(Article begins on next page)

Generic Correlation Increases Noncoherent MIMO Capacity

Günther Koliander[1], Erwin Riegler[1], Giuseppe Durisi[2], and Franz Hlawatsch[1]

[1]Institute of Telecommunications, Vienna University of Technology, 1040 Vienna, Austria
[2]Department of Signals and Systems, Chalmers University of Technology, 41296 Gothenburg, Sweden

***Abstract*—We study the high-SNR capacity of MIMO Rayleigh block-fading channels in the noncoherent setting where neither transmitter nor receiver has *a priori* channel state information. We show that when the number of receive antennas is sufficiently large and the temporal correlation within each block is "generic" (in the sense used in the interference-alignment literature), the capacity pre-log is given by $T(1-1/N)$ for $T<N$, where T denotes the number of transmit antennas and N denotes the block length. A comparison with the widely used constant block-fading channel (where the fading is constant within each block) shows that for a large block length, generic correlation increases the capacity pre-log by a factor of about four.**

I. Introduction

The throughput achievable with multiple-input multiple-output (MIMO) wireless systems is limited by the need to acquire channel state information (CSI) [1]. A fundamental way to assess the corresponding rate penalty is to study capacity in the *noncoherent setting* where neither the transmitter nor the receiver has *a priori* CSI.

We consider a MIMO system with T transmit antennas and R receive antennas. In the widely used *constant block-fading channel model* [2], the fading process takes on independent realizations across blocks of N channel uses ("block-memoryless" assumption), and within each block the fading coefficients are constant. Thus, the N-dimensional channel gain vector describing the channel between antennas t and r (hereafter briefly termed "(t,r) channel") within a block is

$$\mathsf{h}_{r,t} = \mathsf{s}_{r,t}\,\mathbf{1}_N\,. \tag{1}$$

Here, $\mathbf{1}_N$ denotes the N-dimensional all-one vector and $\{\mathsf{s}_{r,t}\}_{r\in\{1,\dots,R\},\,t\in\{1,\dots,T\}}$ are independent $\mathcal{CN}(0,1)$ random variables. Unfortunately, even for this simple channel model, a closed-form expression of noncoherent capacity is unavailable. However, an accurate characterization exists for high signal-to-noise ratio (SNR) values. In [3], it was shown that the capacity pre-log (i.e., the asymptotic ratio between capacity and the logarithm of the SNR as the SNR grows large) for the constant block-fading model is given by

$$\chi_{\text{const}} = M\left(1-\frac{M}{N}\right), \quad \text{with } M=\min\{T,R,\lfloor N/2\rfloor\}. \tag{2}$$

This work was supported by the WWTF under grant ICT10-066 (NOWIRE).

A more detailed high-SNR capacity expansion was obtained in [3] for the case $R+T\le N$; this expansion was recently extended in [4] to the large-MIMO setting $R+T>N$.

One limitation of the constant block-fading model is that it fails to describe a specific setting where block-fading models are of interest, namely, cyclic-prefix orthogonal frequency division multiplexing (CP-OFDM) systems [5]. In such systems, the channel input-output relation is most conveniently described in the frequency domain; the vector of channel gains $\mathsf{h}_{r,t}$ is then equal to the Fourier transform of the discrete-time impulse response of the (t,r) channel. Let us assume that $\mathsf{h}_{r,t}$ changes independently across blocks of length N and that

$$\mathsf{h}_{r,t} = \mathsf{s}_{r,t}\,\boldsymbol{z}_{r,t}\,, \tag{3}$$

where $\boldsymbol{z}_{r,t}$ is a deterministic vector whose squared inverse Fourier transform equals the power-delay profile of the (t,r) channel and, as before, $\{\mathsf{s}_{r,t}\}_{r\in\{1,\dots,R\},\,t\in\{1,\dots,T\}}$ are independent $\mathcal{CN}(0,1)$ random variables. As the vectors $\boldsymbol{z}_{r,t}$ are related to power-delay profiles, it is reasonable to assume that they are different for different (t,r). Note that the constant block-fading model (1) is a special case of (3) in which the impulse response of each (t,r) channel consists of only a single tap, a case for which the use of OFDM is unnecessary.

