

# Multilevel Codes for OFDM-Like Modulation over Underspread Fading Channels

Siddhartha Mallik and Ralf Koetter

*The Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA*

Received 7 June 2005; Revised 3 May 2006; Accepted 12 May 2006

We study the problem of modulation and coding for doubly dispersive, that is, time and frequency selective, fading channels. Using the recent result that underspread linear systems are approximately diagonalized by biorthogonal Weyl-Heisenberg bases, we arrive at a canonical formulation of modulation and code design. For coherent reception with maximum-likelihood decoding, we derive the code design criteria as a function of the channel's scattering function. We use ideas from generalized concatenation to design multilevel codes for this canonical channel model. These codes are based on partitioning a constellation carved out from the integer lattice. Utilizing the block fading interpretation of the doubly dispersive channel, we adapt these partitioning techniques to the richness of the channel. We derive an algebraic framework which enables us to partition in arbitrarily large dimensions.

Copyright © 2006 S. Mallik and R. Koetter. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1. INTRODUCTION

The design of reliable, high data rate mobile wireless communications systems has been an area of tremendous research activity for the last couple of years. New developments in the field of channel modeling, signaling, and code design have enabled technologies that support high data rates in a wireless setting which in turn have fueled consumer interest in adoption and utilization of wireless devices and services.

This paper deals with communication over rapidly time-varying channels, that is, channels which cannot be regarded as time-invariant over a frame. In a typical wireless setting, a signal sent from the transmitter reaches the receiver through multiple paths, collectively termed as *multipath*. Interference among the multiple paths results in a decrease in signal amplitude. Further due to the time-varying nature of the medium, the received signal amplitude varies with time, in other words, the signal undergoes *fading*. The primary means of combating fading is through *diversity*, in which copies of the transmitted message are made available on different dimensions (time, frequency, or space) to the receiver. All wireless communications schemes utilize *temporal diversity* by using sophisticated channel coding in conjunction with interleaving to provide replicas of the transmitted signal in the temporal domain. *Frequency diversity* techniques employ the fact that waves transmitted on different frequencies induce different multipath structure in the

propagation media. In *space* or *antenna diversity* spatially separate antennas are used at the transmitter or the receiver or both. Communication schemes should utilize all available forms of diversity to ensure adequate performance. In this paper we utilize time and frequency diversity by designing an OFDM-like signaling scheme to be used in conjunction with a multilevel coding scheme easily adapted for fading.

To implement an OFDM-like framework over channels that fade in time and frequency, also called doubly dispersive channels, we need signaling waveforms to be well localized in time and frequency. The good localization in frequency is desirable, so that the waveform sees a frequency nonselective channel. At the same time good localization in time is also desirable as it mitigates the effect of temporal variations in the channel. In [1, 2], a class of waveforms known as the Weyl-Heisenberg bases were found to be suitable candidates as signaling waveforms. These biorthogonal bases are obtained by time and frequency shifts of a given prototype pulse. The time shift  $T$  and the frequency shift  $F$  are usually chosen such that  $TF > 1$  so as to minimize the interference at the receiver. On the other hand if maximum spectral efficiency is required, the parameters  $T$  and  $F$  are chosen such that  $TF = 1$  at the expense of interference at the receiver. In this case an interference cancellation technique at the receiver can be used to cancel out the intersymbol interference. Such a scheme is outlined in [3].

Both approaches mentioned above finally lead to an identical canonical vector fading channel model in discrete time given by  $y_k = h_k x_k + n_k$ ,  $k = 1, \dots, D$ , where  $D$  is the number of dimensions we are coding over,  $y_k$ ,  $h_k$ ,  $x_k$ , and  $n_k$  are the received signal, fading realization, transmitted signal, and noise realization in dimension  $k$ . Powerful coding schemes have been proposed for this channel in the literature. In [4], high diversity constellations are constructed by applying the canonical embedding to the ring of integers of an algebraic number field. In [5], higher diversity is obtained by applying rotations to a classical signal constellation so that any two points achieve a maximum number of distinct components. Another approach is taken by bit-interleaved coded modulation (BICM) [6], where bit-wise interleaving at the encoder input is used to improve the performance of coded modulation on fading channels. In this paper, we propose a multilevel coded modulation scheme for the canonical channel model described above. This scheme is reminiscent of Ungerboeck's trellis-coded modulation [7]. We develop new partitioning techniques for integer lattices which are particularly well suited for fading channels.

The main contribution of this paper is as follows. We use results from linear operator theory and harmonic analysis to study coding and modulation design for underspread time-varying fading channels. Using the fact that underspread channels are approximately diagonalized by biorthogonal Weyl-Heisenberg bases, we arrive at a canonical formulation of modulation and code design. For a coherent receiver employing maximum-likelihood decoding, we derive the code-design criteria as a function of the channel's scattering function. We provide expressions for the maximum achievable diversity order as a function of the channel's scattering function. Secondly, for this canonical channel, we propose new multilevel codes based on partitioning a signal constellation carved out from the integer lattice  $\mathbb{Z}^n$ . We use ideas from generalized concatenation to derive new set partitioning techniques for the fading channel. We also provide an algebraic framework which enables us to partition signal constellations in arbitrarily large dimensions.

This paper is organized as follows. In Section 2 we introduce the time-varying fading channel and the OFDM-like modulation scheme. In Section 3 we derive the code design criteria and make certain critical observations on the code-design problem for this channel. In Section 4, we describe our set partitioning techniques for fading channels and use it to construct a multilevel coded modulation scheme. Section 5 contains performance plots and discusses how the coding scheme is adapted to the channel. Section 6 contains some concluding remarks.

## 2. UNDERSPREAD TIME-VARYING FADING CHANNELS

In this section, we introduce the time-frequency selective fading channel model, discuss the consequences of the underspread assumption, introduce our modulation scheme based on biorthogonal Weyl-Heisenberg bases, and provide the canonical channel representation.

### 2.1. Time-frequency selective fading channels

We model the mobile as a linear time-variant system with input-output relationship given by

$$y(t) = (\mathbf{H}x)(t) + n_w(t) = \int_{t'} h(t, t') x(t') dt' + n_w(t), \quad (1)$$

where  $x(t)$  is the transmitted signal,  $y(t)$  is the received signal,  $\mathbf{H}$  is the linear operator describing the effect of the channel,  $h(t, t')$  is the kernel of the channel, and  $n_w(t)$  is zero-mean circularly symmetric complex white Gaussian noise. Throughout this paper, we assume that  $h(t, t')$  is a complex Gaussian process in  $t$  and  $t'$ . The time-varying transfer function of the channel is defined as [8]

$$L_{\mathbf{H}}(t, f) = \int_{\tau} h(t, t - \tau) e^{-j2\pi f \tau} d\tau. \quad (2)$$

Note that in the time-invariant case where  $h(t, t - \tau) = h(\tau)$  the time varying transfer function reduces to the ordinary transfer function, that is,  $L_{\mathbf{H}}(t, f) = \int_{\tau} h(\tau) e^{-j2\pi f \tau} d\tau = H(f)$ . An alternative representation of the input-output relation (1) is

$$y(t) = \int_{\tau} \int_{\nu} S_{\mathbf{H}}(\nu, \tau) x(t - \tau) e^{j2\pi \nu t} d\nu d\tau, \quad (3)$$

where  $S_{\mathbf{H}}(\nu, \tau)$  is the channel's *delay-Doppler spreading function* which is related to the impulse response  $h(t, t - \tau)$  through a Fourier transform as

$$S_{\mathbf{H}}(\nu, \tau) = \int_t h(t, t - \tau) e^{-j2\pi \nu t} dt. \quad (4)$$

We invoke a wide-sense stationary uncorrelated scattering (WSSUS) assumption which is

$$\begin{aligned} E_{\mathbf{H}}[S_{\mathbf{H}}(\nu, \tau)] &= 0, \\ E_{\mathbf{H}}[S_{\mathbf{H}}(\nu, \tau) S_{\mathbf{H}}^*(\nu', \tau')] &= C_{\mathbf{H}}(\nu, \tau) \delta(\nu - \nu') \delta(\tau - \tau'), \end{aligned} \quad (5)$$

where  $C_{\mathbf{H}}(\nu, \tau) \geq 0$  denotes the *scattering function* of the channel [9, Section 14.1]. Equivalently, the WSSUS assumption implies that the autocorrelation function of the impulse response  $h(t, t - \tau)$  has the following structure:

$$E_{\mathbf{H}}[h(t, t - \tau) h^*(t', t' - \tau')] = \phi_{\mathbf{H}}(t - t', \tau) \delta(\tau - \tau'). \quad (6)$$

Thus under this model, the channel taps are uncorrelated (but not necessarily i.i.d), and the temporal variations are wide-sense stationary. Finally, we will need the channel's correlation function defined as

$$E_{\mathbf{H}}[L_{\mathbf{H}}(t, f) L_{\mathbf{H}}^*(t', f')] = R_{\mathbf{H}}(t - t', f - f'), \quad (7)$$

with the Fourier correspondence

$$R_{\mathbf{H}}(\Delta t, \Delta f) = \int_{\tau} \int_{\nu} C_{\mathbf{H}}(\nu, \tau) e^{j2\pi(\nu \Delta t - \tau \Delta f)} d\tau d\nu. \quad (8)$$

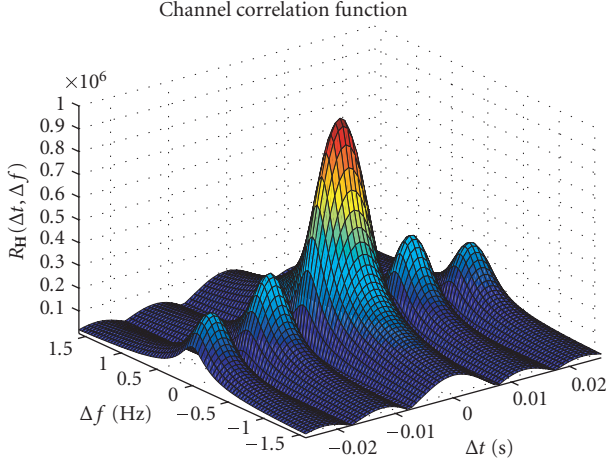


FIGURE 1: Amplitude of the channel correlation function for the Jakes/exponential scattering function. Parameters  $\nu_m = 50$  Hz,  $\tau_0 = 10^{-6}$  Hz.

In literature, it is fairly common to assume that the scattering function has a product form, that is,  $C_H(\nu, \tau) = f(\tau)g(\nu)$ , for example,

$$C_H(\nu, \tau) = \begin{cases} ke^{-\tau/\tau_0} \frac{1}{\pi \nu_m \sqrt{1 - (\nu/\nu_m)^2}} & \text{if } |\nu| \leq \nu_m, \tau \geq 0, \\ 0 & \text{otherwise,} \end{cases} \quad (9)$$

where  $\alpha > 0$ . This particular scattering function is called the *exponential/Jakes scattering function*. Figure 1 is a plot of the above correlation function. The function is normalized, that is,  $R_H(0, 0) = 1$ .

## 2.2. The underspread assumption and its consequences

A fundamental classification of WSSUS channels is into underspread and overspread [9, Section 14.1]. A channel is underspread if its scattering function is highly concentrated around the origin. Note that for simplicity we assume that the scattering function is centered around  $\tau = 0$ , which means that any potential overall delay  $\tau > 0$  has been split off from the channel. A common assumption is that the scattering function is compactly supported within the rectangle  $[-\tau_0, \tau_0] \times [-\nu_0, \nu_0]$  around the origin of the  $(\tau, \nu)$  plane, that is,

$$C_H(\nu, \tau) = 0 \quad \text{for } (\tau, \nu) \notin [-\tau_0, \tau_0] \times [-\nu_0, \nu_0]. \quad (10)$$

Thus the delay spread and Doppler spread are assumed to be bounded. Defining the channel's spread as the area of this rectangle,  $\sigma_H = 4\tau_0\nu_0$ , the channel is said to be underspread if  $\sigma_H \leq 1$  and overspread otherwise. The underspread assumption is relevant as most mobile radio channels are underspread.

As explained in [10] there exist alternative ways to characterize the concentration of the scattering function that

avoid the assumption of compact support. These involve the weighted  $\overline{m}_H^{(\phi)}$  of the scattering function which are defined as

$$\overline{m}_H^{(\phi)} = \frac{\iint_{-\infty}^{\infty} \phi(\tau, \nu) C_H(\nu, \tau) d\tau d\nu}{\iint_{-\infty}^{\infty} C_H(\nu, \tau) d\tau d\nu}, \quad (11)$$

where  $\phi(\tau, \nu) \geq 0$  is a weighting function that satisfies  $\phi(\tau, \nu) \geq \phi(0, 0) = 0$  and penalizes scattering function components lying away from the origin. Special cases are the *moments* obtained with the weighing functions  $\phi_{k,l}(\nu, \tau) = |\nu|^l |\tau|^k$  with  $k, l \in \mathbb{N}$ . Within this framework, a WSSUS channel is called underspread if specific moments and weighted integrals are small.

An important result we are going to build our development on is the fact that underspread systems are approximately diagonalized by biorthogonal Weyl-Heisenberg bases [1, 2]. The Weyl-Heisenberg bases are obtained by time-frequency shifting two normalized functions  $g(t)$  and  $\gamma(t)$  that have good time-frequency localization,

$$g_{k,l}(t) = g(t - kT)e^{j2\pi lFt}, \quad \gamma_{k,l}(t) = \gamma(t - kT)e^{j2\pi lFt}, \quad (12)$$

where  $T$  denotes the time separation and  $F$  denotes the frequency separation between the basis functions. The parameters  $T$  and  $F$  are chosen such that  $TF \geq 1$ . These bases satisfy the *biorthogonality condition*,

$$\langle g_{k,l}, \gamma_{k',l'} \rangle = \int_t g_{k,l}(t) \gamma_{k',l'}^*(t) dt = \delta(k - k') \delta(l - l'). \quad (13)$$

Choosing  $T \leq 1/2\nu_0$  and  $F \leq 1/2\tau_0$ , the kernel  $h(t, t')$  of the underspread fading channel can be well approximated as

$$h(t, t') = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} L_H(kT, lF) g_{k,l}(t) \gamma_{k,l}^*(t'). \quad (14)$$

Details on the choice of  $g(t)$  and  $\gamma(t)$  can be found in [1, 2]. The correlation function of the expansion coefficients  $L_H(kT, lF)$  is given by sampling the channel correlation function

$$E[L_H(kT, lF) L_H^*(k'T, l'F)] = R_H((k - k')T, (l - l')F). \quad (15)$$

## 2.3. Modulation scheme

The diagonalization of underspread systems by the Weyl-Heisenberg bases naturally suggests using an OFDM-like modulation scheme for communication over underspread channels [11]. The transmit signal  $x(t)$  is given by

$$x(t) = \sum_{k=0}^{\infty} \sum_{l=0}^{M-1} \sqrt{E_s} c_{k,l} g_{k,l}(t), \quad (16)$$

where the  $c_{k,l}$  are the information bearing data symbols,  $M$  is the number of OFDM tones, and  $E_s$  is an energy normalization factor. Using (1), (13), and (16), the received signal  $y(t)$

is given by

$$\begin{aligned}
y(t) &= \int_{t'} h(t, t') x(t') dt' + n_w(t) \\
&= \int_{t'} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} L_H(kT, lF) g_{k,l}(t) \gamma_{k,l}^*(t') x(t') dt' + n_w(t) \\
&= \sum_{k=0}^{\infty} \sum_{l=0}^{M-1} L_H(kT, lF) \sqrt{E_s c_{k,l}} g_{k,l}(t) + n_w(t).
\end{aligned} \tag{17}$$

The receiver computes the inner products  $y_{k,l}$ ,

$$y_{k,l} = \int_t y(t) \gamma_{k,l}^*(t) dt = L_H(kT, lF) \sqrt{E_s c_{k,l}} + w_{k,l}, \tag{18}$$

where  $w_{k,l} = \int_t n_w(t) \gamma_{k,l}^*(t) dt$ . Since the signals  $\gamma_{k,l}(t)$  are not orthogonal, there is some correlation between the noise coefficients  $w_{k,l}$ . The noise correlation is ignored and the noise variance is upper bounded using the upper Riesz constant  $B_f$  [11], that is, we assume  $E[w_{k,l} w_{k',l'}^*] = B_f \sigma^2 \delta(k-k') \delta(l-l')$ , where  $\sigma^2$  is the power spectral density of the white Gaussian noise process  $n_w(t)$ . We note that the parameters  $T$  and  $F$  are typically chosen such that  $TF > 1$  is as small as possible in order to maximize the spectral efficiency. Consequently (14) yields an oversampled representation of the channel.

