Noncoherent Space–Frequency Coded MIMO-OFDM

Moritz Borgmann, *Student Member, IEEE,* and Helmut Bölcskei, *Senior Member, IEEE*

***Abstract*—Recently, the use of *coherent* space–frequency coding in orthogonal frequency-division multiplexing (OFDM)-based frequency-selective multiple-input multiple-output (MIMO) fading channels has been proposed. Acquiring knowledge of the fading coefficients in a MIMO channel is already very challenging in the frequency-flat (fast) fading case. In the frequency-selective case, this task becomes significantly more difficult due to the presence of multiple paths, which results in an increased number of parameters to be estimated. In this paper, we address code design for *noncoherent frequency-selective MIMO-OFDM fading links*, where neither the transmitter nor the receiver knows the channel. We derive the *code design criteria*, quantify the maximum achievable diversity gain, and provide *explicit constructions of full-diversity (space and frequency) achieving codes* along with an analytical and numerical performance assessment. We also demonstrate that unlike in the coherent case, noncoherent space–frequency codes designed to achieve full spatial diversity in the frequency-flat fading case can fail completely to exploit not only frequency diversity but also spatial diversity when used in frequency-selective fading environments. We term such codes "catastrophic."**

***Index Terms*—Diversity, frequency-selective fading channels, multiple-input multiple-output (MIMO) systems, noncoherent communications, orthogonal frequency-division multiplexing (OFDM), pairwise error probability (PEP), Rayleigh fading, space–frequency coding.**

I. Introduction and Outline

SPACE–TIME coding has emerged as a promising technique for realizing spatial diversity gains in multiple-input–multiple-output (MIMO) wireless systems. The design criteria for space–time codes (STCs) in the *coherent* case, where the receiver has perfect channel state information (CSI), were derived in [1] and [2].

While the coherent setup is representative for fixed or low mobility wireless systems, future mobile wireless access systems are expected to operate at high vehicle speeds and will, hence, experience fast fading. Estimating the MIMO channel in a fast-fading environment is very challenging and often not possible at all. Addressing this problem, *noncoherent communication*, where neither the transmitter nor the receiver has CSI, was considered recently for the frequency-flat fading case in [3]–[6]. The design criteria for *space–time unitary codebooks* were derived in [4]. A simple systematic code construction was given in [5]. Different code designs were proposed, among others, in [7]–[11]. The decoding rule for unknown deterministic narrowband channels and unknown noise covariance was derived in [12].

Departing from the narrowband channel model, various researchers have recently considered the use of MIMO systems in frequency-selective fading channels in combination with *orthogonal frequency-division multiplexing* (OFDM) and coherent decoding [13]–[19]. Frequency-selective fading channels offer both spatial and frequency diversity. Exploiting these two sources of diversity jointly requires the design of *space–frequency codes* (SFCs), which essentially distribute the transmitted data symbols across space and frequency. Code design criteria for the coherent frequency-selective fading MIMO case were derived in [13], [14], and specific code designs were presented, among others, in [19].

In the frequency-selective case, estimating the channel coefficients is significantly more challenging than in the frequency-flat fading case due to the presence of multiple paths, which requires the estimation of an increased number of fading coefficients. The problem of noncoherent MIMO-OFDM communication was treated in [20], where the design criteria for noncoherent SFCs were given, and systematic code constructions were proposed. The design criteria for the suboptimal generalized likelihood ratio test (GLRT) decoder as well as results of a numerical code search were presented in [21]. A special matrix modulation scheme, with an efficient decoding algorithm based on projections onto convex sets (POCS) type ideas, was considered in [22].

A. Contributions

Throughout this paper, we consider a strategy that consists of coding across antennas and OFDM tones (*space–frequency coding*) within a single OFDM symbol. We emphasize that our focus is on the fundamental aspects of noncoherent MIMO-OFDM communication; differential coding schemes will not be considered in this paper. Our contributions can be summarized as follows.

- We derive the *maximum-likelihood (ML) decoding rule* for noncoherent SFCs. Based on this result, the code design criteria in terms of *pairwise error probability* (PEP) are provided. The resulting design rules are found to be markedly different from the well-known Hochwald–Marzetta criteria for the frequency-flat fading case [4].
- We prove that the maximum achievable diversity order is given by $M_T M_R L$ with M_T and M_R denoting the number of transmit and receive antennas, respectively, and L denoting the length of the complex baseband discrete-time channel impulse response. We furthermore show that employing known noncoherent STCs as SFCs

Manuscript received April 1, 2004; revised February 14, 2005. This work was supported in part by Nokia Research Center Helsinki and in part by ETH under Grant TH-6 02-2. This paper was presented in part at the Allerton Conference on Communication, Control, and Computing, Monticello, IL, October 2002.

The authors are with the Communication Technology Laboratory, Swiss Federal Institute of Technology (ETH) Zurich, 8092 Zurich, Switzerland (e-mail: moriborg@nari.ee.ethz.ch; boelcskei@nari.ee.ethz.ch).

Digital Object Identifier 10.1109/JSAC.2005.853800

by coding across frequency rather than across time will in general not exploit the additionally available frequency diversity. Moreover, we demonstrate the existence of *catastrophic noncoherent codes*, which are noncoherent STCs designed to achieve full spatial diversity in the frequency-flat fading case that fail to exploit *both* spatial *and* frequency diversity in the frequency-selective fading case. This behavior is in stark contrast to the coherent case, where a code designed for a certain diversity order in the frequency-flat fading case is guaranteed to exhibit at least the same diversity order in the frequency-selective fading case [13].

- We present *explicit systematic constructions of full (i.e., spatial and frequency) diversity-achieving* noncoherent SFCs and assess the performance of the proposed codes through a numerical analysis.
- We analyze the impact of delay spread and power delay profile (PDP) on noncoherent SFC performance from an error probability point-of-view and show that unlike in the coherent case, the presence of additional channel taps *does not* monotonically improve performance. Instead, we observe the notion of an *optimum diversity order*, an effect previously found in single-antenna wideband fading channels [23], [24], and recently also reported for the MIMO case from a capacity point-of-view [25].

B. Organization of This Paper

This paper is organized as follows. In Section II, we introduce the signal and channel model. In Section III, we derive the design criteria for noncoherent SFCs and establish corresponding performance bounds. Section IV presents our systematic code constructions and a numerical performance analysis. In Section V, we finally discuss the impact of delay spread and PDP on noncoherent SFC performance.

Notation: Matrices are generally set in boldface capital letters, vectors in boldface lowercase letters. The superscripts ^{T,*,H} stand for transpose, element-wise conjugation, and conjugate transpose, respectively. We write a_i for the ith entry of the $N\times 1$ vector $\mathbf{a}=[a_0\, a_1\, \cdots\, a_{N-1}]^T$. $[\mathbf{A}]_{i,j}$ represents the element in the ith row and jth column of the matrix $\mathbf{A}$. Indices generally start counting at zero. $\mathbf{I}_N$ and $\mathbf{0}_N$ denote the $N\times N$ identity and all zeros matrix, respectively. The notation $\operatorname{diag}_{i=0}^{N-1}\{a_i\}$ stands for the $N\times N$ diagonal matrix with diagonal entries a_i, $i=0,1,\ldots,N-1$. $\|\mathbf{A}\|_{\mathrm{F}}^2=\sum_{i,j}|[\mathbf{A}]_{i,j}|^2$ is the squared Frobenius norm of the matrix $\mathbf{A}$, $\mathbf{A}\otimes\mathbf{B}$ denotes the Kronecker product of the matrices $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{A}^\dagger$ stands for the pseudoinverse of $\mathbf{A}$, $\operatorname{tr}(\mathbf{A})$ denotes the trace of $\mathbf{A}$, and $\lambda_i(\mathbf{A})$ for $i=0,1,\ldots,N-1$ are the N eigenvalues of the $N\times N$ matrix $\mathbf{A}$ sorted in nondecreasing order. Slightly abusing common terminology, we call any $N\times M$ matrix where $N\geq M$ that satisfies $\mathbf{A}^H\mathbf{A}=\mathbf{I}_M$ *unitary*.

The probability of an event is denoted by $\mathrm{P}(\cdot)$, $\mathcal{E}\{\cdot\}$ stands for the expectation operator. An N-variate circularly symmetric zero-mean complex Gaussian random vector is a random vector $\mathbf{z}=\mathbf{x}+j\mathbf{y}\sim\mathcal{CN}_N(\mathbf{0},\mathbf{\Sigma})$, where the real-valued random vectors $\mathbf{x}$ and $\mathbf{y}$ are jointly Gaussian, $\mathcal{E}\{\mathbf{z}\}=\mathbf{0}$, $\mathcal{E}\{\mathbf{z}\mathbf{z}^H\}=\mathbf{\Sigma}$, and $\mathcal{E}\{\mathbf{z}\mathbf{z}^T\}=\mathbf{0}$. For scalar random variables, we omit the subscript and simply write $\mathcal{CN}(0,\sigma^2)$. The discrete Kronecker Delta function $\delta_{i,j}$ is 1 if $i=j$ and 0 else.

For real-valued nonnegative functions $f(t)$ and $\varphi(t)$, we write $f(t)=\mathcal{O}(\varphi(t))$ if $f(t)<A\varphi(t)$ as $t\to\infty$ with some constant A. The notation $f(t)=\Theta(\varphi(t))$ denotes that $A'\varphi(t)<f(t)<A\varphi(t)$ as $t\to\infty$ with some constants A' and A.

II. System Model

A. Discrete-Time Channel Model

In the following, M_T and M_R denote the number of transmit and receive antennas, respectively. We assume that the channel consists of L matrix-valued taps $\mathbf{H}_l$ of dimension $M_R\times M_T$, $l=0,1,\ldots,L-1$, with the matrix-valued transfer function given by

$$\mathbf{H}(e^{j2\pi\theta})=\sum_{l=0}^{L-1}\mathbf{H}_l e^{-j2\pi l\theta},\qquad 0\leq\theta<1. \tag{1}$$

Note that in general there will be a continuum of delays. The channel model (1) is derived from the assumption of having L resolvable paths, where $L=\lfloor B\tau\rfloor+1$ with B and τ denoting the signal bandwidth and delay spread, respectively.

For the sake of simplicity, we restrict ourselves to purely Rayleigh-fading channels. The elements of the $\mathbf{H}_l$ $(l=0,1,\ldots,L-1)$ are circularly symmetric zero-mean uncorrelated complex Gaussian random variables with variance σ_l^2. Moreover, we assume uncorrelated taps $\mathbf{H}_l$. Unless otherwise noted, the receiver will assume that the channel is Gaussian with known PDP (Bayesian approach).

B. Space–Frequency Coded MIMO-OFDM Systems

In an OFDM-based MIMO system, the individual transmit signals corresponding to the M_T transmit antennas are OFDM-modulated before transmission. In the following, N denotes the number of OFDM tones. The OFDM modulator applies an N-point inverse fast Fourier transform (IFFT) to N consecutive data symbols, and then prepends the cyclic prefix of length $L_{\mathrm{CP}}\geq L$ to the parallel-to-serial-converted OFDM symbol. The receiver first discards the cyclic prefix, and then applies an N-point FFT to each of the M_R received signals.

