ADAPTIVE CODED MODULATION WITH RECEIVE ANTENNA DIVERSITY AND IMPERFECT CHANNEL KNOWLEDGE AT RECEIVER AND TRANSMITTER

Duc V. Duong*, Geir E. Øien*, and Kjell J. Hole†

* Dept. of Electronics and Telecommunications, Norwegian University of Science and Technology
O. S. Bragstads pl. 2B, N-7491 Trondheim, Norway. Email: {duong,oien}@iet.ntnu.no
† Coding Theory Group, University of Bergen. Pb. 7800, N-5020 Bergen, Norway. Email: kjell.hole@ii.uib.no

ABSTRACT

We analyze and optimize the performance of a trellis-coded modulation system where receive antenna diversity is implemented. The adaptive coded modulation scheme alone is a promising tool to combat fading and it improves the spectral efficiency in mobile communications. The improvement can be even higher when we have multiple receiver antennas. The analysis is done in the presence of both estimation and prediction errors. Independent Rayleigh fading on each subchannel is considered, and *maximum ratio combining* is employed on the receiver side. As expected, the *average spectral efficiency* increases, while *bit error rate* is always fulfilled.

1. INTRODUCTION

Adaptive coded modulation (ACM) is an efficient tool to counteract the fading in wireless channel communications and to improve the spectral efficiency [1]. The improvement can even be higher when diversity and combining techniques are included [2], [3]. The scheme presented in this paper is a generalization of two recent papers; one by Cai and Giannakis [4], in which an uncoded adaptive modulation system based on M-QAM constellations with single transmit and single receive antenna is investigated, and one by Øien *et al.* [3] 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 [3], coding is included and when MRC is implemented on the receiver side. The pilot spacing and the power allocation to pilot and data symbols are optimized as in [4], for maximal spectral efficiency under an instantaneous *bit error rate* (BER) constraint. Instantaneous here is with respect to predicted channel-signal-to-noise ratio (CSNR).

The estimator and predictor in our adaptive scheme are based on received known pilot symbols which are sent at regular intervals on the transmitter side. The predicted CSI is fed back to the transmitter via a zero-error, non-zero delay feedback channel. Based on this, the transmitter dynamically adapts the rate and the power to a mode which is suitable for the channel to improve the *average spectral efficiency* (ASE) while maintaining the instantaneous BER below some pre-defined target BER$_0$. As in [3], [5], *multi-dimensional* trellis codes designed for additive white Gaussian noise (AWGN) channels are used as component codes.

The reader should be aware that we omit the time indices wherever it is possible, for notational simplicity. $\mathcal{D}(\mathbf{z})$ represents the diagonal matrix with vector $\mathbf{z}$ on its diagonal. $\mathbf{I}_K$ is the $K \times K$ identity matrix. Superscripts $(\cdot)^{\mathrm{T}}$, $(\cdot)^*$ and $(\cdot)^{\mathrm{H}}$ stand for transpose, complex conjugate, and Hermitian transpose, respectively. $\lfloor z \rfloor$ means the integer part of z, and $E[\cdot]$ and $\mathrm{Var}\,(\cdot)$ denotes the expectation and variance operator, respectively.

The outline of the remainder of the paper is as follows. We describe our system in Section 2. The analysis of BER performance and ASE performance can be found in Sections 3 and 4, respectively. Numerical results are given in Section 5 before the conclusions are drawn in Section 6.

2. SYSTEM MODEL

The system we are considering is illustrated in Fig. 1, where we have multiple receive antennas. The adaptive encoder chooses to transmit symbols from the constellation of size M_n that is best suited to the channel state, out of a set of constellations of sizes $\{M_n\}_{n=1}^N$, corresponding to the set of *spectral efficiencies* (SEs) $\{R_n\}_{n=1}^N$, where N is the number of constellations. The SEs are ordered such that $R_1 < R_2 < \cdots < R_N$, and we also have $M_1 < M_2 < \cdots < M_N$. The choice of M_n is based on the CSI fed back from the receiver. Also, the power allocation between data and pilot symbols is also chosen in an optimal way. The criterion for choosing the code and the power used is to maximize ASE while fulfilling a target BER. We include MRC on the receiver side, where the branches are assumed to be mutually uncorrelated.

