invited paper in Proc. 39th Allerton Conf. on Commun., Contr., Comput., Urbana (IL), Oct. 3-5, 2001.

Space-Time Multicarrier Matrix Modulation for Unknown Dispersive MIMO Channels*

Harold Artés, Gerald Matz, Dieter Schafhuber, and Franz Hlawatsch

Institute of Communications and Radio-Frequency Engineering
Vienna University of Technology
Gusshausstrasse 25/389, A-1040 Vienna, Austria
phone: +43 1 58801 38923, fax: +43 1 58801 38999, e-mail: hartes@aurora.nt.tuwien.ac.at
web: http://www.nt.tuwien.ac.at/dspgroup/time.html

Abstract

We propose a space-time multicarrier (OFDM) matrix modulation scheme that is suited for transmission over unknown dispersive MIMO channels. This scheme is flexible in that it allows to trade transmit redundancy (i.e., diversity and channel identification capability) against data rate without code redesign. It is shown that under appropriate conditions, the structure imposed by matrix modulation is sufficient to allow demodulation without knowledge of the channel. We discuss the maximum-likelihood (ML) receiver (joint ML channel estimator/data detector) and present a bound on its pairwise error probability. As an alternative to the computationally intensive ML receiver, we also propose an efficient iterative receiver algorithm. The performance of the proposed MIMO OFDM matrix modulation scheme using the iterative receiver algorithm is assessed via numerical simulations.

1 Introduction

Multiple input/multiple output (MIMO) channels offer the possibility of high data rates (e.g. [1]) or high diversity gains (e.g. [2,3]) or a compromise between the two [4,5]. Most papers on space-time coding/modulation schemes consider flat-fading MIMO channels (i.e., channels without memory or delay spread), only few consider MIMO channels with delay spread [6–8]. Moreover, it is often assumed that the channel is known at the receiver—an assumption that is not always justified.

In this paper, we propose a communications scheme for delay-spread MIMO channels that does not require the channel to be known at the receiver or at the transmitter. The new scheme is flexible in that it allows to adjust data rate versus transmit redundancy (i.e., diversity and channel identification capability) without code redesign. We use MIMO orthogonal frequency-division multiplexing (OFDM) to convert the dispersive MIMO channel into several nondispersive (flat-fading) channels. Furthermore, unlike methods using incoherent receivers (e.g. [9,10]), our method uses the structure of a novel linear space-time modulation scheme—which we call *matrix modulation* [11]—to jointly estimate the channel and the data. Thus, conceptually, our approach lies between coherent detection (e.g. [2]) and incoherent detection (e.g. [9,10]).

*Funding by FWF grants P12228-TEC and P11904-TEC.

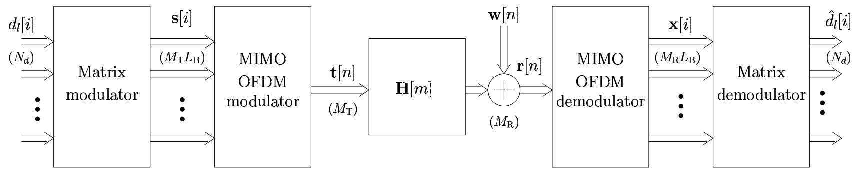

Figure 1: *Block diagram of the space-time multicarrier matrix modulation system. The MIMO OFDM transmission system is surrounded by matrix modulation/demodulation. The dimensions of all vector signals are given in parentheses.*

This paper is organized as follows. Section 2 provides a brief review of MIMO OFDM. In Section 3, we present the novel transmission scheme and an identifiability theorem motivating the joint channel-estimation/data-detection approach. In Section 4, we discuss the maximum-likelihood (ML) estimator/detector and analyze it in terms of pairwise error probability. An efficient iterative receiver algorithm that avoids the high computational cost of the ML receiver is described in Section 5. Finally, simulation results illustrating the performance of our method are presented in Section 6.

2 Review of MIMO OFDM

We consider a MIMO channel with M_{T} transmit antennas and M_{R} receive antennas. The channel features memory (intersymbol interference) and is thus described by

$$\mathbf{r}[n] = \sum_{m=0}^{L-1} \mathbf{H}[m]\, \mathbf{t}[n-m] + \mathbf{w}[n]\,, \tag{1}$$

with the transmitted vector $\mathbf{t}[n] \triangleq \big[t_1[n]\ t_2[n] \cdots t_{M_{\mathrm{T}}}[n]\big]^T$, the received vector $\mathbf{r}[n] \triangleq \big[r_1[n]\ r_2[n] \cdots r_{M_{\mathrm{R}}}[n]\big]^T$, the uncorrelated white noise vector $\mathbf{w}[n] \triangleq \big[w_1[n]\ w_2[n] \cdots w_{M_{\mathrm{R}}}[n]\big]^T$, and the channel impulse response $\mathbf{H}[m]$ ($m = 0, 1, \cdots, L-1$, with $L-1$ the channel's maximum time delay). This channel is shown in the central part of Figure 1. The L impulse response matrices $\mathbf{H}[m]$ of size $M_{\mathrm{R}} \times M_{\mathrm{T}}$ are unknown to both the transmitter and the receiver.