Organizing the transmitted data symbols into vectors $\mathbf{c}_k=[c_{k,0}\; c_{k,1}\; \cdots\; c_{k,M_T-1}]^T$ $(k=0,1,\ldots,N-1)$, with $c_{k,m}$ denoting the data symbol transmitted from the mth antenna on the kth tone, the received signal vectors are given by

$$\mathbf{r}_k=\sqrt{E_s}\mathbf{H}\big(e^{j2\pi\frac{k}{N}}\big)\mathbf{c}_k+\mathbf{w}_k,\qquad k=0,1,\ldots,N-1$$

where E_s is an energy scaling factor, and $\mathbf{w}_k$ is complex-valued circularly symmetric white Gaussian zero-mean noise satisfying $\mathcal{E}\{\mathbf{w}_k\mathbf{w}_{k'}^H\}=\mathbf{I}_{M_R}\delta_{k,k'}$. The transmission of an entire OFDM symbol can now be written as

$$\mathbf{R}=\mathbf{Y}+\mathbf{W} \tag{2}$$

where

$$\mathbf{Y}=\big[\mathbf{y}_0\;\mathbf{y}_1\;\cdots\;\mathbf{y}_{N-1}\big]^T$$
$$\text{with}\quad \mathbf{y}_k=\sqrt{E_s}\mathbf{H}\big(e^{j2\pi\frac{k}{N}}\big)\mathbf{c}_k. \tag{3}$$

Furthermore, $\mathbf{R} = [\mathbf{r}_0\ \mathbf{r}_1\ \cdots\ \mathbf{r}_{N-1}]^T$ is the $N \times M_R$ matrix of received signals, and $\mathbf{W} = [\mathbf{w}_0\ \mathbf{w}_1\ \cdots\ \mathbf{w}_{N-1}]^T$ is the $N \times M_R$ additive noise matrix. The transmitted signal is arranged in the $N \times M_T$ matrix $\mathbf{C} = [\mathbf{c}_0\ \mathbf{c}_1\ \cdots\ \mathbf{c}_{N-1}]^T$.

Throughout this paper, we assume that the frequency-selective channel remains constant for the duration of one OFDM symbol. Our analysis does not assume coding across OFDM symbols. We emphasize furthermore that due to the presence of multipath delay spread, the individual channel gain matrices $\mathbf{H}(e^{j2\pi k/N})$ for $k = 0, 1, \ldots, N-1$ will be correlated. Note that in the narrowband case considered in [3]–[11], where coding is performed across space and time rather than across space and frequency, the channel is flat in the sense that the channel gain matrix $\mathbf{H}(\cdot)$ remains constant within the entire bandwidth used. In this paper, we explicitly exploit the *structure* of the channel correlation across OFDM tones (frequency).

III. Noncoherent Space–Frequency Codes (SFCs): Design Criteria and Performance Bounds

In this section, we derive the design criteria for noncoherent SFCs and compute ultimate performance limits by quantifying the maximum achievable diversity order and coding gain.

A. Conditional Received Signal Density

Assuming that codeword $\mathbf{C}$ was transmitted, we write (3) in terms of the matrix-valued channel impulse response taps $\mathbf{H}_l$ $(l = 0, 1, \ldots, L-1)$ as

$$\mathbf{Y} = \sqrt{E_s} \sum_{l=0}^{L-1} \mathbf{D}^l \mathbf{C} \mathbf{H}_l^T = \sqrt{E_s} \mathbf{E}(\mathbf{C}) \bar{\mathbf{H}} \tag{4}$$

with the $N \times N$ diagonal matrix $\mathbf{D} = \operatorname{diag}_{k=0}^{N-1}\{e^{-j2\pi k/N}\}$, the stacked $N \times M_T L$ matrix

$$\mathbf{E}(\mathbf{C}) = [\mathbf{C} \quad \mathbf{DC} \quad \cdots \quad \mathbf{D}^{L-1}\mathbf{C}]$$

and the stacked matrix $\bar{\mathbf{H}} = [\mathbf{H}_0\ \mathbf{H}_1\ \ldots\ \mathbf{H}_{L-1}]^T$ of size $M_T L \times M_R$. For reasons that will become clear later, we refer to the stacked matrices $\mathbf{E}(\mathbf{C})$ as *pseudocodeword matrices*. Since the columns of $\mathbf{Y}$ are independent identically distributed (i.i.d.) zero-mean circularly symmetric complex Gaussian random vectors, it is sufficient to consider only a single arbitrary column $\mathbf{y} = \sqrt{E_s}\mathbf{E}(\mathbf{C})\bar{\mathbf{h}}$ of $\mathbf{Y}$, where $\bar{\mathbf{h}}$ denotes the corresponding column of $\bar{\mathbf{H}}$ and has covariance matrix $\mathbf{\Sigma}^2 = \mathcal{E}\{\bar{\mathbf{h}}\bar{\mathbf{h}}^H\} = \operatorname{diag}_{l=0}^{L-1}\{\sigma_l^2\} \otimes \mathbf{I}_{M_T}$. Likewise, the columns of $\mathbf{R}$ are i.i.d. with the covariance matrix of an arbitrary column $\bar{\mathbf{r}}_n$ $(n = 0, 1, \ldots, M_R - 1)$ (note that we have previously denoted *rows* of $\mathbf{R}$ by $\mathbf{r}_k$) given by

$$\mathbf{\Lambda}(\mathbf{C}) = \mathbf{I}_N + E_s \mathbf{E}(\mathbf{C}) \mathbf{\Sigma}^2 \mathbf{E}^H(\mathbf{C})$$

which yields the conditional received signal density, over all receive antennas, as

$$p(\mathbf{R}|\mathbf{C}) = \frac{\exp\left(-\operatorname{tr}\left(\mathbf{R}^H \mathbf{\Lambda}^{-1}(\mathbf{C}) \mathbf{R}\right)\right)}{\pi^{NM_R} \left(\det \mathbf{\Lambda}(\mathbf{C})\right)^{M_R}}. \tag{5}$$

B. Maximum-Likelihood (ML) Decoding

If we assume a constellation of K codeword matrices $\mathbf{C}_i \in \mathcal{C}$ (with $\mathcal{C} = \{\mathbf{C}_0, \mathbf{C}_1, \ldots, \mathbf{C}_{K-1}\}$) and abbreviate $\mathbf{\Lambda}(\mathbf{C}_i) = \mathbf{\Lambda}_i$, the ML decoding rule follows from (5) as

$$\hat{\mathbf{C}}_{\mathrm{ML}} = \arg\min_{\mathbf{C}_i \in \mathcal{C}} \left(\operatorname{tr}\left(\mathbf{R}^H \mathbf{\Lambda}_i^{-1} \mathbf{R}\right) + M_R \ln\det \mathbf{\Lambda}_i \right).$$

Furthermore, we assume from now on that $N \geq M_T L$ and that the codeword matrices $\mathbf{C}_i$ are such that the corresponding pseudocodeword matrices $\mathbf{E}_i = \mathbf{E}(\mathbf{C}_i)$ satisfy $\mathbf{E}_i^H \mathbf{E}_i = \mathbf{I}_{M_T L}$ $(i = 0, 1, \ldots, K-1)$. This type of constellation will henceforth be referred to as a *space–frequency unitary codebook* or *constellation*. We note that unitarity of the $\mathbf{E}_i$ has to be ensured explicitly by code design. Examples of such constructions will be given in Section IV. We emphasize that the restriction to space–frequency unitary codebooks is conceptual, allowing us to obtain analytically tractable results. However, since noncoherent codes need to implicitly train the channel, the use of unitary space–frequency constellations can additionally be motivated by the fact that space–frequency unitary signals, which correspond to shift-orthogonal sequences in the time domain, constitute optimum training sequences [26]. In the frequency-flat fading case, isotropically distributed unitary signals were shown to be capacity-achieving for $N \gg M_T$ in [3]. In the frequency-selective fading case, space–frequency unitary codebooks have recently been shown to perform well in terms of achievable rate [25]. Despite the fact that the capacity of the noncoherent frequency-selective MIMO fading channel as well as the properties of capacity-achieving signals are not known, the evidence collected above suggests that starting the analysis of the noncoherent frequency-selective MIMO fading channel under the assumption of space–frequency unitary constellations is sensible.

Using the space–frequency unitarity of the codebook, analogously to the technique employed by Hochwald and Marzetta in [4], we apply the matrix inversion lemma to $\mathbf{\Lambda}_i^{-1}$ and note that $\ln\det \mathbf{\Lambda}_i$ is a constant independent of i. The ML decoding rule therefore simplifies to

$$\hat{\mathbf{C}}_{\mathrm{ML}} = \arg\max_{\mathbf{C}_i \in \mathcal{C}} \operatorname{tr}\left(\mathbf{R}^H \mathbf{E}_i \mathbf{\Sigma}^2 \left(\mathbf{I}_{M_T L} + E_s \mathbf{\Sigma}^2\right)^{-1} \mathbf{E}_i^H \mathbf{R}\right) \tag{6}$$

$$= \arg\max_{\mathbf{C}_i \in \mathcal{C}} \sum_{n=1}^{M_R} \bar{\mathbf{r}}_n^H \left(\sum_{l=0}^{L-1} \frac{\sigma_l^2}{1 + E_s \sigma_l^2} \mathbf{D}^l \mathbf{C}_i \mathbf{C}_i^H \mathbf{D}^{-l} \right) \bar{\mathbf{r}}_n. \tag{7}$$

Equation (7) shows that ML decoding requires knowledge of the σ_l^2 $(l = 0, 1, \ldots, L-1)$. For the special case of a uniform PDP, i.e., $\sigma_l^2 = const$ for $l = 0, 1, \ldots, L-1$, the ML decoding rule simplifies further to

$$\hat{\mathbf{C}}_{\mathrm{ML}} = \arg\max_{\mathbf{C}_i \in \mathcal{C}} \operatorname{tr}\left(\mathbf{R}^H \mathbf{E}_i \mathbf{E}_i^H \mathbf{R}\right). \tag{8}$$

C. Generalized Likelihood Ratio Test (GLRT) Decoding

In scenarios where the receiver does not have knowledge of the PDP, ML decoding [cf. (7)] is not applicable. In this case, one can resort to a non-Bayesian approach, such as the

GLRT [21], which uses the decision rule for the *coherent* decoder with the unknown channel replaced by its ML estimate assuming that the ith codeword was sent. In our case, this ML channel estimate is obtained from (2) and (4) as $\hat{\mathbf{H}}_i = E_s^{-1/2}\mathbf{E}_i^{\dagger}\mathbf{R} = E_s^{-1/2}\mathbf{E}_i^H\mathbf{R}$, where $\mathbf{E}_i^H\mathbf{E}_i = \mathbf{I}_{M_TL}$ was used. The GLRT decoding rule follows after simple manipulations as:

$$\begin{aligned}\hat{\mathbf{C}}_{\mathrm{ML}} &= \arg\min_{\mathbf{C}_i\in\mathcal{C}} \left\|\mathbf{R} - \sqrt{E_s}\mathbf{E}_i\hat{\mathbf{H}}_i\right\|_{\mathrm{F}}^2 \\ &= \arg\min_{\mathbf{C}_i\in\mathcal{C}} \mathrm{tr}\left(\mathbf{R}^H\mathbf{R} - \mathbf{R}^H\mathbf{E}_i\mathbf{E}_i^H\mathbf{R}\right) \\ &= \arg\max_{\mathbf{C}_i\in\mathcal{C}} \mathrm{tr}\left(\mathbf{R}^H\mathbf{E}_i\mathbf{E}_i^H\mathbf{R}\right)\end{aligned}$$

which shows that GLRT decoding under *arbitrary PDP* is equivalent to ML decoding under uniform PDP [cf. (8)].

