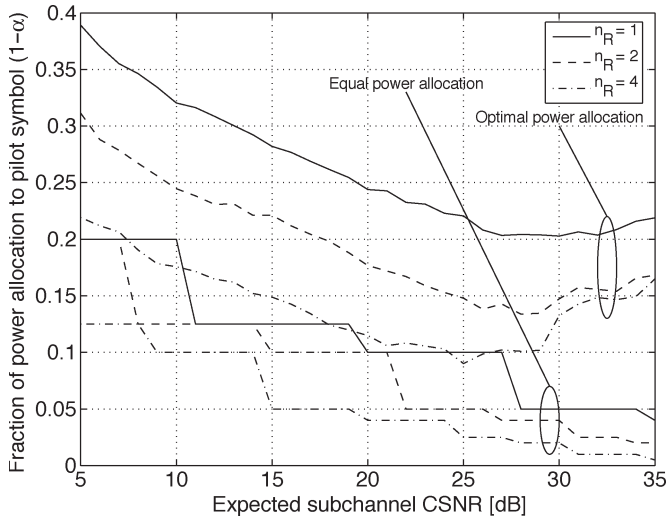


Fig. 4. Fraction of power allocation to pilot symbols (i.e., $1 - \alpha$) when the pilot period L is optimal. The number of receive antennas is $n_R = 1, 2,$ and 4 .

thus, less pilot power is needed when more antenna is available. To correspond to the larger pilot period, the optimal power allocation scheme puts more power on the pilots than when the power is equally allocated.

On the other hand, the pilot power increases again at very high CSNR for the optimal case. This corresponds to the steeply increasing pilot period in the same CSNR region. In order to have a good channel prediction and estimation that the system can rely on—so that the BER requirement is maintained—more power should be allocated to the pilots.

Because of the finite number of codes, the ASE reaches a ceiling when the CSNR grows. As expected, the ASE is higher when the transmit power and the pilot period are optimized and when we have more antennas available to combine. This is due to the array gain as well as to the fact that the pilot period is larger for higher n_R . This is shown clearly in Fig. 5.

As can be seen in Fig. 6, the theoretical average BER curves—produced from (15)—are always lower than

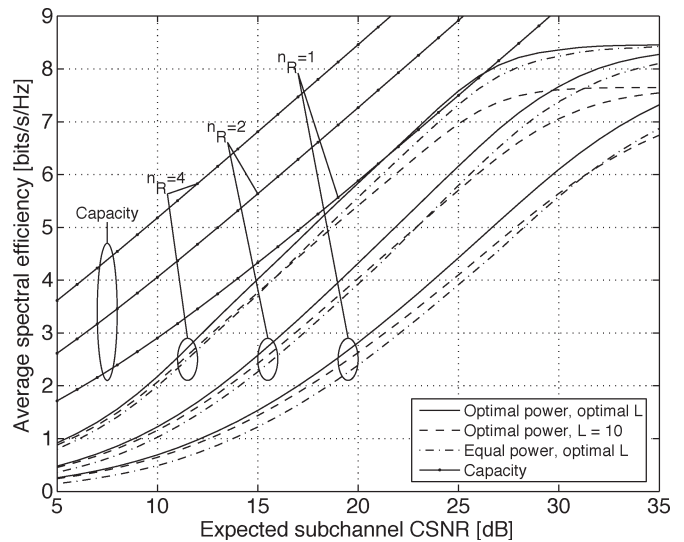


Fig. 5. Average spectral efficiency for both optimal parameters and some fixed parameters together with the channel capacity for Rayleigh fading. n_R is the number of receive antennas and is set to 1, 2, and 4.

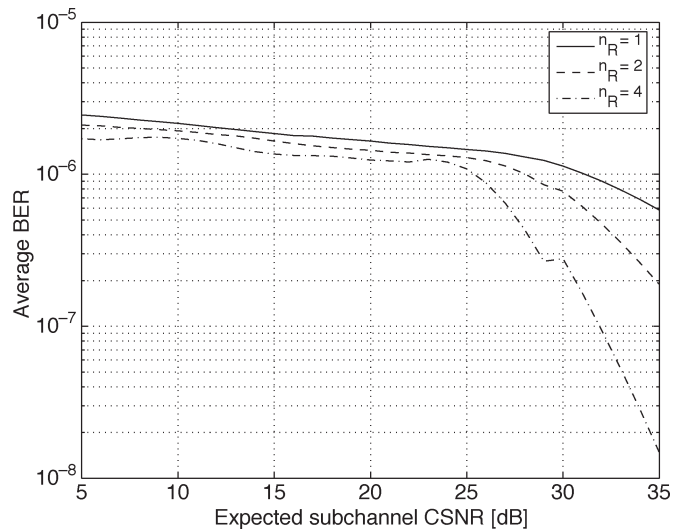


Fig. 6. Average BER for optimal power and optimal L with different numbers of receive antennas ($n_R = 1, 2,$ and 4).

