Impact of Spatial Correlation and Precoding Design in OSTBC MIMO Systems

EMIL BJÖRNSON, EDUARD JORSWIECK,
AND BJÖRN OTTERSTEN

Stockholm 2010

Impact of Spatial Correlation and Precoding Design in OSTBC MIMO Systems

Emil Björnson, *Student Member, IEEE,* Eduard Jorswieck, *Senior Member, IEEE,*
and Björn Ottersten, *Fellow, IEEE*

Abstract—The impact of transmission design and spatial correlation on the symbol error rate (SER) is analyzed for multi-antenna communication links. The receiver has perfect channel state information (CSI), while the transmitter has either statistical or no CSI. The transmission is based on orthogonal space-time block codes (OSTBCs) and linear precoding. The precoding strategy that minimizes the worst-case SER is derived for the case when the transmitter has no CSI. Based on this strategy, the intuitive result that spatial correlation degrades the SER performance is proved mathematically.

In the case when the transmitter knows the channel statistics, the correlation matrix is assumed to be jointly-correlated (a generalization of the Kronecker model). The eigenvectors of the SER-optimal precoding matrix are shown to originate from the correlation matrix and the remaining power allocation is a convex problem. Equal power allocation is SER-optimal at high SNR. Beamforming is SER-optimal at low SNR, or for increasing constellation sizes, and its optimality range is characterized. A heuristic low-complexity power allocation is proposed and evaluated numerically. Finally, it is proved analytically that receive-side correlation always degrades the SER. Transmit-side correlation will however improve the SER at low to medium SNR, while its impact is negligible at high SNR.

Index Terms—Beamforming, channel state information, MIMO systems, orthogonal space-time block codes, power allocation, spatial correlation, symbol error rate.

I. INTRODUCTION

IN wireless communication, the use of antenna arrays at the transmitter and receiver can greatly improve the spectral efficiency and system performance. Under the ideal conditions of uncorrelated antennas and perfect channel state information (CSI), it was shown in [1] and [2] that the ergodic capacity improves linearly as the number of antennas increases at both sides. In practice, this fundamental gain is difficult to obtain. Firstly, the channel fading makes it costly for the transmitter to keep track on the current CSI. Secondly, the scattering is often

spatially limited which leads to a correlated channel [3]–[5], also known as *spatial correlation.*

The impact of CSI and spatial correlation on the ergodic capacity has received much attention. For simplicity, the receiver is usually assumed to have perfect CSI [6], while various types of CSI has been considered at the transmitter [7]–[11]. The impact of spatial correlation on the capacity was evaluated numerically in [9] (among others), but the relationship was first derived analytically in [10]. It was shown that spatial correlation decreases the capacity when the transmitter has no CSI or perfect CSI, which is intuitive since correlated channels have fewer degrees of freedom and thus less suitable for spatial multiplexing. When the transmitter has statistical CSI, this negative effect is however countered by the advantage of having smaller channel variations; in highly correlated channels, the channel direction is in fact given by the statistics. Interestingly, it was proved in [10] that correlation among the transmit antennas improves the capacity in this case.

While most previous work considered the ergodic capacity requiring Gaussian constellations, this paper considers the symbol error rate (SER) with practical symbol constellations. Prior work includes [12] and [13] that made numerical observations on the impact of spatial correlation on error rates. Herein, we derive an analytical solution to the impact of correlation by analyzing a general class of SER-like functions. This class includes the exact SER for Rayleigh fading channels with orthogonal space-time block codes (OSTBCs), linear precoding [14]–[18], and uncoded PAM, PSK, or QAM. We use the jointly-correlated model, proposed in [19] and [20], to analyze transmission design and the impact of spatial correlation under more general conditions than the commonly used Kronecker model [12], [13], [21]. Our main contributions are:

- **Optimal transmission strategies:** When the transmitter has no CSI, it can protect itself against the unknown Rayleigh fading channel by using OSTBCs and equal power allocation in all spatial directions. This precoding strategy minimizes the worst-case SER (Theorem 1). When the transmitter has statistical CSI, the eigenvector structure of the SER minimizing precoder is derived for jointly-correlated systems (Theorem 2). This structure reduces the transmission design to a convex power allocation problem that can be solved numerically or heuristically with low complexity (Strategy 1). At high SNR, the power is allocated equally among the available eigendirections. Single-stream beamforming in

the dominant eigendirection is SER-optimal at low and medium SNR and this is also the case asymptotically as the constellation size grows (Section V). The SNR range where beamforming is SER-optimal is characterized as a function of the constellation size (Theorem 5).

- **Impact of spatial correlation:** When the transmitter has no CSI, it is proven that spatial correlation always degrades the SER in jointly-correlated Rayleigh fading systems with OSTBCs (Theorem 3). In the case with statistical CSI at the transmitter, correlation between eigendirections at the receiver also degrades the performance. Transmit-side correlation will however improve the performance at low and medium SNR (Theorem 4), while the impact at high SNR is negligible (Section V).

The conclusion is that CSI and spatial correlation impacts the SER in a jointly-correlated system with OSTBCs in a similar (but non-identical) manner as the ergodic capacity in Kronecker-structured Rayleigh fading systems [10]; statistical CSI at the transmitter can improve the performance by proper transmission design that adapts to the correlation and turns transmit-side correlation into an advantage.

Notations: We use boldface (lower case) for column vectors, $\mathbf{x}$, and (upper case) for matrices, $\mathbf{X}$. With $\mathbf{X}^T$, $\mathbf{X}^H$, and $\mathbf{X}^*$ we denote the transpose, the conjugate transpose, and the conjugate of $\mathbf{X}$, respectively. The Kronecker and Hadamard products of two matrices $\mathbf{X}$ and $\mathbf{Y}$ are denoted $\mathbf{X} \otimes \mathbf{Y}$ and $\mathbf{X} \odot \mathbf{Y}$, respectively. The column vector obtained by stacking the columns of $\mathbf{X}$ is denoted $\text{vec}(\mathbf{X})$ and the matrix trace is $\text{tr}(\mathbf{X})$. The diagonal matrix $\text{diag}(\mathbf{x})$ has the elements of the vector $\mathbf{x}$ at the main diagonal. $\mathcal{CN}(\bar{\mathbf{x}}, \mathbf{Q})$ is used to denote circularly symmetric complex Gaussian random vectors, where $\bar{\mathbf{x}}$ is the mean and $\mathbf{Q}$ the covariance matrix. The operator $\triangleq$ is used for definitions. The squared Frobenius norm of $\mathbf{X}$ is denoted $\|\mathbf{X}\|^2$ and is defined as the sum of the squared absolute values of all the elements.

II. SYSTEM MODEL

We consider an arbitrarily correlated Rayleigh flat-fading channel with n_T transmit antennas and n_R receive antennas, represented by the channel matrix $\mathbf{H} \in \mathbb{C}^{n_R \times n_T}$. The transmission is based on OSTBCs with linear precoding, where the OSTBC is used for diversity gains and the transmitter achieves antenna gains by CSI-aware precoding. This is a standard form[1] of space-time codes for informed transmitters [24, Chapter 10], for which single-stream beamforming (as assumed in [12] and [13]) appears as a special case when the spatial coding block length B is one.

The OSTBC transmits K symbols over T symbols slots (i.e., the coding rate is K/T). Let $\mathbf{s} = [s_1, \ldots, s_K]^T \in \mathbb{C}^K$ represent these K data symbols, where each symbol $s_i \in \mathcal{A}$ has average power $\mathbb{E}\{|s_i|^2\} = \gamma$ and are uniformly distributed in the constellation set $\mathcal{A}$ (different constellations

[1] In general, non-orthogonal space-time block codes have better performance at the cost of increased decoding complexity, but we limit ourselves to OSTBCs to achieve analytical tractability. In practice, the orthogonality is often a minor restriction as the simple encoding/decoding of OSTBCs has made them popular in standards (i.e., LTE [22] and WLAN [23]). In addition, OSTBCs are rate optimal if $B \leq 2$ and the channel $\mathbf{H}$ is rank one [24, Theorem 7.4], and we show in Section IV that the SER minimizing spatial block length is often that small.

(a) Linear precoded OSTBC MIMO system.

(b) Equivalent parallel single-input single-output (SISO) systems, for $k = 1, \ldots, K$.

Fig. 1. Block model of the original MIMO communication system and its equivalent parallel structure after receive processing.

will be considered). These symbols are coded in an OSTBC matrix $\mathbf{C}(\mathbf{s}) \in \mathbb{C}^{B \times T}$ that fulfills the orthogonality property $\mathbf{C}(\mathbf{s})\mathbf{C}(\mathbf{s})^H = \|\mathbf{s}\|^2 \mathbf{I}$ and has the spatial coding block length B. The linear precoder $\mathbf{W} \in \mathbb{C}^{n_T \times B}$ is used to project the signal into advantageous spatial directions by using the available transmit-side CSI [16]. Its maximal rank is denoted by $m \triangleq \min(n_T, B)$ and the design of $\mathbf{W}$ will be considered in Section III for different CSI. By introducing the power constraint $\|\mathbf{W}\|^2 = 1$, we make sure that the average transmit power allocated per symbol is $\mathbb{E}\{\|\mathbf{W}\mathbf{C}(\mathbf{s})\|^2\}/K = \gamma$.

Observe that OSTBCs only exist for certain combinations of K, T, and B. In the simplest case, $K = T = B = 1$, it corresponds to single-stream beamforming with $\mathbf{C}(\mathbf{s}) = s_1$. Another important case is the Alamouti code, with $K = T = B = 2$ and $\mathbf{C}(\mathbf{s}) = \begin{bmatrix} s_1 & -s_2^* \\ s_2 & s_1^* \end{bmatrix}$, as it also provides full coding rate [14]. In general, the maximum possible coding rate approaches $1/2$ from above as the spatial dimension B increases [17]. For explicit codes and systematic code generation, see for example [15] and [18].

Under these assumptions, we achieve the system in Fig. 1(a). The received signal $\mathbf{Y} \in \mathbb{C}^{n_R \times T}$ is

$$\mathbf{Y} = \mathbf{H}\mathbf{W}\mathbf{C}(\mathbf{s}) + \mathbf{N} \tag{1}$$

where the total power has been normalized such that the elements of the additive noise $\mathbf{N} \in \mathbb{C}^{n_R \times T}$ are independent and identically distributed (i.i.d.) as $\mathcal{CN}(0, 1)$.

The precoding matrix $\mathbf{W}$ is a not part of the OSTBC, but a way of creating an effective channel, $\mathbf{H}\mathbf{W}$, with better properties. The receiver is assumed to know the effective channel perfectly, while separate knowledge of $\mathbf{H}$ and $\mathbf{W}$ is unrequired (this simplifies the channel estimation [6]). Then, the receiver can perform block-wise maximum likelihood detection of the symbols $\mathbf{s} = [s_1, \ldots, s_K]^T$ to find an estimate $\hat{\mathbf{s}} = [\hat{s}_1, \ldots, \hat{s}_K]^T$. As shown in [25], [26], an important property of OSTBCs is that the original system in (1) can be transformed into K independent and virtual single-antenna systems as

$$y_k' = \|\mathbf{H}\mathbf{W}\|s_k + n_k', \quad k = 1, \ldots, K, \tag{2}$$

where $n_k' \in \mathcal{CN}(0, 1)$. Thus, a low-complexity receiver structure is achieved where each symbol can be detected separately, as illustrated in Fig. 1(b). This result is due to the structure of the OSTBCs and the assumption of perfect CSI at the receiver side. We have made no assumptions on the

information available at the transmitter side. In principle, the transmitter can be completely uninformed of the CSI. We will however show that having statistical CSI can greatly improve the performance in certain environments.