Some parallels can be drawn with discrete time channel models. Consider the channel model given  $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$ , where  $\mathbf{w}, \mathbf{y} \in \mathbb{C}^{MN}$  are the noise vector and the received channel vector, respectively,  $\mathbf{x} \in \mathbb{C}^{MN}$  is the transmitted signal vector and  $\mathbf{H}$  is the random channel matrix. Let  $\mathbf{H} = \mathbf{U}\mathbf{D}\mathbf{V}$  be the singular value decomposition of  $\mathbf{H}$ . If the channel is known then the transmitter spreads signals across the right singular vectors  $\mathbf{V}$ , and the receiver correlates across the left singular vectors  $\mathbf{U}$ . This is analogous to transceiver architecture of Figure 2. As mentioned in (14), the underspread assumption implies that a particular choice of  $\mathbf{U}$  and  $\mathbf{V}$ , viz., the Weyl-Heisenberg bases, enables the diagonalization of the channel even when the channel is unknown at the transmitter.

#### 2.4. The canonical channel model

Let  $\bar{\mathbf{y}}_k = (y_{k,0}, y_{k,1}, \dots, y_{k,M-1})^T$ ,  $h_{k,l} = L_H(kT, lF)$ ,  $\bar{\mathbf{h}}_k = (h_{k,0}, h_{k,1}, \dots, h_{k,M-1})^T$ ,  $\bar{\mathbf{c}}_k = (c_{k,0}, c_{k,1}, \dots, c_{k,M-1})^T$ , and  $\bar{\mathbf{w}}_k = (w_{k,0}, w_{k,1}, \dots, w_{k,M-1})^T$ , where  $(\cdot)^T$  and  $(\cdot)^*$  denote the transpose operator and the conjugate transpose operator, respectively. The equivalent complex baseband discrete time vector channel model is then given by

$$\bar{\mathbf{y}}_k = \sqrt{E_s} \bar{\mathbf{h}}_k \odot \bar{\mathbf{c}}_k + \bar{\mathbf{w}}_k, \quad k \in \mathbb{Z}, \bar{\mathbf{c}}_k \in \mathbb{C}^M, \tag{19}$$

where  $\odot$  denotes the component-wise product of two vectors. The noise  $w_{k,l}$  and the channel gains  $h_{k,l}$  are zero mean, circularly symmetric, complex Gaussian random variables with  $E[\bar{\mathbf{w}}_k \bar{\mathbf{w}}_k^*] = 2\sigma^2 \mathbf{I}_{M \times M}$  and  $E[h_{k,l} h_{k',l'}^*] = R_H((k-k')T, (l-l')F)$ .

Equation (19) represents a set of parallel, correlated (in time and frequency) discrete time Rayleigh fading channels.

Thus making use of the important result that underspread time-varying systems are approximately diagonalized by Weyl-Heisenberg bases, the OFDM-like modulation scheme allows us to formulate the code-design problem in a canonical domain.

It may be argued that the use of biorthogonal Weyl-Heisenberg bases is unnecessary. In particular, for extremely underspread channels of the form depicted in Figure 1 (with a spread factor of  $5 \times 10^{-5}$ ), orthogonal basis functions would not suffer much in terms of interference as compared to biorthogonal basis functions [3]. Since the same bases are used at the transmitter and the receiver, the complexity of an orthogonal scheme would be lower. The key point is that, both approaches would result in the same canonical channel model. In particular, an interference cancelling technique mentioned in [3] may be used to cancel out any intersymbol or intercarrier interference resulting due to the use of orthogonal basis functions.

### 3. CODE DESIGN CRITERIA

In this section we consider a block-coded modulation scheme. We derive an expression for the pairwise error probability assuming maximum-likelihood decoding and perfect channel state information at the receiver. Using the expression for the pairwise error probability as a starting point, we develop a framework for designing codes for the canonical channel described by (19).

#### 3.1. The block-coded modulation scheme

We consider a block-coded modulation scheme where a codeword spans  $M$  tones and  $N$  time slots; that is, we code across time and frequency so as to exploit time-frequency diversity. A codeword  $\mathbf{c} = (\bar{c}_1^T, \bar{c}_2^T, \dots, \bar{c}_{M-1}^T)^T$  is an  $NM$ -dimensional vector obtained by stacking  $M$  column vectors  $\bar{\mathbf{c}}_k$ , each of length  $N$ . Similarly, vectors  $\mathbf{y}$ ,  $\mathbf{h}$ , and  $\mathbf{w}$  are given by  $\mathbf{y} = (\bar{\mathbf{y}}_0^T, \bar{\mathbf{y}}_1^T, \dots, \bar{\mathbf{y}}_{M-1}^T)^T$ ,  $\mathbf{h} = (\bar{\mathbf{h}}_0^T, \bar{\mathbf{h}}_1^T, \dots, \bar{\mathbf{h}}_{M-1}^T)^T$ , and  $\mathbf{w} = (\bar{\mathbf{w}}_0^T, \bar{\mathbf{w}}_1^T, \dots, \bar{\mathbf{w}}_{M-1}^T)^T$ . From (19), the received vector  $\mathbf{y}$  is given by

$$\mathbf{y} = \sqrt{E_s} \mathbf{h} \odot \mathbf{c} + \mathbf{w}. \tag{20}$$

Because of assumptions made in Section 2.1,  $\mathbf{h}$  and  $\mathbf{w}$  are zero mean, circularly symmetric, complex Gaussian vectors with correlation matrices  $\mathbf{R} = E[\mathbf{h}\mathbf{h}^*]$  and  $E[\mathbf{w}\mathbf{w}^*] = 2\sigma^2 \mathbf{I}_{NM \times NM}$ . As a result, the received vector  $\mathbf{y}$  is conditioned on the transmitted codeword  $\mathbf{c}$  and the channel state  $\mathbf{h}$  is also complex Gaussian.

The following proposition gives the Chernoff upper bound on the pairwise error probability of this block-coded modulation scheme. In the proposition, the quantity  $n$  equals  $MN$ .

**Proposition 1.** *Let  $\mathbf{h}, \mathbf{w} \in \mathbb{C}^n$  be circularly symmetric, complex Gaussian random vectors with  $\mathbf{R} = E[\mathbf{h}\mathbf{h}^*]$  and  $E[\mathbf{w}\mathbf{w}^*] = 2\sigma^2 \mathbf{I}_{n \times n}$ . Let  $\sqrt{E_s}$  be an energy normalization factor and let  $\rho \triangleq E_s/8\sigma^2$ . Let  $\mathbf{c}^{(i)}$  and  $\mathbf{c}^{(j)}$  be two signal points in signal constellation  $\mathcal{M}$  which consists of points in  $\mathbb{C}^n$ . Let  $\boldsymbol{\alpha}$  be the*

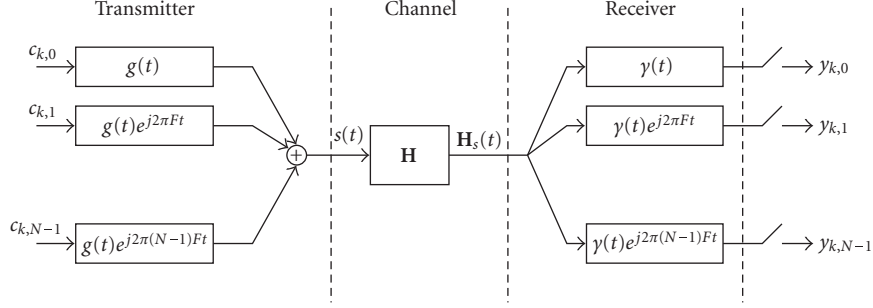


FIGURE 2: The transmitter/receiver structure of the OFDM-like system.

difference vector between these two points, that is,  $\alpha = \mathbf{c}^{(i)} - \mathbf{c}^{(j)}$ . Further,  $\mathbf{Z} = [z_{ij}]$  is an  $n \times n$  diagonal matrix with  $z_{ii} = |\alpha_i|^2$ . The pairwise error probability  $P(\mathbf{c}^{(i)} \rightarrow \mathbf{c}^{(j)})$  for two signal points  $\mathbf{c}^{(i)}, \mathbf{c}^{(j)} \in \mathcal{M}$  transmitted over the correlated Rayleigh fading channel

$$\mathbf{y} = \sqrt{E_s} \mathbf{h} \odot \mathbf{c} + \mathbf{w} \quad (21)$$

is upper bounded by

$$P(\mathbf{c}^{(i)} \rightarrow \mathbf{c}^{(j)}) \leq \frac{1}{\det(\mathbf{I} + \rho \mathbf{RZ})} \quad (22)$$

$$= \prod_{i=1}^n \frac{1}{1 + \rho \lambda_i}, \quad (23)$$

where  $\lambda_i \geq 0$  are the eigenvalues of  $\mathbf{RZ}$ .

*Proof.* The proof is straightforward. See for example [12, the appendix]. A proof appears in the appendix of this paper for the sake of completeness.  $\square$

### 3.2. The role of deep fades in pairwise error probability

We begin by first deriving a lower bound on the pairwise error probability. It is straightforward to show that the pairwise error probability is given by the following expression:

$$P(\mathbf{c}^{(i)} \rightarrow \mathbf{c}^{(j)}) = E_{\mathbf{h}} \left[ Q \left( \frac{E_s}{4\sigma^2} \mathbf{h}^* \mathbf{Z} \mathbf{h} \right) \right], \quad (24)$$

where  $Q(x)$  is the Q function which is defined as  $Q(x) = (1/\sqrt{2\pi}) \int_x^\infty e^{-x^2/2} dx$ .

Consider the following approximation to the Q function.

Let

$$\tilde{Q}(x) = \begin{cases} Q(1), & x \leq 1, \\ 0 & \text{otherwise.} \end{cases} \quad (25)$$

Since  $\tilde{Q}(x) \leq Q(x)$  for all  $x$ , it follows that

$$\begin{aligned} P(\mathbf{c}^{(i)} \rightarrow \mathbf{c}^{(j)}) &\geq E_{\mathbf{h}} \left[ \tilde{Q} \left( \frac{E_s}{4\sigma^2} \mathbf{h}^* \mathbf{Z} \mathbf{h} \right) \right] \\ &= Q(1) P(\mathbf{h}^* \mathbf{Z} \mathbf{h} \leq \frac{2}{\rho}). \end{aligned} \quad (26)$$

We will consider two extreme cases of correlated fading, viz., independent and identically distributed (i.i.d) fading and block fading. A more comprehensive treatment appears in [13] where this idea of behavior at origin and diversity has been generalized to arbitrary fading distributions. The fading is said to be i.i.d if  $h_i$  are independent and identically distributed that is,  $\mathbf{R} = E[\mathbf{h}\mathbf{h}^*] = \mathbf{I}_{NM \times NM}$ . The channel is said to undergo block fading if  $h_i$  are completely correlated, that is,  $h_1 = h_2 = \dots = h_n$ .

We first consider the i.i.d fading scenario. Let  $\beta = (\beta_1, \beta_2, \dots, \beta_n)$  be a permutation of the entries of the vector  $\alpha = (|\alpha_1|^2, |\alpha_2|^2, \dots, |\alpha_n|^2)$  such that the entries of  $\beta$  are arranged in descending order. Let  $L$  be the position of the last nonzero entry in  $\beta$ , that is,  $\beta_i > 0$ , for all  $i \leq L$ . Let  $\Lambda = \sum_{i=1}^L |h_i|^2$ . It follows that

$$P(\mathbf{h}^* \mathbf{Z} \mathbf{h} \leq \frac{2}{\rho}) \geq P(\beta_1 \Lambda \leq \frac{2}{\rho}). \quad (27)$$

If  $\mathbf{R} = \mathbf{I}$ ,  $\Lambda$  is the sum of the squares of  $2L$  Gaussian random variables. Its distribution is known as the Chi-square distribution with  $2L$  degrees of freedom and is given by

$$f_{\Lambda}(x) = \frac{1}{(L-1)!} x^{L-1} e^{-x}, \quad x \geq 0. \quad (28)$$

For small  $x$ , the probability density function of  $\Lambda$  is approximately

$$f_{\Lambda}(x) \approx \frac{1}{(L-1)!} x^{L-1} \quad (29)$$

and hence for i.i.d fading for high SNR, that is, for large  $\rho$ ,

$$P(\Lambda \leq \frac{2}{\rho \beta_L}) \approx \int_0^{\frac{2}{\rho \beta_L}} \frac{1}{(L-1)!} x^{L-1} dx \quad (30)$$

$$= \frac{1}{L!} \left( \frac{2}{\beta_L} \right)^L \frac{1}{\rho^L}. \quad (31)$$

Now let us consider the block-fading scenario. In this case,  $\mathbf{R}$  has rank 1; in fact, all entries of  $\mathbf{R}$  are 1, and the  $\lambda = NM$  is the only nonzero eigenvalue. Thus, from (23)

$$P(\mathbf{c}^{(i)} \rightarrow \mathbf{c}^{(j)}) \leq \frac{1}{1 + \rho NM}. \quad (32)$$

Let  $\beta_i$  and  $\Lambda$  be defined as before. In this case,  $\Lambda = L|h_1|^2$  has an exponential distribution,

$$f_\Lambda(x) = \frac{1}{L}e^{-x/L}, \quad x \geq 0. \quad (33)$$

Thus,

$$P\left(\Lambda \leq \frac{2}{\rho\beta_L}\right) = 1 - e^{-2/\rho L\beta_L} \quad (34)$$

$$\approx \frac{2}{\rho L\beta_L} \quad \text{for large } \rho. \quad (35)$$

Given two functions  $f(x)$  and  $g(x)$  we say  $f(x) \doteq g(x)$  if

$$\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} = k, \quad k \in \mathbb{R}, k \neq 0. \quad (36)$$

For a fixed SNR  $\rho$ , we can say that the  $k$ th channel is in a deep fade if  $|h_k|^2 < 1/\rho$ . From (23) and (31) it follows that in the high SNR regime, for i.i.d fading,

$$\begin{aligned} \frac{\gamma}{\rho^L} &\leq Q(1)P\left(\Lambda \leq \frac{2}{\beta_L\rho}\right) \leq P(\mathbf{c}^{(i)} \rightarrow \mathbf{c}^{(j)}) \\ &\leq \prod_{i=1}^{NM} \frac{1}{1 + \rho\lambda_i} = \prod_{i=1}^L \frac{1}{1 + \rho\beta_i}, \end{aligned} \quad (37)$$

where  $\gamma > 0$  is a constant.

Similarly for block-fading,

$$\frac{\gamma'}{\rho} \leq Q(1)P\left(\Lambda \leq \frac{2}{\beta_L\rho}\right) \leq P(\mathbf{c}^{(i)} \rightarrow \mathbf{c}^{(j)}) \leq \frac{1}{1 + \rho NM}, \quad (38)$$

where  $\gamma' > 0$  is a constant.

In particular, for both i.i.d fading and block fading

$$P\left(\Lambda \leq \frac{1}{\rho}\right) \doteq P(\mathbf{c}^{(i)} \rightarrow \mathbf{c}^{(j)}). \quad (39)$$

The quantity  $P(\Lambda \leq 2/\beta_L\rho)$  is a measure of the probability that the  $L$  parallel Rayleigh channels fade simultaneously. Since the codewords  $\mathbf{c}^{(i)}$  and  $\mathbf{c}^{(j)}$  differ in  $L$  components, we see that the pairwise error probability is dominated by the event that the  $L$  channels  $h_i, i = 1, \dots, L$ , are simultaneously in a deep-fade. Equations (32) and (35) tell the same story for the block fading scenario. For the general case of correlated fading which lies in between these two extreme cases, one would expect  $P(\mathbf{c}^{(i)} \rightarrow \mathbf{c}^{(j)}) \doteq 1/\rho^r$ , where  $1 \leq r = \text{rank}(\mathbf{RZ}) \leq L$ . This will be shown later.