Similar to in [4], [5], we use a frame structure where each frame starts with a pilot symbol and is followed by $L - 1$ information symbols. Both the estimator and predictor are linear and are made optimal in *maximum a posteriori* (MAP) sense [6]. The difference between the two is that the estimator is non-causal and uses information from both past and future symbols, while the predictor is strictly causal. Due to 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 jth branch are written in complex baseband as

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

and the received data symbols as

$$y_{d;j}(k;l) = \sqrt{\mathcal{E}_d} h_j(k;l) s(k;l) + n_j(k;l),\ l \in [1, \cdots, L-1], \quad (2)$$

where k is the frame index, l is the symbol index in that frame, and $j \in [1, \cdots, n_R]$. $\mathcal{E}_{pl}$ 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 kth 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 jth branch, $h_j(k;l)$, is assumed to be a stationary complex Gaussian random process with zero mean and variance $\sigma_{h_j}^2 = 1$. Moreover, the channel 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.

Similarly to in [4], [7], the average power per data symbol and per pilot symbol is $\bar{\mathcal{E}}_d = \alpha L \mathcal{E}/(L-1)$ and $\bar{\mathcal{E}}_{pl} = (1-\alpha) L \mathcal{E}$, respectively, where $\mathcal{E}$ is the average transmit power for both pilot and data symbols. Equal data and pilot power is obtained if $\alpha = 1 - 1/L$. We will later on optimize the system with respect to α and L.

3. BER PERFORMANCE ANALYSIS

3.1 BER Analysis in the Presence of Estimation Errors

The optimally estimated channel in the MAP sense for the complex Gaussian case is a linear combination of the observations [6, p. 741–742]. Let $h_{e;j}(k;l)$ be the optimal linear estimate of the fading
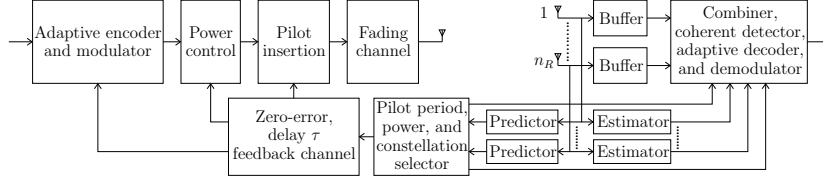
Figure 1: The adaptive pilot-symbol-assisted modulation system model combined with receive diversity combining.

channel of the jth branch, let $\epsilon_{e;j}(k;l) = h_j(k;l) - h_{e;j}(k;l)$ be the corresponding estimation error, and let the mean square error (MSE) be $\sigma_{e;j}^2(l) = E[|\epsilon_{e;j}(k;l)|^2]$. Since $h_{e;j}(k;l)$ is a result of an optimal filtering process and $h_j(k;l)$ is a zero mean Gaussian random variable (RV), $h_{e;j}(k;l)$ and $\epsilon_{e;j}(k;l)$ are also zero mean Gaussian RVs.