A. Preliminaries on Spatial Correlation and Majorization

Herein, we will analyze the average performance of the system in (1) and (2) in terms of the SER. Thus, we need to specify the statistical properties of the channel matrix $\mathbf{H}$. In general, we have that $\text{vec}(\mathbf{H}) \in \mathcal{CN}(\mathbf{0}, \mathbf{R})$ for some arbitrary correlation matrix $\mathbf{R}$, defined on the column stacking of $\mathbf{H}$. To achieve an analytic structure on the statistics, we consider two popular MIMO channel models that have been verified by field measurements in realistic environments: the state-of-the-art *Jointly-correlated model* [19], [20] and the simplified *Kronecker model* [3]–[5]. These can be defined as follows.

Definition 1. The channel matrix $\mathbf{H}$ is jointly-correlated Rayleigh fading if

$$\mathbf{H} = \mathbf{U}_R(\widetilde{\boldsymbol{\Omega}} \odot \mathbf{G})\mathbf{U}_T^H \qquad (3)$$

where $\mathbf{U}_R \in \mathbb{C}^{n_R \times n_R}$ and $\mathbf{U}_T \in \mathbb{C}^{n_T \times n_T}$ are unitary matrices that describe transmit and receive eigendirections, respectively. The elements of $\mathbf{G} \in \mathbb{C}^{n_R \times n_T}$ are i.i.d. as $\mathcal{CN}(0, 1)$ and $\widetilde{\boldsymbol{\Omega}} \in \mathbb{C}^{n_R \times n_T}$ is the element-wise square root of the so-called coupling matrix $\boldsymbol{\Omega}$ (with positive entries) that determines the variance of each element in $\widetilde{\boldsymbol{\Omega}} \odot \mathbf{G}$. Without loss of generality, let the columns of $\boldsymbol{\Omega}$ be ordered with decreasing element sums. In terms of the correlation matrix $\mathbf{R}$, this model corresponds to the eigenvalue decomposition $\mathbf{R} = (\mathbf{U}_T^* \otimes \mathbf{U}_R)\text{diag}(\text{vec}(\boldsymbol{\Omega}))(\mathbf{U}_T^* \otimes \mathbf{U}_R)^H$ with separable eigenvector matrices.

Definition 2. The channel matrix $\mathbf{H}$ follows the Kronecker model if Definition 1 is fulfilled with a rank-one coupling matrix $\boldsymbol{\Omega} = \boldsymbol{\lambda}_R \boldsymbol{\lambda}_T^T$, where $\boldsymbol{\lambda}_R \in \mathbb{C}^{n_R}$ and $\boldsymbol{\lambda}_T \in \mathbb{C}^{n_T}$ are vectors with positive entries.[2]

To summarize, the Kronecker model represents the assumption that the transmit-side and the receive-side correlation can be completely separated, while the jointly-correlated model only assumes that the eigenvectors can be separated in this manner.

The spatial channel correlation can be measured in the eigenvalue distribution of the correlation matrix; weak correlation is represented by almost identical eigenvalues, while strong correlation means that a few eigenvalues dominate. Thus, in a highly correlated system, the channel is approximately confined to a small eigensubspace, while all eigenvectors are equally important in an uncorrelated system. In urban cellular systems, base stations are typically elevated and exposed to little near-field scattering. Thus, their antennas are strongly spatially correlated and the spread in $\boldsymbol{\lambda}_T$ is large. The receiving users will on the other hand be exposed to rich scattering and have weak spatial correlation if the

antenna spacing is sufficiently large [27], which means that the elements of $\boldsymbol{\lambda}_R$ are of similar magnitude.

The notion of majorization [28] provides a useful measure of the spatial correlation [29] and will be used herein for various purposes. Let $\mathbf{x} = [x_1, \ldots, x_N]^T$ and $\mathbf{y} = [y_1, \ldots, y_N]^T$ be two non-negative real-valued vectors of arbitrary length N. We say that $\mathbf{x}$ majorizes $\mathbf{y}$ if

$$\sum_{k=1}^{l} x_{[k]} \geq \sum_{k=1}^{l} y_{[k]}, \text{ for } l = 1, \ldots, N-1,$$
$$\text{and } \sum_{k=1}^{N} x_k = \sum_{k=1}^{N} y_k, \qquad (4)$$

where $x_{[k]}$ and $y_{[k]}$ are the kth largest ordered elements of $\mathbf{x}$ and $\mathbf{y}$, respectively. This majorization property is denoted $\mathbf{x} \succeq \mathbf{y}$. If $\mathbf{x}$ and $\mathbf{y}$ contain eigenvalues of channel correlation matrices, then $\mathbf{x} \succeq \mathbf{y}$ corresponds to that $\mathbf{x}$ is more spatially correlated than $\mathbf{y}$. Majorization only provides a partial order of vectors, but is still very powerful due to its connection to certain order-preserving functions:

A function $f(\cdot) : \mathbb{R}^N \to \mathbb{R}$ is said to be Schur-convex if $f(\mathbf{x}) \geq f(\mathbf{y})$ for all $\mathbf{x}$ and $\mathbf{y}$, such that $\mathbf{x} \succeq \mathbf{y}$. Similarly, $f(\cdot)$ is said to be Schur-concave if $\mathbf{x} \succeq \mathbf{y}$ implies that $f(\mathbf{x}) \leq f(\mathbf{y})$.

B. Expressions for the Symbol Error Rate

Throughout the paper, the performance measure will be the SER; that is, the probability that the receiver makes an error in the detection of source symbols. Since the equivalent channels in (2) are identical for all symbols in the OSTBC, it is clear that the SER only depends on the distribution of the SNR, $\|\mathbf{H}\mathbf{W}\|^2$, and on the type on the symbol constellation set, $\mathcal{A}$. Next, we will present SER expressions for three commonly considered symbol constellations, but first we introduce a general class of functions.

Definition 3. We define the function

$$F_{\mathbf{a},\mathbf{b},\mathbf{c}}(\boldsymbol{\Phi}, x) \triangleq \sum_{k=1}^{n} \frac{c_k}{\pi} \int_{a_k}^{b_k} \frac{d\theta}{\det\left(\mathbf{I} + \frac{x}{\sin^2(\theta)}\boldsymbol{\Phi}\right)} \qquad (5)$$

where $\boldsymbol{\Phi}$ is a positive semi-definite matrix and $x \geq 0$. The vectors $\mathbf{a} = [a_1, \ldots, a_n]^T$, $\mathbf{b} = [b_1, \ldots, b_n]^T$, and $\mathbf{c} = [c_1, \ldots, c_n]^T$ have arbitrary length n and fulfill $a_k \leq b_k$ and $c_k \geq 0$ for all k.

This class of functions is important since the SERs with Pulse Amplitude Modulation (PAM), Phase-Shift Keying (PSK), and Quadrature Amplitude Modulation (QAM) belong to it. The variable x is proportional to the SNR, but the scaling depends on the modulation. Let $g_{\text{PAM}} \triangleq 3/(M^2 - 1)$, $g_{\text{PSK}} \triangleq \sin^2(\pi/M)$, and $g_{\text{QAM}} \triangleq 3/(2M - 2)$, then the exact SER of the system in (2) was derived in [26], [30] as

$$\text{SER}_{\text{PAM}}(\mathbf{R}, \mathbf{W}, \gamma) = F_{0, \frac{\pi}{2}, \frac{2(M-1)}{M}}(\boldsymbol{\Phi}, \gamma g_{\text{PAM}}),$$

$$\text{SER}_{\text{PSK}}(\mathbf{R}, \mathbf{W}, \gamma) = F_{0, \frac{\pi(M-1)}{M}, 1}(\boldsymbol{\Phi}, \gamma g_{\text{PSK}}),$$

$$\text{SER}_{\text{QAM}}(\mathbf{R}, \mathbf{W}, \gamma)$$
$$= F_{[0\ \frac{\pi}{4}]^T, [\frac{\pi}{4}\ \frac{\pi}{2}]^T, [\frac{4(\sqrt{M}-1)}{M}\ \frac{4(\sqrt{M}-1)}{\sqrt{M}}]^T}(\boldsymbol{\Phi}, \gamma g_{\text{QAM}}),$$
$$(6)$$

---

[2]This is equivalent to the more common definition: $\mathbf{H} = \mathbf{R}_R^{1/2}\bar{\mathbf{G}}\mathbf{R}_T^{1/2}$, where the elements of $\bar{\mathbf{G}}$ are i.i.d. as $\mathcal{CN}(0, 1)$. In this formulation, $\mathbf{R}_R = \mathbf{U}_R\text{diag}(\boldsymbol{\lambda}_R)\mathbf{U}_R^H$ and $\mathbf{R}_T = \mathbf{U}_T\text{diag}(\boldsymbol{\lambda}_T)\mathbf{U}_T^H$ are positive semi-definite matrices that represent the receive-side and transmit-side correlation, respectively. In terms of the general correlation matrix, we have $\mathbf{R} = \mathbf{R}_T^T \otimes \mathbf{R}_R$.

for M-PAM, M-PSK, and M-QAM constellations, respectively. For all three constellations, we have $\boldsymbol{\Phi} = (\mathbf{W}^T \otimes \mathbf{I})\mathbf{R}(\mathbf{W}^T \otimes \mathbf{I})^H$, which is the correlation matrix of the effective channel $\mathbf{HW}$. Note that these SER expressions are valid for uncoded systems, while the performance with outer coding behaves differently [31].

Observe that the integrals in Definition 3 are the main building stones in all the SER expressions in (6). The determinant in the integrands can equally be expressed as

$$\det \left(\mathbf{I} + \frac{x}{\sin^2(\theta)}\boldsymbol{\Phi}\right) = \prod_{j=1}^{mn_R}\left(1 + \frac{x}{\sin^2(\theta)}\lambda_j(\boldsymbol{\Phi})\right), \quad (7)$$

where $\lambda_j(\boldsymbol{\Phi})$ denotes the jth largest eigenvalue of $\boldsymbol{\Phi}$. Thus, we conclude that the eigenvalues of $\boldsymbol{\Phi}$ (and not the eigenvectors) determine the SER. Since (7) is a Schur-concave function with respect to the eigenvalues, it is clear that the eigenvalue spread will affect the performance. This brings us back to the notion of spatial correlation discussed in the last section. In Section IV, we will analyze how the SER performance depends on the spatial correlation and we will focus on comparing systems with different eigenvalue distributions. All analytic results will be derived for the class of functions in Definition 3, and the interpretations for PAM, PSK, and QAM will be given as corollaries.