Contributions: We study the capacity pre-log (hereafter briefly termed "pre-log") of MIMO block-fading channels modeled as in (3). We show that when the deterministic vectors $\{\boldsymbol{z}_{r,t}\}$ are *generic*,[1] the pre-log can be larger than the pre-log in the constant block-fading case as given in (2). Specifically, we show that for the *generic block-fading model* (i.e., the model (3) with generic vectors $\{\boldsymbol{z}_{r,t}\}$), when $T<N$ and the number of receive antennas is sufficiently large such that $R\ge T(N-1)/(N-T)$, the pre-log is given by

$$\chi_{\text{gen}} = T\left(1-\frac{1}{N}\right). \tag{4}$$

For large N, the highest achievable χ_{gen} (with appropriately chosen T and R) is about four times as large as the highest achievable χ_{const}. As we will demonstrate, this is because under the generic block-fading model, the received signal vectors in the absence of noise span a subspace of higher dimension than under the constant block-fading model.

[1]We use the term "generic" in the same sense as in the interference-alignment literature [6]. A rigorous definition will be provided in Section II.

To establish (4), we derive an upper bound on the pre-log of the model (3). This upper bound matches asymptotically the pre-log lower bound that was recently developed in [7] in a more general setting (the generic block-fading model considered in this paper is a special case of the system model in [7] for correlation rank $Q=1$). Thus, the combination of the two bounds establishes the pre-log expression (4). As the proof in [7] is rather involved, we also illustrate the main ideas of the proof of the lower bound using an example. In this illustration, we present a new method for bounding the change in differential entropy that occurs when a random variable undergoes a finite-to-one mapping; this method significantly simplifies one step in the proof.

Notation: Sets are denoted by calligraphic letters (e.g., $\mathcal{I}$), and $|\mathcal{I}|$ denotes the cardinality of $\mathcal{I}$. The indicator function of a set $\mathcal{I}$ is denoted by $\mathbb{1}_{\mathcal{I}}$. We use the notation $[M:N] \triangleq \{M, M+1, \dots, N\}$ for $M, N \in \mathbb{N}$. Boldface uppercase (lowercase) letters denote matrices (vectors). Sans serif letters denote random quantities, e.g., A is a random matrix, x is a random vector, and s is a random scalar. The superscripts $^{\mathrm{T}}$ and $^{\mathrm{H}}$ stand for transposition and Hermitian transposition, respectively. The all-zero vector of appropriate size is written as $\mathbf{0}$, and the $M \times M$ identity matrix as $\mathbf{I}_M$. The entry in the ith row and jth column of a matrix $\boldsymbol{A}$ is denoted by $[\boldsymbol{A}]_{i,j}$, and the ith entry of a vector $\boldsymbol{x}$ by $[\boldsymbol{x}]_i$. We denote by $\mathrm{diag}(\boldsymbol{x})$ the diagonal matrix with the entries of $\boldsymbol{x}$ in its main diagonal, and by $|\boldsymbol{A}|$ the modulus of the determinant of a square matrix $\boldsymbol{A}$. For $x \in \mathbb{R}$, we define $\lfloor x \rfloor \triangleq \max\{m \in \mathbb{Z} \,|\, m \leq x\}$. We write $\mathbb{E}[\cdot]$ for the expectation operator, and $\mathsf{x} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma})$ to indicate that x is a circularly symmetric complex Gaussian random vector with covariance matrix $\boldsymbol{\Sigma}$. The Jacobian matrix of a differentiable function ϕ is denoted by $\boldsymbol{J}_{\phi}$.