In MIMO OFDM [6,7], similarly to conventional OFDM (e.g. [12]), a block of N transmit vectors $\mathbf{t}_i[0], \mathbf{t}_i[1], \cdots, \mathbf{t}_i[N-1]$ and a cyclic prefix $\mathbf{t}_i[N-N_{\mathrm{cp}}], \mathbf{t}_i[N-N_{\mathrm{cp}}+1], \cdots, \mathbf{t}_i[N-1]$ of length $N_{\mathrm{cp}} \geq L$ are combined into an *OFDM symbol*. Here, N is the number of subcarriers (frequency bins). The ith OFDM symbol thus corresponds to the sequence of transmit vectors $\mathbf{t}_i[N-N_{\mathrm{cp}}], \mathbf{t}_i[N-N_{\mathrm{cp}}+1], \cdots, \mathbf{t}_i[N-1], \mathbf{t}_i[0], \mathbf{t}_i[1], \cdots, \mathbf{t}_i[N-1]$. The transmit vectors $\mathbf{t}_i[0], \mathbf{t}_i[1], \cdots, \mathbf{t}_i[N-1]$ are computed as the N-point inverse FFT of the M_{T}-dimensional vectors $\mathbf{s}_{i,k}$ (with $k \in \{0, 1, \cdots, N-1\}$ the subcarrier or frequency index) that form the input to the OFDM modulator (see Figure 1), i.e.,

$$\mathbf{t}_i[n] = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \mathbf{s}_{i,k}\, e^{j2\pi \frac{kn}{N}}\,, \qquad n = 0, 1, \cdots, N-1.$$

The channel output corresponding to the ith transmitted OFDM symbol is given by the sequence of received vectors $\mathbf{r}_{\mathrm{cp},i}[0], \mathbf{r}_{\mathrm{cp},i}[1], \cdots, \mathbf{r}_{\mathrm{cp},i}[N_{\mathrm{cp}}-1], \mathbf{r}_i[0], \mathbf{r}_i[1], \cdots, \mathbf{r}_i[N-1]$. The OFDM demodulator discards the first N_{cp} vectors (corresponding to the cyclic prefix) and applies an FFT to the remaining N vectors, i.e.,

$$\mathbf{x}_{i,k} = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \mathbf{r}_i[n]\, e^{-j2\pi \frac{kn}{N}}, \qquad k = 0, 1, \cdots, N-1.$$

It can then be shown that the vectors $\mathbf{s}_{i,k}$ and $\mathbf{x}_{i,k}$ are related as

$$\mathbf{x}_{i,k} = \mathbf{H}_k \mathbf{s}_{i,k} + \mathbf{w}_{i,k}, \qquad k = 0, 1, \cdots, N-1, \tag{2}$$

where $\mathbf{H}_k = \sum_{m=0}^{L-1} \mathbf{H}[m]\, e^{-j2\pi \frac{mk}{N}}$ and the $\mathbf{w}_{i,k}$ are uncorrelated white Gaussian noise vectors. According to (2), the original channel in (1) has been block-diagonalized in that it has been replaced by N separate subchannels that do not interfere with one another; moreover, each subchannel is a flat-fading (nondispersive) MIMO channel.

3 Combining Matrix Modulation with MIMO OFDM

We could use the subchannels in (2) by first estimating the subchannel transfer function matrices $\mathbf{H}_k$ and then transmitting data separately over each subchannel. However, such a scheme has two drawbacks. First, estimating the $\mathbf{H}_k$ in a time-varying environment would require a considerable percentage of transmit symbols to be used as pilot symbols. Second, due to the channel's delay spread, different subchannels $\mathbf{H}_{k_1}$ and $\mathbf{H}_{k_2}$ fade independently if $|k_1 - k_2|$ is larger than the channel's relative coherence bandwidth [13] which is about N/L, resulting in a maximum frequency diversity order of L [6]. With separate transmission over the subchannels, this frequency diversity could not be exploited.

3.1 Matrix Modulation

These drawbacks can be avoided by proper use of a novel linear space-time coding scheme called *matrix modulation* [11]. Matrix modulation imposes a structure on the transmit signal that allows joint channel and data estimation without using pilot symbols. To exploit the available frequency diversity, we apply matrix modulation to sets of $L_\mathrm{B} \geq L$ subchannels corresponding to the frequency bins $k_1, k_1+\Delta k, k_1+2\Delta k, \cdots, k_1+(L_\mathrm{B}-1)\Delta k$, where $k_1 \in \{0, 1, \cdots, \Delta k - 1\}$ and $\Delta k = N/L_\mathrm{B}$. (Here, N is chosen as a multiple of L_B.) Note that the subcarrier frequencies belonging to any given subchannel set are spaced maximally far apart to achieve uncorrelated fading and thus maximum frequency diversity. For $L_\mathrm{B} < L$, we would not exploit the available frequency diversity, whereas for L_B significantly larger than L, the subchannels would be correlated and we would not obtain additional frequency diversity.

We will explain the matrix modulation scheme for a given $k_1 \in \{0, 1, \cdots, \Delta k - 1\}$. Defining the block-diagonal channel matrix $\mathbf{G} \triangleq \mathrm{diag}\big\{\mathbf{H}_{k_1}, \mathbf{H}_{k_1+\Delta k}, \cdots, \mathbf{H}_{k_1+(L_\mathrm{B}-1)\Delta k}\big\}$ of size $L_\mathrm{B} M_\mathrm{R} \times L_\mathrm{B} M_\mathrm{T}$ as well as the input vector $\mathbf{s}[i] \triangleq \big[\mathbf{s}_{i,k_1}^T\, \mathbf{s}_{i,k_1+\Delta k}^T \cdots\, \mathbf{s}_{i,k_1+(L_\mathrm{B}-1)\Delta k}^T\big]^T$, the output vector $\mathbf{x}[i] \triangleq \big[\mathbf{x}_{i,k_1}^T\, \mathbf{x}_{i,k_1+\Delta k}^T \cdots\, \mathbf{x}_{i,k_1+(L_\mathrm{B}-1)\Delta k}^T\big]^T$, and the noise vector $\mathbf{w}[i] \triangleq \big[\mathbf{w}_{i,k_1}^T\, \mathbf{w}_{i,k_1+\Delta k}^T \cdots\, \mathbf{w}_{i,k_1+(L_\mathrm{B}-1)\Delta k}^T\big]^T$, the subchannel input-output relations (2) associated to the frequency bins $k_1, k_1+\Delta k, \cdots, k_1+(L_\mathrm{B}-1)\Delta k$ can be combined as

$$\mathbf{x}[i] = \mathbf{G}\,\mathbf{s}[i] + \mathbf{w}[i]. \tag{3}$$

This corresponds to the central part of Figure 2. We now apply the matrix modulation technique to the vector channel in (3). A block of N_B consecutive OFDM symbols are