III. LINEAR PRECODING WITH DIFFERENT TYPES OF CSI

The purpose of applying linear precoding to OSTBCs is to adapt it to the channel conditions known at the transmitter and thereby improve the system performance. Herein, the performance measure is the SER and thus the precoding matrix should be selected as

$$\mathbf{W} = \underset{\mathbf{W} \in \mathbb{C}^{n_T \times B};\ \|\mathbf{W}\|^2 = 1}{\arg\min} \mathrm{SER}(\mathbf{R}, \mathbf{W}, \gamma). \quad (8)$$

Depending on the type of symbol constellation, the SER expression in this optimization problem will be slightly different. Apart from the constellation, the SER also depends on the precoder $\mathbf{W}$, the channel correlation matrix $\mathbf{R}$, and the SNR γ. Thus, the quality of the precoding design will depend on whether the correlation and SNR is known at the transmitter or not. Next, we will solve (8) assuming that the these statistical parameters are either unknown or perfectly known to the transmitter.

A. Without CSI at the Transmitter

When the transmitter is unaware of the channel correlation matrix, $\mathbf{R}$, and potentially unaware of the SNR, γ, robustness against channel fading can be achieved by minimizing the worst-case SER. This worst case scenario corresponds to that for every precoder we select, the channel conditions always become the worst possible. Formally, the worst-case SER is given by the following optimization problem:

$$\max_{\substack{\mathbf{R} \in \mathbb{C}^{n_T n_R \times n_T n_R};\\ \mathbf{R} \succeq 0,\ \mathrm{tr}(\mathbf{R}) = n_T n_R}} \min_{\substack{\mathbf{W} \in \mathbb{C}^{n_T \times B};\\ \|\mathbf{W}\|^2 = 1}} \mathrm{SER}(\mathbf{R}, \mathbf{W}, \gamma). \quad (9)$$

Next, we solve this problem for the class of SER-like functions in Definition 3 and give the structure of the optimal precoding matrices.

Theorem 1. Consider minimization of the worst-case function value of $F_{\mathbf{a},\mathbf{b},\mathbf{c}}(\boldsymbol{\Phi}, x)$, with $\boldsymbol{\Phi} = (\mathbf{W}^T \otimes \mathbf{I})\mathbf{R}(\mathbf{W}^T \otimes \mathbf{I})^H$, by selection of $\mathbf{W} \in \mathbb{C}^{n_T \times B}$ with $\|\mathbf{W}\|^2 = 1$. For all $x > 0$, we have

$$\max_{\substack{\mathbf{R} \in \mathbb{C}^{n_T n_R \times n_T n_R};\\ \mathbf{R} \succeq 0,\ \mathrm{tr}(\mathbf{R}) = n_T n_R}} \min_{\substack{\mathbf{W} \in \mathbb{C}^{n_T \times B};\\ \|\mathbf{W}\|^2 = 1}} F_{\mathbf{a},\mathbf{b},\mathbf{c}}(\boldsymbol{\Phi}, x)$$

$$= \begin{cases} F_{\mathbf{a},\mathbf{b},\mathbf{c}}(\mathbf{0}, x), & B < n_T, \\ F_{\mathbf{a},\mathbf{b},\mathbf{c}}(\mathrm{diag}([n_R, 0, \ldots, 0]), x), & B \geq n_T. \end{cases} \quad (10)$$

If the dimension $B < n_T$, the minimal value is achieved for any $\mathbf{W}$, while the optimal precoding matrix for $B \geq n_T$ is $\mathbf{W} = \sqrt{1/n_T}\mathbf{V}_1[\mathbf{I}\ \mathbf{0}]\mathbf{V}_2$ for arbitrary unitary matrices $\mathbf{V}_1 \in \mathbb{C}^{n_T \times n_T}$ and $\mathbf{V}_2 \in \mathbb{C}^{B \times B}$.

Proof: The proof is given in Appendix B. ∎

The following corollary interprets the theorem in terms of the SER.

Corollary 1. Consider the worst-case SER in (9) with either M-PAM, M-PSK, or M-QAM. If $B < n_T$, then the worst-case SER is $(M-1)/M$ independently of the structure of the precoding matrix. If $B \geq n_T$, the minimal worst-case SER is strictly smaller than $(M-1)/M$ and is achieved by precoding matrices of the type

$$\mathbf{W} = \sqrt{\frac{1}{n_T}}\mathbf{V}_1[\mathbf{I}\ \mathbf{0}] \quad (11)$$

where $\mathbf{V}_1$ is a unitary matrix.

Two important conclusions can be drawn. Firstly, before data transmission, the probability of falsely predicting the next symbol is $(M-1)/M$. This is also the worst-case SER when $B < n_T$, and thus we need $B \geq n_T$ (i.e., exploiting all spatial directions) to guarantee that useful information is received. Secondly, an example of the optimal precoding matrix, for $B \geq n_T$, is $\mathbf{W} = \sqrt{1/n_T}[\mathbf{I}\ \mathbf{0}]$, which is a scaled $n_T \times n_T$ identity matrix padded by zeros. It is obviously not beneficial to have $B > n_T$, since the $B - n_T$ additional degrees of freedom appear in the null space of the channel. To summarize, when the transmitter is unaware of the channel statistics, the optimal spatial coding block length is $B = n_T$ and power should be allocated isotropically (i.e., $\mathbf{W} = \sqrt{1/n_T}\mathbf{I}$).

B. With Statistical CSI at the Transmitter

When statistical CSI is available at the transmitter, the precoding matrix $\mathbf{W}$ can be adapted to the spatial properties of the $\mathbf{R}$ and to the average SNR of the system. The purpose of this section is to characterize the solution of the SER minimization in (8). First, we show the structure of the optimal precoder under the assumption of jointly-correlated channels. This structure reduces the precoding design to a convex power allocation problem. Explicit asymptotic solutions will be derived at low and high SNRs and for large symbol constellations. In addition, a simple approximate power allocation will be proposed.

We begin with a theorem that derives the general structure and the asymptotic properties of precoding matrices $\mathbf{W}$ that minimize the SER-like class of functions in Definition 3.

Theorem 2. Consider minimization of $F_{\mathbf{a},\mathbf{b},\mathbf{c}}(\boldsymbol{\Phi}, x)$, with $\boldsymbol{\Phi} = (\mathbf{W}^T \otimes \mathbf{I})\mathbf{R}(\mathbf{W}^T \otimes \mathbf{I})^H$, by selection of $\mathbf{W} \in \mathbb{C}^{n_T \times B}$ with $\|\mathbf{W}\|^2 = 1$. If $\mathbf{R}$ is jointly-correlated and known, the solutions to

$$\min_{\mathbf{W} \in \mathbb{C}^{n_T \times B}; \|\mathbf{W}\|^2 = 1} F_{\mathbf{a},\mathbf{b},\mathbf{c}}(\boldsymbol{\Phi}, x), \qquad (12)$$

have the structure $\mathbf{W} = \mathbf{U}_T \boldsymbol{\Pi} \boldsymbol{\Delta} \mathbf{V}$ for some n_T-dimensional permutation matrix $\boldsymbol{\Pi}$ and arbitrary unitary matrix $\mathbf{V} \in \mathbb{C}^{B \times B}$. The rectangular diagonal matrix $\boldsymbol{\Delta} \in \mathbb{C}^{n_T \times B}$ has $\sqrt{p_1}, \ldots, \sqrt{p_m}$ on the main diagonal, and $F_{\mathbf{a},\mathbf{b},\mathbf{c}}(\boldsymbol{\Phi}, x)$ is convex in p_j for all j. The limiting solution at large x is given by $p_1 = \ldots = p_m = 1/m$ and a permutation matrix that selects the m eigendirections with the largest element products in columns of $\boldsymbol{\Omega}$. The limiting solution at small x is given by $\boldsymbol{\Pi} = \mathbf{I}$ and all power is allocated to $p_1, \ldots, p_{\tilde{m}}$, where $\tilde{m}$ is the multiplicity of the largest column sum of $\boldsymbol{\Omega}$.

Under the Kronecker model, the solution has $\boldsymbol{\Pi} = \mathbf{I}$. At small x, the limiting solution performs equal power allocation among the strongest directions: $p_1 = \ldots = p_{\tilde{m}} = 1/\tilde{m}$.

Proof: The proof is given in Appendix B. ∎

The following corollary interprets the theorem in terms of the SER, and is a generalization and correction of [26] (which treats the Kronecker model and has disregarded the eigenvalue ordering).

Corollary 2. The SERs of M-PAM, M-PSK, and M-QAM are minimized by precoding matrices with the structure

$$\mathbf{W} = \begin{cases} \mathbf{U}_T \boldsymbol{\Pi} \begin{bmatrix} \mathbf{D} \\ \mathbf{0} \end{bmatrix} & \text{if } B < n_T \\ \mathbf{U}_T [\mathbf{D} \ \mathbf{0}] & \text{if } B \geq n_T \end{cases} \qquad (13)$$

where $\mathbf{D} \in \mathbb{C}^{m \times m}$ is a diagonal matrix. The limiting solution at high SNR and fixed constellation size M is equal power allocation in $\mathbf{D}$. At low SNR, beamforming in the direction of the first column of $\mathbf{U}_T$ is the SER minimizing solution. This is also the asymptotically optimal solution as the constellation size $M \to \infty$.

The first conclusion is that the structure of the SER minimizing precoding matrix in jointly-correlated channels is similar as under the Kronecker model [26], [32]. Having $B > \mathrm{rank}(\mathbf{D})$ will not improve the performance, and thus there is no reason to have $B > n_T$. At high SNR, it was expected that equal power allocation is the limiting solution [26], but an important result from Theorem 2 is that beamforming is optimal both at low SNR and for large symbol constellations.

The optimal precoding structure derived in Theorem 2 for jointly-correlated systems reduces the precoding optimization to a convex power allocation problem (and selection of the active eigendirections, if $B < n_T$). This power allocation can be solved numerically in an efficient fashion using gradient methods [33]. For low-complexity implementations, we propose the following heuristic power allocation.

Strategy 1. A heuristic solution to the precoding power allocation in Theorem 2 is

$$p_j = \min\left(\frac{n_R}{\alpha} - \frac{n_R}{x\mu_j}, 0\right) \quad \text{for } j = 1, \ldots, m, \qquad (14)$$

where μ_j is the element sum of the jth column of $\boldsymbol{\Omega}$ and we use $\boldsymbol{\Pi} = \mathbf{I}$ as permutation matrix. The parameter α is selected to fulfill the power constraint $\sum_{j=1}^{m} p_j = 1$.

This power allocation behaves similar to the optimal strategy in terms of the waterfilling property that gives beamforming at low x and equal power allocation at large x. The number of active precoding directions increases with x and all directions are used if

$$x > n_R \left(\frac{m}{\mu_m} - \sum_{j=1}^{m} \frac{1}{\mu_j} \right). \qquad (15)$$

Otherwise, the number of active directions is $\tilde{m} = \mathrm{rank}(\mathbf{D}) < m$, where $\tilde{m}$ is the positive integer that fulfills

$$n_R \left(\frac{\tilde{m}}{\mu_{\tilde{m}}} - \sum_{j=1}^{\tilde{m}} \frac{1}{\mu_j} \right) < x \leq n_R \left(\frac{\tilde{m}+1}{\mu_{\tilde{m}+1}} - \sum_{j=1}^{\tilde{m}+1} \frac{1}{\mu_j} \right). \qquad (16)$$

The power allocation in Strategy 1 is derived from the Chernoff bound

$$F_{\mathbf{a},\mathbf{b},\mathbf{c}}(\boldsymbol{\Phi}, x) \leq \sum_{k=1}^{n} \frac{c_k(b_k - a_k)}{\pi \det(\mathbf{I} + x\boldsymbol{\Phi})}, \qquad (17)$$

which is minimized by (14) under the condition that the coupling matrix can be factorized as $\boldsymbol{\Omega} = [1, \ldots, 1]\boldsymbol{\lambda}_T$ (i.e., Kronecker model with uncorrelated receiver). The performance of this power allocation will be evaluated in Section V for a general $\boldsymbol{\Omega}$ and compared with the optimal strategy.