### 3.3. Preferred directions

Unlike the Gaussian channel, the contours of pairwise error probability are not concentric spheres but are star-shaped objects. Consider, for example, the two-dimensional case. Let the channel correlation matrix be denoted as  $\mathbf{R} = \begin{pmatrix} r_0 & r_1^* \\ r_1 & r_0^* \end{pmatrix}$ , where  $r_i = E[h_{k+i}h_k^*]$ ,  $\mathbf{Z}_\alpha = \begin{pmatrix} |\alpha_0|^2 & 0 \\ 0 & |\alpha_1|^2 \end{pmatrix}$ , and

$$\begin{aligned} \det(\mathbf{I} + \rho\mathbf{RZ}_\alpha) &= 1 + \rho r_0(|\alpha_0|^2 + |\alpha_1|^2) \\ &\quad + \rho^2 |\alpha_0|^2 |\alpha_1|^2 (r_0^2 - |r_1|^2). \end{aligned} \quad (40)$$

As a further simplification, consider a signal constellation  $\mathcal{M}$  consisting of points in real space  $\mathbb{R}^2$ . This corresponds to using only the in-phase component in the passband signal constellation. Let  $\boldsymbol{\alpha} \triangleq (x, y)^T \in \mathbb{R}^2$  denote the difference vector. Figure 3 gives a contour plot of  $\det(\mathbf{I} + \rho\mathbf{RZ}_\delta)$  as a function of  $x$  and  $y$ . Such plots for the special case of i.i.d fading and high SNR can also be found in [4]. From the figures, the contours of equal pairwise error probably do not show circular symmetry unless  $\mathbf{R}$  has rank 1. This can also be verified from (40). The lack of circular symmetry leads to the notion of *preferred directions*. Under the norm constraint  $|x|^2 + |y|^2 = 1$ , the pairwise error probability is significantly lower if the difference vector  $\boldsymbol{\alpha} = (x, y)^T$  points in a particular direction, for example, along the unit vector  $(\pm 1/\sqrt{2}, \pm 1/\sqrt{2})^T$  instead of  $(\pm 1, 0)^T$  or  $(0, \pm 1)^T$ .

In the three-dimensional case,  $\mathbf{R}$  can be any three-dimensional toeplitz block toeplitz (TBT) matrix. As special cases, consider the correlation matrices

$$\mathbf{R}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{R}_2 = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{R}_3 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad (41)$$

respectively. The matrix  $\mathbf{R}_1$  represents i.i.d fading,  $\mathbf{R}_2$  refers to the case  $h_1 = h_2$  and independent of  $h_3$ , whereas  $\mathbf{R}_3$  refers to the block fading scenario  $h_1 = h_2 = h_3$ . The contours of equal pairwise error probability are given in Figure 4.

As in the two-dimensional case, when  $\mathbf{R}$  is full rank the locus is star-shaped; in the block fading case where  $\mathbf{R}$  has rank 1, the locus is a sphere. As before, the higher the rank of  $\mathbf{R}$ , the smaller the value of  $|x|$ ,  $|y|$ , and  $|z|$  required to achieve a given PEP at a given  $\rho$ . From the figures, it is clear, that in order to design good signal constellations, the signal points should be arranged in space such that the difference vectors avoid the ‘‘nonpreferred’’ directions.

### 3.4. Key observations

Beyond three dimensions, things become difficult to visualize; the aim of this section is to make some key observations which help us to design signal constellations for correlated fading channels. For the sake of completeness, we begin by proving that the matrix  $\mathbf{RZ}$  has nonnegative eigenvalues.

**Theorem 1.** *The matrices  $\mathbf{Y} = \mathbf{RZ}$  and  $\tilde{\mathbf{Y}} \triangleq E[(\mathbf{h} \circ \boldsymbol{\alpha})(\mathbf{h} \circ \boldsymbol{\alpha})^*]$ , where  $\mathbf{R} = E[\mathbf{h}\mathbf{h}^*]$ ,  $\mathbf{Z} = \text{diag}(|\alpha_1|^2, |\alpha_2|^2, \dots, |\alpha_n|^2)$ , and  $\boldsymbol{\alpha}$  is the column vector  $(\alpha_1, \alpha_2, \dots, \alpha_n)^T \in \mathbb{C}^n$ , have the same eigenvalues.*

*Proof.* Consider an  $n \times n$  matrix  $\mathbf{A}$  and an index set  $\gamma \subseteq \{1, 2, \dots, n\}$  with  $k, k \leq n$  elements. The  $k \times k$  submatrix  $\mathbf{A}(\gamma)$  that lies in the rows and columns of  $\mathbf{A}$  indexed by  $\gamma$  is called a  $k$ -by- $k$  *principal submatrix* of  $\mathbf{A}$ . A  $k$ -by- $k$  *principal minor* is the determinant of such a principal submatrix. There are  $\binom{n}{k}$  different  $k$ -by- $k$  principal minors of  $\mathbf{A}$ , and the sum of these is denoted by  $E_k(\mathbf{A})$ . The characteristic function  $p_{\mathbf{A}}(s) \triangleq \det(s\mathbf{I} - \mathbf{A})$  can be written in terms of  $E_k(\mathbf{A})$

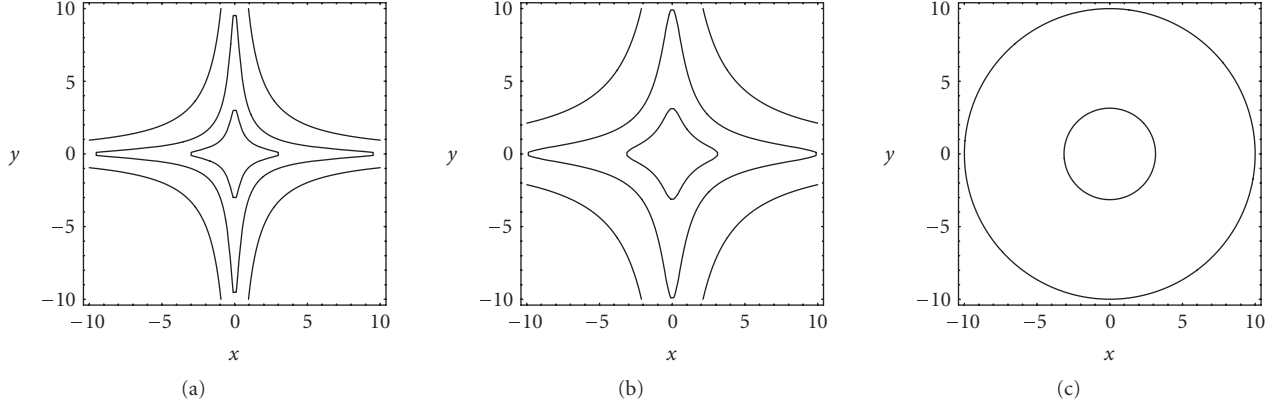


FIGURE 3: Three contours of the pairwise error probability expression in the two-dimensional case,  $\rho = 10$ ,  $r_0 = 1$ . (a)  $r_1 = 0$  i.i.d fading,  $\text{rank}(\mathbf{R}) = 2$ . (b)  $r_1 = 0.8 + j0.4$  correlated fading,  $\text{rank}(\mathbf{R}) = 2$ . (c)  $r_1 = 1$ , correlated fading,  $\text{rank}(\mathbf{R}) = 1$ .

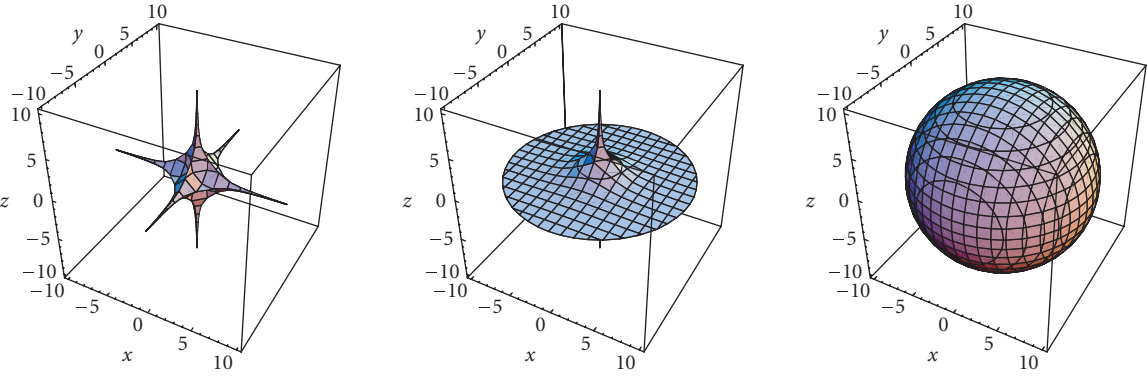


FIGURE 4: Surface of constant pairwise error probability in 3D case for  $\mathbf{R} = \mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3$ , respectively,  $\rho = 10$  and  $P(\mathbf{c}^{(i)} \rightarrow \mathbf{c}^{(j)}) = 10^{-3}$ .

as  $p_{\mathbf{A}}(s) = s^n - E_1(\mathbf{A})t^{n-1} + E_2(\mathbf{A})t^{n-2} - \dots \pm E_n(\mathbf{A})$ . Thus, it is sufficient to show that  $\mathbf{Y}$  and  $\tilde{\mathbf{Y}}$  have the same minors. Let  $\gamma = \{i_1, i_2, \dots, i_k\}$ ,  $1 \leq k \leq n$ , be an index set. But,  $\det(\mathbf{Y}(\gamma)) = (\prod_{l=1}^k |\alpha_{i_l}|^2) \det(\mathbf{R}(\gamma)) = \det(\tilde{\mathbf{Y}}(\gamma))$  which implies  $p_{\tilde{\mathbf{Y}}}(s) = p_{\mathbf{Y}}(s)$ .  $\square$

**Corollary 1.** *The matrix  $\mathbf{Y} = \mathbf{R}\mathbf{Z}$  has nonnegative eigenvalues.*

*Proof.* The matrix  $\mathbf{Y}$  is not Hermitian. However, the matrix  $\tilde{\mathbf{Y}}$  as defined in Theorem 1 is Hermitian and positive semidefinite as  $E[\mathbf{z}^* \tilde{\mathbf{Y}} \mathbf{z}] = E[\sum_{k=1}^n z_k^* \alpha_k h_k]^2] \geq 0$ . The result now follows from Theorem 1.  $\square$

*Definition 1.* The *diversity order* of a signal constellation is the minimum Hamming distance between the coordinate vectors of any two distinct points in the signal constellation.

We will denote the diversity order of a constellation  $\mathcal{M}$  by the symbol  $\mathcal{L}(\mathcal{M})$ . Note that diversity order is a property of the signal constellation and does not depend on the channel model.

*Definition 2.* The  $\ell$ -*product distance* between two signal points  $\mathbf{x}$  and  $\mathbf{y}$  that differ in  $\ell$  components, denoted by

$d_p^{(\ell)}(\mathbf{x}, \mathbf{y})^2$ , is the product of the nonzero components of the difference vector  $\mathbf{e} = \mathbf{x} - \mathbf{y}$ , that is,

$$d_p^{(\ell)}(\mathbf{x}, \mathbf{y})^2 = \prod_{x_i \neq y_i} (x_i - y_i)^2. \quad (42)$$

In the high SNR regime for the i.i.d Rayleigh fading channel, the *diversity order* and the *product distance* of a constellation are important criteria for code design [14]. This is well-known in literature. For the correlated Rayleigh fading channel, the generalization is quite straightforward and involves taking the channel correlation matrix  $\mathbf{R}$  into account. This requires a generalization of the concept of the product distance. See [15] for similar calculations for the multiple antenna space-time codes. The calculations for our OFDM-like scheme on the doubly dispersive channel are similar in spirit.

For i.i.d fading, in the plot of pairwise error probability versus signal-to-noise ratio, the diversity order determines the slope of the curve. In correlated fading, the rank  $r$  of the matrix  $\mathbf{R}\mathbf{Z}$  plays similar role. Note that this quantity is always smaller than the diversity order of the constellation, as  $\text{rank}(\mathbf{R}\mathbf{Z}) \leq \min\{\text{rank}(\mathbf{R}), \text{rank}(\mathbf{Z})\}$ .

The  $k$ th elementary symmetric function of  $n$  numbers  $t_1, t_2, \dots, t_n, k \leq n$ , is

$$S_k(t_1, t_2, \dots, t_n) = \sum_{1 \leq i_1 < \dots < i_k \leq n} \prod_{j=1}^k t_{i_j}. \quad (43)$$

The following elementary theorem helps to generalize the notion of product distance.

**Theorem 2.** Let  $d \geq 1$  be the Hamming weight of the difference vector  $\alpha \in \mathbb{C}^n$ . Let  $r$  be the rank of the correlation matrix  $\mathbf{R}$ . Let  $r_\alpha$  be the rank and let  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{r_\alpha} > \lambda_{r_\alpha+1} = \dots = \lambda_n = 0$  be the eigenvalues of the matrix  $\mathbf{RZ}_\alpha$ . Then,

$$\det(\mathbf{I} + \rho \mathbf{RZ}_\alpha) = 1 + \sum_{k=1}^{r_\alpha} \rho^k S_k(\lambda_1, \lambda_2, \dots, \lambda_n), \quad (44)$$

where  $1 \leq r_\alpha \leq \min\{d, r\}$ .

*Proof.* The proof is straightforward. The eigenvalues are numbered in descending order. Hence,  $\lambda_{r_\alpha+1} = 0$  implies  $S_k(\lambda_1, \dots, \lambda_n) = 0$  for all  $k > r_\alpha$ . Thus,

$$\begin{aligned} \det(\mathbf{I} + \rho \mathbf{RZ}_\alpha) &= \prod_{i=1}^n (1 + \rho \lambda_i) \\ &= 1 + \sum_{k=1}^n \rho^k S_k(\lambda_1, \lambda_2, \dots, \lambda_n) \\ &= 1 + \sum_{k=1}^{r_\alpha} \rho^k S_k(\lambda_1, \lambda_2, \dots, \lambda_n). \end{aligned} \quad (45)$$

The rank of the product of two square matrices can be no greater than the minimum of the ranks of the individual matrices. Since  $\text{rank}(\mathbf{Z}_\alpha) = d$ , we have  $r_\alpha \leq \min\{d, r\}$ .  $\square$

It follows from the previous theorem that, for correlated fading, in the high SNR regime

$$\begin{aligned} P(\mathbf{c}^{(i)} \rightarrow \mathbf{c}^{(j)}) &\leq \frac{1}{1 + \sum_{k=1}^{r_\alpha} \rho^k S_k(\lambda_1, \lambda_2, \dots, \lambda_n)} \\ &\approx \frac{\rho^{-r_\alpha}}{S_{r_\alpha}(\lambda_1, \lambda_2, \dots, \lambda_n)} \quad \text{for large } \rho. \end{aligned} \quad (46)$$

The quantity  $S_{r_\alpha}(\lambda_1, \dots, \lambda_n)$ , where  $\alpha \triangleq \mathbf{x} - \mathbf{y}$ , is the generalization of the notion of product distance between  $\mathbf{x}$  and  $\mathbf{y}$ . Unlike product distance, it depends on the channel statistics since the eigenvalues and the quantity  $r_\alpha$  are functions of the correlation matrix  $\mathbf{R}$ . In i.i.d fading, we have  $\mathbf{R} = \mathbf{I}_{n \times n}$ , which implies  $r_\alpha = d$ . Further,  $|\alpha_i|^2, i = 1, \dots, n$ , are the eigenvalues of the diagonal matrix  $\mathbf{RZ}_\alpha$ . Thus  $S_{r_\alpha}(\lambda_1, \dots, \lambda_n) = \prod_{\alpha_i \neq 0} |\alpha_i|^2 = d_p(\mathbf{x}, \mathbf{y})$ .

### 3.5. Implications for code design for OFDM schemes under the block fading assumption

Consider a signal constellation  $\mathcal{M}$  in  $\mathbb{C}^n$  with diversity order  $\mathcal{L}$  to be used for communication over the canonical channel given by (19). Recall that the diversity order is an intrinsic

property of the signal constellation and does not depend on the channel model. Given a particular channel, we say that  $\mathcal{M}$  achieves a diversity of  $m$  if for every pair of signal points in  $\mathcal{M}$  the pairwise error probability decays at least as fast as  $\rho^{-m}$ . A channel is specified by  $\mathbf{R}$ , the correlation matrix of the fading coefficients. This matrix depends on the channel scattering function  $C_H(\nu, \tau)$  and the grid parameters  $T$  and  $F$  of the OFDM-like modulation scheme.