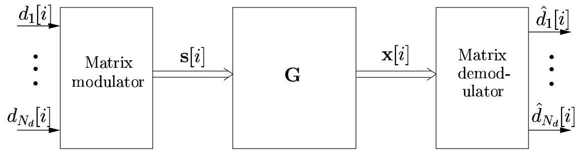

Figure 2: *Matrix modulation for one set of subchannels.*

coded at once; specifically, a "modulation structure" is forced on the $L_\mathrm{B} M_\mathrm{T} \times N_\mathrm{B}$ transmit matrix $\mathbf{S} \triangleq \big[\mathbf{s}[0]\, \mathbf{s}[1] \cdots \mathbf{s}[N_\mathrm{B}-1]\big]$ according to

$$\mathbf{S} = \sum_{l=1}^{N_d} \mathbf{M}_l \mathbf{D}_l. \tag{4}$$

Here, the N_d data matrices $\mathbf{D}_l$ are diagonal $N_\mathrm{B} \times N_\mathrm{B}$ matrices defined as $\mathbf{D}_l \triangleq \mathrm{diag}\{d_l[0], d_l[1], \cdots, d_l[N_\mathrm{B}-1]\}$. The N_d *modulation matrices* $\mathbf{M}_l$ of size $L_\mathrm{B} M_\mathrm{T} \times N_\mathrm{B}$ are assumed to be known to the receiver; some comments regarding their choice will be made presently. The data sequences $d_l[i] \in \mathbb{C}$ $(l = 1, 2, \cdots, N_d)$ form the input of the overall transmission system (cf. Figure 1). Note that for matrix modulation, the $d_l[i]$ need not be taken from a finite alphabet. The number N_d of data streams is bounded as $N_d \leq L_\mathrm{B} M_\mathrm{T} - 1$. The value of N_d determines the redundancy of the transmission scheme and allows to trade off data rate against reliability without redesigning the modulation matrices $\mathbf{M}_l$.

Next, we group N_B received vectors $\mathbf{x}[i]$ into the matrix $\mathbf{X} \triangleq \big[\mathbf{x}[0]\, \mathbf{x}[1] \cdots \mathbf{x}[N_\mathrm{B}-1]\big]$ and define the noise matrix $\mathbf{W}$ analogously. Combining (3) and (4), we then obtain

$$\mathbf{X} = \mathbf{G}\mathbf{S} + \mathbf{W} = \mathbf{G} \sum_{l=1}^{N_d} \mathbf{M}_l \mathbf{D}_l + \mathbf{W}. \tag{5}$$

3.2 Uniqueness of Reconstruction

We now present a uniqueness/identifiability theorem which states that in the absence of noise, $\mathbf{W} = \mathbf{0}$, and under appropriate conditions, the structure of $\mathbf{S}$ defined by (4) allows unique reconstruction (up to a constant factor) of the data $d_l[i]$ from $\mathbf{X}$ without knowledge of $\mathbf{G}$. That is, the received matrix $\mathbf{X} = \mathbf{G} \sum_{l=1}^{N_d'} \mathbf{M}_l \mathbf{D}_l$ (where $N_d' \leq N_d$ is the number of *active* data streams, i.e., we allow some data streams to be zero) does not permit a different representation $\tilde{\mathbf{G}} \sum_{l=1}^{N_d} \mathbf{M}_l \tilde{\mathbf{D}}_l$. This is true even though no finite-alphabet property was assumed for $d_l[i]$. This result provides a motivation of our joint channel-estimation/data-detection approach.

Theorem 1. *Let $\mathbf{D}_l$ and $\tilde{\mathbf{D}}_l$ be N_d' resp. N_d diagonal matrices of size $N_\mathrm{B} \times N_\mathrm{B}$, with $\mathbf{D}_l$ nonsingular, and let $\mathbf{G}$ and $\tilde{\mathbf{G}}$ be block-diagonal matrices of size $L_\mathrm{B} M_\mathrm{R} \times L_\mathrm{B} M_\mathrm{T}$ with full rank and block entries of size $M_\mathrm{R} \times M_\mathrm{T}$. Let $N_d' \leq N_d \leq L_\mathrm{B} M_\mathrm{T} - 1$, $M_\mathrm{R} \geq M_\mathrm{T}$, and $N_\mathrm{B} \geq \left\lceil \frac{L_\mathrm{B} M_\mathrm{T}^2 - 1}{L_\mathrm{B} M_\mathrm{T} - N_d} \right\rceil$. Then there exist N_d matrices $\mathbf{M}_l$ $(l = 1, \cdots, N_d)$ of size $L_\mathrm{B} M_\mathrm{T} \times N_\mathrm{B}$ such that for a given set of data matrices $\mathbf{D}_l$,*

$$\tilde{\mathbf{G}} \sum_{l=1}^{N_d} \mathbf{M}_l \tilde{\mathbf{D}}_l = \mathbf{G} \sum_{l=1}^{N_d'} \mathbf{M}_l \mathbf{D}_l$$

implies $\tilde{\mathbf{G}}^{\#}\mathbf{G} = c\mathbf{I}$ *(here,* $\tilde{\mathbf{G}}^{\#}$ *denotes the pseudo-inverse of* $\tilde{\mathbf{G}}$*) and*

$$\tilde{\mathbf{D}}_l = \begin{cases} c\mathbf{D}_l\,, & l \le N_d' \\ \mathbf{0}\,, & N_d' < l \le N_d\,, \end{cases}$$

where $c \in \mathbb{C}$ *is an unknown factor.*

This theorem extends the theorem in [11] to the case of block-diagonal matrices $\mathbf{G}$. A generalized version of the theorem (allowing rank-deficient channels $\mathbf{G}$ and $M_{\mathrm{R}} < M_{\mathrm{T}}$) and its proof will be presented in a future publication.