We assume that the estimator is of order K_e, which means that it uses a vector of K_e received pilot symbols, $\mathbf{y}(k;0) = [y_{pl;j}(k - \lfloor K_e/2 \rfloor;0), \cdots, y_{pl;j}(k + \lfloor (K_e - 1)/2 \rfloor;0)]^{\mathrm{T}}$, to estimate one sample. We define the pilot symbol vector $\mathbf{s} = [s(k - \lfloor K_e/2 \rfloor;0), \cdots, s(k + \lfloor (K_e - 1)/2 \rfloor;0)]^{\mathrm{T}}$, the channel gain vector $\mathbf{h} = [h_j(k - \lfloor K_e/2 \rfloor;0), \cdots, h_j(k + \lfloor (K_e - 1)/2 \rfloor;0)]^{\mathrm{T}}$, the covariance matrix $\mathbf{R}_e = E[\mathbf{h}\mathbf{h}^{\mathrm{H}}]$, and the covariance vector $\mathbf{r}_e = E[\mathbf{h}h_j^*(k;l)]$. The linear channel estimator is given by [4]: $\mathbf{w}_e = \sqrt{\mathcal{E}_{pl}}(\mathcal{E}_{pl}\boldsymbol{\mathcal{D}}(\mathbf{s})\mathbf{R}_e\boldsymbol{\mathcal{D}}^*(\mathbf{s}) + N_0\mathbf{I}_{K_e})^{-1}\boldsymbol{\mathcal{D}}(\mathbf{s})\mathbf{r}_e$, and the estimated channel is $h_{e;j}(k;l) = \mathbf{w}_e^{\mathrm{H}}\mathbf{y}(k;0)$. As a result, the minimum MSE (MMSE) of the estimation error for the jth branch can be calculated as

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

In (3), $\{\mathbf{u}_\kappa\}$ denote the eigenvectors of $\mathbf{R}$, $\{\lambda_\kappa\}$ are the corresponding eigenvalues, and $\bar{\gamma}_j = E[|h_j(k;l)|^2]\mathcal{E}/N_0 = \mathcal{E}/N_0$ is the average CSNR of one branch.

To reduce the complexity of the receiver, we use the suboptimal symbol-by-symbol *maximum likelihood* (ML) detection on each subchannel; $z(k;l) = y_{d;j}(k;l)/(\sqrt{\mathcal{E}_d}h_{e;j}(k;l))$. Based on that detection rule, the CSNR of a single branch is given in [4]. The total CSNR after MRC is the sum of the individual branch CSNRs,

$$\gamma(k;l) = \sum_{j=1}^{n_R} \frac{\mathcal{E}_d|h_{e;j}(k;l)|^2}{N_0 + g\mathcal{E}_d\sigma_{e;j}^2(l)} = \frac{\mathcal{E}_d\sum_{j=1}^{n_R}|h_{e;j}(k;l)|^2}{N_0 + g\mathcal{E}_d\sigma_e^2(l)}, \tag{4}$$

where the last equality is obtained by assuming that the variance of the estimation error is the same for all branches. The constant $g = 1$ for 4-QAM and $g = 1.3$ for M-QAM when $M > 4$ [4].

Tight approximations for trellis-coded modulation (TCM) BER performance on AWGN channels can be found in [3], [5]. In order to obtain a closed-form and mathematically tractable solution when MRC is considered we use a somewhat looser BER approximation expression which is given by

$$\mathrm{BER}\left(M_n\middle|\gamma\right) = \sum_{\ell=1}^{\mathcal{L}} a_n(\ell)\exp\left(-\frac{b_n(\ell)\gamma}{M_n}\right) \tag{5}$$

where $\mathcal{L}$ is the number of the exponential functions which approximate the simulated BER, M_n is the number of points in the symbol constellation, and $a_n(\ell)$, $b_n(\ell)$ are constellation dependent constants which can be found in [7, Tab. I]. The approximation is illustrated in Fig. 2.

Inserting (4) into (5), the BER becomes

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

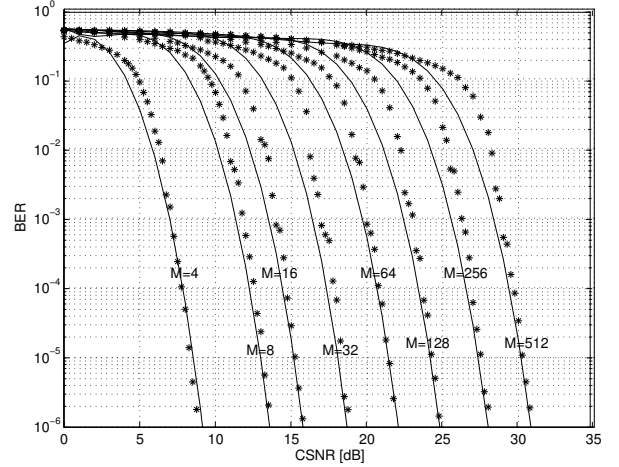where $A_n = b_n(\ell)/(M_n(\mathcal{E}/\bar{\gamma}_j + g\mathcal{E}_d\sigma_e^2(l)))$.