IV. IMPACT OF SPATIAL CORRELATION WITH DIFFERENT TYPES OF CSI

The SER depends on the spatial correlation, as pointed out in Section II-B. Next, we will analyze this dependence in more detail using the tool of majorization. If we can show that the SER is a Schur-convex function, then spatial correlation increases the error rate and thereby degrades the performance. If the SER, on the other hand, is Schur-concave, then spatial correlation improves the performance. In this section, we prove that both properties can apply, depending on the CSI available at the transmitter.

A. Without CSI at the Transmitter

When the transmitter is unaware of the CSI, Theorem 1 showed that equal power allocation in all spatial directions minimizes the worst-case SER. Assuming that such precoding is applied, the following theorem derives the impact of spatial correlation on the class of SER-like functions in Definition 3.

Theorem 3. Consider $F_{\mathbf{a},\mathbf{b},\mathbf{c}}(\boldsymbol{\Phi}, x)$, with $\boldsymbol{\Phi} = (\mathbf{W}^T \otimes \mathbf{I})\mathbf{R}(\mathbf{W}^T \otimes \mathbf{I})^H$, where $B \geq n_T$ and $\mathbf{W} = \sqrt{1/n_T}\mathbf{V}_1[\mathbf{I} \ \mathbf{0}]\mathbf{V}_2$. This function is Schur-convex with respect to any subset of eigenvalues of $\mathbf{R}$, while the other eigenvalues are fixed. Under the Kronecker model, this means that $F_{\mathbf{a},\mathbf{b},\mathbf{c}}(\boldsymbol{\Phi}, x)$ is Schur-convex with respect to $\boldsymbol{\lambda}_T$ when $\boldsymbol{\lambda}_R$ is fixed and Schur-convex with respect to $\boldsymbol{\lambda}_R$ when $\boldsymbol{\lambda}_T$ is fixed.

Proof: The theorem follows directly from Lemma 1 in Appendix A since all non-zero eigenvalues of $\boldsymbol{\Phi}$ also are eigenvalues of $\mathbf{R}$. ∎

The following corollary interprets the theorem in terms of the SER.

Corollary 3. When the precoder minimizes the worst-case SER, spatial correlation always degrades the SER performance with M-PAM, M-PSK, and M-QAM (even if only a certain eigenspace is considered). Under the Kronecker model, both receive and transmit-side correlation degrades the performance.

The intuitive conclusion is that when the transmitter has no CSI and therefore makes an isotropic signal power allocation, the preferred fading environment is isotropic (i.e., all directions should be equally strong a priori). As spatial correlation creates a few dominant directions, isotropic transmission will waste transmission power in other directions which leads to performance degradation.

B. With Statistical CSI at the Transmitter

Next, we consider the case when the transmitter knows the channel correlation matrix and the average SNR of the system. When SER minimizing precoding is applied, according to Theorem 2, we prove that the impact of spatial correlation changes with the SNR.

Theorem 4. Consider $F_{\mathbf{a},\mathbf{b},\mathbf{c}}(\mathbf{\Phi}, x)$, with $\mathbf{\Phi} = (\mathbf{W}^T \otimes \mathbf{I})\mathbf{R}(\mathbf{W}^T \otimes \mathbf{I})^H$, where $\mathbf{W} = \mathbf{U}_T\mathbf{\Pi}\mathbf{\Delta}\mathbf{V}$ minimizes the function as in Theorem 2. If $\mathbf{R}$ is jointly-correlated and known, let the (l,j)th element of the coupling matrix $\mathbf{\Omega}$ be parameterized as $\mu_j \bar{\omega}_{l,j}$, where μ_j is the sum of the jth column and $\sum_{l=1}^{n_R} \bar{\omega}_{l,j} = 1$ for all j. Then, the function is Schur-convex with respect to $\bar{\omega}_{1,j}, \ldots, \bar{\omega}_{n_R,j}$ for each j (when all μ_j are fixed). The function is Schur-convex with respect to $\mu_{\pi(1)}, \ldots, \mu_{\pi(m)}$ (for fixed $\bar{\omega}_{l,j}$) at large x (the bijective permutation function $\pi(\cdot)$ represents $\mathbf{\Pi}$) and Schur-concave with respect to $\mu_1, \ldots, \mu_{n_T}$ at small x.

Under the Kronecker model, this implies that the function is Schur-convex with respect to $\mathbf{\lambda}_R$ (when $\mathbf{\lambda}_T$ is fixed). The function is Schur-convex with respect to the m largest elements of $\mathbf{\lambda}_T$ (when $\mathbf{\lambda}_R$ is fixed) at large x, while it is Schur-concave with respect to the complete vector $\mathbf{\lambda}_T$ at small x.

Proof: The proof is given in Appendix B. ∎

The following corollary interprets the theorem in terms of the SER.

Corollary 4. With SER-optimal precoding for M-PAM, M-PSK, or M-QAM, the impact of spatial correlation depends on the SNR. In jointly-correlated systems, spatial correlation is characterized as the spread of channel gains between different eigendirections at the transmitter and receiver side. Spatial correlation in receive eigendirections always degrades the performance. At high SNR, spatial correlation in transmit eigendirections also degrades performance, while correlation improves the SER at low SNR. Under the Kronecker model, these behaviors decouple; receive-side correlation decreases the performance, while transmit-side correlation improves the performance at low SNR and degrades it at high SNR.

In other words, even if optimal precoding is applied, spatial receive-side correlation will always degrade the performance. For transmit-side correlation, there is however a remarkable change in behavior between low and high SNR, which requires further specification. The low SNR behavior was proved using Theorem 2 which showed that beamforming is optimal in this SNR region. Thus, spatial correlation improves the performance in an SNR region that is at least as large as the beamforming optimality range. This range is characterized by the following theorem.

Theorem 5. When minimizing the function $F_{\mathbf{a},\mathbf{b},\mathbf{c}}(\mathbf{\Phi}, x)$, a necessary and sufficient condition for optimality of beamforming (i.e., $p_1 = 1, p_2 = \ldots = p_m = 0$) is

$$\sum_{k=1}^{n} \frac{c_k}{\pi} \int_{a_k}^{b_k} \left(\sum_{l=1}^{n_R} \frac{\omega_{l,1}}{\sin^2(\theta) + x\omega_{l,1}} - \frac{\omega_{l,2}}{\sin^2(\theta)} \right) \frac{d\theta}{\det\left(\mathbf{I} + \frac{x}{\sin^2(\theta)}\mathbf{A}\right)} \geq 0, \tag{18}$$

where $\omega_{l,j}$ is the (l,j)th element of $\mathbf{\Omega}$ and $\mathbf{A} = \text{diag}(\omega_{1,1}, \ldots, \omega_{n_R,1})$.

Proof: The proof is given in Appendix B. ∎

The following corollary interprets the theorem in terms of the SER.

Corollary 5. The SNR range with optimality for single-stream beamforming is $\gamma \in [0, \upsilon]$, where the upper bound υ solves (18) with equality using $x = \upsilon g_{\text{PAM}}$, $x = \upsilon g_{\text{PSK}}$, and $x = \upsilon g_{\text{QAM}}$ for M-PAM, M-PSK, or M-QAM, respectively. The parameters $\mathbf{a}, \mathbf{b}, \mathbf{c}$ are given in (6) for each modulation scheme.

In general, the beamforming optimality range cannot be derived explicitly. The expression in (18) is however monotonically decreasing in x and thus the x-value that provides equality can be derived by simple line search procedures. An approximate expression for the optimality range can be derived using the low-complexity precoding strategy proposed in Strategy 1 by simply substituting $\tilde{m} = 1$ into (16).

Finally, we stress that in practice the positive impact of transmit-side correlation in Theorem 4 can be observed for an SNR range considerably larger than the optimality range for single-stream beamforming[3]. This analytical result stands in contrast to the numerical conclusion in [13] that the SER increases monotonically with the correlation. This misconception originates from varying the transmit and receive-side correlation simultaneously. Next, we will show numerically that transmit-side correlation improves the performance at both low and medium SNRs, while the correlation impact is negligible at high SNR.

V. NUMERICAL EXAMPLES

In this section, we provide numerical examples that demonstrate the precoding results in Section III and the impact of spatial correlation that was analyzed in Section V. First, the performance loss of the proposed heuristic power allocation strategy will be evaluated along with the size of the beamforming optimality range. Then, we will clarify how the low and high SNR-behaviors derived in Section V affect the performance in the range of practical SNRs.

---

[3]In fact, the beamforming range cannot be used directly to determine the low SNR region since it depends on the spatial correlation, while the low SNR property is valid for any correlation.

Fig. 2. Power allocation among p_1, p_2, p_3 as a function of the SNR in a jointly-correlated system with $n_T = 3$, $n_R = 2$, and 16-QAM. The SER minimizing strategy is compared with the low-complexity approach proposed in Strategy 1.

Fig. 4. Upper bound on the beamforming optimality range (in SNR) as a function of the modulation size. A jointly-correlated system is considered with $n_T = 3$, $n_R = 2$, and different symbol constellations.

Fig. 3. Relative increase in SER, as a function of the SNR, when using the heuristic power allocation as compared to SER minimizing power allocation. A jointly-correlated system is considered with $n_T = 3$, $n_R = 2$, $M = 16$, and different symbol constellations.

A. Precoding Strategies

When the transmitter has statistical CSI, the structure of the SER minimizing precoding matrix was given by Theorem 2. The remaining convex power allocation problem can either be solved optimally using numerical methods or approximately using Strategy 1. Next, we evaluate the difference in performance and behavior between these strategies. To ensure repeatability, we consider the coupling matrix

$$\boldsymbol{\Omega} = \begin{bmatrix} 3.6 & 0.4 & 0.5 \\ 1 & 0.3 & 0.2 \end{bmatrix} \quad (19)$$

that was introduced in [34] and has $n_T = 3$ and $n_R = 2$. This coupling matrix represents a clearly non-Kronecker model scenario with one strong transmit eigendirection and two equally weak transmit directions with different spreads over the receive eigendirections. In Fig. 2, the optimal and heuristic power allocations are shown as functions of the SNR, γ. The symbol constellation is 16-QAM and observe that the total element sum in $\boldsymbol{\Omega}$ is normalized to $n_T n_R$. The difference between the two strategies is clearly visible; the heuristic strategy requires slightly higher SNR before allocating power in more than one direction and always gives $p_2 = p_3$, although the first of these directions is slightly advantageous.

The perceived difference between the optimal and heuristic strategy, in terms of the relative increase in SER when using the latter, is illustrated in Fig. 3. The performance loss is given for 16-PAM, 16-PSK, and 16-QAM as a function of the

SNR. At low and high SNR, the difference is nonexistent or negligible, while there is a peak in the area where the heuristic strategy uses beamforming although higher performance can be achieved by spatial multiplexing. The maximum relative performance loss is around 5 percent, which can be seen as validation of the heuristic strategy.