D. Pairwise Error Probability (PEP)

We shall next derive the PEP $\mathrm{P}(\mathbf{C}_i \to \mathbf{C}_j|\mathbf{C}_i)$ for ML decoding, i.e., the probability that the decoder decides for $\mathbf{C}_j$ instead of $\mathbf{C}_i$ in a binary hypothesis test, given that $\mathbf{C}_i$ was sent. Setting $\mathbf{T} = \mathbf{I}_{M_TL} + E_s\boldsymbol{\Sigma}^2$, the PEP follows from the ML decoding rule (6) as

$$\mathrm{P}(\mathbf{C}_i \to \mathbf{C}_j|\mathbf{C}_i) = \mathrm{P}\left(\sum_{n=0}^{M_R-1} v_n > 0 \;\middle|\; \mathbf{C} = \mathbf{C}_i\right) \tag{9}$$

with

$$v_n = \bar{\mathbf{r}}_n^H\left(\mathbf{E}_j\boldsymbol{\Sigma}^2\mathbf{T}^{-1}\mathbf{E}_j^H - \mathbf{E}_i\boldsymbol{\Sigma}^2\mathbf{T}^{-1}\mathbf{E}_i^H\right)\bar{\mathbf{r}}_n. \tag{10}$$

It is shown in Appendix A that the moment-generating function (MGF) of the random variables v_n for each $n = 0, 1, \ldots, M_R - 1$ is given by

$$\begin{aligned}\phi(s) &= \mathcal{E}\{e^{sv_n}|\mathbf{C} = \mathbf{C}_i\} \qquad (11)\\ &= \det{}^{-1}\left(\mathbf{I}_{M_TL} + s\boldsymbol{\Sigma}^2\right) \\ &\quad\cdot \det{}^{-1}\Big(\mathbf{I}_{M_TL} - s\boldsymbol{\Sigma}^2\mathbf{T}^{-1} - s(E_s - s) \\ &\qquad\cdot \boldsymbol{\Sigma}^2\mathbf{T}^{-1}\mathbf{E}_j^H\mathbf{E}_i\boldsymbol{\Sigma}^2\left(\mathbf{I}_{M_TL} + s\boldsymbol{\Sigma}^2\right)^{-1}\mathbf{E}_i^H\mathbf{E}_j\Big). \qquad (12)\end{aligned}$$

Denoting the region of convergence of the MGF $\phi(s)$ as $\mathcal{S}$, and recalling that the vectors $\bar{\mathbf{r}}_n$ are i.i.d., we can write the Chernoff bound on the PEP as [27]

$$\mathrm{P}(\mathbf{C}_i \to \mathbf{C}_j|\mathbf{C}_i) \le \min_{s\in\mathcal{S}, s>0} \phi^{M_R}(s). \tag{13}$$

In the general case, the optimization over s in (13) is analytically intractable. However, in the following two relevant special cases, we are able to obtain an explicit expression for the minimizing value $s^\star$.

1) Uniform Power Delay Profile (PDP) Case: Although a uniform PDP is not very likely to be encountered in practice, the motivation for the analysis of this case is twofold. First of all, we can develop a basic understanding of the behavior of SFCs in frequency-selective environments. Second, as shown in Section III-C, the GLRT decoder for arbitrary PDP is equivalent to the ML decoder for uniform PDP. In the following, without loss of generality (w.l.o.g.), we set $\sigma_l^2 = 1$ for $l = 0, 1, \ldots, L-1$, i.e., $\boldsymbol{\Sigma}^2 = \mathbf{I}_{M_TL}$. The MGF (12) then simplifies substantially to

$$\phi(s) = \prod_{r=0}^{M_TL-1}\left(1 + \frac{s(E_s - s)}{1 + E_s}\left(1 - d_r^2(i,j)\right)\right)^{-1} \tag{14}$$

where the $d_r(i,j)$ $(r = 0, 1, \ldots, M_TL - 1)$ are the singular values of the matrix

$$\mathbf{E}_j^H\mathbf{E}_i = \begin{bmatrix} \mathbf{C}_j^H\mathbf{C}_i & \mathbf{C}_j^H\mathbf{D}\mathbf{C}_i & \cdots & \mathbf{C}_j^H\mathbf{D}^{L-1}\mathbf{C}_i \\ \mathbf{C}_j^H\mathbf{D}^{-1}\mathbf{C}_i & \mathbf{C}_j^H\mathbf{C}_i & \cdots & \mathbf{C}_j^H\mathbf{D}^{L-2}\mathbf{C}_i \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{C}_j^H\mathbf{D}^{-L+1}\mathbf{C}_i & \mathbf{C}_j^H\mathbf{D}^{-L+2}\mathbf{C}_i & \cdots & \mathbf{C}_j^H\mathbf{C}_i \end{bmatrix}.$$

Each of the M_TL factors in the product (14) attains its minimum at $s^\star = E_s/2$, so that the Chernoff bound reads

$$\mathrm{P}(\mathbf{C}_i \to \mathbf{C}_j|\mathbf{C}_i) \le \prod_{r=0}^{M_TL-1}\left(1 + \frac{E_s^2}{4(1+E_s)}\left(1 - d_r^2(i,j)\right)\right)^{-M_R}. \tag{15}$$

For fixed E_s, the PEP upper bound (15) is minimized and maximized, respectively, for $d_r(i,j) = 0$ and $d_r(i,j) = 1$ $(r = 0, 1, \ldots, M_TL - 1)$. The ideal (in terms of PEP) space–frequency constellation, therefore, has all the columns of $\mathbf{E}_i$ orthogonal to all the columns of $\mathbf{E}_j$ for $i \neq j$. We call such a constellation a *mutually space–frequency orthogonal codebook*.

2) Mutually Space–Frequency Orthogonal Codebook Case: Next, we consider an arbitrary PDP and assume a mutually space–frequency orthogonal codebook, i.e., $\mathbf{E}_j^H\mathbf{E}_i = \delta_{i,j}\mathbf{I}_{M_TL}$. It is easy to verify that mutual space–frequency orthogonality is possible only if $K \le N/(M_TL)$. Satisfying this condition in practice will yield very low code rates and result in a signaling scheme that is analogous to orthogonal schemes for single-antenna transmission such as frequency-shift keying (FSK) or pulse position modulation (PPM), traditionally used in the power-limited regime and known to achieve capacity in the wideband limit [28]. Although mutually space–frequency orthogonal constellations achieve extremely low rate, their use can be motivated through the fact that just as their single-antenna counterparts, they achieve capacity in the wideband limit, where the number of degrees of freedom is infinite, but power is limited [29].

Using $\mathbf{E}_j^H\mathbf{E}_i = \delta_{i,j}\mathbf{I}_{M_TL}$ in (12), the MGF simplifies to

$$\phi(s) = \det{}^{-1}\left(\mathbf{I}_{M_TL} + s\boldsymbol{\Sigma}^2\right)\det{}^{-1}\left(\mathbf{I}_{M_TL} - s\boldsymbol{\Sigma}^2\mathbf{T}^{-1}\right).$$

The minimizing value of s is again found to be $s^\star = E_s/2$, so that

$$\mathrm{P}(\mathbf{C}_i \to \mathbf{C}_j|\mathbf{C}_i) \le \prod_{l=0}^{L-1}\left(1 + \frac{E_s^2\sigma_l^4}{4\left(1 + E_s\sigma_l^2\right)}\right)^{-M_TM_R}. \tag{16}$$

Note that (16) makes the impact of the PDP on error probability explicit. A systematic study of SFC performance as a function of the PDP will be conducted in Section V.

E. Asymptotic PEP Bound

As pointed out above, the Chernoff bound becomes intractable for general PDP and general space–frequency unitary codebooks. The general case can be dealt with, however, by

establishing an analytical upper bound on the Chernoff bound. We note that our approach presented below differs from [30] and [31], as our bound on the Chernoff bound is based on purely algebraic methods. Although our new technique yields looser bounds than those that could be obtained by applying the techniques in [30] and [31], our results still reflect the fundamental behavior of the PEP correctly (this statement will be made more precise below).

In the following, we have to distinguish two cases. The *full-diversity case*, where all the singular values of $\mathbf{E}_j^H\mathbf{E}_i$ satisfy $d_r(i,j) < 1$. We will show later in Section III-F that in this case, the underlying SFC achieves full space–frequency diversity. The *rank-deficient case* is obtained when at least one of the $d_r(i,j) = 1$. Denoting the number of singular values of $\mathbf{E}_j^H\mathbf{E}_i$ that satisfy $d_r(i,j) < 1$ by $R_{i,j}$, we obtain the following bound for the rank-deficient case:

$$\mathrm{P}(\mathbf{C}_i \to \mathbf{C}_j|\mathbf{C}_i) < \left(\frac{E_s}{b}\right)^{-R_{i,j}M_R} \cdot \left(\prod_{r=0}^{R_{i,j}-1} \lambda_r\left(\mathbf{\Sigma}^2\left(\mathbf{I}_{M_TL} - \mathbf{E}_j^H\mathbf{E}_i\mathbf{E}_i^H\mathbf{E}_j\right)\right)\right)^{-M_R} \quad (17)$$

with

$$b = \frac{(R_{i,j} + M_TL)^{1+\frac{M_TL}{R_{i,j}}}}{(M_TL)^{\frac{M_TL}{R_{i,j}}} R_{i,j}}.$$

In the full-diversity case, the bound simplifies to

$$\mathrm{P}(\mathbf{C}_i \to \mathbf{C}_j|\mathbf{C}_i) < \left(\frac{E_s}{4}\right)^{-M_TM_RL} \left(\prod_{l=0}^{L-1}\sigma_l^2\right)^{-M_TM_R} \cdot \left(\prod_{r=0}^{M_TL-1}\left(1 - d_r^2(i,j)\right)\right)^{-M_R}. \quad (18)$$

For a derivation of (17) and (18), the interested reader is referred to Appendix B, which also shows that in the full-diversity case, our bound is asymptotically (as $E_s \to \infty$) tight to the exact Chernoff bound on a double-logarithmic scale. On the other hand, in the rank-deficient case, our bound exhibits an offset on a double-logarithmic scale, but still the same high-SNR slope as the Chernoff bound. This implies that in the rank-deficient case, we can infer the diversity gain [1] from (17), but not the coding gain.

