Analysis and Optimization of SIMO Systems with Adaptive Coded Modulation in Spatially Correlated Rayleigh Fading

Duc V. Duong*, Bengt Holter†, and Geir E. Øien*
*Dept. of Electronics and Telecommunications, Norwegian University of Science and Technology
N-7491 Trondheim, Norway
†SINTEF ICT, Communication Systems, O. S. Bragstads plass 2C, N-7465 Trondheim, Norway
Email: {duong, oien}@iet.ntnu.no, bengt.holter@sintef.no

Abstract—In this paper we investigate the performance of an adaptive coded modulation (ACM) system in single-input multiple-output Rayleigh fading channels with spatial correlation. The considered ACM system has a set of multi-dimensional trellis codes from which it can adaptively select the code best suited to the channel quality. The channel quality is tracked at the receiver by a predictor using a pilot-symbol-assisted modulation scheme. We optimize both the pilot spacing, and the power allocation to pilot and data symbols. The channel predictor is made optimal in the maximum a posteriori sense and the correlation to other subchannels is taken into account. A numerical example is given for flat Rayleigh fading subchannels with Jakes correlation profile.

I. INTRODUCTION AND BACKGROUND

In analysis of antenna diversity systems, independence between the received signals from different antenna branches is usually assumed [1]. However, this assumption might not hold, due to, e.g., insufficient spacing of antenna elements at the mobile terminal (MT) or insufficient scattering around the base station (BS). Thus, this way of thinking, spatial correlation is playing an important role when analyzing the systems performance, since it is well known that correlation degrades performance. However, this degradation can be reduced if the correlation is exploited properly.

Overview of research contributions to the field of adaptive communications can be found in [2]. The adaptive coded modulation (ACM) system in this paper extends the work in [3] to include spatial correlation in the receive antenna array and the optimization is inspired by [4]. Different from [5] and [6]—where the subchannels were predicted independently of each other without considering the spatial correlation—we will predict the subchannels jointly, i.e. the spatial correlation is assumed known so that it can be taken into account. Thus, a “space-time predictor” is needed and must be derived. This predictor is used to track the variations of a single-input multi-output (SIMO) fading channel by means of pilot-symbol-assisted modulation, and the information is fed back to the transmitter to aid the adaptation. Here, only channel prediction is considered since it is necessary for system adaptation and since the channel estimation can be made with high accuracy using an optimal noncausal estimator with certain delay [7]. Throughout in the paper we use (·)* and (·)H to denote complex conjugate and conjugate transpose, respectively, tr(A) is the trace of the matrix A, vec(A) is the vectorize operator, and ⊗ is the kronecker product. Moreover, we denote the expectation, variance, and covariance by E[·], Var(·), and Cov(·, ·), respectively. The notation C denotes the set of complex numbers. In next section, we present the system and space-time correlation model. The space-time predictor and the signal statistics are discussed in Section III, whereas bit error rate (BER) and average spectral efficiency (ASE) performance are analyzed in Section IV. A numerical example is given in Section V and concluding remarks are drawn in Section VI.

II. SYSTEM AND CORRELATION MODEL DESCRIPTION

The system is depicted in Figure 1 and the received data vector \( y(k; l) \in \mathbb{C}^{n_k \times 1} \) is written as

\[
y(k; l) = \sqrt{E_d} h(k; l) s(k; l) + n(k; l), \quad l \in [1, \dots, L],
\]

and the pilot symbol vector \( y(k; 0) \in \mathbb{C}^{n_k \times 1} \) as

\[
y(k; 0) = \sqrt{E_p} h(k; 0) s(k; 0) + n(k; 0).
\]

Note that \( x(k; l) \) is the compact way of writing \( x(k; TS_s + lT_s) \) where \( T_s \) is a channel symbol duration and \( L \) is the pilot spacing. According to the model in (1) and (2) the transmitted signal is divided into frames of length \( L \), where each frame starts with a pilot symbol. \( s(k; l) \) is transmitted data symbols except for when \( l = 0 \) where a pilot is transmitted. Both the received vectors \( y \), the channel gain vector \( h \), and the additive white Gaussian noise (AWGN) vector \( n \) are complex with dimension \( n_R \times 1 \). The channel gains \( h \) are assumed to be stationary complex Gaussian with zero mean and unity variance. The AWGN is assumed uncorrelated both in space and time, and has zero mean and variance \( N_0 \). Furthermore, \( E_d \) and \( E_p \) is the power per data symbol and per pilot symbol, respectively. We assume that \( |s(k; 0)| = 1 \).