$BER_0 = 10^{-5}$. This is due to the requirement that the instantaneous BER is less than or equal to BER_0 .

We observe that our optimum ASE is approximately 0.7 b/s/Hz higher than for uncoded M -QAM for $n_R = 1$ (cf. Fig. 5 and [1, Fig. 5] at 20 dB). This corresponds to about 3 dB gain in average CSNR because of the coding.

With the same delay and system parameters, the system in [2] will not be able to operate acceptably with one or even with two receive antennas. Combining four antennas, the system starts functioning at an average subchannel CSNR = 21 dB for $L = 10$. To conclude, our system operates satisfactorily and superior to the system in [2] for the whole range of CSNRs under consideration and for the considered delay. This holds especially at the high-CSNR region where fixing of pilot spacing is clearly a disadvantage, because the channel can also be satisfactorily tracked with larger pilot period.

VII. CONCLUSION

We have investigated an ACM system where receive diversity is implemented by means of MRC. The ASE is a result of a joint optimization of the pilot symbol period and the power distribution between pilot and data symbols. The ASE is considerably increased compared to when the pilot spacing is fixed and when the power is equally allocated to pilot and data symbols. The increased ASE is obtained without losing BER performance, especially in the high-CSNR regions. This gain is due to the fact that the rate of the pilot symbols is substantially reduced in that region; thus, fixing of the pilot symbol rate is clearly a disadvantage.

The optimal power allocation strategy is the one that puts more power on pilot symbols as a response to the decreased pilot symbol rate. The pilot spacing increases and the pilot power decreases for higher diversity orders and for both optimal and nonoptimal power distribution schemes.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their careful reading, comments, and feedback on the manuscript, which significantly improved the quality of this paper.