Let  $\gamma(\mathcal{M})$  be defined as the minimum of the rank of the matrix  $\mathbf{RZ}_\alpha$  over all choices of the difference vector  $\alpha$ . Hence, for a signal constellation  $\mathcal{M}$  of diversity order  $\mathcal{L}$  to achieve a diversity of  $m$  on a channel with correlation matrix  $\mathbf{R}$ , we need

- (i)  $m \leq \gamma(\mathcal{M}) \leq \min\{\text{rank}(\mathbf{R}), \mathcal{L}\}$ ,
- (ii) for high signal-to-noise ratios, the pairwise error probability is smallest for the constellation with greatest  $\gamma(\mathcal{M})$ . For two constellations with the same  $\gamma(\mathcal{M})$ , the one with greater  $S_\gamma(\lambda_1, \lambda_2, \dots, \lambda_n)$  has a smaller pairwise error probability.

Until now, we have allowed arbitrary correlation between the time-frequency channel coefficients in (19). The level of time-frequency diversity is captured in the number of nonzero eigenvalues of the channel correlation matrix  $\mathbf{R} = E[\mathbf{h}\mathbf{h}^*]$ . As shown in [3], the level of delay-Doppler diversity can be estimated via the delay and Doppler spreads and signaling duration of the signaling scheme. The maximum available delay-Doppler diversity, that is, the number of nonzero channel eigenvalues, can be accurately estimated as  $D = \lceil T_m W \rceil \lceil B_d T_s \rceil$ , where  $T_m$  and  $B_d$  are the delay and Doppler spreads of the channel, and  $T_s = NT$  and  $W = MF$  are the signaling duration and bandwidth, respectively. This delay-Doppler diversity leads to the notion of time-frequency coherence subspaces as argued in [3], resulting in a block fading interpretation of the doubly dispersive channel in the short-time Fourier domain. In other words, the number of signal space dimensions  $NM$ , can be partitioned into  $D$  coherence subspaces, each with dimension  $NM/D$ . In the block fading approximation, the channel coefficients are assumed identical in each time-frequency coherence subspace, whereas the coefficients in different subspaces are statistically independent. The number of independent coherence subspaces,  $D$ , which also equals the delay-Doppler diversity in the channel, then represents the maximum number of nonzero eigenvalues of the channel correlation matrix  $\mathbf{R}$ . This means that the matrix  $\mathbf{R}$  is a block-diagonal matrix with  $D$  blocks.

In the next section, we use constellation partitioning ideas to design codes with any desired diversity order and then use the block fading interpretation to adapt the codes to the channel structure.

So far we have been exclusively concerned with the pairwise error probability  $P(\mathbf{c}^{(i)} \rightarrow \mathbf{c}^{(j)})$ . Using the union bound, the probability of decoding error when  $\mathbf{c}^{(i)}$  is transmitted  $P(\text{error} | \mathbf{c}^{(i)})$  is upper bounded as

$$P(\text{error} | \mathbf{c}^{(i)}) \leq \sum_{\mathbf{c}^{(j)} \in \mathcal{M}, \mathbf{c}^{(j)} \neq \mathbf{c}^{(i)}} P(\mathbf{c}^{(i)} \rightarrow \mathbf{c}^{(j)}). \quad (47)$$



Let  $M$  denote the number of signal points in constellation  $\mathcal{M}$ . Assuming all codewords have the same a priori probability, that is,  $P(\mathbf{c}^{(i)}) = 1/M$  for all  $i$ ,

$$\begin{aligned} P(\text{error}) &= \sum_{i=1}^M P(\text{error} \mid \mathbf{c}^{(i)})P(\mathbf{c}^{(i)}) \\ &\leq \frac{1}{M} \sum_{i=1}^M \sum_{j=1, j \neq i}^M P(\mathbf{c}^{(i)} \rightarrow \mathbf{c}^{(j)}). \end{aligned} \quad (48)$$

The above analysis is based on the pairwise error probability and yields a good approximation to the overall probability of error if the union bound is tight. This approach has its limitations, in particular in the design of capacity approaching schemes.

#### 4. CODE DESIGN BY SET PARTITIONING

In 1977, Imai and Hirakawa [16] presented their multilevel method for constructing binary block codes. Codewords from the component codes, also called as outer codes, form rows of a binary array, and the columns of this array are used as information symbols for another code called the inner code. If on the other hand, each column of this array of outer codes is used to label a signal point in a signal constellation, we obtain a coded-modulation scheme. Such techniques were also used in [7, 17] for the design of effective coded-modulation schemes for the AWGN channel. Nowadays, multilevel techniques, also called generalized concatenation, are well recognized as a powerful tool for designing new codes in Hamming and Euclidean spaces [18]. In this section we use the technique of generalized concatenation to design signal constellations with high diversity order.

##### 4.1. An example in two dimensions

Our idea to partition signal constellations is inspired by Ungerboeck's trellis coded-modulation schemes. Recognizing that the Euclidean distance is an important design parameter for minimizing pairwise error probability, in [7] standard QAM constellations were partitioned such that subconstellations had greater Euclidean distance. For fading channels, we design partitioning schemes to ensure that subconstellations have a greater diversity order. We illustrate this by means of an example. We will generalize this scheme in Section 4.3.

Consider the signal constellation  $\mathcal{M}^1$  shown in Figure 5. It can be defined as

$$\mathcal{M}^1 \triangleq \left\{ (x_1, x_2)^T \mid x_i \in \left\{ \pm \frac{1}{2}, \pm \frac{3}{2} \right\} \right\}. \quad (49)$$

We partition it into four subconstellations  $\mathcal{M}_0^2$ ,  $\mathcal{M}_\alpha^2$ ,  $\mathcal{M}_1^2$ , and  $\mathcal{M}_{\bar{\alpha}}^2$  as shown in Figure 5. The primary objective of the partitioning scheme is to ensure that the subsets  $\mathcal{M}_i^2$  have a larger diversity order  $\mathcal{L}$  than the parent constellation  $\mathcal{M}^1$ . For this particular partitioning scheme, we have  $\mathcal{L}(\mathcal{M}_i^2) = 2\mathcal{L}(\mathcal{M}^1) = 2$ .

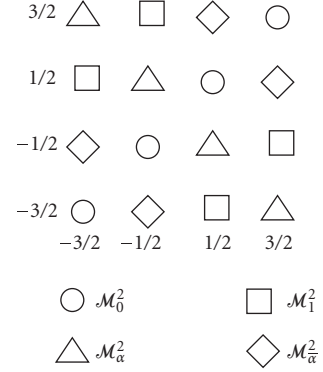


FIGURE 5: Algebraic description of partitioning scheme A.

##### 4.2. Algebraic description of partitioning scheme A

To generalize scheme A to  $m$  dimensions we first need to give it an algebraic description. This is done as follows. Let  $\mathbb{F}_4$  denote the finite field of cardinality 4. Let  $\alpha$  be a primitive element of  $\mathbb{F}_4$ . Let the elements of  $\mathbb{F}_4$  be given by  $\{0, 1, \alpha, \bar{\alpha}\}$ , where  $\bar{\alpha}$  denotes the element  $\alpha^2$ . Consider the bijective map  $\phi_\alpha : \mathbb{F}_4 \rightarrow \{-3/2, -1/2, 1/2, 3/2\}$  given by

$$\phi_\alpha(\gamma) = \begin{cases} -\frac{3}{2} & \text{if } \gamma = 0, \\ i - \frac{3}{2} & \text{if } \gamma = \alpha^i, 1 \leq i \leq 3. \end{cases} \quad (50)$$

Let  $\Phi_\alpha$  be the vector map corresponding to component-wise scalar maps  $\phi_\alpha$ . Given a set  $S$ , let  $\Phi_\alpha(S)$  denote the set of all values the map  $\Phi_\alpha$  can take as its argument varies over  $S$ .

As shown in Figure 5, the partitions are now identified by labels over  $\mathbb{F}_4$ . The partition  $\mathcal{M}_\alpha^2$  consists of the four points  $(3/2, -3/2)$ ,  $(1/2, -1/2)$ ,  $(-1/2, 1/2)$ ,  $(-3/2, 3/2)$  in  $\mathbb{R}^2$ . We say that this partitioning scheme is defined by its generator matrix  $\mathcal{P}_A = \begin{pmatrix} 1 & \alpha \\ 1 & \bar{\alpha} \end{pmatrix}$ , since the partitions  $\mathcal{M}_0^2$ ,  $\mathcal{M}_1^2$ ,  $\mathcal{M}_\alpha^2$ , and  $\mathcal{M}_{\bar{\alpha}}^2$  can then be defined as

$$\begin{aligned} \mathcal{M}_0^2 &= \Phi_\alpha(\{(\gamma, 0)\mathcal{P}_A \mid \gamma \in \mathbb{F}_4\}), \\ \mathcal{M}_1^2 &= \Phi_\alpha(\{(\gamma, 1)\mathcal{P}_A \mid \gamma \in \mathbb{F}_4\}), \\ \mathcal{M}_\alpha^2 &= \Phi_\alpha(\{(\gamma, \alpha)\mathcal{P}_A \mid \gamma \in \mathbb{F}_4\}), \\ \mathcal{M}_{\bar{\alpha}}^2 &= \Phi_\alpha(\{(\gamma, \bar{\alpha})\mathcal{P}_A \mid \gamma \in \mathbb{F}_4\}). \end{aligned} \quad (51)$$

It is easy to see that each of these partitions has diversity order 2. This is because, if  $\mathbf{s}_1, \mathbf{s}_2 \in \mathcal{M}_i^2$  and  $\mathbf{s}_1 \neq \mathbf{s}_2$ , then  $\mathbf{s}_1 - \mathbf{s}_2$  is a multiple of  $\Phi_\alpha((1, 1))$ . Thus  $\mathbf{s}_1$  and  $\mathbf{s}_2$  differ in two coordinates.

We now use the idea of generalized concatenation to combine the constellation  $\mathcal{M}^1$  in  $\mathbb{R}^2$  with suitably chosen outer codes of length  $n$  to construct constellations in  $\mathbb{R}^{2n}$  with desired diversity order. Consider two outer codes  $\mathcal{C}^i[n, k_i, d_i]_4$ ,  $i = 1, 2$ , over  $\mathbb{F}_4$  of length  $n$ , dimension  $k_i$ , and minimum distance  $d_i$  where  $d_1 > d_2$ . Code  $\mathcal{C}^i$  contains  $M_i = 4^{k_i}$  codewords. Each point in  $\mathcal{M}^1$  can be uniquely determined by the label  $(\omega_1, \omega_2)$ , where  $\omega_1, \omega_2 \in \mathbb{F}_4$ . In particular, the

pair  $(c_k^1, c_k^2)$  of the  $k$ th coordinate,  $1 \leq k \leq n$ , of the two code-words  $\mathbf{c}^1 = (c_1^1, c_2^1, \dots, c_n^1) \in \mathcal{C}^1$  and  $\mathbf{c}^2 = (c_1^2, c_2^2, \dots, c_n^2) \in \mathcal{C}^2$  can be used to label signal points in  $m^1$ . Thus, a pair of code-words, one from each outer code, labels a signal point in  $\mathbb{R}^{2n}$ . We thus have a construction for a signal constellation  $\mathcal{M}_{CM}$  in  $2n$ -dimensional real space.

We now show that  $\mathcal{M}_{CM}$  has  $M_1 M_2$  signal points and a diversity order of at least  $\min_i \{d_i \mathcal{L}(\mathcal{M}^i)\}$ , where  $\mathcal{M}^i$  stands for any one of the four subconstellations  $\mathcal{M}_\omega^i$ ,  $\omega \in \mathbb{F}_4$ . Note that  $\mathcal{L}(\mathcal{M}^i)$  is well defined since all of these subconstellations have the same diversity order of 2. Fixing a codeword  $\mathbf{c}^1 \in \mathcal{C}^1$ ,  $M_2$  different signal points can be labeled with code-words of  $\mathcal{C}^2$ . Thus the cardinality of  $\mathcal{M}_{CM}$  is  $M_1 M_2$ . A signal point  $\mathbf{s}$  in  $\mathcal{M}_{CM}$  is uniquely identified by a pair of code-words, one each from  $\mathcal{C}^1$  and  $\mathcal{C}^2$ . Consider two distinct signal points  $\mathbf{s}_1$  and  $\mathbf{s}_2$  in  $\mathcal{M}_{CM}$ . Since  $\mathbf{s}_1 \neq \mathbf{s}_2$  we have two possibilities.

- (1) The signal points correspond to distinct codewords from  $\mathcal{C}^1$ . Since  $\mathcal{C}^1$  has a Hamming distance  $d_1$ , it follows that  $\mathbf{s}_1$  and  $\mathbf{s}_2$  differ in at least  $d_1$  times  $\mathcal{L}(\mathcal{M}^1)$  coordinates. Note that this holds true independent of whether the two signal points correspond to the same or different codewords from  $\mathcal{C}^2$ .
- (2) The signal points correspond to the same codeword from  $\mathcal{C}^1$  but different codewords from  $\mathcal{C}^2$ . Hence, arguing as above, since two codewords from  $\mathcal{C}^2$  differ in at least  $d_2$  positions,  $\mathbf{s}_1$  and  $\mathbf{s}_2$  differ in at least  $d_2$  times  $\mathcal{L}(\mathcal{M}^2)$  coordinates.

We conclude this subsection with some terminology that will be helpful in subsequent sections. We partition the constellation  $\mathcal{M}^1$  once to create four constellations at level 1, viz.,  $\mathcal{M}_\omega^2$ ,  $\omega \in \mathbb{F}_4$ . We partition a second time to create 16 constellations at level 2, viz.,  $\mathcal{M}_{\omega_1, \omega_2}^3$ ,  $\omega_1, \omega_2 \in \mathbb{F}_4$ . The partitioning is stopped when each constellation consists of a single point. In other words, the parent constellation is at level 0 and the constellations at the last level consist of a single point each. The *order* of a partitioning scheme is defined as the number of levels in the scheme. This should not be confused with the term *diversity order*. In subsequent sections, the term  $\mathcal{M}^1$  will refer to any signal constellation that we wish to partition. It will not refer to the particular constellation given by (49) unless it is explicitly mentioned to be so.

### 4.3. Generalizing partitioning scheme A

Scheme A described in the previous subsection has order 2. In general an  $L \times m$  partition generator matrix  $\mathcal{P}$  whose entries are elements in  $\mathbb{F}_q$  represents a scheme of order  $L$  in  $m$ -dimensional real space with less than or equal to  $q$  signal points per dimension.