As in [3], the feedback information is the overall predicted channel-signal-to-noise ratio (CSNR), \( \hat{\gamma} \), which is computed once per frame. Based on this, the transmission mode is dynamically adapted such that a code with higher rate will be transmitted when the channel quality is good (high predicted
CSNR) and the rate will be reduced as the channel is getting worse (lower predicted CSNR). Moreover, no data is transmitted when the channel quality is below a certain threshold. The overall goal is to transmit as much information as possible, i.e., to maximize the ASE without sacrificing BER performance.

Similarly to [3] and [4], the actual data power can be configured as \(E_d = E_{d\ell} / \int_0^\infty p(\gamma)\,d\gamma\), where \(\gamma\) is the threshold below which only pilots are transmitted in order to track the channel variations and data is buffered, and \(p(\gamma)\) is the probability density function (pdf) of the predicted CSNR. Using the frame-based data stream the average data power is \(\bar{E}_{d\ell} = \alpha E_{d\ell} / (L - 1)\). Hence, the pilot power is \(E_p = (1 - \alpha)E_{d\ell}\). Note that \(\alpha\) is the variable which determines how much power should be allocated to pilot and data symbols, respectively, and \(E_{d\ell}\) is the average power for both pilot and data symbol.

When the MT moves away from (or toward) the BS—i.e., when the mobile moving angle is the same as the mobile position angle (or \(+\pi\)) related to the antenna array (see [8, Fig. 1])—the space-time correlation of the fading gain is approximately the product of the spatial and temporal correlation function [8]. According to [9], such a model is adequate for gauging average system behavior. On the other hand, the exact and general space-time correlation function can be found in [10] and [11]. Here we are interested only in average behavior and thus, with reference to [8], we assume that the space-time correlation between antenna branch \(\mu\) and \(\nu\) at lag \(m\) is

\[
E[h_\mu(n)h^*_\nu(n + m)] \approx E[h_\mu(n)h^*_\mu(n)] E[h_\nu(n)h^*_\nu(n + m)] = \rho_{h,\mu}(m) \cdot \rho_{h,\nu}(m),
\]

where \(\rho_{h,\mu}\) and \(\rho_{h,\nu}\) denote the spatial and temporal correlation of the fading gain, respectively.

### III. CHANNEL PREDICTION AND PDFS OF THE COMBINED SIGNALS

We buffer the pilot symbols received at all the antennas as

\[
\mathbf{y} = \begin{bmatrix}
y_1(k - D; 0) & \ldots & y_{m_1}(k - D; 0) \\
\vdots & \ddots & \vdots \\
y_1(k - D - K_p + 1; 0) & \ldots & y_{n_K}(k - D - K_p + 1; 0)
\end{bmatrix},
\]

where \(K_p\) is the predictor order and \(D\) (a positive integer) is the prediction horizon measured in the number of frames. The corresponding channel matrix \(\mathbf{H}\) and noise matrix \(\mathbf{N}\) are organized in a similar way. Assuming the same pilot symbol is transmitted, the maximum likelihood estimate of the received vector is \(\hat{\mathbf{y}} = \text{vec}(\mathbf{y}) / (s\sqrt{E_p})\).

Let the vector of the predicted channels be \(\hat{\mathbf{h}} = \mathbf{G}^H\hat{\mathbf{y}}\). Then the optimal predictor \(\mathbf{G}\) in the maximum a posteriori sense—taking into account the spatial correlation—can be found as

\[
\mathbf{G} = \mathbf{R}_{\hat{\mathbf{y}}\hat{\mathbf{y}}}^{-1}\mathbf{R}_{\hat{\mathbf{y}}\mathbf{x}} = \left(\mathbf{R}_{\hat{\mathbf{h}}} + \frac{N_0}{(1 - \alpha)L}E_{\mathbf{I}_n}\right)^{-1}\mathbf{R}_{\hat{\mathbf{y}}\mathbf{x}}
\]