The modulation matrices $\mathbf{M}_l$ must be chosen such that for any fixed $i \in \{0, 1, \cdots, N_{\mathrm{B}}-1\}$, the ith columns of all $\mathbf{M}_l$ are linearly independent; this guarantees a one-to-one relation between $\mathbf{S} = \sum_{l=1}^{N_d} \mathbf{M}_l \mathbf{D}_l$ and the data involved in $\mathbf{S}$ and thus allows the data to be recovered from $\mathbf{S}$. In addition, the $\mathbf{M}_l$ must satisfy other properties that are more difficult to characterize. In our experiments, we observed that taking realizations of iid Gaussian random variables as matrix entries always resulted in suitable $\mathbf{M}_l$.

The minimum block length N_{B} according to the theorem, $N_{\mathrm{B,min}} = \left\lceil \frac{L_{\mathrm{B}} M_{\mathrm{T}}^2 - 1}{L_{\mathrm{B}} M_{\mathrm{T}} - N_d} \right\rceil$, is approximately independent of the number of channels over which we code, L_{B}, as long as the redundancy $\frac{N_d}{N_{d,\max}} = \frac{N_d}{L_{\mathrm{B}} M_{\mathrm{T}} - 1}$ is unchanged. To illustrate this point, let us set $N_d = K L_{\mathrm{B}}$ where $K \le M_{\mathrm{T}} - 1$. Then, both the redundancy $\frac{N_d}{N_{d,\max}} = \frac{K L_{\mathrm{B}}}{L_{\mathrm{B}} M_{\mathrm{T}} - 1} \approx \frac{K}{M_{\mathrm{T}}}$ and the minimum block length $N_{\mathrm{B,min}} = \left\lceil \frac{L_{\mathrm{B}} M_{\mathrm{T}}^2 - 1}{L_{\mathrm{B}} (M_{\mathrm{T}} - K)} \right\rceil \approx \left\lceil \frac{M_{\mathrm{T}}^2}{M_{\mathrm{T}} - K} \right\rceil$ are approximately independent of L_{B}. Thus, for fixed redundancy $N_d / N_{d,\max}$, coding over multiple subchannels to exploit frequency diversity does not increase the minimum block length $N_{\mathrm{B,min}}$.

4 Maximum-Likelihood Receiver

In this section, we discuss the maximum-likelihood (ML) receiver for the matrix modulation scheme. For ML detection, we assume that $d_l[i] \in \mathcal{D}$ with $\mathcal{D}$ being a finite data alphabet. Let $\left(\mathbf{D}_1^j, \mathbf{D}_2^j, \cdots, \mathbf{D}_{N_d}^j\right)$ denote the jth data matrix sequence with $d_l^j[i] \in \mathcal{D}$, and let $\mathbf{S}_j = \sum_{l=1}^{N_d} \mathbf{M}_l \mathbf{D}_l^j \in \mathcal{S}$ denote the associated jth transmit matrix (here, $\mathcal{S}$ denotes the finite alphabet of all possible $\mathbf{S}_j$). If the ith columns of all modulation matrices $\mathbf{M}_l$ are linearly independent so that there is a one-to-one relation between $\mathbf{S}_j$ and the data $d_l^j[i]$ contained in $\mathbf{S}_j$, then data detection is equivalent to detection of $\mathbf{S}_j$.

Allowing additive channel noise, the received matrix $\mathbf{X}$ corresponding to the transmit matrix $\mathbf{S}_j$ is $\mathbf{X} = \mathbf{G}\mathbf{S}_j + \mathbf{W} = \mathbf{G} \sum_{l=1}^{N_d} \mathbf{M}_l \mathbf{D}_l^j + \mathbf{W}$ (cf. (5)). Assuming the noise vector sequence $\mathbf{w}[n]$ in (1) to be temporally and spatially white, i.e., $\mathbb{E}\{\mathbf{w}[n]\mathbf{w}^H[n']\} = \sigma_w^2 \, \mathbf{I}\, \delta_{n,n'}$, the entries of $\mathbf{W}$ are iid Gaussian with zero mean and variance σ_w^2.

4.1 ML Detector

With these assumptions, it can be shown that the ML receiver performing joint ML data detection and ML channel estimation is given by

$$(\hat{\mathbf{S}}_{\mathrm{ML}}, \hat{\mathbf{G}}_{\mathrm{ML}}) = \operatorname*{argmin}_{\mathbf{S}_j \in \mathcal{S}, \mathbf{G}} \left\| \mathbf{X} - \mathbf{G}\mathbf{S}_j \right\|^2, \qquad (6)$$

where $\|\cdot\|$ denotes the Frobenius norm. For a given (fixed) $\mathbf{S}_j \in \mathcal{S}$, it can be shown that $\hat{\mathbf{G}}_{\mathrm{ML}} = \mathbf{X}\mathbf{S}_j^{\#}$. Inserting in (6) then yields the ML data detector as

$$\hat{\mathbf{S}}_{\mathrm{ML}} = \operatorname*{argmin}_{\mathbf{S}_j \in \mathcal{S}} \left\| \mathbf{X} - \mathbf{X}\mathbf{S}_j^{\#}\mathbf{S}_j \right\|^2 = \operatorname*{argmin}_{\mathbf{P}_j \in \mathcal{P}} \left\| \mathbf{X} - \mathbf{X}\mathbf{P}_j \right\|^2 = \operatorname*{argmax}_{\mathbf{P}_j \in \mathcal{P}} \left\| \mathbf{X}\mathbf{P}_j \right\|^2, \qquad (7)$$

where the $N_{\mathrm{B}} \times N_{\mathrm{B}}$ matrix $\mathbf{P}_j \triangleq \mathbf{S}_j^{\#}\mathbf{S}_j$ denotes the orthogonal projector onto the space spanned by the rows of the transmit matrix $\mathbf{S}_j$ and $\mathcal{P}$ is the finite alphabet of all possible $\mathbf{P}_j$. From (7), it can be seen that in the special case of unitary $\mathbf{S}_j$, our joint estimator/detector receiver is equivalent to the incoherent receiver from [9].