Figure 2: BER performance of TCM codes on AWGN channels for different M-QAM constellations. The solid lines denote the approximations, while the stars are the simulated values.

3.2 BER Analysis in the Presence of both Estimation and Prediction Errors

The predictor is strictly causal, and 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 kth frame. Let the predicted subchannel be $h_{p;j}(k;l)$, the prediction error is $\epsilon_{p;j}(k;l) = h_j(k;l) - h_{p;j}(k;l)$, and the MSE of the prediction error is $\sigma_{p;j}^2(l) = E[|\epsilon_{p;j}(k;l)|^2]$. The optimally predicted channel of one branch $h_{p;j}(k;l)$ is also a Gaussian distributed random variable with zero mean and variance $\sigma_{h_{p;j}}^2(l) = \sigma_{h_j}^2 - \sigma_{p;j}^2 = 1 - \sigma_{p;j}^2$.

We assume now that the system feedback delay[1] $\tau = DLT_s$, where D is a positive integer (for notational simplicity) and T_s is the duration of a channel symbol. Defining the transmitted pilot symbol vector $\mathbf{s} = [s(k - D;0), \cdots, s(k - D - K_p + 1;0)]^{\mathrm{T}}$—corresponding to the receive pilot vector $\mathbf{y}(k;0)$—the channel gain vector is $\mathbf{h} = [h_j(k - D;0), \cdots, h_j(k - D - K_p + 1;0)]^{\mathrm{T}}$, the covariance matrix is $\mathbf{R}_p = E[\mathbf{h}\mathbf{h}^{\mathrm{H}}]$, and the covariance vector is $\mathbf{r}_p(l) = E[\mathbf{h}h_j^*(k;l)]$. Similar to the estimation case, the predicted channel is given as $h_{p;j}(k;l) = \mathbf{w}_p^{\mathrm{H}}\mathbf{y}$ where $\mathbf{w}_p = \sqrt{\mathcal{E}_{pl}}(\mathcal{E}_{pl}\boldsymbol{\mathcal{D}}(\mathbf{s})\mathbf{R}_p\boldsymbol{\mathcal{D}}^*(\mathbf{s}) + N_0\mathbf{I}_{K_p})^{-1}\boldsymbol{\mathcal{D}}(\mathbf{s})\mathbf{r}_p$ is the MAP-optimal predictor, and the MMSE of the predicted channel is

$$\sigma_{p;j}^2(l) = 1 - \sum_{\kappa=1}^{K_p} \frac{|\mathbf{u}_\kappa^{\mathrm{H}}\mathbf{r}_p|^2(1-\alpha)L\bar{\gamma}_j}{(1-\alpha)L\bar{\gamma}_j\lambda_\kappa + 1}, \tag{6}$$

where $\{\mathbf{u}_\kappa\}$ and $\{\lambda_\kappa\}$ are the sets of eigenvectors and eigenvalues of $\mathbf{R}$, respectively.