F. Diversity Order and Coding Gain

Based on the results in Section III-E, we are now in a position to establish the achievable diversity order and coding gain of noncoherent SFCs. It follows from (17) that in the most general case (arbitrary PDP and no assumptions on the SFC except for unitarity of pseudocodewords), the PEP behaves as $\mathrm{P}(\mathbf{C}_i \to \mathbf{C}_j|\mathbf{C}_i) = \mathcal{O}(E_s^{-R_{i,j}M_R})$ (note that we show in Appendix B that (17) correctly reflects the high-SNR slope of the Chernoff bound in all cases). Setting $R = \min_{0\le i<j\le K-1} R_{i,j}$, we conclude that the diversity order [1] achieved by the code is given by RM_R. Since $R_{i,j} \le M_TL$, the maximum achievable diversity order is given by M_TM_RL, which demonstrates that noncoherent SFCs can potentially achieve the same maximum diversity order as coherent SFCs [13]. Codes achieving full diversity order will be presented in Section IV. Since (17) is not asymptotically (as $E_s \to \infty$) tight to the exact Chernoff bound, it is not possible to infer the coding gain (which is usually defined based on the Chernoff bound [1]) from (17). However, for the uniform PDP case, the coding gain follows from (15) as the minimum (taken over all distinct pairs of codewords that satisfy $R_{i,j} = R$) of

$$\left(\prod_{r=0}^{R_{i,j}-1}\left(1 - d_r^2(i,j)\right)\right)^{1/R_{i,j}} \quad (19)$$

and for the full diversity, general PDP, case from (18) as the minimum of

$$\left(\prod_{l=0}^{L-1}\sigma_l^2\right)^{1/L} \left(\prod_{r=0}^{M_TL-1}\left(1 - d_r^2(i,j)\right)\right)^{1/(M_TL)}. \quad (20)$$

The latter case includes mutually orthogonal space–frequency codebooks, where $d_r(i,j) = 0$ for all $r = 0, 1, \ldots, M_TL-1$. In both cases, the coding gain is maximized if $d_r(i,j) = 0$ for $r = 0, 1, \ldots, R_{i,j} - 1$ for all distinct pairs of codewords with $R_{i,j} = R$. It can be shown that a code with $N < 2M_TL$ cannot achieve full diversity gain; this result parallels [9], where it was shown that in the frequency-flat $(L = 1)$ fading case, a full-diversity achieving code must satisfy $N \ge 2M_T$. Equation (20) furthermore shows that for fixed $\sum_{l=0}^{L-1}\sigma_l^2$, the coding gain is maximized for a uniform PDP.

G. PEP Time-Domain Formulation

We shall next provide a time-domain interpretation of the PEP results above, which provides additional insight into the fundamental difference between noncoherent communication over frequency-flat fading and wideband channels, respectively. Set $\mathbf{C} = \mathbf{F}\mathbf{C}_t$ with $[\mathbf{F}]_{m,n} = (1/\sqrt{N})e^{-j2\pi mn/N}$ denoting the $N \times N$ FFT matrix, and $\mathbf{C}_t$ denoting the time-domain version of $\mathbf{C}$. Next, note that $\mathbf{F}^H\mathbf{D}^l\mathbf{F}\mathbf{C}_t = \mathbf{C}_{t-l}$, where $\mathbf{C}_{t-l}$ denotes the matrix obtained by cyclically shifting the columns of $\mathbf{C}_t$ by l positions downwards. It then follows that:

$$\mathbf{E}_j^H\mathbf{E}_i = \mathbf{E}_{j,t}^H\mathbf{E}_{i,t}$$

with

$$\mathbf{E}_{i,t} = [\mathbf{C}_{i,t} \quad \mathbf{C}_{i,t-1} \quad \cdots \quad \mathbf{C}_{i,t-L+1}]. \quad (21)$$

Due to this unitary equivalence, we can replace $\mathbf{E}_j^H\mathbf{E}_i$ by the time-domain version $\mathbf{E}_{j,t}^H\mathbf{E}_{i,t}$ throughout this paper and obtain an interpretation of the additional taps inducing "virtual antennas," which transmit delayed versions of $\mathbf{C}_{i,t}$ [cf. (21)].

H. Code Design Criteria

Noncoherent SFC design can be carried out in essentially the same fashion as coherent STC design in [1]. Based on the results established so far, the procedure amounts to: 1) ensuring that for a given target diversity order D all pairs of distinct codewords satisfy $R_{i,j}M_R \ge D$ and 2) to maximizing the coding gain (19) or (20) over all codeword pairs satisfying $R_{i,j}M_R = D$. The second criterion is applicable only to the uniform PDP case and

the full-diversity, general PDP case. In the rank-deficient, general PDP case, we do not have an expression for the coding gain as the corresponding PEP bound (17) is loose (cf. Section III-E).

In the following, we shall discuss alternative pragmatic design criteria. We start by defining the *diversity product*

$$\gamma = \min_{0\le i<j\le K-1} \prod_{r=0}^{M_T L-1} \left(1 - d_r^2(i,j)\right). \tag{22}$$

Since codebooks with $\gamma > 0$ trivially achieve full space–frequency diversity, maximizing γ ensures full diversity gain and maximizes the coding gain, independent of the PDP.

In [5], a simplified design criterion for the frequency-flat fading case is derived, which aims at minimizing the so-called *correlation* $\delta = \max_{0\le i<j\le K-1} \|\mathbf{C}_j^H \mathbf{C}_i\|_{\mathrm{F}}^2$. In the frequency-selective case (uniform PDP), this simplified criterion generalizes to a minimization of the corresponding metric in terms of the pseudocodewords

$$\delta = \max_{0\le i<j\le K-1} \left\|\mathbf{E}_j^H \mathbf{E}_i\right\|_{\mathrm{F}}^2. \tag{23}$$

Similarly to [5], we can interpret $\|\mathbf{E}_j^H \mathbf{E}_i\|_{\mathrm{F}}^2$ as a measure related to the *chordal distance* between the "induced" subspaces spanned by the columns of $\mathbf{E}_j$ and $\mathbf{E}_i$ or, equivalently, $\mathbf{E}_{j,t}$ and $\mathbf{E}_{i,t}$, which consist of the codewords along with their cyclically shifted versions [cf. (21)]. In contrast, in the frequency-flat fading case treated in [5], only the "distance" between the subspaces spanned by the columns of the codewords $\mathbf{C}_j$ and $\mathbf{C}_i$ themselves determines δ. In terms of packings in Grassmannian manifolds [8], the design problem considered in this paper results in devising packings of subspaces induced by the individual codewords along with their cyclically shifted versions. Analogously to [5], the correlation δ can be shown to have operational meaning as it governs a certain bound on the block error rate for the entire constellation, cf. [5, eq. (12)]. As a drawback of this particular criterion, it must be noted that although it is generally representative of the code performance, minimization of δ does not always lead to full-diversity codes, as exemplified in [7].

It is instructive to consider a simple example with $M_T = 1$ and $L = 2$, where (in terms of the equivalent time-domain code vectors)

$$\left\|\mathbf{E}_{j,t}^H \mathbf{E}_{i,t}\right\|_{\mathrm{F}}^2 = 2\left|\mathbf{C}_{j,t}^H \mathbf{C}_{i,t}\right|^2 + \left|\mathbf{C}_{j,t}^H \mathbf{C}_{i,t-1}\right|^2 + \left|\mathbf{C}_{j,t-1}^H \mathbf{C}_{i,t}\right|^2.$$

We can see that in the frequency-selective fading case, minimizing the maximum of $|\mathbf{C}_{j,t}^H \mathbf{C}_{i,t}|$ alone (the relevant metric in the frequency-flat fading case) will not minimize δ defined in (23). The frequency-selective nature of the channel has to be taken into account *explicitly*. We conclude by noting that this statement holds true for other performance metrics such as the diversity product as well.

IV. Design of Noncoherent SFCs

In this section, we propose systematic SFC constructions that are *specifically tailored* to frequency-selective channels. We first strengthen the case for SFC design by demonstrating the existence of what we call *catastrophic codes*, i.e., codes designed for frequency-flat fading channels whose performance breaks down (in a sense to be made precise later) when used in a frequency-selective environment. We will then propose a class of codes that guarantees full space–frequency diversity and present a generalization thereof that provides additional degrees of freedom in the code design. The code design procedures outlined below assume that the tap delays are previously known and remain fixed.

A. Catastrophic Codes

Note that designing a code such that the singular values of $\mathbf{C}_j^H \mathbf{C}_i$ are small will in general not guarantee that the singular values of $\mathbf{E}_j^H \mathbf{E}_i$ are small as well. In other words, a code that performs well in the frequency-flat fading case (i.e., $L = 1$) will in general not exploit the additionally available frequency diversity. The situation can be even worse, as the following example illustrates. Assume we have a code for $M_T = M_R = 1$ and $L = 1$ such that all the code vectors $\mathbf{C}_i$ are mutually orthogonal, which minimizes the upper bound on PEP in the presence of frequency-flat fading. With this assumption, the PEP upper bound in (15) exhibits inversely proportional behavior to E_s at high SNR, i.e., the code achieves first-order diversity. Using this code in a scenario where $M_T = M_R = 1$ and $L = 2$ (i.e., frequency-selective fading) with uniform PDP satisfying $\sigma_0^2 = \sigma_1^2 = 1$, the PEP upper bound is determined by the singular values of (for the sake of clarity of exposition we work in the time domain)

$$\mathbf{E}_{j,t}^H \mathbf{E}_{i,t} = \begin{bmatrix} \mathbf{C}_{j,t}^H \mathbf{C}_{i,t} & \mathbf{C}_{j,t}^H \mathbf{C}_{i,t-1} \\ \mathbf{C}_{j,t-1}^H \mathbf{C}_{i,t} & \mathbf{C}_{j,t}^H \mathbf{C}_{i,t} \end{bmatrix} \tag{24}$$

where we exploited $\mathbf{C}_{j,t-1}^H \mathbf{C}_{i,t-1} = \mathbf{C}_{j,t}^H \mathbf{C}_{i,t}$. Since the $\mathbf{C}_i$ were assumed to be mutually orthogonal, (24) simplifies to

$$\mathbf{E}_{j,t}^H \mathbf{E}_{i,t} = \begin{bmatrix} 0 & \mathbf{C}_{j,t}^H \mathbf{C}_{i,t-1} \\ \mathbf{C}_{j,t-1}^H \mathbf{C}_{i,t} & 0 \end{bmatrix}.$$