Let  $\alpha$  be a primitive element in  $\mathbb{F}_q$ , the finite field with  $q$  elements. Consider the map  $\phi_\alpha : \mathbb{F}_q \rightarrow \{(-q+1)/2, (-q+3)/2, \dots, (q-1)/2\}$  given by  $\phi_\alpha(\gamma) = i - (q-1)/2$ , if  $\gamma = \alpha^i$ ,  $1 \leq i \leq q-1$ , and  $\phi_\alpha(0) = (-q+1)/2$ . Let  $\Phi_\alpha$  be the vector map corresponding to component-wise scalar maps  $\phi_\alpha$ . Let  $\mathcal{M}^1$  be a constellation carved out from the

integer lattice  $\mathbb{Z}^m$ . Consider the partitioning matrix

$$\mathcal{P} \triangleq \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \alpha & \alpha^2 & \cdots & \alpha^{m-1} \\ 1 & \alpha^2 & \alpha^4 & \cdots & \alpha^{2(m-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha^{L-1} & \alpha^{2(L-1)} & \cdots & \alpha^{(L-1)(m-1)} \end{pmatrix}, \quad (52)$$

where  $L \leq q-1$ ,  $m \leq q-1$ , and the set

$$\begin{aligned} \mathcal{M}_{\omega_1, \omega_2, \dots, \omega_k}^{k+1} \\ \triangleq \Phi_\alpha(\{\beta \mathcal{P} \mid \beta = (\beta_{L-k}, \dots, \beta_1, \omega_k, \omega_{k-1}, \dots, \omega_1) \in \mathbb{F}_q^L\}). \end{aligned} \quad (53)$$

In the above equation the vector  $\beta$  takes all possible values in  $\mathbb{F}_q^{L-k}$ . Constellation  $\mathcal{M}_{\omega_1, \omega_2, \dots, \omega_k}^{k+1}$  consists of  $q^{L-k}$  points each labeled by a distinct vector  $\beta$ . Further, it will be clear from Theorem 3 that we need  $L \leq m$  for the diversity order of the constellation at level  $l$  to be a strictly increasing function of  $l$ . We take a moment to clarify the notation. In the above equation,  $\alpha$  is a primitive element in  $\mathbb{F}_q$ , whereas  $\omega_j$ s represent arbitrary (not necessarily primitive) elements  $\mathbb{F}_q$ . We thus have a partitioning scheme of order  $L$  in an  $m$ -dimensional Euclidean space indexed by labels  $\omega_k \in \mathbb{F}_q$  given by

$$\begin{aligned} \mathcal{M}^1 &= \bigcup_{\omega_1} \mathcal{M}_{\omega_1}^2, \\ \mathcal{M}_{\omega_1}^2 &= \bigcup_{\omega_2} \mathcal{M}_{\omega_1, \omega_2}^3, \\ &\vdots \\ \mathcal{M}_{\omega_1, \omega_2, \dots, \omega_{L-1}}^L &= \bigcup_{\omega_L} \mathcal{M}_{\omega_1, \omega_2, \dots, \omega_{L-1}, \omega_L}^{L+1}. \end{aligned} \quad (54)$$

The parameter  $\omega_1 \in \mathbb{F}_q$  labels the subconstellation  $\mathcal{M}_{\omega_1}^1$  of  $\mathcal{M}^1$ ,  $\omega_2$  labels the subconstellation  $\mathcal{M}_{\omega_1, \omega_2}^3$  of  $\mathcal{M}_{\omega_1}^2$ , and so on. Note that  $\mathcal{M}_{\omega_1, \omega_2, \dots, \omega_{L-1}}^L$  consists of a set of  $q$  points given by  $\mathcal{M}_{\omega_1, \omega_2, \dots, \omega_L}^{L+1}$ ,  $\omega_L \in \mathbb{F}_q$ . For the example given in Section 4.1, we have

$$\begin{aligned} \mathcal{M}^1 &= \mathcal{M}_0^2 \cup \mathcal{M}_\alpha^2 \cup \mathcal{M}_{\bar{\alpha}}^2 \cup \mathcal{M}_1^2, \\ \mathcal{M}_\omega^2 &= \mathcal{M}_{\omega 0}^3 \cup \mathcal{M}_{\omega \alpha}^3 \cup \mathcal{M}_{\omega \bar{\alpha}}^3 \cup \mathcal{M}_{\omega 1}^3 \quad \forall \omega \in \mathbb{F}_4. \end{aligned} \quad (55)$$

**Theorem 3.**  $\mathcal{L}(\mathcal{M}_{\omega_1, \omega_2, \dots, \omega_{l-1}}^l) = (l+m-L)_+$ , for all  $l$  such that  $1 \leq l \leq L$ , where  $x_+ \triangleq \max\{x, 0\}$ .

*Proof.* Consider that the two distinct points, that is,  $\mathbf{s}_1, \mathbf{s}_2 \in \mathcal{M}_{\omega_1, \omega_2, \dots, \omega_{l-1}}^l$ ,  $\mathbf{s}_1 \neq \mathbf{s}_2$ , have the following identification labels:  $(\beta_{L-l+1} \cdots \beta_1 \omega_{l-1} \cdots \omega_1)$  and  $(\gamma_{L-l+1} \cdots \gamma_1 \omega_{l-1} \cdots \omega_1)$ , respectively. Further assume that  $\mathbf{s}_1, \mathbf{s}_2$  are chosen such that  $\beta_1 \neq \gamma_1$ . Let  $\zeta_k \triangleq \beta_k - \gamma_k$ ,  $k = 1, \dots, L-l+1$ . Consider the polynomial  $g(x) = \zeta_{L-l+1} + \zeta_{L-l}x + \cdots + \zeta_1 x^{L-l}$ . Since  $\zeta_1 \neq 0$ ,  $g(x)$  is a polynomial of degree of  $L-l$  and can have at most  $L-l$  roots in  $\mathbb{F}_q$ . But

$$\mathbf{s}_1 - \mathbf{s}_2 = \Phi_\alpha(g(1), g(\alpha), g(\alpha^2), \dots, g(\alpha^{m-1})), \quad (56)$$

where  $\alpha$  is a primitive element in  $\mathbb{F}_4$ . This implies that signal points  $\mathbf{s}_1$  and  $\mathbf{s}_2$  differ in at least  $m - (L - l)$  positions which implies

$$\mathcal{L}(\mathcal{M}_{\omega_1, \omega_2, \dots, \omega_{l-1}}^l) \geq (l + m - L)_+. \quad (57)$$

We now show that we have equality in (57). Consider the polynomial  $h(x) = (x - 1)(x - \alpha) \cdots (x - \alpha^{L-l-1})$ . Let  $\mathbf{s}_1$  denote the signal point  $\mathcal{M}_{0,0,\dots,0}^{L+1}$ .

$g_1(x) = \sum_i^L \omega_i x^{i-1}$  and let  $g_2(x) = g_1(x) + h(x)$ . Let  $g_2(x)$  be of the form  $\sum_i^L \omega_i x^{i-1}$ . Let  $\mathbf{s}_2$  be a point with identification label  $(\gamma_L, \dots, \gamma_2, \gamma_1)$ . It follows that the difference vector  $\mathbf{s}_1 - \mathbf{s}_2 = \Phi_\omega((h(1) \ h(\omega) \ h(\omega^2) \ \cdots \ h(\omega^{m-1})))$  has Hamming weight  $(m - L + l)_+$ . Since the polynomial  $h(x)$  has degree  $L - l$ , we have  $\gamma_i = \omega_i$  for  $1 \leq i \leq l - 1$ , which implies  $\mathbf{s}_2 \in \mathcal{M}_{\omega_1, \omega_2, \dots, \omega_{l-1}}^l$ .  $\square$

Theorem 3 shows that the diversity order increases as we go down the partition chain. It will be strictly increasing if  $L \leq m$ . The reason for this partitioning will be clear from Theorem 4 where we will combine this partitioning scheme with outer codes to create a signal constellation in higher dimensions with higher diversity. Figure 6 shows the partitioning scheme A in three dimensions. The constellation  $\mathcal{M}^1$  is carved from a shifted version of  $\mathbb{Z}^3$ , the integer lattice in three dimensions, and has  $q = 4$  points per dimension. The partitioning scheme of order 3 can be represented by (53) and (54) with  $L = 3$ ,  $m = 3$ ,  $q = 4$ . Figure 6(a) shows the partition  $\mathcal{M}_0^2$  which is further divided into 4 subpartitions  $\mathcal{M}_{00}^3$ ,  $\mathcal{M}_{01}^3$ ,  $\mathcal{M}_{0\alpha}^3$ , and  $\mathcal{M}_{0\bar{\alpha}}^3$ , as shown in Figure 6(b). As expected,  $\mathcal{M}_0^2$  has diversity 2 and  $\mathcal{M}_{0\alpha}^3$  has diversity order 3. Figure 7 illustrates this three-dimensional example in more detail.

#### 4.4. Outer codes

In the example considered in Section 4.1, we needed two outer codes. In general, we need  $L$  outer codes, where  $L$  is the depth of the partitioning scheme. There are three parameters that have to be chosen for each outer code  $\mathcal{C}^i[N, k_i, d_i]_{p_i}$ .

- (1) The *finite field* over which the outer code is defined. This is dependent on the partitioning scheme. Consider a partitioning scheme of order  $k$ . Let  $t_j$  denote the number of partitions at level  $j$  and  $\mathbb{F}_{p_j}$  denote the finite field over which the  $j$ th outer code is defined. To label each of the  $t_j$  partitions  $\mathcal{M}_1^j, \dots, \mathcal{M}_{t_j}^j$ , it is necessary that  $p_j \geq t_j$ . For the particular partitioning scheme described in the previous subsection,  $t_j$  equals  $q$ , hence  $p_j \geq q$  suffices. We choose  $p_j = q$  for all  $j$ .
- (2) The *block length*  $N$  of the outer code. This is dependent on a number of factors like design constraints and decoding complexity. If ergodic capacity-achieving schemes are desired, it is necessary to consider long block lengths.
- (3) The *rate*  $R_i$  of the outer codes. This is related to the desired error performance. If pairwise error probability is the criterion we wish to optimize, then the outer codes are chosen such that each subpartition has the same pairwise error probability. This will be elaborated in Section 5.

We now describe the multilevel encoder and the multistage decoding (MSD) algorithm, first presented by Imai and Hirakawa in [16]. Figure 8 shows a multilevel encoder for a partitioning scheme of order  $L$ . This figure appears in [18]. For simplicity, assume that  $p_1 = p_2 = \cdots = p_L = p$ , that is, all outer codes are defined on the same field  $\mathbb{F}_p$ . In the encoder, a block of  $K$  source data symbols  $\mathbf{q} = (q_1, \dots, q_K)$ ,  $q_i \in \mathbb{F}_p$ , is partitioned into  $L$  blocks

$$\mathbf{q}^i = (q_1^i, \dots, q_{k_i}^i), \quad i = 1, \dots, L, \quad (58)$$

of length  $k_i$  with  $\sum_{i=1}^L k_i = K$ . Each data block  $\mathbf{q}^i$  is fed into an individual  $p$ -ary encoder  $E_i$  generating codewords  $\mathbf{c}^i = (c_1^i, \dots, c_N^i)$  of the component code  $\mathcal{C}^i$ . For simplicity, we assume here equal code lengths  $N$  at all levels, but the choice of code lengths can be arbitrary. For example, block codes, convolutional codes, or turbo codes can be used. The codeword symbols  $c_t^i$ ,  $t = 1, \dots, N$ , of the codewords  $\mathbf{c}^i$ ,  $i = 1, \dots, L$ , at one time instant  $t$  form the  $p$ -ary label  $c_t = (c_t^1, \dots, c_t^L)$ , which is mapped to the signal point  $s_{c_t}$ .

Let  $\mathcal{M}_{CM}$  be the constellation in  $\mathbb{R}^{mn}$  obtained by concatenating the partition scheme of order  $L$  as given by (53) and (54) with  $L$  outer codes  $\mathcal{C}^i[N, k_i, d_i]_q$ ,  $1 \leq i \leq L$ . Theorem 4 proves that  $\mathcal{M}_{CM}$  has cardinality  $\prod_i q^{k_i}$ . Let  $\eta$  denote the spectral efficiency in bits per real dimension of  $\mathcal{M}_{CM}$ . It follows that

$$\eta = \frac{\log_2 \prod_i q^{k_i}}{nm} = \frac{\log_2 q}{m} \sum R_i, \quad (59)$$

where  $R_i$  is the rate of the  $i$ th outer code.

**Theorem 4.** *The set  $\mathcal{M}_{CM}$  has cardinality  $\prod_i q^{k_i}$  and diversity order of at least  $\min_i \{d_i(l + m - L)_+\}$ , where  $x_+ \triangleq \max\{x, 0\}$ .*

*Proof.* Let  $\mathbf{c}^{(i)}$  be a codeword in the outer code  $\mathcal{C}^i[N, k_i, d_i]_q$ ,  $1 \leq i \leq L$ . Further, let  $(c_1^i, c_2^i, \dots, c_N^i)$  be the representation of the codeword  $\mathbf{c}^i$ . The  $L \times N$  codeword matrix

$$\begin{pmatrix} c_1^1 & c_2^1 & \cdots & c_N^1 \\ c_1^2 & c_2^2 & \cdots & c_N^2 \\ \vdots & \vdots & \ddots & \vdots \\ c_1^L & c_2^L & \cdots & c_N^L \end{pmatrix} \quad (60)$$

uniquely identifies a signal point, say  $\mathbf{s}$ , in  $\mathcal{M}_{CM}$ . Since there are  $\prod_i q^{k_i}$  such distinct matrices, it implies that  $\mathcal{M}_{CM}$  has cardinality  $\prod_i q^{k_i}$ . The  $i$ th column of (60) identifies a signal point  $s_i$  in  $\mathcal{M}^1$ . Similarly, let  $\mathbf{t} \in \mathcal{M}_{CM}$  be the point corresponding to codewords  $\mathbf{d}^{(i)}$ ,  $1 \leq i \leq L$ . Let the quantity  $l$  be defined as  $l = \min\{k \mid \mathbf{c}^{(k)} \neq \mathbf{d}^{(k)}\}$ . This implies that  $\mathbf{c}^{(l)}$  and  $\mathbf{d}^{(l)}$  differ in at least  $d_l$  positions. Let  $i$  be one such position and let  $\omega_j = c_j^{(i)}$ . This implies that  $s_i, t_i \in \mathcal{M}_{\omega_1, \omega_2, \omega_{l-1}}^l$  and there exists no  $\gamma \in \mathbb{F}_q$  such that  $s_i, t_i \in \mathcal{M}_{\omega_1, \omega_2, \omega_{l-1}, \gamma}^{l+1}$ . This implies that  $s_i$  and  $t_i$  differ in  $(m - L + l)_+$  positions. This is true for at least  $d_l$  such positions in the outer code. This implies that  $\mathbf{s}$  and  $\mathbf{t}$  differ in at least  $d_l(m - L + l)_+$  positions. The claim now follows by taking a minimum over  $l = 1, 2, \dots, L$ .  $\square$

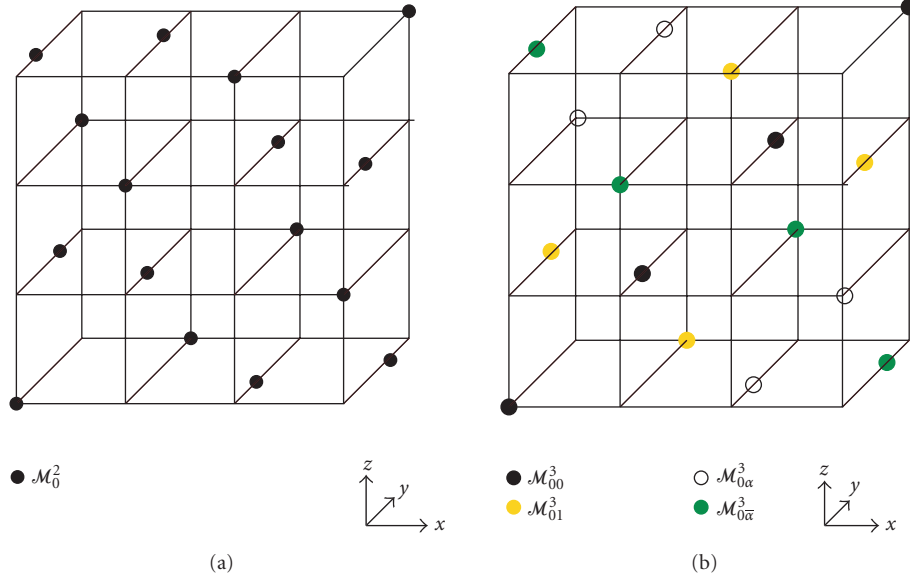


FIGURE 6: Partitioning scheme A in three dimensions  $q = 4, m = 3, L = 3$ . (a) Partition  $\mathcal{M}_0^2$ . (b) Subpartitions  $\mathcal{M}_{00}^3, \mathcal{M}_{01}^3, \mathcal{M}_{0\alpha}^3, \mathcal{M}_{0\bar{\alpha}}^3$ .