As observed in Fig. 2, the SER is minimized by allocating all power to the strongest eigendirection for a large range of SNRs. Theorem 2 proved that this type of single-stream beamforming is optimal at low SNR and the optimality range was characterized in Theorem 5. Next, we illustrate the upper bound of this range. In Fig. 4, the largest SNR that gives beamforming as the SER minimizing precoding is shown as a function of the modulation size, M, for PAM, PSK and QAM. As noted in Corollary 2, the upper bound increases with M and they will approach infinity together. In general, the beamforming optimality range is much wider for PAM and PSK, than for QAM. As we move towards modulations as 64-QAM and 128-QAM, the optimal precoding strategy is beamforming for most practical SNRs.

B. Impact of Spatial Correlation

Next, we illustrate the impact of spatial correlation when the transmitter has either no CSI or statistical CSI. As proved in Section IV, the power distribution in columns and rows of the coupling matrix affects the SER in different ways. To show this in a simple way, we consider a system with $n_T = n_R = 4$ that satisfies the Kronecker model. The antenna correlation follows the exponential model [35], which in principle models a uniform linear array (ULA) with the correlation between adjacent antenna elements as a parameter. The symbol constellation is 16-QAM and the coupling matrix is normalized such that the total element sum is $n_T n_R$.

In Fig. 5, we keep the transmit correlation fixed at 0.5, while the correlation between adjacent receive antennas changes between 0 and 1 (i.e., from completely uncorrelated to completely correlated). As expected from Theorem 3 and 4, the SER is a Schur-convex function with respect to the spatial correlation at all SNRs. Having statistical CSI improves the SER, especially at low and medium SNR, but the overall conclusion is that receive-side correlation always degrades the performance.

Fig. 5. The SER as a function of the correlation between adjacent receive antennas with either uniform (no transmit side CSI) or optimal precoding (statistical transmit-side CSI). The system uses 16-QAM and follows the Kronecker model with $n_T = n_R = 4$ and a transmit antenna correlation of 0.5.

Fig. 6. The SER as a function of the correlation between adjacent transmit antennas with either uniform (no transmit-side CSI) or optimal precoding (statistical transmit-side CSI). The system uses 16-QAM and follows the Kronecker model with $n_T = n_R = 4$ and a receive antenna correlation of 0.5.

Next, we keep the receive correlation fixed at 0.5 and vary the correlation between adjacent transmit antennas. This case is of special interest since Theorem 3 and 4 showed different behaviors depending on the SNR and available CSI. Without transmit-side CSI, Fig. 6 shows that the SER becomes a Schur-convex function at all SNRs. With statistical CSI, we observe the opposite behavior; the SER is a Schur-concave function, and thereby improves the performance with increasing correlation, in an SNR range that reaches up to 14 dB. For larger SNRs, there is a transition range where the SER is neither Schur-convex nor Schur-concave. At very high SNR, the SER becomes Schur-convex, but observe that it has already reached such low values (below 10^{-6}) that the dependence on the spatial correlation in principle is negligible. Thus, we conclude that with statistical transmit-side CSI, the SER is Schur-concave at low to medium SNRs and approximately Schur-concave at high SNRs.

VI. CONCLUSION

The optimal precoder and the impact of spatial correlation on the symbol error rate have been shown to depend strongly on the CSI available at the transmitter. The considered system was Rayleigh fading with OSTBC transmission, perfect CSI at the receiver side, and the SER was used as performance measure. If the transmitter has no CSI, then the optimal precoding strategy is to allocate the power equally over all eigendirections and thereby protect the system against the worst-case conditions. For this type of open-loop system, the intuitive result that spatial correlation degrades the performance was proven.

When the transmitter has statistical CSI, the transmission strategy can be adapted to the spatial correlation and exploit its advantages. While correlation increases the channel knowledge at the transmitter, it also decreases the degrees of freedom, and thus it is not intuitively clear how spatial correlation affects the SER. The analysis herein was based on the assumption of jointly-correlated channel statistics, which better complies with measurements than the Kronecker model previously used in this area. The optimal precoding strategy

exploits the channel eigendirections at the transmitter side and allocates power according to a convex optimization problem. While equal power allocation is optimal at very high SNR, it was shown that single-stream beamforming in general is optimal at low SNR and often at most practical SNRs. The beamforming optimality range was characterized and shown to increase with the size of the symbol constellation. Furthermore, a low-complexity algorithm for power allocation was proposed and it was illustrated that its performance loss compared with power optimal allocation is small.

For this type of closed-loop systems, it was proven that receive side correlation always degrades the performance, while transmit side correlation is favorable at low and medium SNR (including the beamforming optimality range) and has negligible impact at high SNR. In practice, these results impose difficult design considerations as the uplink-downlink channel reciprocity means that we cannot achieve an optimal fading environment in both the uplink and downlink.

APPENDIX A

The following lemma shows some behaviors of the class of functions introduced in Definition 3.

Lemma 1. The function $F_{\mathbf{a},\mathbf{b},\mathbf{c}}(\mathbf{\Phi}, x)$ is Schur-convex with respect to the eigenvalues of $\mathbf{\Phi}$ and it is a decreasing function in x. The function is also decreasing in $\mathrm{tr}(\mathbf{\Phi})$ for fixed eigenvalue spread.

Proof: First, by inserting (7) into (5), and multiplying with $\sin^2(\theta)/\sin^2(\theta)$, we achieve

$$F_{\mathbf{a},\mathbf{b},\mathbf{c}}(\mathbf{\Phi}, x) = \sum_{k=1}^{n} \frac{c_k}{\pi} \int_{a_k}^{b_k} \underbrace{\left(\prod_{i=1}^{mn_R} \frac{\sin^2(\theta)}{\sin^2(\theta) + x\lambda_i(\mathbf{\Phi})} \right)}_{\triangleq g(\theta, x, \lambda_1(\mathbf{\Phi}), \dots, \lambda_{mn_R}(\mathbf{\Phi}))} d\theta.$$
(20)

This expression resolves the ambiguity in intervals $[a_k, b_k]$ that contain θ such that $\sin^2(\theta) = 0$ and keeps the strict equality since these points have zero measure. Next, observe that the

integrand $g(\theta, x, \lambda_1(\boldsymbol{\Phi}), \ldots, \lambda_{mn_R}(\boldsymbol{\Phi}))$ is continuous for $x \geq 0$ and $\lambda_j(\boldsymbol{\Phi}) \geq 0$ for all j. The partial derivatives of $g(\cdot)$ are

$$
\begin{aligned}
\frac{\partial}{\partial x} g(\cdot) &= -g(\cdot) \sum_{j=1}^{mn_R} \frac{\lambda_j(\boldsymbol{\Phi})}{\sin^2(\theta) + x\lambda_j(\boldsymbol{\Phi})}, \\
\frac{\partial}{\partial \lambda_j(\boldsymbol{\Phi})} g(\cdot) &= -g(\cdot) \frac{x}{\sin^2(\theta) + x\lambda_j(\boldsymbol{\Phi})},
\end{aligned}
\tag{21}
$$

which both are continuous and negative. Hence, we use [36, Theorem 9.42] that states that differentiation of $F_{\mathbf{a},\mathbf{b},\mathbf{c}}(\boldsymbol{\Phi}, x)$ with respect to x and $\lambda_j(\boldsymbol{\Phi})$ can be determined by differentiation of the integrand (i.e., interchanging the order of integration and differentiation).

We should prove that the function is Schur-convex. According to Schur's condition [28, Theorem 3.A.4], a function $f(x_1, \ldots, x_N)$ is Schur-convex if and only if f is symmetric in its arguments and if

$$
(x_1 - x_2) \left(\frac{\partial f}{\partial x_1} - \frac{\partial f}{\partial x_2} \right) \geq 0.
\tag{22}
$$

Similarly, the function is Schur-concave if and only if it is symmetric in its arguments and (22) is fulfilled with the opposite inequality. Using (21), we have

$$
\begin{aligned}
&\frac{\partial}{\partial \lambda_j(\boldsymbol{\Phi})} F_{\mathbf{a},\mathbf{b},\mathbf{c}}(\boldsymbol{\Phi}, x) \\
&= -\sum_{k=1}^{n} \frac{c_k}{\pi} \int_{a_k}^{b_k} \frac{g(\theta, x, \lambda_1(\boldsymbol{\Phi}), \ldots, \lambda_{mn_R}(\boldsymbol{\Phi}))x}{\sin^2(\theta) + x\lambda_j(\boldsymbol{\Phi})} d\theta
\end{aligned}
\tag{23}
$$

which is negative since all components of the sum are positive (recall that $c_k \geq 0$ and $b_k \geq a_k$ by definition). Except from within the function $g(\cdot)$, the eigenvalue $\lambda_j(\boldsymbol{\Phi})$ only appears in numerators and hence the derivative with respect to $\lambda_1(\boldsymbol{\Phi})$ will be larger than for $\lambda_2(\boldsymbol{\Phi})$ if $\lambda_1(\boldsymbol{\Phi}) > \lambda_2(\boldsymbol{\Phi})$. Thus, Schur's condition in (22) is fulfilled and $F_{\mathbf{a},\mathbf{b},\mathbf{c}}(\boldsymbol{\Phi}, x)$ is Schur-convex with respect to the eigenvalues of $\boldsymbol{\Phi}$.

Next, by interchanging integration and differentiation order and using (21), we have that

$$
\begin{aligned}
&\frac{\partial}{\partial x} F_{\mathbf{a},\mathbf{b},\mathbf{c}}(\boldsymbol{\Phi}, x) \\
&= -\sum_{k=1}^{n} \frac{c_k}{\pi} \int_{a_k}^{b_k} \sum_{j=1}^{mn_R} \frac{g(\theta, x, \lambda_1(\boldsymbol{\Phi}), \ldots, \lambda_{mn_R}(\boldsymbol{\Phi}))\lambda_j(\boldsymbol{\Phi})}{\sin^2(\theta) + x\lambda_j(\boldsymbol{\Phi})} \\
&\leq 0
\end{aligned}
\tag{24}
$$

since the sum coefficients and the integrands are all positive. Hence, the function is decreasing in x.

Finally, let $\beta \triangleq \operatorname{tr}(\boldsymbol{\Phi})$ and define the normalized matrix $\widetilde{\boldsymbol{\Phi}} \triangleq \frac{\boldsymbol{\Phi}}{\beta}$ (with constant unit trace). We want to show how $F_{\mathbf{a},\mathbf{b},\mathbf{c}}(\boldsymbol{\Phi}, x)$ depends on the trace β. Observe that the function identically can be expressed as $F_{\mathbf{a},\mathbf{b},\mathbf{c}}(\beta\widetilde{\boldsymbol{\Phi}}, x) = F_{\mathbf{a},\mathbf{b},\mathbf{c}}(\widetilde{\boldsymbol{\Phi}}, \beta x)$, and thus increasing β is equivalent to increasing x. We conclude that the function is decreasing in $\operatorname{tr}(\boldsymbol{\Phi})$ for fixed eigenvalue spread in $\boldsymbol{\Phi}$. ∎

The concepts of the next lemma have previously been used to prove that optimal precoders diagonalize the channel statistics [6], [26], [32]. Herein, it is generalized for non-Kronecker model systems.