4.2 Pairwise Error Probability

Let us model the diagonal blocks $\mathbf{H}_{k_1+(\mu-1)\Delta k}$ of $\mathbf{G} = \mathrm{diag}\left\{\mathbf{H}_{k_1}, \mathbf{H}_{k_1+\Delta k}, \cdots, \mathbf{H}_{k_1+(L_{\mathrm{B}}-1)\Delta k}\right\}$ as full-rank random matrices whose entries are iid Gaussian random variables (i.e., we assume that the subchannels fade independently). We suppose that a specific transmit matrix $\mathbf{S}_1$ was sent and consider the pairwise error probability (PEP) $P\{\mathbf{S}_1 \!\to\! \mathbf{S}_2 | \mathbf{S}_1\}$, i.e., the probability that the ML detector erroneously decides on a specific transmit matrix $\mathbf{S}_2 \neq \mathbf{S}_1$. From (7), we have

$$P\{\mathbf{S}_1 \!\to\! \mathbf{S}_2 | \mathbf{S}_1\} = P\left\{ \|\mathbf{X}\mathbf{P}_1\|^2 < \|\mathbf{X}\mathbf{P}_2\|^2 \,\middle|\, \mathbf{S}_1 \right\}.$$

Using the Chernoff bound, it can be shown that $P\{\mathbf{S}_1 \!\to\! \mathbf{S}_2 | \mathbf{S}_1\}$ is upper bounded as

$$P\{\mathbf{S}_1 \!\to\! \mathbf{S}_2 | \mathbf{S}_1\} \le \left[\prod_{m=1}^{M_{\mathrm{T}} L_{\mathrm{B}}} \left(1 - \alpha^2 \sigma_w^4 \lambda_m^2\right) \prod_{\mu=1}^{L_{\mathrm{B}}} \left| \mathbf{I} + \mathbf{I}_\mu \mathbf{S}_1 \left(\mathbf{I} - \alpha \sigma_w^2 \boldsymbol{\Delta}\right)^{-1} \boldsymbol{\Delta} \mathbf{S}_1^H \right| \right]^{-M_{\mathrm{R}}}, \qquad (8)$$

where λ_m is the mth positive eigenvalue[1] of $\boldsymbol{\Delta} \triangleq \mathbf{P}_2 - \mathbf{P}_1$, $|\cdot|$ denotes the determinant, $\mathbf{I}_\mu$ denotes a diagonal selection matrix of size $L_{\mathrm{B}} M_{\mathrm{T}} \times L_{\mathrm{B}} M_{\mathrm{T}}$ that has ones as its $[(\mu-1)M_{\mathrm{T}} + 1]$st through μM_{T}th diagonal entries and zeros elsewhere, and α is the Chernoff parameter bounded as $0 < \alpha < 1/\sigma_w^2$. Using the partitioning $\mathbf{S}_1 = [\mathbf{S}_{1,1}^T \, \mathbf{S}_{1,2}^T \cdots \mathbf{S}_{1,L_{\mathrm{B}}}^T]^T$, with submatrices $\mathbf{S}_{1,\mu}$ of size $M_{\mathrm{T}} \times N_{\mathrm{B}}$, and setting $\alpha = 1/(2\sigma_w^2)$, we can simplify (8) if we assume that all $\mathbf{S}_{1,\mu}$ have the same row span:

$$\begin{aligned} P\{\mathbf{S}_1 \!\to\! \mathbf{S}_2 | \mathbf{S}_1\} &\le \left[\prod_{m=1}^{r} \left(\frac{\lambda_m^2}{4\sigma_w^2} \right)^{L_{\mathrm{B}}} \prod_{\mu=1}^{L_{\mathrm{B}}} \left| \mathbf{S}_{1,\mu} \mathbf{S}_{1,\mu}^H \right| \right]^{-M_{\mathrm{R}}} \\ &= \left(\frac{1}{4\sigma_w^2} \right)^{-L_{\mathrm{B}} M_{\mathrm{R}} r} \left(\sqrt[r]{\prod_{m=1}^{r} \lambda_m^2} \right)^{-L_{\mathrm{B}} M_{\mathrm{R}} r} \left[\prod_{\mu=1}^{L_{\mathrm{B}}} \left| \mathbf{S}_{1,\mu} \mathbf{S}_{1,\mu}^H \right| \right]^{-M_{\mathrm{R}}}. \end{aligned} \qquad (9)$$

Here, r denotes the rank of $\boldsymbol{\Delta}$. The first and second term in this upper bound correspond to the well-known rank criterion and determinant criterion [2], respectively, albeit with squared eigenvalues. The third term involves the Gram matrices $\mathbf{S}_{1,\mu}\mathbf{S}_{1,\mu}^H$ and causes the PEP bound to be asymmetric, i.e., the bound on $P\{\mathbf{S}_1 \to \mathbf{S}_2 | \mathbf{S}_1\}$ is not equal to the bound on $P\{\mathbf{S}_2 \!\to\! \mathbf{S}_1 | \mathbf{S}_2\}$ in general. For unitary matrices $\mathbf{S}_j$, the bound in (9) can be shown to be similar to the bound in [9, Eq. (18)]. We can conclude from (9) that the $\mathbf{S}_{1,\mu}$ should be unitary since this can be shown to maximize the determinant $\left|\mathbf{S}_{1,\mu}\mathbf{S}_{1,\mu}^H\right|$ under the constraint $\|\mathbf{S}_{1,\mu}\| = \mathrm{const}$. Furthermore, we want as many eigenvalues λ_m as possible to be as large as possible. Ideally, $\mathbf{S}_1$ and $\mathbf{S}_2$ would be orthogonal, i.e., $\mathbf{S}_1\mathbf{S}_2^H = \mathbf{0}$, since this would lead to the maximal eigenvalues $\lambda_m = 1$ for $m = 1, 2, \cdots, M_{\mathrm{T}}$.