Since $h_j = h_{p;j} + \epsilon_{p;j}$, we can express the estimated channel as

---

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

$$h_{e;j}(k;l) = h_{p;j}(k;l) + \epsilon_{p;j}(k;l) - \epsilon_{e;j}(k;l). \quad (7)$$

For a single antenna system $p(|h_{e;j}||h_{p;j})$ is Rice distributed with $K = |(1-\rho)h_{p;j}(k;l)|^2/\tilde{\sigma}^2_{h_{e;j}}$, where ρ is the correlation between $\epsilon_{e;j}(k;l)$ and $h_{p;j}(k;l)$. It is shown in [4] that ρ typically takes on very small values and that it can be set to zero. On the other hand, in contrary to [4], the estimation error $\epsilon_{e;j}(k;l)$ is *correlated* with the prediction error $\epsilon_{p;j}(k;l)$. Hence, $\tilde{\sigma}^2_{h_{e;j}} = \text{Var}\big(|h_{e;j}||h_{p;j}\big) = \sigma^2_{p;j} - \sigma^2_{e;j}$.

Now we consider the MRC technique with the branches assumed uncorrelated. Thus, $p(\{|h_e|\}|\{h_p\}) = \prod_{j=1}^{n_R} p(|h_{e;j}||h_{p;j})$. As a result, $\text{BER}\big(M_n|\{h_{p;j}\}\big)$ can be calculated as:

$$\text{BER}\big(M_n|\{h_{p;j}\}\big) = \underbrace{\int_0^\infty \cdots \int_0^\infty}_{n_R-\text{fold}} \text{BER}\big(M_n|\{|h_{e;j}|\}\big)$$
$$\times p\big(\{|h_{e;j}|\}|\{h_{p;j}\}\big)\, d|h_{e;1}| \cdots d|h_{e;n_R}|, \quad (8)$$

and the result can be written as

$$\text{BER}\big(M_n|\{h_{p;j}\}\big) = \sum_{i=1}^{\mathcal{L}} a_n(\ell) d_n^{n_R} \exp\left(-A_n d_n \mathcal{E}_d \sum_{j=1}^{n_R} |h_{p;j}|^2\right) \quad (9)$$

where we have assumed that $\tilde{\sigma}^2_{h_{e;j}} = \tilde{\sigma}^2_{h_e}\ \forall j$, and have let $d_n = 1/(A_n \mathcal{E}_d \tilde{\sigma}^2_{h_e} + 1)$. The predicted CSNR on each branch is defined as in [3] 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 $\hat{\gamma} = \bar{\mathcal{E}}_d \sum_{j=1}^{n_R} |h_{p;j}(k;l)|^2/N_0$. Inserting this into (9) gives

$$\text{BER}(M_n|\hat{\gamma}) = \sum_{i=1}^{\mathcal{L}} a_n(\ell) d_n^{n_R} \exp\left(-\frac{\hat{\gamma} A_n d_n \mathcal{E} \mathcal{E}_d}{\bar{\gamma}_j \bar{\mathcal{E}}_d}\right). \quad (10)$$

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 (constellation of size M_n) is used. 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 (10) equal to BER_0 and solve for $\hat{\gamma}_n$. Here we have to use a numerical approach to obtain the solutions and this will be explained in Section 4.

We can find the average predicted CSNR as $\bar{\hat{\gamma}} = (\bar{\mathcal{E}}_d/N_0) \sum_{j=1}^{n_R} E[|h_{p;j}(k;l)|^2] = r\bar{\gamma}_j n_R$ where $r = \bar{\mathcal{E}}_d(1 - \sigma^2_p)/\mathcal{E}$ and we have assumed that the prediction error is the same for all branches. The overall predicted CSNR with MRC of n_R branches will follow a Gamma distribution, i.e. $\hat{\gamma} \sim \mathcal{G}(n_R, r\bar{\gamma}_j)$ [3].