The singular values of $\mathbf{E}_{j,t}^H \mathbf{E}_{i,t}$ are consequently given by $d_1 = |\mathbf{C}_{j,t}^H \mathbf{C}_{i,t-1}|$ and $d_2 = |\mathbf{C}_{j,t-1}^H \mathbf{C}_{i,t}|$, which, using (15), results in

$$\mathrm{P}(\mathbf{C}_i \to \mathbf{C}_j | \mathbf{C}_i) \le \left(1 + \frac{E_s}{4}\left(1 - \left|\mathbf{C}_{j,t}^H \mathbf{C}_{i,t-1}\right|^2\right)\right)^{-1} \cdot \left(1 + \frac{E_s}{4}\left(1 - \left|\mathbf{C}_{j,t-1}^H \mathbf{C}_{i,t}\right|^2\right)\right)^{-1}.$$

If $\mathbf{C}_{j,t}$ and $\mathbf{C}_{i,t}$ are such that $d_1 < 1$ and $d_2 < 1$, the PEP decays as E_s^{-2} when $E_s \to \infty$, and we achieve second-order diversity. If only one singular value is smaller than 1, the PEP is proportional to E_s^{-1} and, hence, we obtain the same diversity order as in the frequency-flat fading case. Finally, if $d_1 = d_2 = 1$, the diversity order is zero, i.e., the performance is worse than in the frequency-flat fading case. We call any code exhibiting reduced diversity order when employed in the frequency-selective case (as compared with the frequency-flat case) *catastrophic*.[1] A simple example of a catastrophic code-vector pair are the two-periodic vectors $\mathbf{C}_{i,t} = [0\ c\ 0\ c\ \cdots\ c]^T$ and $\mathbf{C}_{j,t} = \mathbf{C}_{i,t-1}$ with $|c| = \sqrt{2/N}$. It is easily verified that this choice implies $d_1 = d_2 = 1$. While we have shown the existence of catastrophic codes only for the simple example of

[1]Note that our use of the term *catastrophic* is not in the sense of catastrophic convolutional codes as described in [32].

mutually orthogonal codewords and $M_T = 1$, it is obvious that catastrophic behavior can also occur in the general codebook, multiantenna case, since nonoverlapping subspaces spanned by the codewords $\mathbf{C}_{j,t}$ and $\mathbf{C}_{i,t}$ can well lead to (at least partially) overlapping subspaces spanned by the pseudocodewords $\mathbf{E}_j$ and $\mathbf{E}_i$. Note that the existence of *one* catastrophic codeword pair is sufficient to degrade the performance of the whole codebook, as this pair is bound to dominate the error rate at high SNR.

The existence of catastrophic codes is striking as it shows a fundamental difference between the coherent and the noncoherent frequency-selective fading cases. In the *coherent case*, it was shown in [13] that a code designed to achieve a certain diversity order in the frequency-flat fading case achieves at least the same diversity order in the frequency-selective fading case. In other words, using a code designed for the frequency-flat fading case does not degrade diversity order in the frequency-selective fading case (the coding gain may still be smaller though). In the *noncoherent case*, codes designed to achieve a certain diversity order under frequency-flat fading conditions are not even guaranteed to achieve at least the same diversity order in a frequency-selective environment.

B. Simple Construction

We shall now present a very simple noncoherent SFC construction that is capable of exploiting both spatial and frequency diversity. Our design is heavily inspired by a construction first proposed in [5]. We emphasize, however, that the construction in [5] does not take into account frequency-selectivity and does, therefore, neither ensure that the $\mathbf{E}_i$ are unitary nor guarantee full space–frequency diversity. For the sake of simplicity of exposition, in the following, we consider a scenario with $M_T = 2$ transmit antennas and $L = 2$ taps. The extension to more antennas and/or taps is straightforward. The codeword matrices $\mathbf{C}_i$ $(i = 0, 1, \ldots, K-1)$ are chosen as

$$\mathbf{C}_i = \mathbf{\Phi}^i \begin{bmatrix} \mathbf{f}_p & \mathbf{f}_{p'} \end{bmatrix}, \qquad i = 0, 1, \ldots, K-1 \tag{25}$$

where $\mathbf{f}_p$ denotes the pth column of the $N \times N$ FFT matrix $\mathbf{F}$, and

$$\mathbf{\Phi} = \operatorname{diag}_{k=0}^{N-1} \left\{ e^{j\frac{2\pi}{K}u_k} \right\}$$

where w.l.o.g. $0 \leq u_k \leq K-1$ for $k = 0, 1, \ldots, N-1$. Our construction contains two key elements, namely, the choice of the parameters p and p' and the choice of the u_k $(k = 0, 1, \ldots, N-1)$. The parameters p and p' have to be chosen such that the stacked matrix

$$\mathbf{B}_{p,p'} = \begin{bmatrix} \mathbf{f}_p & \mathbf{f}_{p'} & \mathbf{D}\mathbf{f}_p & \mathbf{D}\mathbf{f}_{p'} \end{bmatrix} = \begin{bmatrix} \mathbf{f}_p & \mathbf{f}_{p'} & \mathbf{f}_{p+1} & \mathbf{f}_{p'+1} \end{bmatrix}$$

is unitary. Here, we exploited the fact that $\mathbf{D}\mathbf{f}_p = \mathbf{f}_{p+1}$. Noting that $\mathbf{D}$ and $\mathbf{\Phi}^i$ $(i = 0, 1, \ldots, K-1)$ commute, it then follows that choosing p and p' such that $\mathbf{B}_{p,p'}$ is unitary ensures that $\mathbf{E}_i$ is unitary. For the example under consideration, setting $p = 0$ and $p' = 2$, we obtain $\mathbf{B}_{p,p'} = [\mathbf{f}_0\ \mathbf{f}_2\ \mathbf{f}_1\ \mathbf{f}_3]$, which is simply a permutation of the first four columns of the FFT matrix $\mathbf{F}$, and hence $\mathbf{B}_{p,p'}^H \mathbf{B}_{p,p'} = \mathbf{I}_4$.

TABLE I
TABLE OF BEST FOUND $M_T = 1$ ANTENNA CONSTELLATIONS FOR $N = 8$ TONES, ACHIEVING DIVERSITY ORDER LM_R. THE NUMBER OF SIGNALS IN THE CONSTELLATION IS K. THE CODING GAIN IS DENOTED BY CG

K	L	CG	$[u_1, u_2, \ldots, u_7]$
8	1	1.0000	[0 4 2 6 3 5 7]
8	2	0.8461	[7 5 4 3 0 5 6]
8	3	0.6300	[2 4 7 3 6 4 3]
64	2	0.6283	[32 38 30 51 52 5 41]
256	2	0.4182	[272 254 145 252 55 195 137]

TABLE II
TABLE OF BEST FOUND $M_T = 2$ ANTENNA CONSTELLATIONS FOR $N = 8$ TONES, ACHIEVING DIVERSITY ORDER $2LM_R$, WITH $p = 0$ AND $p' = 2$. THE NUMBER OF SIGNALS IN THE CONSTELLATION IS K. THE CODING GAIN IS DENOTED BY CG

K	L	CG	$[u_1, u_2, \ldots, u_7]$
8	1	0.8364	[0 5 4 6 2 1 3]
8	2	0.5000	[0 3 4 1 0 3 4]
16	1	0.7603	[3 10 15 8 13 11 14]
16	2	0.3663	[4 3 0 1 8 3 12]
64	1	0.6261	[52 33 35 61 32 41 58]
64	2	0.2101	[2 31 61 4 58 10 15]
256	2	0.1051	[69 234 230 121 40 59 84]
512	2	0.0830	[75 42 476 326 49 353 66]

The parameters u_k in $\mathbf{\Phi}$ have to be chosen to optimize one of the design criteria discussed in Section III-H and can be determined by random search. We shall next provide some explicit full space–frequency diversity achieving designs. Owing to the special structure of our codes we do not have to perform an optimization involving all $K(K-1)/2$ pairs of codeword matrices. Noting that the codebook structure is *circulant*, i.e., $\mathbf{E}_j^H \mathbf{E}_i$ is a function of $i - j$ only, it follows that we need to take into account $\mathbf{E}_0^H \mathbf{E}_i$ with $i = 1, 2, \ldots, K-1$ only. This observation drastically simplifies the process of finding the parameters u_k. Tables I and II list the best $M_T = 1$ and $M_T = 2$, $N = 8$ constellations we could find for different values of L and K, employing the maximum diversity product (22) as design criterion and assuming a uniform PDP. W.l.o.g., we have set $u_0 = 1$ in our random code search. Recall that $\gamma > 0$ implies that the SFC achieves full space–frequency diversity in the general case.

C. Generalized Construction

An alternative and more general construction than the one described in Section IV-B is obtained by replacing $\mathbf{f}_p$ and $\mathbf{f}_p'$ in (25) by more general vectors $\mathbf{g}_0$ and $\mathbf{g}_1$ (recall that $\mathbf{f}_p$ denotes the pth column of the FFT matrix) to obtain

$$\mathbf{C}_i = \mathbf{\Phi}^i \begin{bmatrix} \mathbf{g}_0 & \mathbf{g}_1 \end{bmatrix}, \qquad i = 0, 1, \ldots, K-1.$$

This construction exhibits more degrees of freedom as code optimization involves both finding the u_k as well as the vectors $\mathbf{g}_0$ and $\mathbf{g}_1$, which have to be chosen such that the stacked matrix $\mathbf{B} = [\mathbf{g}_0\ \mathbf{g}_1\ \mathbf{D}\mathbf{g}_0\ \mathbf{D}\mathbf{g}_1]$ satisfies $\mathbf{B}^H\mathbf{B} = \mathbf{I}_4$.

Unfortunately, it appears to be rather difficult to obtain analytical results on the performance of generalized constructions compared with the simple constructions described in Section IV-B. Random code searches suggest that the generalized constructions yield only slight improvements in terms of error probability performance. However, the additional degrees of

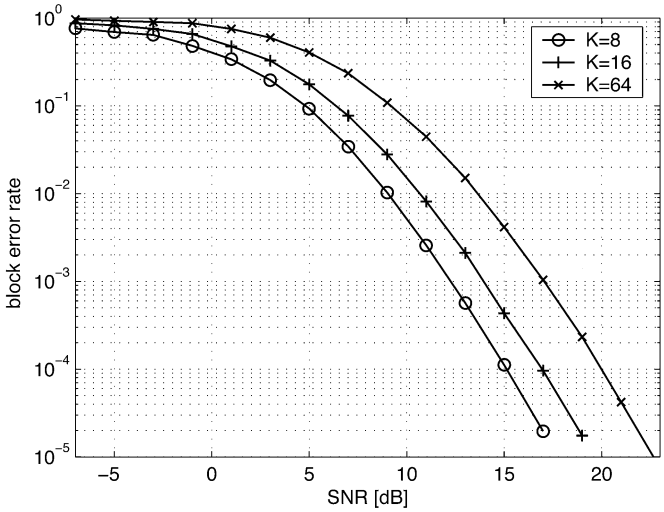


Fig. 1. Performance of noncoherent SFC constructions from Table II for $M_T = 2$ and different transmission rates.