Note that  $(l + m - L)_+$  is the diversity order of a subconstellation  $\mathcal{M}_{\omega_1, \omega_2, \dots, \omega_{l-1}}^l$  at level  $l$ . This is an increasing function of  $l$ . Since the diversity order of  $\mathcal{M}_{CM}$  is the minimum over  $l$  of the product  $d_l \mathcal{L}(\mathcal{M}^l)$ , the distance  $d_l$  of outer code  $\mathcal{C}^l$ , needed to attain a particular value of  $\mathcal{M}_{CM}$ , decreases as  $l$  increases. This enables higher rate codes to be used at higher levels, that is, levels corresponding to a larger value of  $l$ . Theorem 4 also illustrates how by lowering the minimum distance  $d_l$  of all the outer codes, thereby increasing their rate, we can trade off the diversity order of the constellation  $\mathcal{M}_{CM}$  for the rate of the code. The significance of Theorem 3 also now becomes clear. In particular if the fading is i.i.d, then the outer code  $\mathcal{C}^l$  sees an *equivalent channel* with diversity  $(l + m - L)_+$ . The notion of equivalent channels is described in detail in [18]. Since this is a better channel than the channel seen by the outer code  $\mathcal{C}^{l-1}$ ,  $\mathcal{C}^l$  needs a lower correction capability than  $\mathcal{C}^{l-1}$ . If the fading is not i.i.d but correlated,  $\mathcal{C}^l$  may not see a channel with diversity as high as  $(l + m - L)_+$ , but the channel will be better than that seen by  $\mathcal{C}^{l-1}$ .

We now take a look at the decoding algorithm for multilevel codes. Figure 9 shows a multistage decoder. This figure also appears in [18]. In this low-complexity decoding algorithm, the component codes  $\mathcal{C}^i$  are successively decoded by the corresponding decoders  $D_i$ . At stage  $i$ , decoder  $D_i$  processes not only the block  $\mathbf{y} = (y_1, \dots, y_N), y_k \in \mathbb{R}^m$ , where  $m$  is the dimension of the signal space, but also decisions  $\hat{\mathbf{c}}^j, j = 1, \dots, i - 1$ , of the previous decoding stages  $j$ . Let  $P_{e,j}$  denote the word error rate of code  $\mathcal{C}^j$  given that the previous  $j - 1$  stages have been decoded correctly, that is,

$$P_{e,j} \triangleq P(\hat{\mathbf{c}}^j \neq \mathbf{c}^j \mid \hat{\mathbf{c}}^1 = \mathbf{c}^1, \dots, \hat{\mathbf{c}}^{j-1} = \mathbf{c}^{j-1}). \quad (61)$$

It follows from the union bound that the overall probability

of error  $P_e$  is upper bounded by

$$P_e \leq \sum_{j=1}^L P_{e,j}. \quad (62)$$

Let  $R = \sum R_i$  denote the sum of the rates of the outer codes. As mentioned in [18], if error propagation in MSD is neglected, the bit-error  $P_b$  probability for multilevel coded transmissions is given by

$$P_b = \sum_{l=1}^L \frac{R_l}{R} P_{b,i}, \quad (63)$$

where  $P_{b,i}$  denotes the bit-error probability for decoding at level  $i$  when error-free decisions are assumed at the decoding stages of the previous levels.

#### 4.5. Adaptation of the partitioning scheme to the block fading channel

So far, we have seen how the partitioning scheme can be used with outer codes to construct codes of any desired diversity order. We now adapt these codes to a block fading channel.

Consider a coding scheme over  $M$  tones and  $N$  time slots. The underlying channel structure results in a block fading channel with  $D$  coherent subspaces or blocks each of size  $b = NM/D$ . To design a coded-modulation scheme with spectral efficiency of  $\eta$  bits per dimension, start with an integer lattice in  $m \leq D$  dimensions, and carve out a constellation  $\mathcal{M}^1$  consisting of  $q^m$  points. The parameter  $m$  is chosen to be quite smaller than  $D$ . This is explained in detail in Section 5.1. This signal constellation has an uncoded spectral efficiency of  $\log_2 q$  bits per dimension. The parameter  $q$  is chosen so as to ensure a constellation expansion factor of at

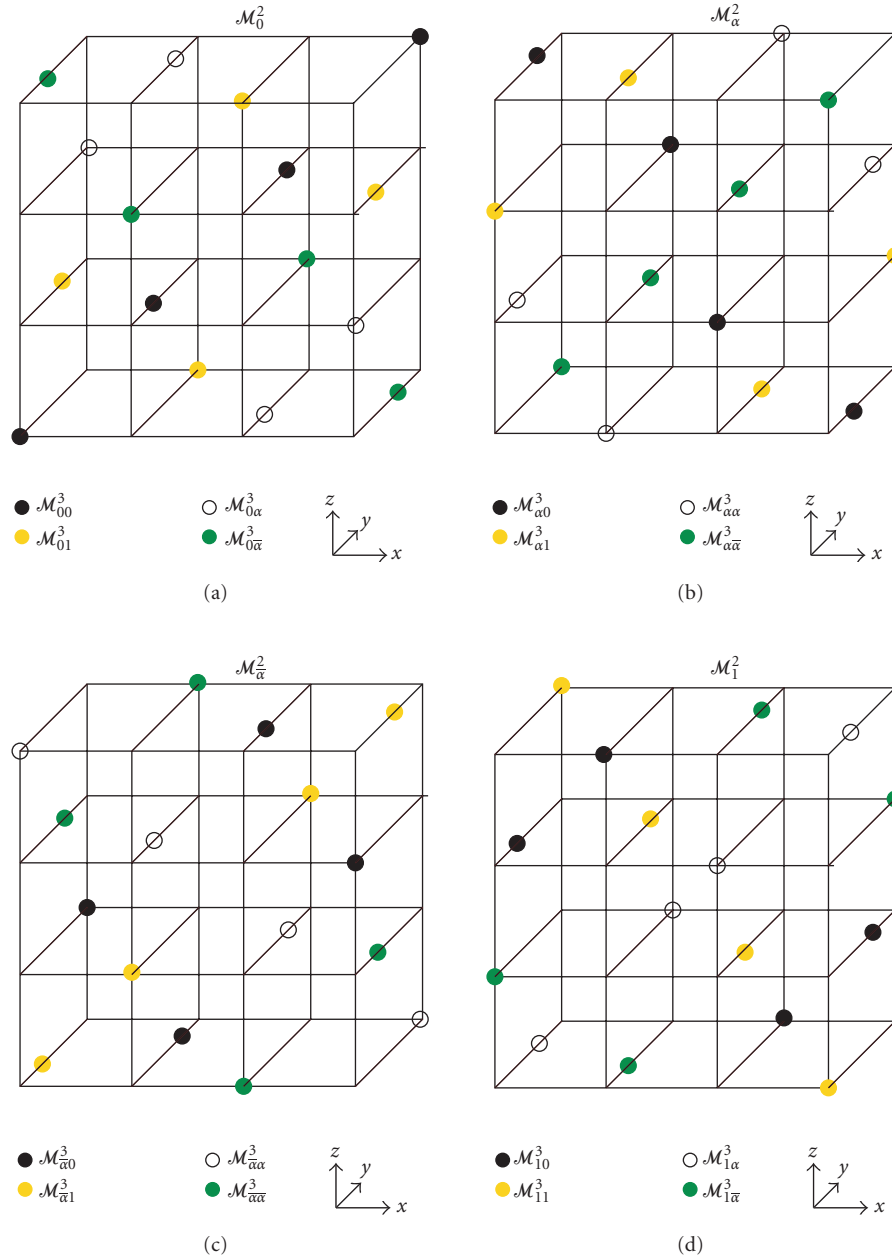


FIGURE 7: Partitioning scheme A in three dimensions  $q = 4, m = 3, l = 3$ .  $\mathcal{M}^1$  decomposes into three partitions  $\mathcal{M}_0^2, \mathcal{M}_\alpha^2, \mathcal{M}_{\bar{\alpha}}^2$  and  $\mathcal{M}_1^2$ . (a) Partition  $\mathcal{M}_0^2$  and its subpartitions  $\mathcal{M}_{00}^3, \mathcal{M}_{01}^3, \mathcal{M}_{0\alpha}^3, \mathcal{M}_{0\bar{\alpha}}^3$ . (b) Partition  $\mathcal{M}_\alpha^2$  and its subpartitions  $\mathcal{M}_{\alpha 0}^3, \mathcal{M}_{\alpha 1}^3, \mathcal{M}_{\alpha\alpha}^3, \mathcal{M}_{\alpha\bar{\alpha}}^3$ . (c) Partition  $\mathcal{M}_{\bar{\alpha}}^2$  and its subpartitions  $\mathcal{M}_{\bar{\alpha} 0}^3, \mathcal{M}_{\bar{\alpha} 1}^3, \mathcal{M}_{\bar{\alpha}\alpha}^3, \mathcal{M}_{\bar{\alpha}\bar{\alpha}}^3$ . (d) Partition  $\mathcal{M}_1^2$  and its subpartitions  $\mathcal{M}_{10}^3, \mathcal{M}_{11}^3, \mathcal{M}_{1\alpha}^3, \mathcal{M}_{1\bar{\alpha}}^3$ .

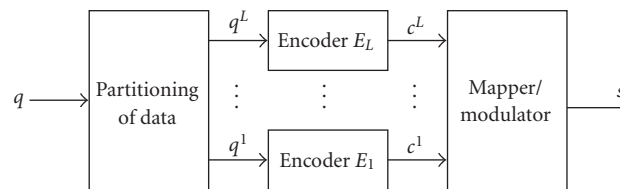


FIGURE 8: Multilevel encoder.

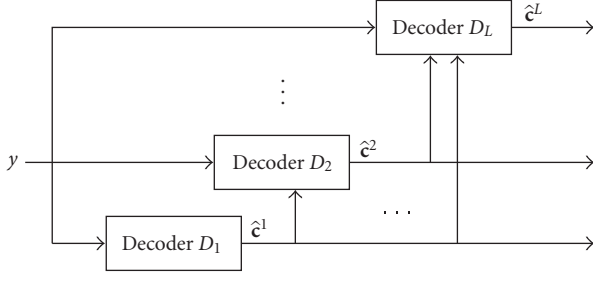


FIGURE 9: Multistage decoding (MSD) algorithm.

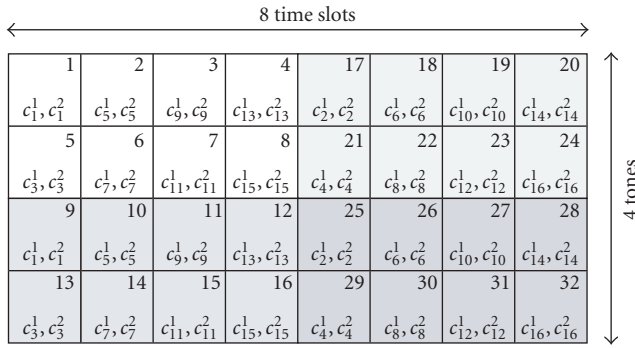


FIGURE 10: Interleaved channel.

least 2, that is,  $\lceil \log_2 q/\eta \rceil \geq 2$ . Fixing  $q$  constrains how large  $m$  can be as  $m \leq q - 1$ . Now partition  $\mathcal{M}^1 L$  times thereby assigning each signal point a  $q$ -ary label of length  $L$ . Use  $L$  outer codes over  $\mathbb{F}_q$ . In our partitioning scheme, we fix  $L = m$ , that is, we partition  $m$  times and hence need to choose  $m$  outer codes. The rates  $R_i$  of the outer codes should be chosen to satisfy the desired spectral efficiency as per (59). Finally the coded bits of the outer codes are passed through an interleaver before being modulated and transmitted. We choose short constraint length convolutional codes as outer codes. The interleaver is designed so that successive code symbols of each outer code see independent fades [19].

Consider the following example of how the interleaver is designed. It also helps to explain how the coded symbols are mapped onto the different time frequency slots in OFDM modulation. Consider a channel with  $M = 4$  tones. Let the coherence time correspond to 4 time slots as shown in Figure 10. Let the tone-spacing be such that the coherence bandwidth of the channel corresponds to 2 tones. A block of 8 time-frequency slots corresponds to a time-frequency coherence subspace and is indicated in the figure by a particular shade of the grey color. All time-frequency slots in a coherence subspace see the same fade. Each time-frequency slot corresponds to one complex dimension or equivalently 2 real dimensions. We code jointly over 8 time slots and 4 tones, which corresponds to 2 sets of 32 real dimensions corresponding to the inphase and quadrature components. Thus we are coding across  $D = 4$  coherence subspaces each of size  $b = 8$ . The desired spectral efficiency  $\eta$  is 1 bit per real dimension.

We choose  $\mathcal{M}^1$  as defined by (49) as the signal constellation with 16 points. This fixes  $m = 2$  and hence we partition twice and need two outer codes over  $\mathbb{F}_4$ . We choose convolutional codes as outer codes. Thus we need a rate pair  $(R_1, R_2)$  such that  $R_1 + R_2 = 1$ . Let  $\mathbf{c}^1$  and  $\mathbf{c}^2$  be codewords corresponding to outer codes  $\mathcal{C}^1$  and  $\mathcal{C}^2$ , respectively. We first modulate over the inphase components and then over the quadrature components. A pair of code symbols  $(c_t^1, c_t^2)$ ,  $1 \leq t \leq 16$ , uniquely identifies a signal  $\mathbf{s}_t = (s_{t,1}, s_{t,2}) \in \mathbb{R}^2$ . Thus  $(c_t^1, c_t^2)$  determines signal point  $\mathbf{s}_1 = (s_{1,1}, s_{1,2})$  and we send  $s_{1,1}$  in over time-frequency slot 1 and  $s_{1,2}$  over time-frequency slot 9 which is that first slot that fades independently of slot 1. We indicate this in Figure 10 by noting down in  $c_1^1, c_1^2$  in the time-frequency slots numbered 1 and 9. Similarly, we send  $s_{2,1}$  and  $s_{2,2}$  over time-frequency slots 17 and 25, respectively, which fade independently of each other and slots 1 and 3. We now have run out of independently fading slots, so the next signal point corresponding to  $t = 3$  is sent over slots 5 and 13. We continue till  $t = 16$  at which point the inphase components of the 32 time-frequency slots have been exhausted. We then modulate for the quadrature components. Thus the primary objective of interleaving the code symbols is to guarantee that successive code symbols see independent fades. This helps to combat slow or block fading by creating an implicit time-frequency diversity effect. This trick is well known, see, for example [19]. We fixed  $M = 4$  and  $N = 8$ , but the procedure to interleave for larger values of  $M$  and  $N$  is a natural extension of the above technique. Let  $\alpha$  be the difference vector between two signal points in  $\mathcal{M}_{CM}$ . Let  $\mathbf{E}_n$  denote the square matrix of all ones of size  $n$ . Let  $\mathbf{I}_n$  denote an identity matrix of size  $n$ . In the uninterleaved block fading channel with  $D$  blocks of length  $b$  each, the matrix  $\mathbf{R}$  is given by  $\mathbf{R} = \mathbf{E}_b \otimes \mathbf{I}_D$ . In the interleaved channel it is given by  $\mathbf{R} = \mathbf{I}_D \otimes \mathbf{E}_b$ , where  $\otimes$  denotes the tensor product between two matrices. The amount of delay that can be tolerated influences the value of  $N$  and hence the number of coherence subspaces  $D$ . In the subsequent sections, when we mention interleaver we mean the interleaver designed above.

In Section 3.3 the notion of preferred directions was introduced. A direction  $\alpha$  is a preferred direction if the quantity  $\det(\mathbf{I} + \rho \mathbf{RZ}) = 1 + \sum_{k=1}^{r_\alpha} \rho^k S_k(\lambda_1, \lambda_2, \dots, \lambda_n)$  is large. Here  $r_\alpha$  denotes the rank of the matrix  $\mathbf{RZ}_\alpha$ . Since the constellation  $\mathcal{M}_{CM}$  is carved from an integer lattice it follows that if

$$S_k(\lambda_1, \lambda_2, \dots, \lambda_n) > 0 \implies S_k(\lambda_1, \lambda_2, \dots, \lambda_n) \geq 1, \quad (64)$$

where any constant scaling factor corresponding to the desired SNR has been absorbed in the quantity  $\rho$ . Further, choosing rates  $R_i$  of the outer codes so as to maximize the diversity order of the constellation  $\mathcal{M}_{CM}$ , and using an interleaver as described above to ensure that consecutive code symbols of the convolutional code see independent fades ensures the rank of the matrix  $\mathbf{RZ}_\alpha$  is large. The codes that we design do not maximize the quantity  $S_{r_\alpha}(\lambda_1, \lambda_2, \dots, \lambda_n)$ . However as mentioned above, for our codes, the quantity  $S_{r_\alpha}(\lambda_1, \lambda_2, \dots, \lambda_n)$  is never less than 1. Thus for a block fading channel, our choice of a constellation carved from an integer lattice and the interleaver described above helps to approximate the problem of designing constellations with difference

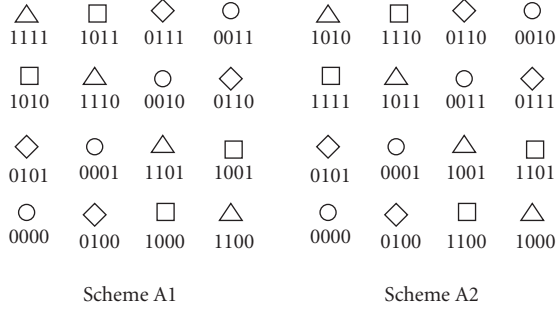


FIGURE 11: Binary labels for partitioning scheme A.

vectors along preferred directions to the problem of partitioning an integer lattice so as to ensure a high diversity order.