3.3 Overall Average BER Performance Analysis

The average BER for the nth constellation is found by averaging (10) over the Gamma pdf of $\hat{\gamma}$. That is,

$$\text{BER}(M_n) = \int_{\hat{\gamma}_n}^{\hat{\gamma}_{n+1}} \text{BER}(M_n|\hat{\gamma}) p(\hat{\gamma}) d\hat{\gamma}$$
$$= \sum_{i=1}^{\mathcal{L}} a_n(\ell) \left(\frac{d_n \bar{\mathcal{E}}_d}{r A_n d_n \mathcal{E} \mathcal{E}_d + \bar{\mathcal{E}}_d}\right)^{n_R} \times \Psi \quad (11)$$

where $\Psi = \bar{\Gamma}\left(n_R, \hat{\gamma}_n \frac{r A_n d_n \mathcal{E} \mathcal{E}_d + \bar{\mathcal{E}}_d}{r\bar{\gamma}_j \bar{\mathcal{E}}_d}\right) - \bar{\Gamma}\left(n_R, \hat{\gamma}_{n+1} \frac{r A_n d_n \mathcal{E} \mathcal{E}_d + \bar{\mathcal{E}}_d}{r\bar{\gamma}_j \bar{\mathcal{E}}_d}\right)$ and $\bar{\Gamma}(\cdot, \cdot)$ is the normalized incomplete gamma function [7].

The overall average BER is defined as the ratio between the average number of bits in error, and the number of bits transmitted in total [3], [4]:

$$\text{BER} = \frac{\sum_{n=1}^N \text{BER}(M_n) R_n}{\sum_{n=1}^N P_n R_n}, \quad (12)$$

where R_n is the SE of the nth constellation, and P_n is the probability that code no. n is utilized. I.e. $P_n = \int_{\hat{\gamma}_n}^{\hat{\gamma}_{n+1}} p(\hat{\gamma}) d\hat{\gamma} = \bar{\Gamma}(n_R, \hat{\gamma}_n/r\bar{\gamma}_j) - \bar{\Gamma}(n_R, \hat{\gamma}_{n+1}/r\bar{\gamma}_j)$.

4. OPTIMIZATION OF ASE

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 when the order of the estimator is of order $K_e \geq 20$ [4]. Thus, we use the variance of the estimation error $\sigma^2_{e;j}(L-1)$ and the conservative choice of the variance of the prediction error $\sigma^2_{p;j}(L-1)$ when deriving the optimal switching thresholds $\{\hat{\gamma}_n\}_{n=1}^N$ as well as in the further optimization process.

4.1 ASE Performance Analysis

The system is experiencing an *outage* when the predicted CSNR falls below $\hat{\gamma}_1$, since there is no code in our code set which then 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—while the data is buffered at the transmitter. And since 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 $\mathcal{E}_d = \bar{\mathcal{E}}_d / \int_{\hat{\gamma}_1}^\infty p(\hat{\gamma}) d\hat{\gamma} = \bar{\mathcal{E}}_d / \bar{\Gamma}(n_R, \hat{\gamma}_1/r\bar{\gamma}_j)$.

The SE of the nth constellation used by the $2G$-dimensional trellis code is $R_n = (1 - 1/L)(\log_2(M_n) - 1/G)$ where $G = 2$ [3]. Hence, the ASE is given by

$$\text{ASE} = \sum_{n=1}^N R_n P_n$$
$$= \frac{L-1}{L} \sum_{n=1}^N \left(\log_2(M_n) - \frac{1}{G}\right)$$
$$\times \left\{\bar{\Gamma}\left(n_R, \frac{\hat{\gamma}_n}{r\bar{\gamma}_j}\right) - \bar{\Gamma}\left(n_R, \frac{\hat{\gamma}_{n+1}}{r\bar{\gamma}_j}\right)\right\} \quad (13)$$

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

When using Nyquist sampling, L must be less than $L_{\max} = \lfloor 1/(2f_d T_s) \rfloor$ [3], [4] where f_d is the maximum Doppler shift. Thus, for $L \in [2, \cdots, L_{\max}]$ we have the following optimization problem:

$$\max_{\alpha, L} \text{ASE}(\alpha) \quad \text{subject to } 0 < \alpha < 1. \quad (14)$$

The optimization algorithm used is described in [5], [7], (a more complicated algorithm is also used in [4]); thus we choose not to go into details about it here and, instead, refer to those papers.