II. System Model

For the block-fading channel defined by (3), the input-output relation for a given block of length N is

$$\mathsf{y}_r = \sqrt{\frac{\rho}{T}} \sum_{t \in [1:T]} \mathsf{s}_{r,t} \boldsymbol{Z}_{r,t} \mathsf{x}_t + \mathsf{w}_r, \quad r \in [1:R]. \tag{5}$$

Here, $\mathsf{x}_t \in \mathbb{C}^N$ is the signal vector transmitted by the tth transmit antenna; $\mathsf{y}_r \in \mathbb{C}^N$ is the vector received by the rth receive antenna; $\mathsf{s}_{r,t} \sim \mathcal{CN}(0,1)$ is a random variable describing the (t,r) channel; $\boldsymbol{Z}_{r,t} \triangleq \mathrm{diag}(\boldsymbol{z}_{r,t})$, where $\boldsymbol{z}_{r,t}$ is a deterministic vector; $\mathsf{w}_r \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$ is the noise vector at the rth receive antenna; and $\rho \in \mathbb{R}^+$ is the SNR. If $\boldsymbol{Z}_{r,t} = \mathbf{I}_N$ for all $r \in [1:R]$ and $t \in [1:T]$, then (5) reduces to the constant block-fading model. We assume that all $\mathsf{s}_{r,t}$ and w_r are mutually independent and independent across different blocks, and that the vectors x_t are independent of all $\mathsf{s}_{r,t}$ and w_r.

For later use, we define the vectors $\mathsf{x} \triangleq (\mathsf{x}_1^{\mathrm{T}} \cdots \mathsf{x}_T^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{C}^{TN}$, $\mathsf{y} \triangleq (\mathsf{y}_1^{\mathrm{T}} \cdots \mathsf{y}_R^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{C}^{RN}$, and $\mathsf{w} \triangleq (\mathsf{w}_1^{\mathrm{T}} \cdots \mathsf{w}_R^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{C}^{RN}$ and the matrix $\boldsymbol{Z} \triangleq (\boldsymbol{z}_{r,t})_{r \in [1:R], t \in [1:T]} \in \mathbb{C}^{RN \times T}$. We will use the phrase "for a generic correlation" or "for a generic $\boldsymbol{Z}$" to indicate that a property holds for almost every matrix $\boldsymbol{Z}$, which means more specifically that the set of all $\boldsymbol{Z}$ for which the property does *not* hold has Lebesgue measure zero.

III. Pre-log Characterization

A. Main Result

Because of the block-memoryless assumption, the coding theorem in [8, Section 7.3] implies that the capacity of the channel (5) is given by

$$C(\rho) = \frac{1}{N} \sup I(\mathsf{x}; \mathsf{y}). \tag{6}$$

Here, $I(\mathsf{x}; \mathsf{y})$ denotes mutual information [9, p. 251] and the supremum is taken over all input distributions on $\mathbb{C}^{TN}$ that satisfy the average power constraint

$$\mathbb{E}[\|\mathsf{x}\|^2] \leq TN.$$

The pre-log is then defined as

$$\chi \triangleq \lim_{\rho \to \infty} \frac{C(\rho)}{\log(\rho)}. \tag{7}$$

Our main result is the following theorem.

Theorem 1: Let $T < N$ and $R \geq T(N-1)/(N-T)$. For a generic correlation, the pre-log of the channel (5) is given by (4), i.e., $\chi_{\text{gen}} = T(1 - 1/N)$.

Proof: In Section IV, we will show that the pre-log is upper-bounded by $T(1-1/N)$. For $T < N$, $R \geq T(N-1)/(N-T)$, and a generic correlation, this pre-log is achievable as a consequence of the lower bound in [7, Theorem 1]. ■

B. Pre-log Gain

For the constant block-fading model (1), it follows from (2) that the pre-log is maximized for $T = R = \lfloor N/2 \rfloor$, which yields $\chi_{\text{const}} = \lfloor N^2/2 \rfloor / (2N) \leq N/4$. In contrast, for the generic block-fading model (3) with $T < N$, it follows from (4) that the pre-log is maximized for $T = N-1$ and $R = (N-1)^2$, which results in $\chi_{\text{gen}} = (N-1)^2/N$. For large N, this is about four times as large as the highest achievable χ_{const}. We will now provide some intuition regarding this pre-log gain. For concreteness and simplicity, we consider the case $T=2, R=3, N=4$.