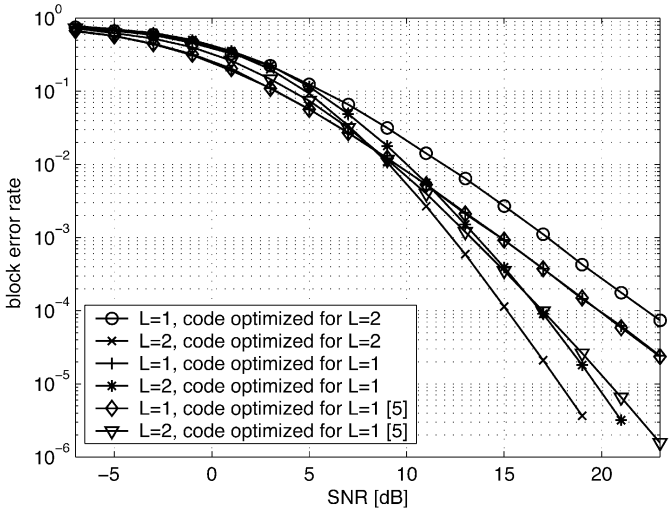


Fig. 2. Performance of noncoherent SFC constructions from Table II and [5] for $M_T = 2$ under frequency-flat and frequency-selective fading.

freedom offered by relaxing the condition $|c_{k,m}|^2 = 1/N$ ($k = 0, 1, \ldots, N-1, m = 0, 1, \ldots, M_T - 1$) may be useful for reducing the peak-to-average power ratio in a practical OFDM system.

D. Simulation Results

We present simulation examples for the simple construction according to (25) with parameters $M_T = 2$, $M_R = 1$, and $N = 8$. The PDP is assumed to be uniform with $\sigma_l^2 = 1/L$ ($l = 0, 1, \ldots, L-1$); we define $\mathrm{SNR} = E_s M_T / N$ and use ML decoding. Our first example in Fig. 1 shows the performance of the $K = 8$, 16, and 64 codes (with corresponding rates of 0.375, 0.5, and 0.75 bits/channel use, respectively) optimized for $L = 2$ from Table II in an $L = 2$ channel. It can be seen that all codes achieve the maximum achievable diversity order of $M_T M_R L = 4$, but that an increase in rate leads to a loss in coding gain.

In our second example, we employ the $K = 8$ codes from Table II optimized for $L = 1$ and $L = 2$, respectively, and the $K = 8$ code optimized for $L = 1$ (flat fading) from [5]. Fig. 2 again shows that for $L = 2$, the code optimized for $L = 2$ indeed achieves the maximum possible diversity order of $M_T M_R L = 4$. In the frequency-flat fading case (i.e., $L = 1$), the same code yields diversity order two and, hence, again achieves the maximum possible diversity order. Our code optimized for $L = 1$, as well as the code from [5], achieve virtually the same performance for $L = 1$ (the corresponding curves in Fig. 2 lie on top of each other). Both yield second-order diversity. However, the code from [5], optimized for the frequency-flat fading case ($L = 1$), does not exploit (all of) the additionally available frequency diversity when used in an $L = 2$ channel. It can be shown that, since the corresponding matrix $\mathbf{B}_{p,p'}$ is rank-deficient, the diversity order in this case is inherently limited to three. Numerical analysis confirms that the code achieves only third-order diversity. However, the code does not exhibit catastrophic behavior, either. On the other hand, our code optimized for $L = 1$ achieves full (fourth-order) diversity in the $L = 2$ case, but suffers from a coding gain loss when compared with the code optimized for $L = 2$. It must be stressed, though, that the fact that our code optimized for $L = 1$ achieves full diversity is a mere coincidence. Especially for higher rate designs, it becomes more likely to arrive at a diversity-deficient code when the correct number of channel taps is not taken into account in the code search.

V. Impact of the Wideband Channel on Space–Frequency Coded Systems

In this section, we discuss the impact of delay spread and shape of the PDP on the performance of noncoherent SFCs. It is interesting to note at the outset that the effects observed are significantly different from those obtained in coherent SFC schemes.

A. Impact of Delay Spread

It is instructive to investigate the error rate performance with increasing number of resolvable channel taps. We start by assuming that the PDP is uniform with $\sigma_l^2 = 1/L$ ($l = 0, 1, \ldots, L-1$). Furthermore, we consider a mutually space–frequency orthogonal codebook, as discussed in Section III-D2. Since all error events are equally likely, we immediately have the union bound on average error rate as $P_e \leq (K-1)\,\mathrm{P}(\mathbf{C}_i \to \mathbf{C}_j | \mathbf{C}_i)$, which implies that the PEP is directly representative of the overall code performance. Using the PEP expression for the coherent [13] and for the noncoherent case provided in (16), respectively, we obtain[2]

$$\lim_{L\to\infty} \mathrm{PEP}_{\mathrm{coh}} \leq \lim_{L\to\infty} \left(1 + \frac{E_s}{2L}\right)^{-M_T M_R L} = e^{-\frac{1}{2} M_T M_R E_s}$$

$$\lim_{L\to\infty} \mathrm{PEP}_{\mathrm{nch}} \leq \lim_{L\to\infty} \left(1 + \frac{E_s^2}{4L(L+E_s)}\right)^{-M_T M_R L} = 1.$$

For $M_T = 2$, $M_R = 1$, and $E_s = 12$ dB, Fig. 3 illustrates the PEP behavior for the coherent and noncoherent cases as a function of L: In the case of coherent decoding, additional

[2]Note that as $L \to \infty$, we need to ensure that the codewords are mutually space–frequency orthogonal, which requires that the block length $N \to \infty$.

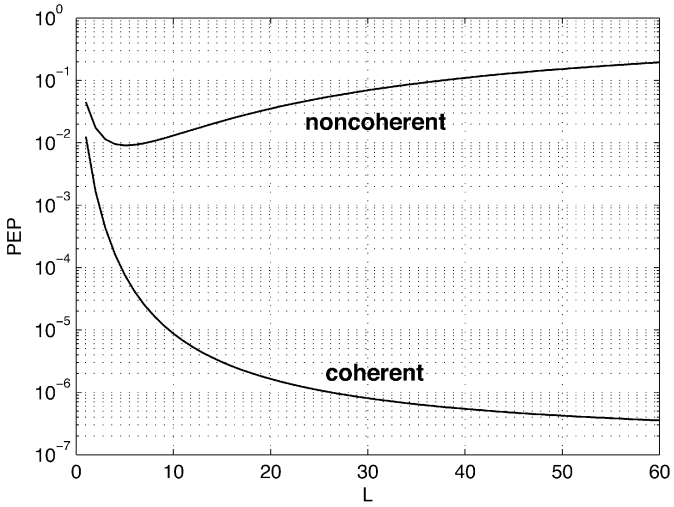


Fig. 3. Comparison of PEP for coherent and noncoherent decoding for a mutually space–frequency orthogonal codebook and uniform PDP with $\sigma_l^2 = 1/L$ $(l = 0, 1, \ldots, L-1)$ as a function of the number of channel taps L at $E_s = 12$ dB.

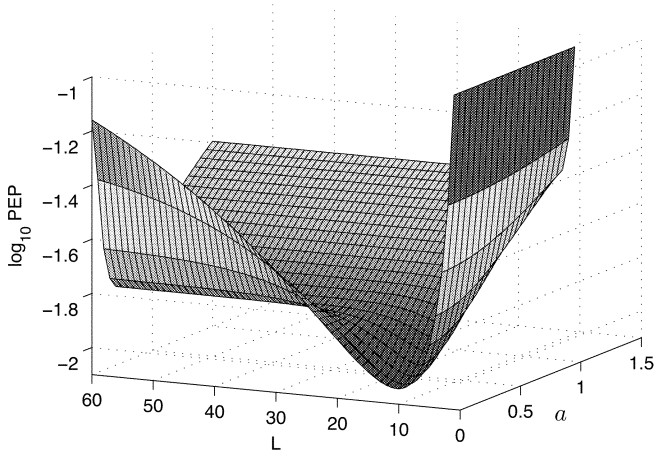


Fig. 4. Impact of delay spread and shape of the PDP on PEP performance of noncoherent space–frequency orthogonal signaling for $M_T = 1$ at $E_s = 15$ dB.

taps monotonically improve performance, although the error probability converges to a constant (the error probability of the AWGN channel, which is obtained in the large diversity order limit) with increasing number of taps. In the noncoherent case, additional channel taps improve performance by increasing the diversity gain; at the same time, channel uncertainty increases due to an increasing number of taps that need to be (implicitly) estimated, which results in a performance degradation. As a consequence of these two competing effects, the PEP exhibits a minimum for finite L. Our considerations clearly show that an *optimum diversity order* previously described in [23] and, more recently, in [24] also exists in the wideband multiple-antennas case. From the capacity analysis in [25], the existence of an optimum number of transmit antennas in the multiple-antennas case also follows.

B. Impact of the Power Delay Profile (PDP)

In the previous section, we showed that the number of channel taps can have a crucial impact on the performance of SFCs with noncoherent decoding. It is, therefore, natural to ask how (if at all) the shape of the PDP influences performance. To this end, we no longer assume a uniform PDP, but a truncated L-tap exponential PDP according to

$$\sigma_l^2 = \frac{1-e^{-a}}{1-e^{-aL}} e^{-al}, \qquad l = 0, 1, \ldots, L-1 \qquad (26)$$

parameterized by the coefficient a, which determines the decay rate of the PDP ($a \to 0$ results in a uniform PDP). It is easily verified that σ_l^2 in (26) satisfies $\sum_{l=0}^{L-1} \sigma_l^2 = 1$, independent of a. Again, assuming mutually space–frequency orthogonal codebooks, we show the PEP (16) as a function of the number of channel taps and the coefficient a in Fig. 4, assuming $M_T = M_R = 1$ (the results are qualitatively equivalent for multiple transmit and/or receive antennas) and $E_s = 15$ dB. The trend exhibited in Fig. 4 is representative for low and intermediate SNRs. For small a and hence a PDP close to uniform, we see exactly the same behavior as observed in the previous section—performance first improves with an increasing number of taps, but then quickly deteriorates as the receiver is increasingly unable to cope with the channel uncertainty. For higher values of a, the PDP is "peaky," and the total channel energy will be concentrated in the first few taps, which implies that the performance deterioration for increasing L is much less pronounced. Inspection of Fig. 4 also allows us to draw the following additional conclusions: When the channel has only a few taps, a uniform PDP is optimal in terms of PEP. With increasing number of total taps, the optimum channel shape will move from a uniform PDP to an ever more peaky shape. However, it can be deduced from the asymptotic bound (18) that in the high-SNR limit, a uniform PDP is always PEP-optimal for fixed total channel energy, even regardless of the codebook used (as long as it achieves full diversity).