### 5. SIMULATION RESULTS

Consider a transmission scheme over  $M = 16$  tones. We choose to code across 16 tones and  $N = 128$  time slots to exploit the time-frequency diversity. The underlying time-frequency coherence structure results in a block fading channel with say  $D = 8$  coherence subspaces each of size  $b = 256$ . We also consider other scenarios, viz.,  $D = 16$ ,  $b = 128$ , and  $D = 32$ ,  $b = 64$ . In particular, larger the value of  $D$ , the “richer” the channel and better is the error performance of a given coding scheme. We desire a coding scheme with spectral efficiency of  $\eta = 1$  bit per real dimension. We code over the inphase and quadrature components separately. We choose  $\mathcal{M}^1$  as the 16-point constellation over  $m = 2$  dimensions as specified as by (49). Since  $q = 4$ , we need two outer codes over  $\mathbb{F}_4$ .

As mentioned earlier, we choose convolutional codes as outer codes. This is primarily because we will use decode using the BCJR algorithm to minimize the symbol-(bit-) error rate. For convolutional codes, the complexity of the code is determined by a parameter  $\nu$  called the *total memory* of the encoder for the code. An encoder for a convolutional code, by design, corresponds to a  $k$ -input,  $n$ -output finite state machine. A convolutional code is said to have total memory of  $\nu$  if  $2^\nu$  represents the total number of states of the state machine.

Instead of working with outer codes over  $\mathbb{F}_4$  we choose to work with binary outer codes. As a result we have to map the 4-ary labels of the signal points to binary labels. As illustrated in Figure 11, there are two distinct ways of doing this. At each partition level, the neighbors of signal points in scheme A2 differ in fewer bit positions than those of signal points in A1. Hence for working with binary outer codes we choose scheme A2. It is important to remark here that if 4-ary outer codes are used we do not need to make this distinction.

Let  $R_1$  and  $R_2$  denote the rates of the first and second outer codes, respectively. Since the uncoded scheme has a spectral efficiency of 2 bits per dimension, this means that we

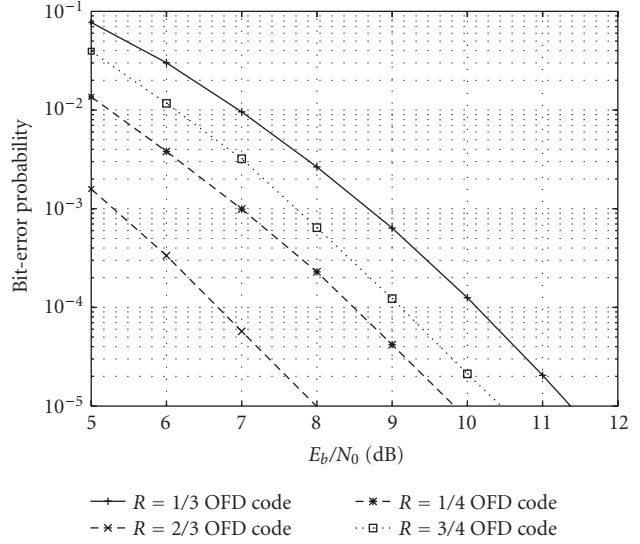


FIGURE 12: Performance of  $\mathcal{M}^1$  with the rate pairs  $(1/3, 2/3)$  and  $(1/4, 3/4)$  using optimal free distance (OFD) convolutional codes on a block fading channel of 32 blocks of size 64 each.  $\eta = 1$  bit/real dimension. All codes have  $\nu = 5$ , except the rate 1/4 code which has  $\nu = 4$ .

have to choose the rate pair  $(R_1, R_2)$  such that  $R_1 + R_2 = 1$ . We consider two such rate pairs:  $(1/3, 2/3)$  and  $(1/4, 3/4)$ .

Figure 12 shows the performance of the rate pairs  $(1/3, 2/3)$  and  $(1/4, 3/4)$  under multistage decoding. The outer codes were decoded by the BCJR algorithm. The BCJR algorithm was run on a window of size 2048 bits. The  $X$  axis refers to the energy per bit for the combined modulation scheme. Further, while calculating the bit-error probability for the second outer code, we assume that the first code has been decoded correctly. The aim is to choose rates of the outer codes so that their individual bit-error rate (BER) curves are matched as closely as possible. In order words, we choose the outer codes using the *equal error probability rule* of [18].

As can be seen from Figure 12, for the rate pair  $(1/3, 2/3)$ , for a given trellis complexity of both outer codes, the performance of the overall code is determined by the first outer code. For the rate pair  $(1/4, 3/4)$ , the BER curves are well matched. The rate pair  $(1/4, 3/4)$  has a better performance than the  $(1/3, 2/3)$  pair.

Figure 13 compares the performance of the same rate pair  $(1/4, 3/4)$  on a block fading channel with  $D = 32$  and  $b = 64$  with that on a block fading channel with  $D = 16$  and  $b = 128$ . As before the rate 1/4 code and rate 3/4 have *total memory* of 4 and 5, respectively. As expected the plots for the channel with greater  $D$  have a steeper slope. The “richer” channel gains about 2 dB at a bit-error rate of  $10^{-4}$ .

#### 5.1. Further guidelines on the adaptation of the partitioning scheme and outer codes to the block fading channel

In this section, we show how to adapt the parameter  $m$  and the *total memory* of the convolutional outer codes to the

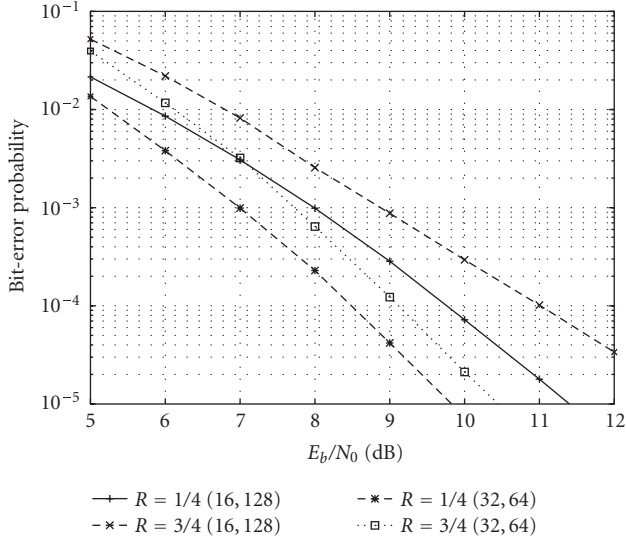


FIGURE 13: Performance improvement of  $\mathcal{M}^1$  with the rate pair (1/4, 3/4) on the block fading channel with parameters  $D = 32$ ,  $b = 64$  as compared to the block fading channel with parameters  $D = 16$ ,  $b = 32$ .  $\eta = 1$  bit/dim.

channel parameter  $D$ . The relationship between  $m$  and  $D$  is best illustrated by means of an example. For a target spectral efficiency of  $\eta = 1$  bit per real dimension, we consider two schemes corresponding to choosing  $m = 1$  and 2, respectively. If  $m = 1$ , then we use 4 PAM as our signal constellation in 1-dimensional space and combine it with a rate 1/2 outer code. Since  $m = 1$  we need only one outer code and there is no partitioning involved. We compare the performance of this scheme with a scheme corresponding to  $m = 2$ , that is, the constellation  $\mathcal{M}^1$  coupled with the rate pair (1/4, 3/4). Note that both 4-PAM and  $\mathcal{M}^1$  have an uncoded spectral efficiency of 2 bits per dimension. We simulate the performance of these two competing schemes over 3 different block-fading channels characterized by  $D = 16, 32$ , and 2048, respectively, with the product  $Db$  kept constant at 2048. We constrain the total number of states in the encoder for each scheme to be not greater than 64.

Figure 14 shows that for low values of  $D$  the 4-PAM scheme beats the multilevel scheme. But as the diversity in the channel, characterized by the parameter  $D$ , increases, the multilevel scheme performs better. The convolutional encoder for the multilevel scheme has  $16 + 32 = 48$  states. For the channel with  $D = 32$ ,  $b = 64$ , it beats the 4-PAM with a 32 state encoder by 1.5 dB at a BER of  $10^{-5}$ . It performs as well as the 4-PAM with a 64 state encoder. As shown in Figure 15, the performance gains are even higher on an i.i.d channel, here the multilevel code gains over 1 dB at a BER of  $10^{-5}$ .

Let  $\mathbf{R} = \mathbf{I}_D \otimes \mathbf{E}_b$  denote the channel correlation matrix of the interleaved block fading channel. We use short constraint length convolutional codes as outer codes. Let  $\alpha$  denote the difference between two signal points  $\mathbf{s}_1, \mathbf{s}_2 \in \mathcal{M}_{CM}$  corresponding to codewords  $(\mathbf{c}_1^1, \mathbf{c}_1^2, \dots, \mathbf{c}_1^L)$  and  $(\mathbf{c}_2^1, \mathbf{c}_2^2, \dots, \mathbf{c}_2^L)$ ,

respectively. We say that  $\alpha$  is of type  $i$ ,  $1 \leq i \leq L$ , and if  $\mathbf{c}_1^j = \mathbf{c}_2^j$  for all  $j$  such that  $1 \leq j < i$  and  $\mathbf{c}_1^i \neq \mathbf{c}_2^i$ . If  $\alpha$  is of type  $i$  then its Hamming weight cannot be less than  $\mathcal{L}(\mathcal{M}^i)d_i$ , where  $d_i$  is the free distance of convolutional outer code  $\mathcal{C}^i$ . Suppose  $\alpha$  is of type  $i$ . It follows that

$$\text{rank}(\mathbf{RZ}_\alpha) \leq \left( \min \left\{ \left\lceil \frac{D}{m} \right\rceil, d_i \right\} \right) \mathcal{L}(\mathcal{M}^i). \quad (65)$$

The  $d_i$ 's are a decreasing function of  $i$ , hence  $m$  should be chosen such that  $\lceil D/m \rceil$  is not smaller than  $d_L$ .

For a given total memory, which is a measure of the complexity of the encoder and decoder of the convolutional code, and spectral efficiency  $\eta$ , increasing  $m$ , the dimensionality of the uncoded signal constellation  $\mathcal{M}^1$  increases the diversity order of the coded superconstellation  $\mathcal{M}_{CM}$  provided the  $L$  outer codes are chosen as OFD codes for the given total memory  $\sum v_i = \nu$ ,  $I = 1, \dots, L$ . However, if the channel is poor, that is,  $D$  is low or equivalently the rank of channel matrix  $\mathbf{R}$  is low, the extra diversity order is of no use as indicated above. As  $D$  increases, or equivalently, as the rank of the matrix  $\mathbf{R}$  increases, the extra diversity order gained by partitioning in higher dimensions comes into play and there is a corresponding-increase in performance as illustrated in Figures 14 and 15.

## 6. CONCLUDING REMARKS

This paper dealt with a framework for communication over doubly dispersive channels. Using the fact that Weyl-Heisenberg bases approximately diagonalize an underspread linear system we arrived at a canonical formulation of modulation and code design. We derived the code design criteria and characterized the maximum achievable diversity in terms of the scattering function of the channel. We then introduced new set partitioning techniques for multilevel coding schemes for the canonical fading channel model. We used these partitioning schemes to partition a signal constellation  $\mathcal{M}^1$  in  $m$  dimensions and combined it with  $L$  outer codes  $\mathcal{C}^l[N, k_l, d_l]_q$ ,  $1 \leq l \leq L$ , to design a coded signal constellation  $\mathcal{M}_{CM}$  in  $nm$  dimensions. To a first-order approximation, the performance of this scheme is determined by its diversity order  $\mathcal{L}^* = \min_l d_l(l + m - L)_+$ . The constellation  $\mathcal{M}_{CM}$  has  $\prod_{i=1}^L q^{k_i}$  points. This implies that it is straightforward to trade constellation size for diversity order by adjusting the rate of the outer codes. The algebraic description through a generator matrix enables partitioning in large dimensions. This ability to partition in arbitrarily large dimensions and change to the rate of the outer codes gives us the flexibility to adjust the scheme to the "richness" of the fading channel, that is, the number of non-zero eigenvalues of  $\mathbf{R}$ . In other words, if the channel offers more diversity, then one can increase the rate of the outer codes while maintaining the same error probability. We described a procedure to adapt these codes to the block fading channel thereby making them suitable for coded modulation schemes over doubly dispersive channels. Finally we illustrated the performance of these codes through simulations.



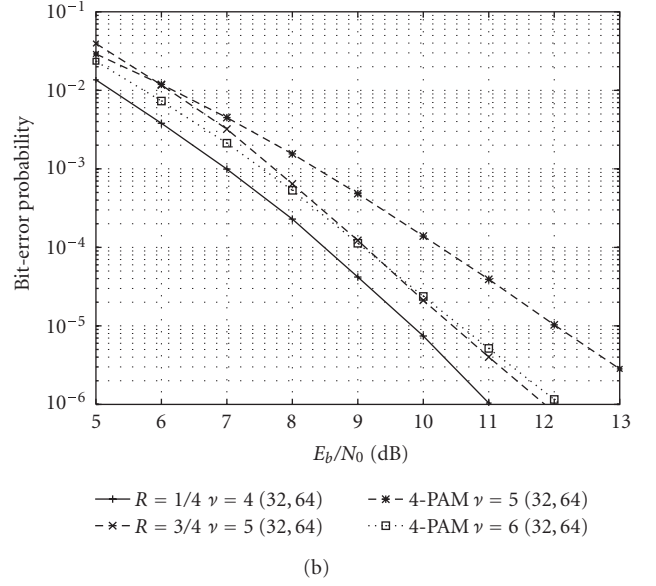
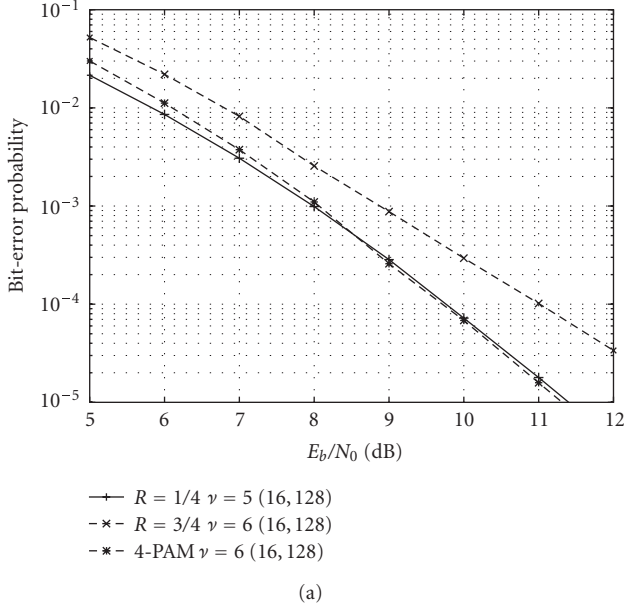


FIGURE 14: Performance comparison of the 16-point constellation  $\mathcal{M}^1$  with rate pair  $(1/4, 3/4)$  versus 4-PAM with rate  $1/2$  outer code on two different block fading channels. (a)  $D = 16, b = 128$ , (b)  $D = 32, b = 64$ .