Lemma 2. Let $f(\cdot)$ be a Schur-convex function and consider the matrix $\mathbf{A} = \boldsymbol{\Lambda}_1 + \boldsymbol{\Lambda}_2(\mathbf{F} \otimes \mathbf{I})\boldsymbol{\Lambda}_3$, where $\boldsymbol{\Lambda}_j$ are positive semi-definite diagonal matrices for all j and $\mathbf{W}$ is positive semi-definite. Then,

$$
\min_{\mathbf{F}; \operatorname{tr}(\mathbf{F})=1} f([\lambda_1(\mathbf{A}), \ldots, \lambda_N(\mathbf{A})]^T)
\tag{25}
$$

can only be solved by matrices $\mathbf{F}$ that are diagonal.

Proof: Assume, for the purpose of contradiction, that there exist a non-diagonal optimal solution $\mathbf{F}_{\text{opt}}$. It can be expressed as $\mathbf{F}_{\text{opt}} = \mathbf{D} + \mathbf{B}$, where $\mathbf{D}$ is diagonal and $\mathbf{B} \neq \mathbf{0}$ has zero diagonal elements. Observe that

$$
\begin{aligned}
\mathbf{A} &= \boldsymbol{\Lambda}_1 + \boldsymbol{\Lambda}_2((\mathbf{D} + \mathbf{B}) \otimes \mathbf{I})\boldsymbol{\Lambda}_3 \\
&= \underbrace{\boldsymbol{\Lambda}_1 + \boldsymbol{\Lambda}_2(\mathbf{D} \otimes \mathbf{I})\boldsymbol{\Lambda}_3}_{\triangleq \mathbf{C}, \text{ diagonal}} + \underbrace{\boldsymbol{\Lambda}_2(\mathbf{B} \otimes \mathbf{I})\boldsymbol{\Lambda}_3}_{\text{zero diagonal}}.
\end{aligned}
\tag{26}
$$

Next, [28, Theorem 9.B.1] says that the eigenvalues of $\mathbf{A}$ majorizes the vector with diagonal elements of $\mathbf{A}$. Thus, the eigenvalues of $\mathbf{A}$ majorizes the eigenvalues of $\mathbf{C}$, which means that

$$
f([\lambda_1(\mathbf{A}), \ldots, \lambda_N(\mathbf{A})]^T) \geq f([\lambda_1(\mathbf{C}), \ldots, \lambda_N(\mathbf{C})]^T)
\tag{27}
$$

with equality if and only if $\mathbf{A} = \mathbf{C}$. Hence, $\mathbf{F}_{\text{opt}}$ can be replaced by $\mathbf{D}$ (with identical trace) that gives a lower function value. This optimality contradiction means that the solution must be diagonal. ∎

APPENDIX B

Proof of Theorem 1: In the case $B < n_T$, every $\mathbf{W}$ can be described by the singular value decomposition

$$
\mathbf{W} = \mathbf{V}_1 \begin{bmatrix} \mathbf{D} \\ \mathbf{0} \end{bmatrix} \mathbf{V}_2
\tag{28}
$$

for some diagonal $\mathbf{D} \in \mathbb{C}^{B \times B}$ and unitary $\mathbf{V}_1 \in \mathbb{C}^{n_T \times n_T}$ and $\mathbf{V}_2 \in \mathbb{C}^{B \times B}$. By selecting

$$
\mathbf{R} = (\mathbf{V}_1^* \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A} \end{bmatrix} \mathbf{V}_1^T \otimes \mathbf{I})
\tag{29}
$$

for some arbitrary $\mathbf{A} \in \mathbb{C}^{n_T - B \times n_T - B}$ that fulfills $\operatorname{tr}(\mathbf{A}) = n_T$, we achieve

$$
\begin{aligned}
\boldsymbol{\Phi} &= (\mathbf{W}^T \otimes \mathbf{I})\mathbf{R}(\mathbf{W}^T \otimes \mathbf{I})^H \\
&= (\mathbf{V}_2^T[\mathbf{D}^T \; \mathbf{0}] \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A} \end{bmatrix} \begin{bmatrix} \mathbf{D}^* \\ \mathbf{0} \end{bmatrix} \mathbf{V}_2^* \otimes \mathbf{I}) = \mathbf{0}
\end{aligned}
\tag{30}
$$

which is the global maximum of the outer optimization in (10) since Lemma 1 states that $F_{\mathbf{a},\mathbf{b},\mathbf{c}}(\boldsymbol{\Phi}, x)$ increases with decreasing $\operatorname{tr}(\boldsymbol{\Phi})$. If $B < n_T$, we can thus achieve the global maximum of $F_{\mathbf{a},\mathbf{b},\mathbf{c}}(\boldsymbol{\Phi}, x)$ by worst-case selection of $\mathbf{R}$ for any choice of $\mathbf{W}$.

In the case $B \geq n_T$, we first analyze the maximal function value with respect to $\mathbf{R}$ for a given $\mathbf{W}$. Then, we identify the matrix $\mathbf{W}$ that gives the smallest maximized value. Let the singular value decomposition of $\mathbf{W}$ be denoted $\mathbf{W} = \mathbf{V}_1[\mathbf{D} \; \mathbf{0}]\mathbf{V}_2$, where $\mathbf{D} = \operatorname{diag}(d_1, \ldots, d_{n_T})$ has elements ordered with decreasing magnitude and $\mathbf{V}_1 \in \mathbb{C}^{n_T \times n_T}$, $\mathbf{V}_2 \in \mathbb{C}^{B \times B}$ are unitary matrices. Lemma 1 showed that $F_{\mathbf{a},\mathbf{b},\mathbf{c}}(\boldsymbol{\Phi}, x)$ increases with decreasing $\operatorname{tr}(\boldsymbol{\Phi})$, and thus we want to maximize

$$
\begin{aligned}
\operatorname{tr}(\boldsymbol{\Phi}) &= \operatorname{tr}((\mathbf{W}^T \otimes \mathbf{I})\mathbf{R}(\mathbf{W}^T \otimes \mathbf{I})^H) \\
&= \operatorname{tr}(\widetilde{\mathbf{R}}(\mathbf{D}^*\mathbf{D}^T \otimes \mathbf{I}))
\end{aligned}
\tag{31}
$$

where $\widetilde{\mathbf{R}} = (\mathbf{V}_1^T \otimes \mathbf{I})\mathbf{R}(\mathbf{V}_1^* \otimes \mathbf{I})$ and we used Schur determinant lemma [37]. According to [28, Theorem 20.A.4], (31) is minimized when $\widetilde{\mathbf{R}}$ is diagonal and ordered such that its smallest elements are multiplied with the largest elements of $\mathbf{D}^*\mathbf{D}^T \otimes \mathbf{I}$, and vice versa. Thus, the remaining optimization is

$$\min_{\mathbf{R};\,\mathrm{tr}(\mathbf{R})=n_T n_R} \sum_{j=1}^{n_T} |d_{n_T-j+1}|^2 \left(\sum_{l=1}^{n_R} \lambda_{l+(j-1)n_T}(\mathbf{R}) \right) \tag{32}$$
$$= n_T n_R |d_{n_T}|^2$$

which is clearly minimized when all power of the correlation matrix is in direction of the weakest $|d_j|^2$. Finally, we want to select the precoding matrix $\mathbf{W}$ that yields the best worst-case performance, or in other words maximizes the worst-case expression for $\mathrm{tr}(\mathbf{\Phi})$ in (32). We observe that this precoding matrix should fulfill that all diagonal elements d_j have the same magnitude; that is, $\mathbf{D} = \sqrt{1/n_T}\mathbf{I}$.

Proof of Theorem 2: For the jointly-correlated model, the correlation matrix becomes $\mathbf{R} = (\mathbf{U}_T^* \otimes \mathbf{U}_R)\mathrm{diag}(\mathrm{vec}(\mathbf{\Omega}))(\mathbf{U}_T^* \otimes \mathbf{U}_R)^H$. For simplicity, we will use the notation $\mathbf{\Lambda} \triangleq \mathrm{diag}(\mathrm{vec}(\mathbf{\Omega}))$. Suppose that the solution to (12) is denoted $\mathbf{W}_{\mathrm{opt}}$ and let its singular value decomposition be $\mathbf{W}_{\mathrm{opt}} = \mathbf{U}\mathbf{\Delta}\mathbf{V}$ for some rectangular diagonal matrix $\mathbf{\Delta}$ and unitary matrices $\mathbf{U}$, $\mathbf{V}$. We will first show that $\mathbf{U} = \mathbf{U}_T$.

If we find a $\mathbf{W}$ that minimizes the integrand of $F_{\mathbf{a},\mathbf{b},\mathbf{c}}(\mathbf{\Phi}, x)$ in (5) for all x and θ, then we have reached the global minimum of the function. Let $\mathbf{A} = \mathbf{I} + (x/\sin^2(\theta))\mathbf{\Phi}$, then the integrand becomes $1/\det(\mathbf{A}) = 1/(\prod_{j=1}^{mn_R} \lambda_j(\mathbf{A}))$, which is a Schur-convex function with respect to the eigenvalues of $\mathbf{A}$ since the partial derivative

$$\frac{\partial}{\partial \lambda_l(\mathbf{A})} \prod_{j=1}^{mn_R} \frac{1}{\lambda_j(\mathbf{A})} = -\frac{1}{\lambda_l(\mathbf{A})} \prod_{j=1}^{mn_R} \frac{1}{\lambda_j(\mathbf{A})} \tag{33}$$

satisfies Schur's condition in (22) for Schur-convexity. Now, we simplify the integrand as

$$\frac{1}{\det\left(\mathbf{I} + \frac{x}{\sin^2(\theta)}\mathbf{\Phi}\right)}$$
$$= \det^{-1}\left(\mathbf{I} + \frac{x}{\sin^2(\theta)}(\mathbf{W}_{\mathrm{opt}}^T \otimes \mathbf{I})(\mathbf{U}_T^* \otimes \mathbf{U}_R)\mathbf{\Lambda} \right.$$
$$\left. \times (\mathbf{U}_T^* \otimes \mathbf{U}_R)^H(\mathbf{W}_{\mathrm{opt}}^T \otimes \mathbf{I})^H \right) \tag{34}$$
$$= \det^{-1}\left(\mathbf{I} + \frac{x}{\sin^2(\theta)}\mathbf{\Lambda}(\underbrace{\mathbf{U}_T^T \mathbf{U}^* \mathbf{\Delta}^* \mathbf{\Delta}^T \mathbf{U}^T \mathbf{U}_T^*}_{=\mathbf{F}} \otimes \mathbf{I})\right),$$

where the second equality follows from the singular value decomposition of $\mathbf{W}_{\mathrm{opt}}$ and from using Schur's determinant lemma [37], $\det(\mathbf{I} + \mathbf{B}\mathbf{C}) = \det(\mathbf{I} + \mathbf{C}\mathbf{B})$. Since we have shown that the expression in (34) is Schur-convex, we can apply Lemma 2 on the last expression in (34) and conclude that $\mathbf{F}$ needs to be diagonal in order for $\mathbf{W}_{\mathrm{opt}}$ to be an optimal solution. Since $\mathbf{\Delta}$ is a rectangular diagonal matrix, $\mathbf{\Delta}\mathbf{\Delta}^H$ will be diagonal, and therefore $\mathbf{U}_T^H \mathbf{U}$ needs to be diagonal. Thus, $\mathbf{U} = \mathbf{U}_T \mathbf{\Pi}$, where the permutation matrix $\mathbf{\Pi}$ decides which singular value of $\mathbf{\Delta}$ that belongs to each of the eigenvectors in $\mathbf{U}_T$.