where \(\mathbf{R}_{\hat{\mathbf{y}}\mathbf{x}} = \mathbb{E}[\hat{\mathbf{y}}\mathbf{h}^H]\) and \(\mathbf{R}_{\hat{\mathbf{y}}} = \mathbb{E}[\hat{\mathbf{y}}\hat{\mathbf{y}}^H] = \mathbb{E}[\mathbf{h}\mathbf{h}^H] + \mathbb{E}[\mathbf{m}\mathbf{m}^H]/\{\mathbb{E}[\mathbf{p}^2]\}.\) In that way, the correlation between any branches of any lags is completely described in \(\mathbf{R}_{\hat{\mathbf{h}}}\). Using the assumption in (3), \(\mathbf{R}_{\hat{\mathbf{h}}} = \mathbf{R}^s \otimes \mathbf{R}^t\) and \(\mathbf{R}_{\hat{\mathbf{y}}\mathbf{x}} = \mathbf{R}^s \otimes \mathbf{r}\). Here the notations \(\mathbf{R}^s = \mathbb{E}[\mathbf{h}_\mu h^*_\mu]\), \(\mathbf{R}^t = \bar{\mathbb{E}}[\mathbf{h}_\mu h^*_\mu]\), and \(\mathbf{r} = \mathbb{E}[\mathbf{h}_\mu h^*_\mu(k; l)]\) are used to denote the spatial, temporal correlation matrices, and covariance vector, respectively, and \(\otimes\) is the Kronecker product. As a result, the total prediction error variance is given by \(\sigma^2_p = \mathbf{r}^H \mathbf{G}\).

We will only consider linear array antenna elements where the correlation between such elements decreases exponentially. Thus, the power spatial correlation between the antenna branches \(\mu\) and \(\nu\) is \(\rho_{\mu\nu} = \rho^{|\mu - \nu|}\), where \(\rho_s = |\rho_{h,\nu}(1)|^2 = |(c - j\delta)|^2\) [12], is the absolute square of the envelope spatial correlation between any two adjacent subchannels. For simplicity we assume that the real and imaginary parts are equally correlated, i.e., \(c = d = \sqrt{\rho_s / 2}\).

Using maximum ratio combining (MRC) the instantaneous and predicted CSNR are defined as

\[
\gamma_d = \frac{E_d[||\mathbf{h}(k; l)||^2]}{N_0} \quad \text{and} \quad \hat{\gamma} = \frac{\bar{E}_{\mathbf{y}}[||\hat{\mathbf{h}}(k; l)||^2]}{N_0},
\]

respectively. Thus average CSNRs can be calculated as

\[
\tilde{\gamma}_d = E[\gamma_d], \quad \hat{\gamma} = E[\hat{\gamma}] = \tilde{\gamma} r_{dJ} n_{R},
\]

respectively, where \(\tilde{\gamma}_d = \mathbb{E}[N_0]\) is the average received CSNR on one branch, \(\tilde{\gamma} = \mathbb{E}[\mathbf{I}_n] = \tilde{\gamma} r_{dJ} n_{R}\), and \(\mathbb{E}[\hat{\gamma}] = \tilde{\gamma} r_{dJ} n_{R}\). Both \(\tilde{\gamma}_d\) and \(\tilde{\gamma}\) are introduced in order to use the same notations as in [5]. As apposed to in [3] the combined CSNR, both instantaneous and predicted CSNR, at the receiver is not gamma distributed anymore due to the existence of the spatial correlation between different antenna elements. While the exact pdf for an arbitrarily spatial correlation model can be found in, e.g., [13] there exists no standard bivariate pdf corresponding to the joint pdf of instantaneous and predicted CSNR. Thus we approximate the CSNR pdfs to be gamma distributed\(^1\) with the first two moments equal to the exact

\(^1\)\(x\) is said to be gamma distributed with shape parameter \(\varphi\) and scale parameter \(\lambda\)—denoted by \(x \sim \mathcal{G}(\varphi, \lambda)\)—if the pdf of \(x\) is \(p(x) = \frac{x^{\varphi - 1}}{\Gamma(\varphi)\lambda^\varphi} \exp\left(-\frac{x}{\lambda}\right), x \geq 0.\)
pdf [14]. As a result, the joint pdf is bivariate gamma. Using the notation \( \gamma_d \sim G(m_d, \theta_1) \) and \( \hat{\gamma} \sim G(m_d, \theta_2, \rho) \) to denote instantaneous and predicted CSNR following the gamma distribution, the joint pdf is \( \sim G(m_d, \theta_1, \theta_2, \rho) \) [6] where \( \theta_1 = rv \theta \) and \( \theta_2 = r \theta \). Note that \( \gamma_d = m_d \theta_1 \) and \( \hat{\gamma} = m_d \theta_2 \), and both \( m_d \) and \( \theta_1 \) are listed in [5] and [6]. \( \rho \) is the space-time correlation of \( \gamma_d \) and \( \hat{\gamma} \). Mathematically it is written as:

\[
\rho = \frac{\text{Cov}(\gamma_d, \hat{\gamma})}{\sqrt{\text{Var}(\gamma_d) \text{Var}(\hat{\gamma})}} = \frac{\text{Cov}(\beta^2, \hat{\beta}^2)}{\sqrt{\text{Var}(\beta^2) \text{Var}(\hat{\beta}^2)}}, \tag{7}
\]

where \( \beta^2 = \sum_{\mu} \beta_{\mu}^2 = \sum_{\mu} |h_{\mu}|^2 \), and \( \hat{\beta}^2 = \sum_{\mu} |\hat{h}_{\mu}|^2 \). It can be shown that \( \rho \) can be expressed as (8) in the bottom of this page, where \( \mathbf{g}_m \) is the \( m \)th column of the matrix \( \mathbf{G} \), and \( \mathbf{R}(\alpha) \) is the \( \alpha \)th column of the matrix \( \mathbf{R} \). \( \mathbb{R}(\{z\}) \) and \( \Im(\{z\}) \) denotes real and imaginary part of \( z \), respectively. It is noted that \( \rho \) in [5], [6] is independent of the spatial correlation \( \rho_s \) since the channel is predicted independently on each subchannel. Contrary to that, \( \rho \) in this paper also contains \( \rho_s \). When using the same predictor on all branches and predicting the branches independently (i.e., the effect of spatial correlation is not considered in the prediction process), (8) reduces to the results obtained in [6].

IV. BER AND ASE Performance Analysis

The overall goal of the system is to strive for a maximal information rate at any given time. At the same time the BER performance must be guaranteed. Since the ACM system has a set of codes to switch between, the selected code alone must satisfy the BER requirement. The set of codes used in this paper is based on multidimensional trellis codes which are originally designed for AWGN channel. The actual symbols transmitted over the channel are based on quadrature amplitude modulation (QAM). The BER performance of these codes over a Gaussian channel is given in [3, Eq. (6)]. The adaptivity is relying on the predicted CSNR, and, as a consequence, the switching thresholds must be determined from instantaneous (with respect to predicted CSNR) BER. Therefore, we need an expression for BER as a function of predicted CSNR.

Let \( a_n(\ell), b_n(\ell) \) be code-dependent constants [3, Tab. I], \( M_n \) be the constellation size, and \( \Phi = M_n + \theta_1 b_n(\ell)(1 - \rho) \). Similarly to in [5] the instantaneous BER is given by

\[
\text{BER}(M_n|\gamma) = \int_0^\infty \text{BER}(M_n|\gamma) p(\gamma|\hat{\gamma}) d\gamma \text{ and is}
\]

\[
\text{BER}(M_n|\gamma) = \sum_{\ell=1}^L a_n(\ell) \left( \frac{M_n}{\Phi} \right)^{n_a} \exp \left( -\gamma \rho \theta_1 b_n(\ell) \right).
\]

Thus the switching thresholds \( \{ \gamma_n \}_{n=1}^N \) are found by solving \( \text{BER}(M_n|\gamma) = \text{BER}_0 \) for the corresponding constellation \( M_n \), and the solutions is found by numerical search.

Averaging \( \text{BER}(M_n|\gamma) \) over \( p(\hat{\gamma}) \) would result in the average \( \text{BER}(M_n) \); which is obtained in closed-form [5]. Then the overall average BER is found as [15]

\[
\text{BER} = \frac{\sum_{n=1}^N \text{BER}(M_n) R_n}{\sum_{n=1}^N P_n R_n}, \tag{9}
\]

where \( R_n = (1 - 1/L)(\log_2(M_n) - 1/2) \) is the spectral efficiency of the constellation \( M_n \) [3] and \( P_n = \int_{\gamma_n}^\infty p(\hat{\gamma}) d\hat{\gamma} = \Gamma(m_d, \gamma_n/\theta_2) - \Gamma(m_d, \gamma_n/\theta_2) \) is the probability that constellation \( M_n \) is used. Moreover, \( \Gamma(a, b) = \Gamma(a)/\Gamma(a) \) is the normalized incomplete gamma function.

The overall ASE (the average number of bits transmitted in total) is given by

\[
\text{ASE} = \sum_{n=1}^N R_n P_n. \tag{10}
\]

To avoid aliasing, \( L \) must be less than \( L_{max} = \left\lceil 1/(2f_d T_s) \right\rceil \) where \( f_d \) is the maximum Doppler shift. Thus, for \( L \in [2, \cdots, L_{max}] \) we have the following optimization problem:

\[
\max_{\alpha} \text{ASE} \text{ subject to } 0 < \alpha < 1. \tag{11}
\]

The numerical search optimization algorithm used here is described in [3]; thus we do not go into details about it here and, instead, refer to that paper.