5. NUMERICAL RESULTS

At this point, we consider an example ACM system which 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 code and decode eight 4-dimensional trellis codes. The carrier frequency is 2 GHz and the length of a channel symbol is 5 μs (corresponding to a channel bandwidth of 200 kHz when using Nyquist sampling). The system delay considered is $\tau = DLT_s = 1$ ms (or $\tau f_d = 0.2$) and the mobile velocity $v = 30$ m/s. With those parameters, the Doppler frequency $f_d = 200$ Hz. We require the system to tolerate a $\text{BER}_0 = 10^{-5}$, and choose the order of the estimator and predictor to be $K_e = 20$ and $K_p = 250$, respectively. Contrary to the results given in [4], for $\tau = 1$ ms, L can only take on certain discrete values. This is due to the fact that $D = 200/L$ and it must be an integer. As a result, $L \in \{2, 4, 5, 8, 10, 20, 25, 40, 50, 100, 200\}$.

Fig. 3 shows how the optimal pilot spacing is distributed with the average CSNR per receive branch. As can be seen, the pilot period increases with increasing CSNR and is larger for the optimal power case. It increases faster with more receive antennas.

Compared to the equal power scheme, more power is put on the pilots when it is optimized (cf. Fig. 4). This is natural since
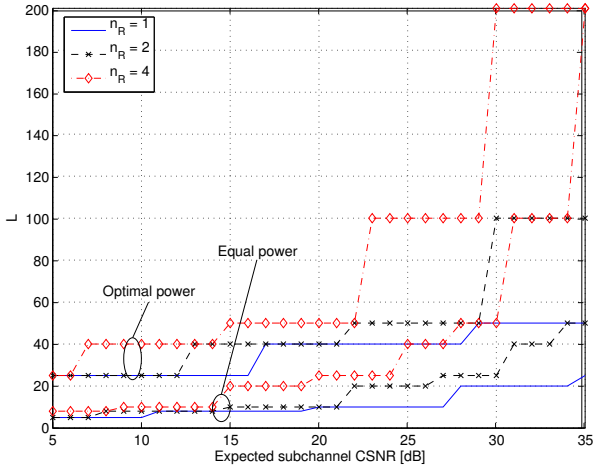
Figure 3: Optimal pilot spacing L when the power is optimally allocated between pilot and data symbols. The number of receive antennas is $n_R = 1, 2$, and 4.
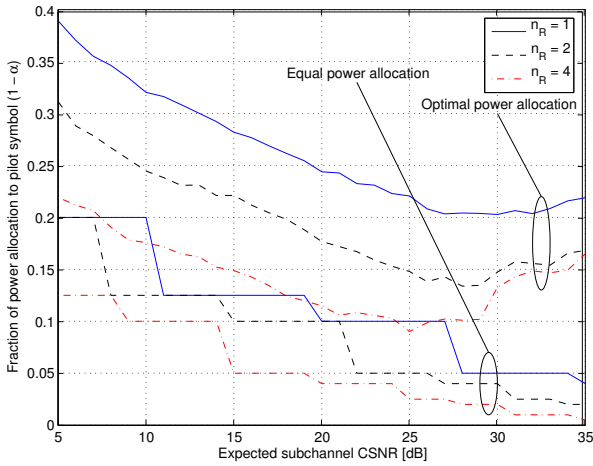

Figure 4: Fraction of power allocation to pilot symbols (i.e. 1α) when the pilot period L is optimal. The number of receive antennas is $n_R = 1, 2$, and 4.

data is already coded and since the distance between two pilot symbols is larger. The power allocated to pilot symbols is lower with higher diversity order for both schemes. It is due to the array gain and the fact that MRC is optimal in the sense that it maximizes the output CSNR, thus less power is needed when more antennas are available. The pilot power decreases with CSNR, but it increases again at about 28 dB for the optimal case. This effect corresponds to the steeply increased pilot spacing in the same CSNR region, and makes sure that the system still has a good channel prediction which it can rely on. Hence, the BER is maintained.