## APPENDIX

We now derive an expression for the pairwise error probability of the block-coded modulation scheme. Let  $\mathbf{c}$  be a codeword chosen with equal probability from a codebook  $\mathcal{C}$ .  $\mathcal{C}$  can also be interpreted as a set of points in  $NM$ -dimensional complex space  $\mathbb{C}^{NM}$ . Let  $\mathbf{y}$  be the received vector. Assuming perfect channel state information at the receiver, the maximum-likelihood decoder output  $\hat{\mathbf{c}}$  is given by

$$\begin{aligned} \hat{\mathbf{c}} &= \arg \max_{\mathbf{c} \in \mathcal{C}} f_{\mathbf{N}}(\mathbf{y} - \mathbf{h} \odot \sqrt{E_s} \mathbf{c}) \\ &= \arg \min_{\mathbf{c}^{(i)} \in \mathcal{C}} \sum_{k=0}^{NM-1} \left| y_k - \sqrt{E_s} h_k c_k^{(i)} \right|^2, \end{aligned} \quad (\text{A.1})$$

where  $f_{\mathbf{N}}(\mathbf{n}) = (1/(2\pi\sigma^2))^{NM} \exp(-\mathbf{n}^* \mathbf{n} / 2\sigma^2)$  is the probability density function of the complex Gaussian vector  $\mathbf{n}$ . Let  $\mathbf{c}^{(i)}, \mathbf{c}^{(j)}$  be two codewords in  $\mathcal{C}$ . The conditional probability of mistaking  $\mathbf{c}^{(i)}$  for another codeword, say  $\mathbf{c}^{(j)}$ , is given by

$$\begin{aligned} P(\mathbf{c}^{(i)} \rightarrow \mathbf{c}^{(j)} | \mathbf{h}) &= P\left( \sum_{k=0}^{NM-1} \left| y_k - \sqrt{E_s} h_k c_k^{(i)} \right|^2 \geq \sum_{k=0}^{NM-1} \left| y_k - \sqrt{E_s} h_k c_k^{(j)} \right|^2 \right) \\ &= P\left( \sum_{k=0}^{NM-1} |n_k|^2 \geq \sum_{k=0}^{NM-1} \left| n_k + \sqrt{E_s} h_k (c_k^{(i)} - c_k^{(j)}) \right|^2 \right). \end{aligned} \quad (\text{A.2})$$

Define

$$A = \frac{E_s}{2} \sum_{k=0}^{NM-1} |h_k|^2 |c_k^{(j)} - c_k^{(i)}|^2, \quad (\text{A.3})$$

$$\beta = \sum_{k=0}^{NM-1} \text{Re} \left[ \sqrt{E_s} h_k (c_k^{(j)} - c_k^{(i)}) n_k^* \right],$$

where  $A$  is a constant and  $\beta$  is a real-valued Gaussian random variable with zero mean and variance  $2A\sigma^2$ . Let  $\boldsymbol{\alpha} \in \mathbb{C}^{NM}$  be the difference vector, that is,  $\boldsymbol{\alpha} = \mathbf{c}^{(i)} - \mathbf{c}^{(j)}$ . Let  $\mathbf{Z}_{\boldsymbol{\alpha}}$  be an  $NM \times NM$  diagonal matrix with  $k$ th diagonal entry given by  $|\alpha_k|^2$ . We drop the subscript  $\boldsymbol{\alpha}$  where there is no chance of confusion. Equation (A.2) can be rewritten as

$$\begin{aligned} P(\mathbf{c}^{(i)} \rightarrow \mathbf{c}^{(j)} | \mathbf{h}) &= P(\beta \geq A) = Q\left(\sqrt{\frac{A}{2\sigma^2}}\right) \\ &\leq e^{-A/4\sigma^2} = e^{-(E_s/8\sigma^2) \mathbf{h}^* \mathbf{Z} \mathbf{h}}, \end{aligned} \quad (\text{A.4})$$

where  $Q(x) = (1/\sqrt{2\pi}) \int_x^\infty e^{-y^2/2} dy$  and we have used the upper bound  $Q(x) \leq (1/2)e^{-x^2/2}$  which is asymptotically tight.

Under the assumption that the matrix  $\mathbf{R}$  has full rank, the probability density function of  $\mathbf{h}$  is well defined and is given by  $f_{\mathbf{H}}(\mathbf{h}) = (1/\pi^n \det(\mathbf{R})) \exp(-\mathbf{h}^* \mathbf{R}^{-1} \mathbf{h})$ . Further for simplicity, assume  $\mathbf{Z}$  to be invertible. We will show shortly that this assumption is not necessary. The pairwise error probability averaged over the channel realizations is given

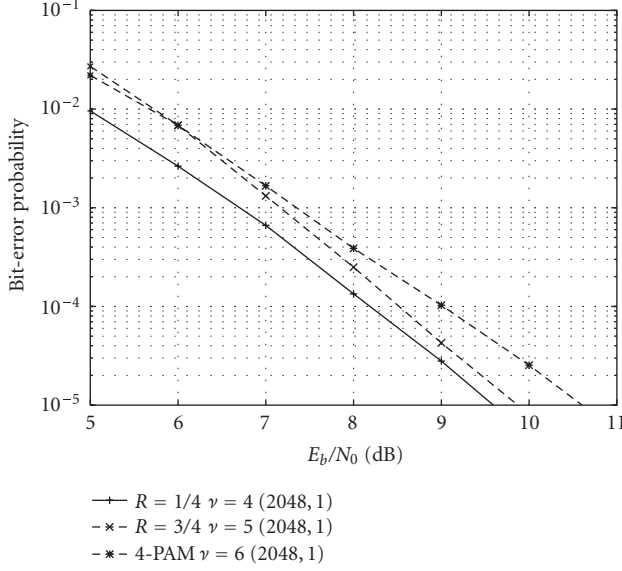


FIGURE 15: Performance comparison of the 16-point constellation  $\mathcal{M}^1$  with rate pair  $(1/4, 3/4)$  versus 4-PAM with rate  $1/2$  outer code on the i.i.d channel, that is,  $D = 2048$ ,  $b = 1$ .

by

$$P(\mathbf{c}^{(i)} \rightarrow \mathbf{c}^{(j)}) \leq \int e^{-(E_s/8\sigma^2)\mathbf{h}^* \mathbf{Z} \mathbf{h}} f_{\mathbf{H}}(\mathbf{h}) d\mathbf{h} \quad (\text{A.5})$$

$$= \frac{\det \mathbf{R}^{-1}}{(\pi)^{NM}} \int e^{-\mathbf{h}^* ((E_s/8\sigma^2)\mathbf{Z} + \mathbf{R}^{-1}) \mathbf{h}} d\mathbf{h} \quad (\text{A.6})$$

$$= \frac{\det(\mathbf{R}^{-1})}{\det(\mathbf{R}^{-1} + (E_s/8\sigma^2)\mathbf{Z})} \times \left[ \frac{\det(\mathbf{R}^{-1} + (E_s/8\sigma^2)\mathbf{Z})}{(\pi)^{NM}} \int e^{-\mathbf{h}^* (E_s/8\sigma^2)\mathbf{Z} + \mathbf{R}^{-1}) \mathbf{h}} d\mathbf{h} \right] \quad (\text{A.7})$$

$$= \frac{1}{\det(\mathbf{I} + (E_s/8\sigma^2)\mathbf{R}\mathbf{Z})}. \quad (\text{A.8})$$

Since  $\mathbf{Z}$  and  $\mathbf{R}$  are positive definite,  $(\mathbf{R} + (E_s/8\sigma^2)\mathbf{Z}^{-1})^{-1}$  is positive definite and hence a valid autocorrelation matrix. As a result, the term in square brackets in (A.7) integrates out to 1.

If  $\mathbf{R}$  does not have full rank, the probability density function  $f_{\mathbf{H}}(\mathbf{h})$  is not defined. We show that even in this case the upper bound on the pairwise error probability given by (A.8) holds. Let  $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$  and  $\mathbf{v} = (v_1, v_2, \dots, v_n)^T$  be random vectors defined on the probability space  $(\Omega, \mathcal{F}, \mathcal{P})$ . Define  $\mathbf{y}_m = \mathbf{x} + (1/m)\mathbf{v}$ . Thus,

$$\lim_{m \rightarrow \infty} \mathbf{y}_m = \mathbf{x} \quad \text{almost surely.} \quad (\text{A.9})$$

Let  $f : \mathbb{C}^n \rightarrow \mathbb{R}$  and suppose that there is a real number  $M$  such that  $|f(\mathbf{s})| \leq M$  for all  $\mathbf{s} \in \mathbb{C}^n$ . From the bounded

convergence theorem [20, Section 4.2], we have

$$E_{\mathbf{y}_m}[f(\mathbf{y}_m)] \rightarrow E_{\mathbf{x}}[f(\mathbf{x})], \quad (\text{A.10})$$

where  $E[\cdot]$  denotes the expectation operator.

If we define  $\mathbf{y}_m = \mathbf{h} + (1/m)\mathbf{v}$  where  $\mathbf{v}$  is a zero-mean circularly symmetric complex Gaussian with  $E[\mathbf{v}\mathbf{v}^*] = \mathbf{I}$ , then  $\mathbf{y}_m$  is also zero-mean circularly symmetric complex Gaussian with positive definite correlation matrix  $\mathbf{R}_m = E[\mathbf{y}_m\mathbf{y}_m^*] = \mathbf{R} + (1/m^2)\mathbf{I}$ . Thus  $\mathbf{R}_m$  has full rank irrespective of the rank of  $\mathbf{R}$  and hence  $\mathbf{y}_m$  has a well-defined probability density function. Define  $f(\mathbf{x}) = e^{-(E_s/8\sigma^2)\mathbf{h}^* \mathbf{Z} \mathbf{x}} \leq 1$  for all  $\mathbf{x} \in \mathbb{C}^{NM}$ . It follows from (A.4), (A.8), and (A.10) that

$$E[f(\mathbf{y}_m)] = \frac{1}{\det(\mathbf{I} + (E_s/8\sigma^2)(\mathbf{R} + (1/m^2)\mathbf{I})\mathbf{Z})} \quad (\text{A.11})$$

and hence

$$\begin{aligned} P(\mathbf{c}^{(i)} \rightarrow \mathbf{c}^{(j)}) &\leq E_{\mathbf{h}}[f(\mathbf{h})] = \lim_{m \rightarrow \infty} E[f(\mathbf{y}_m)] \\ &= \frac{1}{\det(\mathbf{I} + (E_s/8\sigma^2)\mathbf{R}\mathbf{Z})}. \end{aligned} \quad (\text{A.12})$$

Similarly, since the determinant of a matrix is a continuous function of its entries, a limiting argument can be used to show that (A.8) holds even if  $\mathbf{Z}$  does not have full rank.

## ACKNOWLEDGMENTS

The first author would like to thank Dr. Helmut Boelcskei and Dr. Joseph Boutros for their help during various stages of preparing this manuscript. We also would like to thank the anonymous reviewers whose comments helped to improve the quality of this manuscript. This work was supported in part by the National Science Foundation under Grant NSF-CCF 0325924 and a Vodafone-US Foundation Graduate Fellowship. The material in this paper was presented in part at the 2002 and 2004 International Symposium on Information Theory (ISIT).

## REFERENCES

- [1] W. Kozek, *Matched Weyl-Heisenberg expansions of nonstationary environments*, Ph.D. thesis, Vienna University of Technology, Vienna, Austria, March 1997.
- [2] W. Kozek, "Adaptation of Weyl-Heisenberg frames to under-spread environments," in *Gabor Analysis and Algorithms: Theory and Applications*, H. G. Feichtinger and T. Strohmer, Eds., pp. 323–352, Birkhäuser, Boston, Mass, USA, 1998.
- [3] K. Liu, T. Kadous, and A. M. Sayeed, "Orthogonal time-frequency signaling over doubly dispersive channels," *IEEE Transactions on Information Theory*, vol. 50, no. 11, pp. 2583–2603, 2004.
- [4] X. Giraud, E. Boutillon, and J. C. Belfiore, "Algebraic tools to build modulation schemes for fading channels," *IEEE Transactions on Information Theory*, vol. 43, no. 3, pp. 938–952, 1997.
- [5] J. Boutros and E. Viterbo, "Signal space diversity: a power- and bandwidth-efficient diversity technique for the Rayleigh fading channel," *IEEE Transactions on Information Theory*, vol. 44, no. 4, pp. 1453–1467, 1998.

- [6] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Transactions on Information Theory*, vol. 44, no. 3, pp. 927–946, 1998.
- [7] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Transactions on Information Theory*, vol. 28, no. 1, pp. 55–66, 1982.
- [8] L. Zadeh, "Time-varying networks, I," *Proceedings of IRE*, vol. 49, pp. 1488–1503, 1961.
- [9] J. G. Proakis, *Digital Communications*, chapter 14, McGraw-Hill, New York, NY, USA, 4th edition, 2001.
- [10] G. Matz and F. Hlawatsch, "Time-frequency transfer function calculus of linear time-varying systems," in *Time-Frequency Signal Analysis and Processing*, B. Boashash, Ed., Prentice-Hall, Englewood Cliffs, NJ, USA, 2003.
- [11] W. Kozek and A. F. Molisch, "Nonorthogonal pulseshapes for multicarrier communications in doubly dispersive channels," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1579–1589, 1998.
- [12] K. Leeuwijn-Bouille and J. C. Belfiore, "The cutoff rate of time correlated fading channels," *IEEE Transactions on Information Theory*, vol. 39, no. 2, pp. 612–617, 1993.
- [13] Z. Wang and G. B. Giannakis, "A simple and general parameterization quantifying performance in fading channels," *IEEE Transactions on Communications*, vol. 51, no. 8, pp. 1389–1398, 2003.
- [14] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: information-theoretic and communications aspects," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2619–2692, 1998.
- [15] S. Siwamogsatham and M. P. Fitz, "Robust space-time codes for correlated Rayleigh fading channels," *IEEE Transactions on Signal Processing*, vol. 50, no. 10, pp. 2408–2416, 2002.
- [16] H. Imai and S. Hiraikawa, "A new multilevel coding method using error-correcting codes," *IEEE Transactions on Information Theory*, vol. 23, no. 3, pp. 371–377, 1977.
- [17] G. D. Forney Jr., R. G. Gallager, G. R. Lang, F. M. Longstaff, and S. U. Qureshi, "Efficient modulation for band-limited channels," *IEEE Journal on Selected Areas in Communications*, vol. 2, no. 5, pp. 632–647, 1984.
- [18] U. Wachsmann, R. F. H. Fischer, and J. B. Huber, "Multilevel codes: theoretical concepts and practical design rules," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1361–1391, 1999.
- [19] J. W. Modestino and S. Y. Mui, "Convolutional code performance in the rician fading channel," *IEEE Transactions on Communications*, vol. 24, no. 6, pp. 592–606, 1976.
- [20] H. L. Royden, *Real Analysis*, chapter 4, Prentice-Hall, Englewood Cliffs, NJ, USA, 3rd edition, 1988.

**Ralf Koetter** received a Diploma in electrical engineering from the Technical University Darmstadt, Germany, in 1990 and a Ph.D. degree from the Department of Electrical Engineering at Linköping University, Sweden. From 1996/1998, he was a Visiting Scientist at the IBM Almaden Research Lab., San Jose, California. He was a Visiting Assistant Professor at the University of Illinois at Urbana/Champaign and Visiting Scientist at CNRS in Sophia Antipolis, France. He joined the faculty of the University of Illinois at Urbana-Champaign in 1999 and is currently an Associate Professor at the Coordinated Science Laboratory at the University. His research interests include coding and information theories and their application to communication systems. In the years 1999–2001, he served as an Associate Editor for coding theory & techniques for the *IEEE Transactions on Communications*. In 2003, he concluded a term as an Associate Editor for coding theory of the *IEEE Transactions on Information Theory*. He received an IBM Invention Achievement Award in 1997, an NSF CAREER Award in 2000, and an IBM Partnership Award in 2001. He is the co-recipient of the 2004 Paper Award of the Information Theory Society. Since 2003, he has been a Member of the Board of Governors of the IEEE Information Theory Society.



**Siddhartha Mallik** received a B.Tech. degree in electrical engineering from the Indian Institute of Technology, Bombay, in 2001. He received an M.S. degree in electrical and computer engineering from the University of Illinois at Urbana-Champaign in 2004, where he is currently a Ph.D. candidate. His research interests include coding and information theories and their applications in communications systems and networks.