REFERENCES

- [1] X. Cai and G. B. Giannakis, "Adaptive PSAM accounting for channel estimation and prediction errors," *IEEE Trans. Wireless Commun.*, vol. 4, no. 1, pp. 246–256, Jan. 2005.
- [2] G. E. Øien, H. Holm, and K. J. Hole, "Impact of channel prediction on adaptive coded modulation performance in Rayleigh fading," *IEEE Trans. Veh. Technol.*, vol. 53, no. 3, pp. 758–769, May 2004.
- [3] 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.
- [4] 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.
- [5] D. V. Duong and G. E. Øien, "Adaptive trellis-coded modulation with imperfect channel state information at the receiver and transmitter," in *Proc. Nordic Radio Symp.*, Oulu, Finland, Aug. 2004.
- [6] S. Zhou and G. B. Giannakis, "Adaptive modulation for multi-antenna transmissions with channel mean feedback," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1626–1636, Sep. 2004.
- [7] Y. Ko and C. Tepedelenlioglu, "Space-time block coded rate-adaptive modulation with uncertain SNR feedback," in *Proc. Asilomar Conf. Systems and Computers*, Pacific Grove, CA, 2003, pp. 1032–1036.
- [8] —, "Optimal switching thresholds for space-time block coded rate-adaptive M-QAM," in *Proc. Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Montreal, QC, Canada, May 2004, pp. IV-477–IV-480.
- [9] T. S. Rappaport, *Wireless Communications Principles & Practice*. Upper Saddle River, NJ: Prentice-Hall, 1999.
- [10] J. Winters, "Smart antennas for wireless systems," *IEEE Pers. Commun.*, vol. 5, no. 1, pp. 23–27, Feb. 1998.
- [11] B. Holter and G. E. Øien, "Impact of spatial correlation on adaptive coded modulation performance in Rayleigh fading," *IEEE Trans. Veh. Technol.* to be published.
- [12] D. V. Duong, M. Å. Wingar, and G. E. Øien, "Link adaptation in correlated antenna diversity environments," in *Proc. Norwegian Signal Processing Symp. (NORSIG)*, Stavanger, Norway, Sep. 2005.
- [13] D. V. Duong, B. Holter, and G. E. Øien, "Optimal pilot spacing and power in rate-adaptive MIMO diversity systems with imperfect transmitter CSI," in *Proc. IEEE Workshop Signal Processing Advances Wireless Communications (SPAWC)*, New York, Jun. 2005, pp. 47–51.
- [14] S. Zhou and G. B. Giannakis, "How accurate channel prediction needs to be for transmit-beamforming with adaptive modulation over Rayleigh MIMO channels?," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1285–1294, Jul. 2004.
- [15] X. Tang, M. S. Alouini, and A. J. 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.
- [16] H. Meyr, M. Moeneclaey, and S. A. Fechtel, *Digital Communication Receivers: Synchronization, Channel Estimation and Signal Processing*. New York: Wiley, 1998.
- [17] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*, 4th ed. New York: McGraw-Hill, 2002.
- [18] C. W. Therrien, *Discrete Random Signals and Statistical Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1992.
- [19] A. J. Goldsmith and S.-G. Chua, "Adaptive coded modulation for fading channels," *IEEE Trans. Commun.*, vol. 46, no. 5, pp. 595–602, May 1998.
- [20] K. J. Hole, H. Holm, and G. E. Øien, "Adaptive multi-dimensional coded modulation over flat fading channels," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 7, pp. 1153–1158, Jul. 2000.
- [21] S. Falahati, A. Svensson, M. Sternad, and H. Mei, "Adaptive trellis-coded modulation over predicted flat fading channels," in *Proc. IEEE Vehicular Technology Conf. (VTC)*, Orlando, FL, Oct. 2003, pp. 1532–1536.
- [22] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*, 6th ed. San Diego, CA: Academic, 2000.
- [23] D. G. Brennan, "Linear diversity combining techniques," *Proc. IEEE*, vol. 91, no. 2, pp. 331–356, Feb. 2003.
- [24] N. M. Temme, *Special Functions: An Introduction to the Classical Functions of Mathematical Physics*. New York: Wiley, 1996.
- [25] M.-S. Alouini and A. J. Goldsmith, "Adaptive modulation over Nakagami fading channels," *Kluwer J. Wireless Commun.*, vol. 13, no. 1/2, pp. 119–143, May 2000.
- [26] S. T. Chung and A. J. Goldsmith, "Degrees of freedom in adaptive modulation: A unified view," *IEEE Trans. Commun.*, vol. 49, no. 9, pp. 1561–1571, Sep. 2001.
- [27] G. E. 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 Symp. (NORSIG)*, Hurtigruten, Norway, Oct. 2002.



Duc V. Duong (S'03) was born in Qui Nhon, Vietnam, in 1977. He received the M.Sc. degree in electrical engineering from the Department of Telecommunications, Norwegian University of Science and Technology (NTNU), Trondheim, Norway, in 2002, where he is working toward the Ph.D. degree at the Department of Electronics and Telecommunications.

His research interests include the area of wireless communications, especially analysis of fading channel, channel estimation, and channel prediction with application to link adaptation.



Geir E. Øien (S'90–M'01) was born in Trondheim, Norway, in 1965. He received the M.Sc. and the Ph.D. degrees from the Norwegian Institute of Technology (NTH), Trondheim, in 1989 and 1993, respectively, both in electrical engineering.

From 1994 to 1996, he was an Associate Professor at Stavanger University College, Stavanger, Norway. In 1996, he joined the Norwegian University of Science and Technology, Trondheim, as an Associate Professor, and in 2001, he was promoted to Full Professor. His current research interests include the

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 multiple-input multiple-output systems; and orthogonal-frequency-division-multiplexing-based wireless systems.



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, Bergen, Norway, in 1984, 1987, and 1991, respectively.

From August 1988 to May 1990, he was a Visiting Scholar with the Center for Magnetic Recording Research, University of California, San Diego. In 1993, he was with the IBM Almaden Research Center, San Jose, CA. From 1995 to 2005, he was a Research Scientist at the University of Bergen and at the Norwegian University of Science and Technology (NTNU), Trondheim, with funding from the Norwegian Research Council. In 2005, he became a Full Professor at the University of Bergen. His current research interests are in the areas of wireless communication and data security.