The pre-log can be interpreted as the number of entries of $\mathsf{x} \in \mathbb{C}^8$ that can be deduced from a received $\mathsf{y} \in \mathbb{C}^{12}$ in the absence of noise, divided by the block length (coherence length) $N = 4$. In the constant block-fading model, the noiseless received vectors $\bar{\mathsf{y}}_r = \mathsf{s}_{r,1}\mathsf{x}_1 + \mathsf{s}_{r,2}\mathsf{x}_2$, $r = 1,2,3$ belong to the two-dimensional subspace spanned by $\{\mathsf{x}_1, \mathsf{x}_2\}$. Hence, the received vectors $\bar{\mathsf{y}}_1, \bar{\mathsf{y}}_2, \bar{\mathsf{y}}_3$ are linearly dependent, and any two of them contain all the information available about x. From, e.g., $\bar{\mathsf{y}}_1$ and $\bar{\mathsf{y}}_2$, we obtain $2 \cdot 4$ equations in the $8+4$ variables $(\mathsf{x}, \mathsf{s}_{1,1}, \mathsf{s}_{1,2}, \mathsf{s}_{2,1}, \mathsf{s}_{2,2})$. Since we do not have control of the variables $\mathsf{s}_{r,t}$, one way to reconstruct x is to fix four of its entries (or, equivalently, to transmit four pilot symbols) to obtain eight equations in eight variables. By solving this system of equations, we obtain four entries of x, which corresponds to a pre-log of $4/4 = 1$.

In the generic block-fading model, on the other hand, the noiseless received vectors $\bar{\mathsf{y}}_r = \mathsf{s}_{r,1}\boldsymbol{Z}_{r,1}\mathsf{x}_1 + \mathsf{s}_{r,2}\boldsymbol{Z}_{r,2}\mathsf{x}_2$, $r = 1, 2, 3$ can span a three-dimensional subspace. Hence, we obtain a system of $3\cdot 4$ equations in the $8+6$ variables $(\mathbf{x}, \mathsf{s}_{1,1}, \mathsf{s}_{1,2}, \mathsf{s}_{2,1}, \mathsf{s}_{2,2}, \mathsf{s}_{3,1}, \mathsf{s}_{3,2})$. Fixing two entries of $\mathbf{x}$, we are able to recover the remaining six entries. Hence, the pre-log is $6/4 = 3/2$. These arguments suggest that the reason why the generic block-fading model yields a larger pre-log than the constant block-fading model is that the noiseless received vectors span a subspace of $\mathbb{C}^N$ of higher dimension.

IV. Upper bound

The following upper bound on the pre-log of the channel (5) holds for arbitrary T, R, N, and $\boldsymbol{Z}$.

Theorem 2: The pre-log of the channel (5) satisfies

$$\chi \le T\left(1-\frac{1}{N}\right). \tag{8}$$

Proof: We will show that the pre-log is upper-bounded by T times the pre-log of a constant block-fading single-input multiple-output (SIMO) channel. The result then follows from (2).

From (5), the input-output relation at time $n\in[1:N]$ is

$$[\mathbf{y}_r]_n = \sqrt{\frac{\rho}{T}} \sum_{t\in[1:T]} \mathsf{s}_{r,t}\,[\boldsymbol{z}_{r,t}]_n\,[\mathbf{x}_t]_n + [\mathbf{w}_r]_n\,, \quad r\in[1:R]\,. \tag{9}$$

Consider now T constant block-fading SIMO channels with R receive antennas and SNR equal to $K\rho$, where K is any finite constant satisfying $K > \max_{r\in[1:R],\, n\in[1:N]} \sum_{t\in[1:T]} |[\boldsymbol{z}_{r,t}]_n|^2$. The input-output relation of the tth SIMO channel, with $t\in[1:T]$, is

$$[\tilde{\mathbf{y}}_{r,t}]_n = \sqrt{K\rho}\, \mathsf{s}_{r,t}\,[\mathbf{x}_t]_n + [\tilde{\mathbf{w}}_{r,t}]_n\,, \quad r\in[1:R]\,. \tag{10}$$