Using the optimal precoder structure, each integral in (12) can be expressed as

$$\int_{a_k}^{b_k} \frac{d\theta}{\det\left(\mathbf{I} + \frac{x}{\sin^2(\theta)}\mathbf{\Phi}\right)} = \int_{a_k}^{b_k} \prod_{l=1}^{n_R} \prod_{j=1}^{m} \frac{1}{1 + \frac{x}{\sin^2(\theta)}\omega_{l,\pi(j)}p_j}d\theta, \tag{35}$$

where $\omega_{l,j}$ is the (l,j)th element of $\mathbf{\Omega}$. The integral is a convex function of each p_j. At large x, we have

$$\int_{a_k}^{b_k} \prod_{l=1}^{n_R} \prod_{j=1}^{m} \frac{1}{1 + \frac{x}{\sin^2(\theta)}\omega_{l,\pi(j)}p_j}d\theta$$
$$\approx \prod_{j=1}^{m} \frac{1}{p_j^{n_R}}\left(\prod_{l=1}^{n_R}\prod_{j=1}^{m}\frac{1}{\omega_{l,\pi(j)}}\right)\int_{a_k}^{b_k}\left(\frac{\sin^2(\theta)}{x}\right)^{mn_R}d\theta. \tag{36}$$

The second factor is minimized by a permutation matrix that orders the columns of $\mathbf{\Omega}$ with decreasing element product. Since $1/(\prod_j p_j^{n_R})$ is a Schur-convex function, it is minimized by equal power allocation [29, Theorem 2.21]. Thus, each term of $F_{\mathbf{a},\mathbf{b},\mathbf{c}}(\mathbf{\Phi}, x)$ is minimized by the same power allocation and permutation matrix and we have reached the global minimum.

To prove the behavior at small x, observe that $\min_{\mathbf{W};\, \|\mathbf{W}\|^2=1} F_{\mathbf{a},\mathbf{b},\mathbf{c}}(\mathbf{\Phi}, x) = \min_{\mathbf{W};\, \|\mathbf{W}\|^2=x} F_{\mathbf{a},\mathbf{b},\mathbf{c}}(\mathbf{\Phi}, 1)$. Now, suppose that $\pi(j) = j$ for $j \leq \tilde{m}$. Let the power allocation be parameterized as $p_j = \beta_j(x - t)$ for $j \leq \tilde{m}$ and $p_j = \alpha_j t$ for $j > \tilde{m}$, where $\alpha_j \geq 0$ and $\beta_j \geq 0$ are arbitrary coefficients that fulfill $\sum_{j>\tilde{m}} \alpha_j = 1$ and $\sum_{j\leq\tilde{m}} \beta_j = 1$, respectively. It is straightforward to show that $F_{\mathbf{a},\mathbf{b},\mathbf{c}}(\mathbf{\Phi}, 1)$ is convex with respect to t. Hence, in order to show that $t = 0$ (i.e., selective power allocation to strongest direction, with multiplicity) minimizes the SER for small values on x, it is sufficient to show that $(\partial/\partial t)F_{\mathbf{a},\mathbf{b},\mathbf{c}}(\mathbf{\Phi}, 1) > 0$ at $t = 0$ for such x. Recall from the proof of Lemma 1 that we can interchange differentiation and integration. Define

$$g_{l,j}(p) \triangleq \frac{1}{1 + \frac{\omega_{l,j}p}{\sin^2(\theta)}}. \tag{37}$$

Using the parametrization in t, the derivative of (41) at $t = 0$ is

$$\frac{\partial}{\partial t}\int_{a_k}^{b_k}\prod_{l=1}^{n_R}\left(\prod_{j=1}^{\tilde{m}}g_{l,j}\left(\beta_j(x-t)\right)\prod_{j=\tilde{m}+1}^{m}g_{l,\pi(j)}(\alpha_j t)\right)d\theta\Big|_{t=0}$$
$$= \sum_{\tilde{l}=1}^{n_R}\int_{a_k}^{b_k}\left(\sum_{\tilde{j}=1}^{\tilde{m}}\frac{\beta_{\tilde{j}}\omega_{\tilde{l},\tilde{j}}g_{\tilde{l},\tilde{j}}(\beta_{\tilde{j}}x)}{\sin^2(\theta)} - \sum_{\tilde{j}=\tilde{m}+1}^{m}\frac{\alpha_{\tilde{j}}\omega_{\tilde{l},\pi(\tilde{j})}}{\sin^2(\theta)}\right)$$
$$\times \prod_{l=1}^{n_R}\prod_{j=1}^{\tilde{m}}g_{l,j}(\beta_j x)d\theta$$
$$\geq \sum_{\tilde{l}=1}^{n_R}\int_{a_k}^{b_k}\left(\frac{\sum_{\tilde{j}=1}^{\tilde{m}}\beta_{\tilde{j}}\omega_{\tilde{l},\tilde{j}}g_{\tilde{l},\tilde{j}}(\beta_{\tilde{j}}x) - \omega_{\tilde{l},\tilde{m}+1}}{\sin^2(\theta)}\right)$$
$$\times \prod_{l=1}^{n_R}\prod_{j=1}^{\tilde{m}}g_{l,j}(\beta_j x)d\theta, \tag{38}$$

where the inequality follows from selecting $\alpha_{\tilde{m}+1} = 1$ and $\alpha_j = 0$ for $j > \tilde{m} + 1$ (i.e., placing all power in t on the

column among $j = \tilde{m} + 1, \ldots, n_T$ that maximizes $\sum_l \omega_{l,j}$). At $x = 0$, the derivative is strictly positive for all $\beta_1, \ldots, \beta_{\tilde{m}}$ and since it is a continuous function, it will remain positive for a certain interval of small values on x. In other words, each sum component of $F_{\mathbf{a},\mathbf{b},\mathbf{c}}(\mathbf{\Phi}, x)$ will increase with t for small x and thus the function is minimized by $t = 0$.

Finally, under the Kronecker model it holds for $j_1 < j_2$ that $\omega_{l,j_1} \geq \omega_{l,j_2}$ for all l. Hence, each integral in (41) is minimized when $\mathbf{\Pi} = \mathbf{I}$. In addition, if two columns in $\mathbf{\Omega}$ have an identical element sum, then the columns will be identical. Thus, Lemma 1 gives equal power allocation among them.

Proof of Theorem 4: First, we consider the Schur-convexity properties with respect to $\bar{\omega}_{1,j}, \ldots, \bar{\omega}_{n_R,j}$ for a given j. Using [6, Lemma 2] and that $F_{\mathbf{a},\mathbf{b},\mathbf{c}}(\mathbf{\Phi}, x)$ is continuous and twice continuously differentiable, we know that partial derivatives of the optimization problem in (12) can be calculated by evaluating the derivative of $F_{\mathbf{a},\mathbf{b},\mathbf{c}}(\mathbf{\Phi}, x)$ at the optimal solution. Thus, we have

$$\frac{\partial}{\partial \bar{\omega}_{l,\pi(j)}} \min_{\mathbf{W} \in \mathbb{C}^{n_T \times B}; \, \|\mathbf{W}\|^2 = 1} F_{\mathbf{a},\mathbf{b},\mathbf{c}}(\mathbf{\Phi}, x)$$
$$= -\sum_{k=1}^{n} \frac{c_k}{\pi} \int_{a_k}^{b_k} \frac{1}{\det\left(\mathbf{I} + \frac{x}{\sin^2(\theta)}\mathbf{\Phi}\right)} \frac{\frac{x}{\sin^2(\theta)}\mu_{\pi(j)}}{1 + \frac{x}{\sin^2(\theta)}\bar{\omega}_{l,\pi(j)}\mu_{\pi(j)}} d\theta \tag{39}$$

for $j = 1, \ldots, m$, which is negative and only contains $\bar{\omega}_{l,\pi(j)}$ in a denominator (and within the determinant). Hence, the derivative is smaller for $\bar{\omega}_{l_2,\pi(j)}$ than for $\bar{\omega}_{l_1,\pi(j)}$ if $\bar{\omega}_{l_1,\pi(j)} \geq \bar{\omega}_{l_2,\pi(j)}$. Recall from (22) that a function is Schur-convex with respect to $\bar{\omega}_{1,\pi(j)}, \ldots, \bar{\omega}_{n_R,\pi(j)}$ if $\bar{\omega}_{l_1,\pi(j)} \geq \bar{\omega}_{l_2,\pi(j)}$ implies that the corresponding function derivatives satisfies the same inequality. Thus, $F_{\mathbf{a},\mathbf{b},\mathbf{c}}(\mathbf{\Phi}, x)$ is Schur-convex.

At large x, equal power allocation is the optimal precoding strategy and each integral in $F_{\mathbf{a},\mathbf{b},\mathbf{c}}(\mathbf{\Phi}, x)$ can be approximated according to (36) as

$$\int_{a_k}^{b_k} \frac{d\theta}{\det\left(\mathbf{I} + \frac{x}{\sin^2(\theta)}\mathbf{\Phi}\right)}$$
$$\approx m^{n_R m}\left(\prod_{l=1}^{n_R}\prod_{j=1}^{m} \frac{1}{\bar{\omega}_{l,\pi(j)}}\right)\left(\prod_{j=1}^{m} \frac{1}{\mu_{\pi(j)}^{n_R}}\right)\int_{a_k}^{b_k} \frac{\sin^{2mn_R}(\theta)}{x^{mn_R}} d\theta. \tag{40}$$

This is a Schur-convex function with respect to $\mu_{\pi(1)}, \ldots, \mu_{\pi(m)}$ [28, Theorem 3.A.4].

At small x, all power is allocated in the direction with the largest μ_j. In general, this direction is distinct and each integral in $F_{\mathbf{a},\mathbf{b},\mathbf{c}}(\mathbf{\Phi}, x)$ can be expressed as

$$\int_{a_k}^{b_k} \frac{d\theta}{\det\left(\mathbf{I} + \frac{x}{\sin^2(\theta)}\mathbf{\Phi}\right)} = \int_{a_k}^{b_k} \prod_{l=1}^{n_R} \frac{1}{1 + \frac{x}{\sin^2(\theta)}\bar{\omega}_{l,1}\mu_1} d\theta. \tag{41}$$

This is a decreasing function with respect to μ_1 and independent of all other μ_j. Thus, it is a Schur-concave function with respect to the ordered vector with $\mu_1, \ldots, \mu_{n_T}$.

Finally, under the Kronecker model, observe that $\bar{\omega}_{l,1} = \ldots = \bar{\omega}_{l,n_T}$ for all l. Since $\mathbf{\Omega} = \boldsymbol{\lambda}_R \boldsymbol{\lambda}_T^T$, this implies that $\boldsymbol{\lambda}_R$ and $[\bar{\omega}_{1,j}, \ldots, \bar{\omega}_{n_R,j}]^T$ are identical (up to a scaling factor) and it is straightforward to show that the function is also Schur-convex with respect to $\boldsymbol{\lambda}_R$. In the same way, we conclude that

$\boldsymbol{\lambda}_T$ and $\mu_1, \ldots, \mu_{n_T}$ are identical (up to a scaling factor) and share the same Schur-convexity properties.