Because of the finite number of codes, the ASE reaches a ceiling when the CSNR grows large. As expected, the ASE is higher when we have more antennas available to combine, and when the power and pilot period are optimal. This is shown clearly in Fig. 5. At low CSNRs, the gain is not remarkable compared to when, e.g., L is fixed to 10, but in the high CSNR region, when the pilot symbols are dropped out significant, the gain is substantially increased.

We do not include the figure for the average BER here due to space limitations. However, the BER constraint is satisfied (below $BER_0 = 10^{-5}$) for all the values of n_R, and CSNRs.
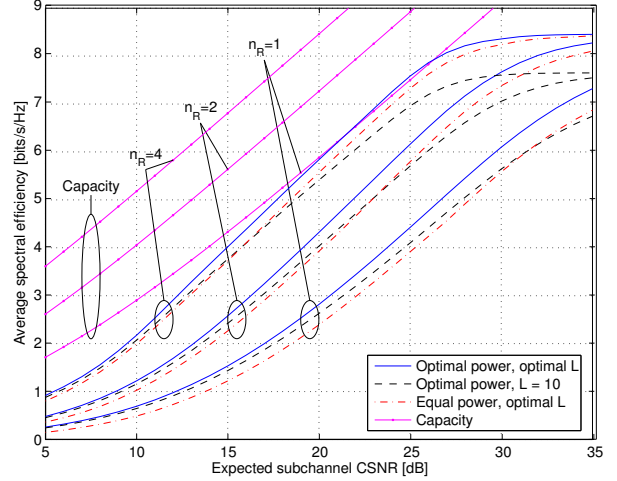

Figure 5: Average spectral efficiency for both optimal and non-optimal parameters. $n_R = 1, 2$, and 4 is the number of receive antennas.

6. CONCLUSIONS

We have investigated an ACM system where receive diversity by means of MRC is implemented. The ASE is a result of an optimization of the power distribution between pilot and data symbols, and of the frequency of pilot symbols. Numerical results are in strong agreement with our expectations. The ASE is considerably increased; especially at high CSNRs. This is due to the fact that the rate of the pilot symbols is significantly reduced in that region. More or less a constant gain is achieved when the power is equally distributed compared to when it is optimal; both for optimal L. In addition, the gain becomes more significant when the number of receive antennas is increased. These improvements are obtained without any sacrifice of BER performance.

REFERENCES

[1] A. J. Goldsmith and S.-G. Chua, "Variable-rate variable-power MQAM for fading channels," *IEEE Transactions on Communications*, vol. 45, no. 10, pp. 1218–1230, Oct 1997.

[2] M.-S. Alouini and A. J. Goldsmith, "Capacity of rayleigh fading channels under different adaptive transmission and diversity-combining techniques," *IEEE Transactions on Vehicular Technology*, vol. 48, no. 4, pp. 1165–1181, July 1999.

[3] G. E. Øien, H. Holm, and K. J. Hole, "Impact of channel prediction on adaptive coded modulation performance in Rayleigh fading," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 3, pp. 758–769, May 2004.

[4] X. Cai and G. B. Giannakis, "Adaptive PSAM accounting for channel estimation and prediction errors," *IEEE Transactions on Wireless Communications*, vol. 4, no. 1, pp. 246–256, Jan 2005.

[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 Symposium*, Oulu, Finland, Aug 2004.

[6] H. Meyr, M. Moeneclaey, and S. A. Fechtel, *Digital Communication Receivers: Synchronization, Channel Estimation and Signal Processing*. John Wiley & Sons, 1998.

[7] D. V. Duong, G. E. Øien, and K. J. Hole, "Adaptive coded modulation with receive antenna diversity and imperfect channel knowledge at receiver and transmitter," submitted to *IEEE Transacstions on Vehicular Technology*.