[1] It can be shown that the eigenvalues of $\boldsymbol{\Delta}$ are bounded as $-1 \le \lambda_m \le 1$ and occur pairwise such that for every λ_m there exists a $\lambda_{m'} = -\lambda_m$.

5 POCS Demodulation Algorithm

As a computationally efficient though suboptimal alternative to the ML detector given by (7), we propose a demodulation technique that is inspired by [11]. This algorithm consits of two stages: first, based on the matrix modulation structure, the channel is equalized using an iterative algorithm; second, the data is detected.

5.1 Equalization

For the formulation of the equalization method, we assume $d_l[i] \in \mathbb{C}$ (conversion to the finite data alphabet $\mathcal{D}$ will be performed by the subsequent detection step) and we initially assume the noise-free case. We are given a received matrix $\mathbf{X} = \mathbf{GS}$ where $\mathbf{G}$ and $\mathbf{S}$ are unknown but the modulation matrices $\mathbf{M}_l$ used at the transmitter are known. It can then be shown using Theorem 1 that the matrix $\mathbf{S} = \sum_{l=1}^{N_d} \mathbf{M}_l \mathbf{D}_l$ and, thus, the N_d data matrices $\mathbf{D}_l$ are uniquely determined (up to a scalar factor) by the following three properties:

1. $\mathbf{S} = \sum_{l=1}^{N_d} \mathbf{M}_l \mathbf{D}_l$ with $\mathbf{D}_l$ diagonal;
2. the row span of $\mathbf{S}$ equals the row span of $\mathbf{X}$;
3. $\mathbf{G}$ is block-diagonal with blocks of size $M_{\mathrm{R}} \times M_{\mathrm{T}}$.

(The second property follows from $\mathbf{X} = \mathbf{GS}$ with $\mathbf{G}$ full rank). It can be shown that these properties imply that $\mathbf{S} \in \mathcal{A} \cap \mathcal{B}$, where $\mathcal{A}$ is the linear subspace of all matrices $\sum_{l=1}^{N_d} \mathbf{M}_l \mathbf{D}_l$ with $\mathbf{M}_l$ given and $\mathbf{D}_l$ diagonal ($\mathcal{A}$ is a linear space since for the equalization step we assume $d_l[i] \in \mathbb{C}$) and $\mathcal{B}$ is the linear subspace of all matrices of the form $\mathbf{BX}$ with some $L_{\mathrm{B}} M_{\mathrm{T}} \times L_{\mathrm{B}} M_{\mathrm{R}}$ block-diagonal matrix $\mathbf{B}$. Since both $\mathcal{A}$ and $\mathcal{B}$ are linear subspaces and thus convex, the formulation $\mathbf{S} \in \mathcal{A} \cap \mathcal{B}$ suggests a POCS (projections onto convex sets) algorithm [14] for calculating and demodulating $\mathbf{S}$. This algorithm is iterative and consists in alternately projecting the iterated version of $\mathbf{S}$ onto $\mathcal{A}$ and $\mathcal{B}$.

Projection onto $\mathcal{A}$: The projection onto $\mathcal{A}$ amounts to forming $\mathbf{S}^{(\nu)} = \sum_{l=1}^{N_d} \mathbf{M}_l \mathbf{D}_l^{(\nu)}$, where the nonzero (diagonal) elements of $\mathbf{D}_l^{(\nu)}$ can be shown to be given by

$$\big(\mathbf{D}_l^{(\nu)}\big)_{i,i} = \frac{1}{L_{\mathrm{B}} M_{\mathrm{T}}} \sum_{m=1}^{L_{\mathrm{B}} M_{\mathrm{T}}} \big(\mathbf{S}^{(\nu-1)}\big)_{m,i} \big(\mathbf{M}_l^{+}\big)_{m,i} .$$

Here, $\mathbf{S}^{(\nu-1)}$ is the result of the previous iteration (i.e., the projection onto $\mathcal{B}$, see below) and the $M_{\mathrm{T}} \times N$ matrices $\mathbf{M}_l^{+}$ are defined such that $(\mathbf{m}_l^{+}[i])^T$, the transpose of the ith column of $\mathbf{M}_l^{+}$, equals the lth row of the $N_d \times L_{\mathrm{B}} M_{\mathrm{T}}$ matrix $\big[\mathbf{m}_1[i]\ \mathbf{m}_2[i] \cdots \mathbf{m}_{N_d}[i]\big]^{\#}$ (here, $\mathbf{m}_l[i]$ denotes the ith column of $\mathbf{M}_l$). If the vectors $\mathbf{m}_1[i], \mathbf{m}_2[i], \cdots, \mathbf{m}_{N_d}[i]$ are orthonormal, then there is simply $\mathbf{M}_l^{+} = \mathbf{M}_l^{*}$ with $\mathbf{M}_l^{*}$ the complex conjugate of $\mathbf{M}_l$.