V. Numerical Example

We consider an example ACM system having a discrete set of \( N = 8 \) QAM signal constellations of sizes \( \{ M_n \} = \{ 4, 8, 16, 32, 64, 128, 256, 512 \} \) available to switch between. These constellations are used to code and decode eight 4-dimensional trellis codes [3], [6]. The carrier frequency is 2 GHz and the length of a channel symbol is 5 \( \mu s \)—corresponding to a channel bandwidth of 200 kHz using

\[
\rho = \frac{1}{n_R \text{tr}(\mathbf{R}^H \mathbf{G} \mathbf{G}^H \mathbf{R})} \times m_d \tag{8}
\]
Nyquist sampling. The system delay (or prediction horizon) considered is $DLT_s = 1$ ms$^2$ (corresponding to the normalized delay $f_dT_s = 0.2$). With the mobile velocity $v = 30$ m/s and the given carrier frequency, the Doppler frequency is $f_d = 200$ Hz. We require the system to tolerate a BER$_0 = 10^{-5}$, and the prediction of one subchannel is based on 250 pilot symbols from itself and 250 from each of the other subchannels when spatial correlation exists. Furthermore, we assume that the expected subchannel CSNR is the same for all the branches—i.e., $\gamma_j = \bar{\gamma}$ $\forall j \in \{1, \cdots, n_R\}$—and $\rho_s \in \{0.2, 0.7\}$. Assuming isotropic scattering these correlation corresponds to an antenna separation of 0.33 and 0.17 wavelength, respectively.

The results for ASE when both pilot spacing and power allocation are optimal are plotted in Figure 2. It is clear that the ASE loss due to the existence of spatial correlation is reduced when all the subchannels are jointly predicted. The gain is larger when there are many antennas available to combine. However, we can see that the curves corresponding to joint prediction of the correlated branches are higher compared to the uncorrelated case at low CSNR. Also, we expect the curves to move closer to the curves for $\rho_s = 0$ at high CSNR, but the results show the opposite. Both effects are a result of using the approximated pdfs for both instantaneous and predicted CSNR instead of using the exact ones. The same behavior is observed in [6] under perfect conditions such as perfect channel estimation and zero delay.

Intuitively, it is expected that the diversity gain disappears when the branches become more correlated. Thus the effective (combined) channel will vary more like the one-branch case and, therefore, we need the same amount of pilot overhead to predict the channel at a given accuracy. This effect is confirmed in Figure 3 where it is clear that, with increasing spatial correlation (from left to right panel), the curves for $L$ when $n_R > 1$ approach the one when $n_R = 1$. However, the combined channel will approach a no diversity scenario with higher average CSNR, due to the array gain. Because of the array gain, the power allocated to pilot symbols does not vary much from the uncorrelated case, which is illustrated in Figure 4. At very high CSNR the pilot power increases again to compensate for the large pilot spacing. The variations of the curves in Figure 4 in that region is due to the numerical instability of the optimization process because the ASE curves saturate.

Looking at Figs. 3 and 4 together we see that it might be sufficient to decrease the pilot spacing only. The statement is confirmed when we plot (not included here) and compare the ASE curves in Figure 2 when $\rho_s = 0.7$ with the ASE performance for the same spatial correlation; using the optimal values of $L$ but the power scheme for the uncorrelated case. The difference in performances is hardly noticeable.

The average BER performance is plotted in Figure 5 where it can be seen that the requirement of BER$_0 = 10^{-5}$ is always satisfied. In presence of spatial correlation and using the approximate pdfs the slope indicating the diversity order at high CSNR is not correctly depicted. As documented in [16], the error rate performance at high CSNR is strongly dependent on the behavior of the lower tail of a pdf, and this is exactly where approximate pdf deviates from the exact one. However, this does not put any limitations on our work since we are more interested in the low and medium CSNR regions. For high CSNR values, the ACM system behaves like a fixed-rate system anyway, since only the largest constellation

---

3Since we allow the system to reconfigure the transmit mode only once per frame, we predict the last symbol in the frame which is 1 ms ahead in time because it will give us the largest prediction error. Moreover, since $D$ must be an integer, $L$ can only take on the values in the set \{2, 4, 5, 8, 10, 20, 25, 40, 50, 100, 200\}.
channels are well documented in the literature [1, Chap. 9.7].

ACKNOWLEDGMENTS

The authors would like to thank Prof. Lajos Hanzo for discussion and comments. Also, we would like to thank the project BEATS and SINTEF for financial support.

REFERENCES