We can make another interesting observation by approximating the individual factors in the right-hand side of (16) for large or small per-tap power $E_s\sigma_l^2$. Taps with $E_s\sigma_l^2 \ll 1$ result in a factor of approximately $(1 + E_s^2\sigma_l^4/4)^{-M_T M_R} \approx 1$. On the other hand, for $E_s\sigma_l^2 \gg 1$, the corresponding term in (16) is given by $(E_s\sigma_l^2/4)^{-M_T M_R}$. We can, therefore, see that only taps with $\sigma_l^2 \gg 1/E_s$ contribute significantly to lowering the error probability. Taps with $\sigma_l^2 \ll 1/E_s$ do not have enough energy to be estimated properly (of course in an implicit fashion).

We conclude by noting that (almost all) the results presented in this section hold in a rigorous fashion only for mutually space–frequency orthogonal codebooks. However, the fundamental effects described here are most likely also observable in the general case.

VI. Conclusion

We considered noncoherent space–frequency coded MIMO-OFDM systems in frequency-selective fading channels. For space–frequency unitary constellations, we derived the code design criteria, quantified the maximum achievable diversity order and coding gain, and provided two alternative code constructions capable of extracting full space–frequency diversity. We showed that the lack of CSI at the receiver does not result in

a reduced maximum achievable diversity order, which is given by $M_T M_R L$ in both the coherent and the noncoherent cases. We used our design criteria to study the impact of channel delay spread and PDP on SFC performance. Furthermore, we showed that noncoherent STCs designed for the frequency-flat fading case will in general not exploit the additionally available frequency diversity when used in frequency-selective fading environments. Finally, we demonstrated the existence of "catastrophic" SFCs, a class of codes not encountered in the coherent case.

The decoding complexity of our code constructions is, unfortunately, exponential in the code rate. Consequently, the design of noncoherent SFCs with reduced complexity algorithms remains an important open problem. Further challenges include the derivation of the design criteria for the general nonunitary space–frequency codebook case, the relation to packings in Grassmannian manifolds, and insights into the capacity of the noncoherent frequency-selective MIMO fading channel.

Appendix A
Moment-Generating Function $\phi(s)$

We can write (10) as

$$v_n = \bar{\mathbf{r}}_n^H \mathbf{V}_{i,j} \boldsymbol{\Delta} \mathbf{V}_{i,j}^H \bar{\mathbf{r}}_n$$

with

$$\mathbf{V}_{i,j} = \begin{bmatrix} \mathbf{E}_i & \mathbf{E}_j \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Delta} = \begin{bmatrix} -\boldsymbol{\Sigma}^2 \mathbf{T}^{-1} & \\ & \boldsymbol{\Sigma}^2 \mathbf{T}^{-1} \end{bmatrix}.$$

It follows from [33] that the MGF of the random variables v_n, as defined in (11), is given by

$$\phi(s) = \det{}^{-1}\left(\mathbf{I}_N - s\boldsymbol{\Lambda}_i \mathbf{V}_{i,j} \boldsymbol{\Delta} \mathbf{V}_{i,j}^H\right) \quad \text{for} \quad s \in \mathcal{S} \tag{27}$$

where $\mathcal{S}$ is a neighborhood of 0. Next, we factor the covariance matrix $\boldsymbol{\Lambda}_i$ according to $\boldsymbol{\Lambda}_i = \mathbf{I}_N + E_s \mathbf{E}_i \boldsymbol{\Sigma}^2 \mathbf{E}_i^H = \tilde{\mathbf{E}}_i \boldsymbol{\Gamma} \tilde{\mathbf{E}}_i^H$ with the $N \times N$ matrices

$$\tilde{\mathbf{E}}_i = \begin{bmatrix} \mathbf{E}_i & \mathbf{E}_i^{\perp} \end{bmatrix} \quad \text{such that} \quad \tilde{\mathbf{E}}_i^H \tilde{\mathbf{E}}_i = \mathbf{I}_N$$

and

$$\boldsymbol{\Gamma} = \begin{bmatrix} \mathbf{T} & \\ & \mathbf{I}_{N - M_T L} \end{bmatrix}.$$

Now, applying $\det(\mathbf{I}_N + \mathbf{AB}) = \det(\mathbf{I}_M + \mathbf{BA})$ with $N \times M$ and $M \times N$ matrices $\mathbf{A}$ and $\mathbf{B}$, respectively, to (27), we obtain

$$\begin{aligned} &\phi(s) \\ &= \det{}^{-1}\left(\mathbf{I}_N - s\tilde{\mathbf{E}}_i \boldsymbol{\Gamma} \tilde{\mathbf{E}}_i^H \mathbf{V}_{i,j} \boldsymbol{\Delta} \mathbf{V}_{i,j}^H\right) \\ &= \det{}^{-1}\left(\mathbf{I}_N - s\boldsymbol{\Delta} \mathbf{V}_{i,j}^H \tilde{\mathbf{E}}_i \boldsymbol{\Gamma} \tilde{\mathbf{E}}_i^H \mathbf{V}_{i,j}\right) \\ &= \det{}^{-1}\begin{bmatrix} \mathbf{I}_{M_T L} + s\boldsymbol{\Sigma}^2 & s\boldsymbol{\Sigma}^2 \mathbf{E}_i^H \mathbf{E}_j \\ -s\boldsymbol{\Sigma}^2 \mathbf{T}^{-1} \mathbf{E}_j^H \mathbf{E}_i \mathbf{T} & \mathbf{I}_{M_T L} - s\boldsymbol{\Sigma}^2 \mathbf{T}^{-1} \mathbf{E}_j^H \boldsymbol{\Lambda}_i \mathbf{E}_j \end{bmatrix}. \end{aligned}$$

Application of Schur's determinant condensation formula [34, 0.8.5] finally yields (12) and completes the proof.

Appendix B
Proof of Asymptotic Bounds

In this appendix, we provide proofs for the bounds (17) and (18), as well as a discussion of their asymptotic tightness. Throughout this appendix, we say that a bound $\varphi(t)$ on a function $f(t)$ is *asymptotically tight* when

$$\lim_{t \to \infty} \left| \log f(t) - \log \varphi(t) \right| = 0.$$

We start with the derivation of (17) and impose the purely technical constraint that $\boldsymbol{\Sigma}$ is nonsingular. Performing simple algebraic manipulations on the MGF (12), we arrive at the alternative representation

$$\phi(s) = \phi_1(s) \phi_2(s) \tag{28}$$

where

$$\begin{aligned} \phi_1(s) &= \det{}^{-1}\left(\mathbf{I}_{M_T L} + s\boldsymbol{\Sigma}^2\right) \det\left(\mathbf{I}_{M_T L} + E_s \boldsymbol{\Sigma}^2\right) \\ \phi_2(s) &= \det{}^{-1}\left(\mathbf{I}_{M_T L} + (E_s - s)\boldsymbol{\Sigma}^2 \boldsymbol{\Psi}\right) \end{aligned}$$

with

$$\begin{aligned} \boldsymbol{\Psi} &= \mathbf{I}_{M_T L} - s\mathbf{E}_j^H \mathbf{E}_i \boldsymbol{\Sigma}^2 \left(\mathbf{I}_{M_T L} + s\boldsymbol{\Sigma}^2\right)^{-1} \mathbf{E}_i^H \mathbf{E}_j \\ &= \mathbf{I}_{M_T L} - \mathbf{E}_j^H \mathbf{E}_i \left(\mathbf{I}_{M_T L} - \left(\mathbf{I}_{M_T L} + s\boldsymbol{\Sigma}^2\right)^{-1}\right) \mathbf{E}_i^H \mathbf{E}_j \\ &= \mathbf{I}_{M_T L} - \mathbf{E}_j^H \mathbf{E}_i \mathbf{E}_i^H \mathbf{E}_j + \mathbf{E}_j^H \mathbf{E}_i \left(\mathbf{I}_{M_T L} + s\boldsymbol{\Sigma}^2\right)^{-1} \mathbf{E}_i^H \mathbf{E}_j. \end{aligned}$$

Later, we will need to minimize $\phi(s)$ over s, and the parameter s will become a function of E_s. Assume for now that $s < E_s$, $s = \Theta(E_s)$, and $E_s - s = \Theta(E_s)$ as $E_s \to \infty$. The validity of these assumptions will be shown below. We continue by noting that

$$\phi_1(s) = \left(\frac{E_s}{s}\right)^{M_T L} - \mathcal{O}\left(E_s^{-1}\right). \tag{29}$$

Since $E_s/s = \Theta(1)$, the bound $\phi_1(s) < (E_s/s)^{M_T L}$ is asymptotically tight.

We proceed by writing $\phi_2(s)$ in terms of the eigenvalues of $\Sigma^2 \boldsymbol{\Psi}$

$$\begin{aligned} \phi_2(s) &= \prod_{r=0}^{M_T L - 1} \left(1 + (E_s - s)\, \lambda_r\left(\boldsymbol{\Sigma}^2 \boldsymbol{\Psi}\right)\right)^{-1} \\ &= \prod_{r=0}^{M_T L - 1} \left(1 + (E_s - s)\, \lambda_r\left(\boldsymbol{\Sigma} \boldsymbol{\Psi} \boldsymbol{\Sigma}\right)\right)^{-1} \\ &= \prod_{r=0}^{M_T L - 1} \left(1 + (E_s - s)\, \lambda_r\left(\boldsymbol{\Sigma} \mathbf{Q} \boldsymbol{\Sigma} + \mathbf{P}\right)\right)^{-1} \end{aligned}$$

where we have defined $\mathbf{Q} = \mathbf{I}_{M_T L} - \mathbf{E}_j^H \mathbf{E}_i \mathbf{E}_i^H \mathbf{E}_j$ and $\mathbf{P} = \boldsymbol{\Sigma} \mathbf{E}_j^H \mathbf{E}_i (\mathbf{I}_{M_T L} + s\boldsymbol{\Sigma}^2)^{-1} \mathbf{E}_i^H \mathbf{E}_j \boldsymbol{\Sigma}$. Observe that $\mathbf{P}$ is semidefinite; hence, all its eigenvalues are nonnegative. Moreover, since $\boldsymbol{\Sigma}^2$ is full rank, it can easily be shown using Ostrowski's Theorem [34, 4.5.9] that $E_s \to \infty$ (and hence, $s \to \infty$) results in $\lambda_r(\mathbf{P}) = \mathcal{O}(E_s^{-1})$ for *all* $r = 0, 1, \ldots, M_T L - 1$. Next,

we apply Weyl's Theorem [34, 4.3.1] for bounding the eigenvalues of a sum of Hermitian matrices, which combined with $\lambda_r(\mathbf{P}) = \mathcal{O}(E_s^{-1})$ yields

$$\prod_{r=0}^{M_TL-1}\Big(1+(E_s-s)\big(\lambda_r\big(\boldsymbol{\Sigma}^2\mathbf{Q}\big)+\mathcal{O}\big(E_s^{-1}\big)\big)\Big)^{-1} \le \phi_2(s) \le \prod_{r=0}^{M_TL-1}\Big(1+(E_s-s)\big(\lambda_r\big(\boldsymbol{\Sigma}^2\mathbf{Q}\big)+\mathcal{O}\big(E_s^{-1}\big)\big)\Big)^{-1}. \tag{30}$$

In the full-diversity case, the matrix $\mathbf{Q}$ is nonsingular, implying that all the $\lambda_r(\boldsymbol{\Sigma}^2\mathbf{Q})$ are nonzero constants, which establishes the asymptotic (as $E_s \to \infty$) tightness of

$$\phi_2(s) \le \prod_{r=0}^{M_TL-1}\Big(1+(E_s-s)\,\lambda_r\big(\boldsymbol{\Sigma}^2\mathbf{Q}\big)\Big)^{-1}. \tag{31}$$

In the rank-deficient case, the matrix $\mathbf{Q}$ is singular, which implies that one or more of the $\lambda_r(\boldsymbol{\Sigma}^2\mathbf{Q})$ is zero. Hence, it can be inferred from (30) that the bound (31) is no longer asymptotically tight. Note, however, that even though the bound (31) is not asymptotically tight anymore, using $(E_s-s)(0+\mathcal{O}(E_s^{-1})) = \mathcal{O}(1)$ it can still be concluded from (30) that the bound (31) correctly reflects the exponent of E_s^{-1} in $\phi_2(s)$. This observation will be used in Section III-F to argue that the bound (17) represents the diversity order correctly.