Projection onto $\mathcal{B}$: The projection onto $\mathcal{B}$ amounts to forming $\mathbf{S}^{(\nu)} = \mathbf{B}^{(\nu)} \mathbf{X}$, where $\mathbf{B}^{(\nu)}$ is a block-diagonal matrix of size $L_{\mathrm{B}} M_{\mathrm{T}} \times L_{\mathrm{B}} M_{\mathrm{R}}$ whose mth diagonal block is given by $\mathbf{B}_m^{(\nu)} = \mathbf{S}_m^{(\nu-1)} \mathbf{X}_m^{\#}$. Here, $\mathbf{S}_m^{(\nu-1)}$ denotes the matrix consisting of the $[(m-1)M_{\mathrm{T}}+1]$st through $m M_{\mathrm{T}}$th row of $\mathbf{S}^{(\nu-1)}$, which in turn denotes the result of the previous iteration (i.e., the projection onto $\mathcal{A}$, see above). Furthermore, $\mathbf{X}_m^{\#}$ is the pseudo-inverse of a matrix consisting of the $[(m-1)M_{\mathrm{T}}+1]$st through $m M_{\mathrm{T}}$th row of $\mathbf{X}$. The matrices $\mathbf{X}_m^{\#}$ need to be computed only once at the start of the iterative equalization procedure.

The POCS algorithm is guaranteed to converge to an intersection point, i.e., $\mathbf{S}^{(\infty)} \in \mathcal{A} \cap \mathcal{B}$ [14]. Thus, in the noise-free case we have $\mathbf{S}^{(\infty)} = c\,\mathbf{S}$ and $\mathbf{D}_l^{(\infty)} = c\mathbf{D}_l$, where the $\mathbf{D}_l$ are the true data matrices and $c \in \mathbb{C}$. There exist various methods to speed up the convergence of the POCS algorithm [11, 14].

When noise is present, i.e., $\mathbf{X} = \mathbf{GS} + \mathbf{W}$ rather than $\mathbf{X} = \mathbf{GS}$, the space $\mathcal{A}$ is unchanged and there still is $\mathbf{S} \in \mathcal{A}$; however, the space $\mathcal{B}$ will be perturbed and $\mathbf{S} \notin \mathcal{B}$ in general. In fact, $\mathcal{A}$ and $\mathcal{B}$ must be expected to be incompatible in the sense that $\mathcal{A} \cap \mathcal{B} = \{0\}$; therefore, the above POCS algorithm would converge to the zero matrix. In this situation, we can use a modified POCS algorithm where the $\mathbf{S}^{(\nu)}$ resulting from the projection on $\mathcal{A}$ is scaled such that $\|\mathbf{S}^{(\nu)}\| = \gamma > 0$, where γ is an arbitrary but fixed positive constant. (The norm γ is arbitrary because $\mathbf{S}$ can be obtained only up to a constant factor anyway). The resulting $\mathbf{S}^{(\nu)}$ then converges to the matrix $\mathbf{S}^{(\infty)}$ that yields the minimum of $\|\mathbf{BX} - \mathbf{S}\|$ among all $\mathbf{S} \in \mathcal{A}$ with $\|\mathbf{S}\| = \gamma$ and all $\mathbf{B}$ that are block-diagonal matrices of size $L_{\mathrm{B}} M_{\mathrm{T}} \times L_{\mathrm{B}} M_{\mathrm{R}}$. Thus, $\mathbf{S}^{(\infty)}$ is the matrix $\mathbf{S} \in \mathcal{A}$ with prescribed Frobenius norm $\|\mathbf{S}\| = \gamma$ that has minimum distance from the space $\mathcal{B}$.

5.2 Detection

The second step (data detection) consists in "rounding" the result $\mathbf{S}^{(\infty)}$ of the POCS equalization algorithm such that it complies with the finite alphabet $\mathcal{S}$ underlying the transmit matrices $\mathbf{S}_j$. For this, the factor $c \in \mathbb{C}$ introduced by the POCS equalization algorithm must first be compensated (it is assumed that this factor has been estimated during a brief training phase). Thus, the data detection step calculates

$$\hat{\mathbf{S}} = \operatorname*{argmin}_{\mathbf{S}_j \in \mathcal{S}} \left\| \frac{1}{c} \mathbf{S}^{(\infty)} - \mathbf{S}_j \right\|^2 . \tag{10}$$

Because of (4), the ith column of $\mathbf{S}_j$ is given by $\mathbf{s}_j[i] = \sum_{l=1}^{N_d} \mathbf{m}_l[i] d_l^j[i]$, where $\mathbf{m}_l[i]$ denotes the ith column of $\mathbf{M}_l$. Using this structure and the one-to-one correspondence between $\mathbf{S}_j$ and the data matrices $\mathbf{D}_1^j, \mathbf{D}_2^j, \cdots, \mathbf{D}_{N_d}^j$, it follows that the minimization in (10) can be performed column by column according to

$$\big(\hat{d}_1[i], \hat{d}_2[i], \cdots, \hat{d}_{N_d}[i]\big) = \operatorname*{argmin}_{\{d_l^j[i]\} \in \mathcal{D}^{N_d}} \left\| \frac{1}{c} \mathbf{s}^{(\infty)}[i] - \sum_{l=1}^{N_d} \mathbf{m}_l[i]\, d_l^j[i] \right\| , \tag{11}$$

where $\mathbf{s}^{(\infty)}[i]$ is the ith column of $\mathbf{S}^{(\infty)}$. If all columns $\mathbf{m}_l[i]$ (for fixed i) are orthonormal, then (11) is equivalent to a simple componentwise rounding operation

$$\hat{d}_l[i] = \operatorname*{argmin}_{d \in \mathcal{D}} \big| s_l[i] - d \big| , \qquad l = 1, 2, \cdots, N_d ,$$

where $s_l[i]$ is the inner product of $\frac{1}{c}\mathbf{s}^{(\infty)}[i]$ with $\mathbf{m}_l[i]$.