Proof of Theorem 5: The proof uses the same type of parametrization as in the "small x"-part of the proof of Theorem 2. First, we make sure that the eigendirection with the largest column sum in $\mathbf{\Omega}$ (or one of them) is among the active directions by selecting $\pi(1) = 1$. We parameterize as $p_1 = x - t$ and $p_j = \alpha_j t$ for $j > 1$, where $\alpha_j \geq 0$ are arbitrary coefficients that fulfill $\sum_{j>1} \alpha_j = 1$. The function $F_{\mathbf{a},\mathbf{b},\mathbf{c}}(\mathbf{\Phi}, x)$ is convex with respect to t and thus a necessary and sufficient condition for beamforming optimality is $\frac{\partial}{\partial t} F_{\mathbf{a},\mathbf{b},\mathbf{c}}(\mathbf{\Phi}, x)|_{t=0} \geq 0$. Similar to (38), straightforward differentiation of $F_{\mathbf{a},\mathbf{b},\mathbf{c}}(\mathbf{\Phi}, x)$ with respect to t yields

$$\sum_{k=1}^{n} \frac{c_k}{\pi} \int_{a_k}^{b_k} \left(\sum_{l=1}^{n_R} \frac{\omega_{l,\pi(1)}}{\sin^2(\theta) + x\omega_{l,\pi(1)}} - \sum_{j=2}^{m} \frac{\alpha_j \omega_{l,\pi(j)}}{\sin^2(\theta)}\right)$$
$$\times \frac{d\theta}{\det\left(\mathbf{I} + \frac{x}{\sin^2(\theta)}\mathbf{A}\right)} \geq 0. \tag{42}$$

This condition needs to be fulfilled for any set of α_j. Therefore, the expression in (18) is achieved by replacing $\sum_{j=2}^{m}(\alpha_j \omega_{l,\pi(j)})/\sin^2(\theta)$ by its maximum, achieved by $\pi(2) = 2$, $\alpha_2 = 1$ and $\alpha_j = 0$ for $j > 2$. Finally, consider the initial assumption that $\pi(1) = 1$. If this is not case, then the second sum within the integral in (42) is larger than the first sum and the beamforming optimality condition can never be fulfilled.

REFERENCES

[1] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, pp. 311–335, 1998.

[2] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, 1999.

[3] J. Kermoal, L. Schumacher, K. Pedersen, P. Mogensen, and F. Frederiksen, "A stochastic MIMO radio channel model with experimental validation," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 6, pp. 1211–1226, 2002.

[4] D. Chizhik, J. Ling, P. Wolniansky, R. Valenzuela, N. Costa, and K. Huber, "Multiple-input-multiple-output measurements and modeling in Manhattan," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 3, pp. 321–331, 2003.

[5] K. Yu, M. Bengtsson, B. Ottersten, D. McNamara, P. Karlsson, and M. Beach, "Modeling of wide-band MIMO radio channels based on NLoS indoor measurements," *IEEE Trans. Veh. Technol.*, vol. 53, no. 3, pp. 655–665, 2004.

[6] E. Björnson and B. Ottersten, "A framework for training-based estimation in arbitrarily correlated Rician MIMO channels with Rician disturbance," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1807–1820, 2010.

[7] A. Narula, M. Lopez, M. Trott, and G. Wornell, "Efficient use of side information in multiple-antenna data transmission over fading channels," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1423–1436, 1998.

[8] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2632–2639, 2001.

[9] M. Ivrlac, W. Utschick, and J. Nossek, "Fading correlations in wireless MIMO communication systems," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 819–828, 2003.

[10] E. Jorswieck and H. Boche, "Optimal transmission strategies and impact of correlation in multiantenna systems with different types of channel state information," *IEEE Trans. Signal Process.*, vol. 52, no. 12, pp. 3440–3453, 2004.

[11] E. Björnson, D. Hammarwall, and B. Ottersten, "Exploiting quantized channel norm feedback through conditional statistics in arbitrarily correlated MIMO systems," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 4027–4041, 2009.

[12] A. Zanella, M. Chiani, and M. Win, "Performance of MIMO MRC in correlated Rayleigh fading environments," in *Proc. IEEE VTC'05-Spring*, 2005, pp. 1633–1637.

[13] M. McKay, A. Grant, and I. Collings, "Performance analysis of MIMO-MRC in double-correlated Rayleigh environments," *IEEE Trans. Commun.*, vol. 55, no. 3, pp. 497–507, 2007.

[14] S. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, 1998.

[15] V. Tarokh, H. Jafarkhani, and A. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1456–1467, 1999.

[16] G. Jöngren, M. Skoglund, and B. Ottersten, "Combining beamforming and orthogonal space-time block coding," *IEEE Trans. Inf. Theory*, vol. 48, no. 3, pp. 611–627, 2002.

[17] X.-B. Liang, "Orthogonal designs with maximal rates," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2468–2503, 2003.

[18] K. Lu, S. Fu, and X.-G. Xia, "Closed-form designs of complex orthogonal space-time block codes of rates $(k + 1)/(2k)$ for $2k - 1$ or $2k$ transmit antennas," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4340–4347, 2005.

[19] W. Weichselberger, M. Herdin, H. Özcelik, and E. Bonek, "A stochastic MIMO channel model with joint correlation of both link ends," *IEEE Trans. Wireless Commun.*, vol. 5, no. 1, pp. 90–100, 2006.

[20] X. Gao, B. Jiang, X. Li, A. Gershman, and M. McKay, "Statistical eigenmode transmission over jointly-correlated MIMO channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3735–3750, 2009.

[21] E. Björnson, B. Ottersten, and E. Jorswieck, "On the impact of spatial correlation and precoder design in MIMO systems with space-time block coding," in *Proc. IEEE ICASSP'09*, 2009, pp. 2741–2744.

[22] Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation (Release 8), 3GPP TS 36.211, Dec. 2008.

[23] IEEE Std 802.11n-2009; Amendment 5: Enhancements for Higher Throughput, IEEE-SA Standards Board, Oct. 2009.

[24] E. Larsson and P. Stoica, *Space-Time Block Coding for Wireless Communications*. Cambridge University Press, 2003.

[25] S. Sandhu and A. Paulraj, "Space-time block codes: a capacity perspective," *IEEE Commun. Lett.*, vol. 4, no. 12, pp. 384–386, 2000.

[26] A. Hjørungnes and D. Gesbert, "Precoding of orthogonal space-time block codes in arbitrarily correlated MIMO channels: iterative and closed-form solutions," *IEEE Trans. Wireless Commun.*, vol. 6, no. 3, pp. 1072–1082, 2007.

[27] R. Ertel, P. Cardieri, K. Sowerby, T. Rappaport, and J. Reed, "Overview of spatial channel models for antenna array communication systems," *IEEE Personal Commun. Mag.*, vol. 5, no. 1, pp. 10–22, 1998.

[28] A. Marshall and I. Olkin, *Inequalities: Theory of Majorization and its Applications*. Academic Press, 1979.

[29] E. Jorswieck and H. Boche, "Majorization and matrix-monotone functions in wireless communications," *Foundations and Trends in Commun. and Inf. Theory*, vol. 3, no. 6, pp. 553–701, 2007.

[30] M. Simon and M. Alouini, "A unified approach to the performance analysis of digital communication over generalized fading channels," *Proc. IEEE*, vol. 86, no. 9, pp. 1860–1877, 1998.

[31] A. Lozano and N. Jindal, "Transmit diversity vs. spatial multiplexing in modern MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 9, no. 1, pp. 186–197, 2010.

[32] K. Phan, S. Vorobyov, and C. Tellambura, "Precoder design for space-time coded MIMO systems with correlated Rayleigh fading channels," in *Proc. CCECE'07*, 2007.

[33] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[34] A. Tulino, A. Lozano, and S. Verdú, "Impact of antenna correlation on the capacity of multiantenna channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2491–2509, 2005.

[35] S. Loyka, "Channel capacity of MIMO architecture using the exponential correlation matrix," *IEEE Commun. Lett.*, vol. 5, no. 9, pp. 369–371, 2001.

[36] W. Rudin, *Principles of Mathematical Analysis*. McGraw-Hill, 1976.

[37] F. Zhang, *The Schur Complement and its Applications*. Springer, 2005.

Emil Björnson (S'07) was born in Malmö, Sweden, in 1983. He received the M.S. degree in engineering mathematics from Lund University, Lund, Sweden, in 2007. He is currently working towards the Ph.D. degree in telecommunications at the Signal Processing Laboratory, KTH Royal Institute of Technology, Stockholm, Sweden. His research interests include wireless communications, resource allocation, estimation theory, stochastic signal processing, and mathematical optimization. For his work on MIMO communications, he received a Best Paper Award at the 2009 International Conference on Wireless Communications and Signal Processing (WCSP 2009).

Eduard A. Jorswieck (S'01-M'05-SM'08) received his Diplom-Ingenieur degree and Doktor-Ingenieur (Ph.D.) degree, both in electrical engineering and computer science from the Berlin University of Technology (TUB), Germany, in 2000 and 2004, respectively. He was with the Fraunhofer Institute for Telecommunications, Heinrich-Hertz-Institute (HHI) Berlin, from 2001 to 2006. In 2006, he joined the Signal Processing Department at the KTH Royal Institute of Technology as a post-doc and became a Assistant Professor in 2007.

Since February 2008, he has been the head of the Chair of Communications Theory and Full Professor at Dresden University of Technology (TUD), Germany. His research interests are within the areas of applied information theory, signal processing and wireless communications. He is senior member of IEEE and elected member of the IEEE SPCOM Technical Committee. From 2008-2011 he serves as an Associate Editor for IEEE SIGNAL PROCESSING LETTERS. In 2006, he was co-recipient of the IEEE Signal Processing Society Best Paper Award.

Björn Ottersten (S'87-M'89-SM'99-F'04) was born in Stockholm, Sweden, in 1961. He received the M.S. degree in electrical engineering and applied physics from Linköping University, Linköping, Sweden, in 1986 and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1989.

He has held research positions at the Department of Electrical Engineering, Linköping University; the Information Systems Laboratory, Stanford University; and the Katholieke Universiteit Leuven, Leuven, Belgium. During 1996–1997, he was Director of Research at ArrayComm Inc., San Jose, CA, a start-up company based on Otterstens patented technology. In 1991, he was appointed Professor of Signal Processing at the KTH Royal Institute of Technology, Stockholm, Sweden. From 2004 to 2008, he was Dean of the School of Electrical Engineering at KTH, and from 1992 to 2004 he was head of the Department for Signals, Sensors, and Systems at KTH. He is also Director of security and trust at the University of Luxembourg. His research interests include wireless communications, stochastic signal processing, sensor array processing, and time-series analysis.

Dr. Ottersten has coauthored papers that received an IEEE Signal Processing Society Best Paper Award in 1993, 2001, and 2006. He has served as Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and on the Editorial Board of the *IEEE Signal Processing Magazine*. He is currently Editor-in-Chief of the *EURASIP Signal Processing Journal* and a member of the Editorial Board of the *EURASIP Journal of Advances in Signal Processing*. He is a Fellow of IEEE and EURASIP. He is one of the first recipients of the European Research Council advanced research grant.