We can now further loosen (31), preserving the asymptotic tightness properties, by writing

$$\phi_2(s) < (E_s-s)^{-R_{i,j}} \prod_{r=0}^{R_{i,j}-1}\lambda_r^{-1}\big(\boldsymbol{\Sigma}^2\mathbf{Q}\big). \tag{32}$$

Combining (29) and (32), we finally obtain with (28)

$$\phi(s) < \left(\frac{E_s}{s}\right)^{M_TL}(E_s-s)^{-R_{i,j}} \cdot \prod_{r=0}^{R_{i,j}-1}\lambda_r^{-1}\big(\boldsymbol{\Sigma}^2\left(\mathbf{I}_{M_TL}-\mathbf{E}_j^H\mathbf{E}_i\mathbf{E}_i^H\mathbf{E}_j\right)\big) \tag{33}$$

which can easily be shown to be asymptotically tight in the full-diversity case and to correctly reflect the diversity order in the general case. Performing the minimization over s in (33), we obtain $s^\star = E_sM_TL/(R_{i,j}+M_TL) < E_s$, which confirms our initial assumptions on the asymptotic behavior of s. This concludes the derivation of (17). In the full-diversity case, (17) is easily seen to simplify to (18).

ACKNOWLEDGMENT

The authors would like to thank U. Schuster for many inspiring discussions and helpful comments on this paper.

REFERENCES

[1] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space–time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.

[2] J. Guey, M. Fitz, M. Bell, and W. Kuo, "Signal design for transmitter diversity wireless communication systems over Rayleigh fading channels," in *Proc. IEEE Veh. Technol. Conf.*, Atlanta, GA, 1996, pp. 136–140.

[3] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 139–157, Jan. 1999.

[4] B. M. Hochwald and T. L. Marzetta, "Unitary space–time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 543–564, Mar. 2000.

[5] B. M. Hochwald, T. L. Marzetta, T. J. Richardson, W. Sweldens, and R. Urbanke, "Systematic design of unitary space–time constellations," *IEEE Trans. Inf. Theory*, vol. 46, no. 6, pp. 1962–1973, Sep. 2000.

[6] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.

[7] M. L. McCloud, M. Brehler, and M. K. Varanasi, "Signal constellations for noncoherent space–time communications," in *Proc. 38th Allerton Conf. Commun., Contr., Comput.*, Monticello, IL, Oct. 2000.

[8] D. Agrawal, T. J. Richardson, and R. L. Urbanke, "Multiple-antenna signal constellations for fading channels," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2618–2626, Sep. 2001.

[9] V. Tarokh and I.-M. Kim, "Existence and construction of noncoherent unitary space–time codes," *IEEE Trans. Inf. Theory*, vol. 48, no. 12, pp. 3112–3117, Dec. 2002.

[10] I. Kammoun and J.-C. Belfiore, "A new family of Grassmann space–time codes for noncoherent MIMO systems," *IEEE Commun. Lett.*, vol. 7, no. 11, pp. 528–530, Nov. 2003.

[11] W. Zhao, G. Leus, and G. B. Giannakis, "Algebraic design of unitary space–time constellations," in *Proc. IEEE Int. Conf. Commun.*, Anchorage, AK, May 2003, pp. 3180–3184.

[12] E. Larsson, P. Stoica, and J. Li, "Orthogonal space–time block codes: Maximum-likelihood detection for unknown channels and unstructured interferences," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 362–372, Feb. 2003.

[13] H. Bölcskei and A. J. Paulraj, "Space–frequency coded broadband OFDM systems," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Chicago, IL, Sep. 2000, pp. 1–6.

[14] B. Lu and X. Wang, "Space–time code design in OFDM systems," in *Proc. IEEE GLOBECOM*, vol. 2, San Francisco, CA, Nov. 2000, pp. 1000–1004.

[15] H. Bölcskei, M. Borgmann, and A. J. Paulraj, "Impact of the propagation environment on the performance of space–frequency coded MIMO-OFDM," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 3, pp. 427–439, Apr. 2003.

[16] D. Agrawal, V. Tarokh, A. Naguib, and N. Seshadri, "Space–time coded OFDM for high data rate wireless communication over wideband channels," in *Proc. IEEE Veh. Technol. Conf.*, vol. 3, Ottawa, ON, Canada, May 1998, pp. 2232–2236.

[17] G. G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Trans. Commun.*, vol. 46, no. 3, pp. 357–366, Mar. 1998.

[18] H. Bölcskei, D. Gesbert, and A. J. Paulraj, "On the capacity of OFDM-based spatial multiplexing systems," *IEEE Trans. Commun.*, vol. 50, no. 2, pp. 225–234, Feb. 2002.

[19] H. Bölcskei, M. Borgmann, and A. J. Paulraj, "Space–frequency coded MIMO-OFDM with variable multiplexing-diversity tradeoff," in *Proc. IEEE Int. Conf. Commun.*, Anchorage, AK, May 2003, pp. 2837–2841.

[20] H. Bölcskei and M. Borgmann, "Code design for noncoherent MIMO-OFDM systems," in *Proc. 40th Allerton Conf. Commun., Contr., Comput.*, Monticello, IL, Oct. 2002, pp. 237–246.

[21] J. Giese and M. Skoglund, "Space–time constellations for unknown frequency-selective channels," in *Proc. IEEE Int. Conf. Commun.*, Anchorage, AK, May 2003, pp. 2583–2587.

[22] H. Artés, G. Matz, D. Schafhuber, and F. Hlawatsch, "Space–time multicarrier matrix modulation for unknown dispersive MIMO channels," in *Proc. 39th Allerton Conf. Commun., Contr., Comput.*, Monticello, IL, Oct. 2001, pp. 466–475.

[23] R. S. Kennedy and I. L. Lebow, "Signal design for dispersive channels," *IEEE Spectr.*, vol. 1, no. 3, pp. 231–237, Mar. 1964.

[24] I. E. Telatar and D. N. C. Tse, "Capacity and mutual information of wideband multipath fading channels," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 1384–1400, Jul. 2000.

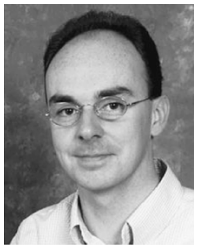
[25] M. Borgmann and H. Bölcskei, "On the capacity of noncoherent wideband MIMO-OFDM channels," presented at the IEEE Int. Symp. Inf. Theory, Adelaide, Australia, Sep. 2005.

[26] I. Barhumi, G. Leus, and M. Moonen, "Optimal training design for MIMO OFDM systems in mobile wireless channels," *IEEE Trans. Signal Process.*, vol. 51, no. 6, pp. 1615–1624, Jun. 2003.
[27] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
[28] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1319–1343, Jun. 2002.
[29] M. Borgmann and H. Bölcskei, "MIMO-OFDM communication in the wideband regime," in preparation, 2005.
[30] M. Brehler and M. K. Varanasi, "Asymptotic error probability analysis of quadratic receivers in Rayleigh fading channels with applications to a unified analysis of coherent and noncoherent space–time receivers," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2383–2399, Sep. 2000.
[31] S. Siwamogsatham, M. P. Fitz, and J. H. Grimm, "A new view of performance analysis of transmit diversity schemes in correlated Rayleigh fading," *IEEE Trans. Inf. Theory*, vol. 48, no. 4, pp. 950–956, Apr. 2002.
[32] J. L. Massey and M. K. Sain, "Inverses of linear sequential circuits," *IEEE Trans. Comput.*, vol. C-17, pp. 330–337, Apr. 1968.
[33] G. L. Turin, "The characteristic function of Hermitian quadratic forms in complex normal variables," *Biometrika*, vol. 47, pp. 199–201, 1960.
[34] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York: Cambridge Press, 1985.

Moritz Borgmann (S'99) was born in Hamburg, Germany. After undergraduate studies at Technische Universität München, Munich, Germany, he received the M.S. degree in electrical engineering from Stanford University, Stanford, CA, in 2000. Since 2002, he has been working towards the Dr. sc. techn. degree at the Swiss Federal Institute of Technology (ETH) Zurich, Zurich, Switzerland.

He was a Visiting Researcher with the Coordinated Science Laboratory, University of Illinois at Urbana–Champaign, Urbana, in 2001. His research interests include coding and modulation as well as efficient receiver algorithms for wideband wireless MIMO communications and corresponding performance limits.

Mr. Borgmann held a permanent scholarship from the German National Academic Foundation (1996–2000).



Helmut Bölcskei (M'98–SM'02) was born in Austria on May 29, 1970. He received the Dr. techn. degree in electrical engineering from Vienna University of Technology, Vienna, Austria, in 1997.

In 1998, he was with Vienna University of Technology. From 1999 to 2001, he was a Postdoctoral Researcher with the Information Systems Laboratory, Department of Electrical Engineering, Stanford University, Stanford, CA. During that period, he was also a Consultant for Iospan Wireless, Inc., San Jose, CA. From 2001 to 2002, he was an Assistant Professor of Electrical Engineering at the University of Illinois at Urbana–Champaign, Urbana. Since February 2002, he has been an Assistant Professor of Communication Theory at the Swiss Federal Institute of Technology (ETH) Zurich, Zurich, Switzerland. He was a Visiting Researcher at Philips Research Laboratories, Eindhoven, The Netherlands, ENST Paris, France, and the Heinrich Hertz Institute, Berlin, Germany. His research interests include communication and information theory with special emphasis on wireless communications and signal processing.

Prof. Bölcskei received a 2001 IEEE Signal Processing Society Young Author Best Paper Award and was an Erwin Schrödinger Fellow (1999–2001) of the Austrian National Science Foundation (FWF). He served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the *EURASIP Journal on Applied Signal Processing*, is currently an Associate Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and serves on the Editorial Board of *Foundations and Trends in Networking*.