6 Simulation Results

We studied the performance of the MIMO OFDM matrix modulation scheme presented in Section 3 using the equalization/detection algorithm described in the last section.

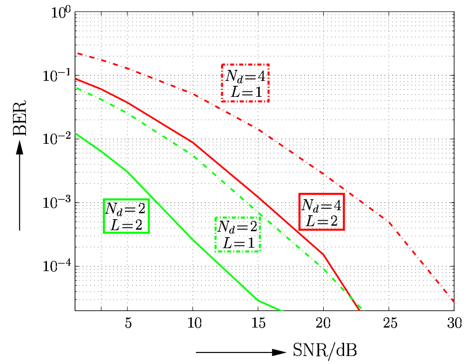

Figure 3: *BER versus SNR.*

The MIMO channels we considered featured $M_{\mathrm{T}} = 4$ transmit antennas, $M_{\mathrm{R}} = 6$ receive antennas, and impulse response lengths $L = 1$ or $L = 2$. We used $N_d = 2$ and $N_d = 4$ data streams with uncoded QPSK data symbols $d_l[i]$. The number of OFDM subcarriers (subchannels) was $N = 8$. The block length used for matrix modulation was $N_{\mathrm{B}} = 100$. The modulation matrices $\mathbf{M}_l$ were randomly generated using iid Gaussian random variables as matrix entries and subsequently orthonormalizing the ith columns of all $\mathbf{M}_l$. The number of subchannels combined for matrix modulation was $L_{\mathrm{B}} = 2$. For each simulation run, the channel impulse responses were randomly generated with iid Gaussian channel taps. The channel output signals were corrupted by white Gaussian noise with variance σ_w^2.

Fig. 3 shows the bit error rate (BER) versus the SNR[2] for a flat-fading channel ($L = 1$) and a dispersive channel with $L = 2$. In each case, we considered both $N_d = 2$ and $N_d = 4$ QPSK data streams, corresponding to 2 and 4 bits per subchannel use, respectively. It is seen from Fig. 3 that decreasing the data rate by decreasing N_d and, thus, increasing redundancy improves BER performance without code redesign. Furthermore, delay spread (i.e., $L = 2$ instead of $L = 1$) also leads to significant inprovements in BER performance. In fact, the results obtained with $L = 2$ and $N_d = 4$ (4 bits per subchannel use) are similar to those obtained with $L = 1$ and $N_d = 2$ (2 bits per subchannel use, i.e., half the data rate).

7 Conclusion

We proposed a transmission scheme for MIMO channels that combines MIMO OFDM with a novel space-time modulation technique termed *matrix modulation*. Under appropriate conditions, the structure imposed by matrix modulation is strong enough to allow demodulation without knowledge of the channel. Furthermore, the proposed transmission scheme is flexible in that it allows to trade transmit redundancy (i.e., diversity and channel identification capability) against data rate without code redesign. We also presented an efficient iterative demodulation method that exploits the subspace structure of matrix modulation and is far less computationally intensive than the ML receiver.

The optimum design of the modulation matrices used by the matrix modulation scheme (optimum with respect to error-rate performance in the presence of noise and/or convergence speed of the iterative POCS equalization algorithm) is an interesting problem for future research.

[2]As usual in the context of MIMO systems, the SNR is defined as $\frac{\mathbb{E}\{\|\mathbf{t}[n]\|^2\}}{M_{\mathrm{T}}\sigma_w^2}$.

Acknowledgment

The authors would like to thank Boris Dortschy for helpful discussions.

References

[1] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.

[2] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communications," *IEEE Trans. Inf. Theory*, vol. 44, pp. 744–765, March 1998.

[3] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Comm.*, vol. 16, pp. 1451–1458, Oct. 1998.

[4] R. W. Heath, H. Bölcskei, and A. J. Paulraj, "Space-time signalling and frame theory," in *Proc. IEEE ICASSP-2001*, (Salt Lake City, UT), pp. 2445–2448, May 2001.

[5] R. W. Heath and A. J. Paulraj, "Switching between multiplexing and diversity based on constellation distance," in *Asilomar Conf. Signals, Systems, Computers*, (Pacific Grove (CA)), Oct. 2000.

[6] H. Bölcskei and A. J. Paulraj, "Space-frequency coded broadband OFDM systems," in *Proc. IEEE WCNC-2000*, (Chicago, IL), pp. 1–6, Sept. 2000.

[7] D. Agrawal, V. Tarokh, A. Naguib, and N. Seshadri, "Space-time coded OFDM for high data rate wireless communications over wideband channels," in *Proc. IEEE VTC-98*, pp. 2232–2236, May 1998.

[8] Z. Liu, G. Giannakis, S. Bararossa, and A. Scaglione, "Transmit-antennae space-time block coding for generalized OFDM in the presence of unknown multipath," *IEEE J. Sel. Areas Comm.*, vol. 19, pp. 1352–1364, July 2001.

[9] B. Hochwald and T. Marzetta, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 46, pp. 543–564, March 2000.

[10] B. Hochwald and W. Sweldens, "Differential unitary space-time modulation," *IEEE Trans. Comm.*, vol. 48, pp. 2041–2052, Dec. 2000.

[11] H. Artés and F. Hlawatsch, "Blind equalization of MIMO channels using deterministic precoding," in *Proc. IEEE ICASSP-2001*, (Salt Lake City, UT), pp. 2153–2156, May 2001.

[12] J. A. C. Bingham, "Multicarrier modulation for data transmission: An idea whose time has come," *IEEE Comm. Mag.*, vol. 28, pp. 5–14, May 1990.

[13] R. S. Kennedy, *Fading Dispersive Communication Channels*. New York: Wiley, 1969.

[14] P. L. Combettes, "The foundations of set theoretic estimation," *Proc. IEEE*, vol. 81, pp. 181–208, Feb. 1993.