We can rewrite (9) using (10) as follows:

$$[\mathbf{y}_r]_n = \frac{1}{\sqrt{KT}} \sum_{t\in[1:T]} [\boldsymbol{z}_{r,t}]_n\,[\tilde{\mathbf{y}}_{r,t}]_n + [\mathbf{w}'_r]_n\,, \tag{11}$$

where the $[\mathbf{w}'_r]_n \sim [\mathbf{w}_r]_n - \sum_{t\in[1:T]}[\boldsymbol{z}_{r,t}]_n\,[\tilde{\mathbf{w}}_{r,t}]_n/\sqrt{KT} \sim \mathcal{CN}\big(0, 1-\sum_{t\in[1:T]}|[\boldsymbol{z}_{r,t}]_n|^2/(KT)\big)$ are mutually independent and independent of all $\mathbf{x}_t, \mathsf{s}_{r,t}$, and $\tilde{\mathbf{w}}_{r,t}$. The additional noise terms $[\mathbf{w}'_r]_n$ ensure that the total noise in (11) has unit variance. The data-processing inequality applied to (11) yields

$$I(\mathbf{x};\mathbf{y}) \le I(\mathbf{x};\tilde{\mathbf{y}}_1,\dots,\tilde{\mathbf{y}}_T)\,, \tag{12}$$

with $\tilde{\mathbf{y}}_t \triangleq (\tilde{\mathbf{y}}_{1,t}^{\mathrm{T}} \cdots \tilde{\mathbf{y}}_{R,t}^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{C}^{RN}$. The right-hand side of (12) can be upper-bounded as follows:

$$\begin{aligned}
I(\mathbf{x};\tilde{\mathbf{y}}_1,\dots,\tilde{\mathbf{y}}_T) &= h(\tilde{\mathbf{y}}_1,\dots,\tilde{\mathbf{y}}_T) - h(\tilde{\mathbf{y}}_1,\dots,\tilde{\mathbf{y}}_T|\mathbf{x}) \\
&\overset{(a)}{=} h(\tilde{\mathbf{y}}_1,\dots,\tilde{\mathbf{y}}_T) - \sum_{t\in[1:T]} h(\tilde{\mathbf{y}}_t|\mathbf{x}_t) \\
&\le \sum_{t\in[1:T]} \big[h(\tilde{\mathbf{y}}_t) - h(\tilde{\mathbf{y}}_t|\mathbf{x}_t)\big] \\
&= \sum_{t\in[1:T]} I(\mathbf{x}_t;\tilde{\mathbf{y}}_t) \\
&\overset{(b)}{\le} TN C_{\text{const}}(K\rho) \\
&\overset{(c)}{=} T(N-1)\log(K\rho) + o(\log(\rho)) \\
&= T(N-1)\log(\rho) + o(\log(\rho))\,.
\end{aligned} \tag{13}$$

Here, h denotes differential entropy, (a) holds because $\tilde{\mathbf{y}}_1,\dots,\tilde{\mathbf{y}}_T$ are conditionally independent given $\mathbf{x}$, (b) follows from (6) (note that $C_{\text{const}}(K\rho)$ refers to the capacity of constant block-fading SIMO channels), and (c) follows from (7) and (2) for $M=1$. Inserting (13) into (12) and using (6) yields

$$C(\rho) \le T\frac{N-1}{N}\log(\rho) + o(\log(\rho))\,,$$

from which (8) follows via (7). ∎

V. Lower bound

According to [7, Theorem 1], for $T<N$ and $R\ge T(N-1)/(N-T)$, the pre-log of the generic block-fading channel (5) is lower-bounded by $\chi_{\text{gen}} \ge T(1-1/N)$. We will now illustrate the main ideas of the proof of this lower bound and present a new method for bounding the change in differential entropy under a finite-to-one mapping (Lemma 1 in Section VI), which significantly simplifies one of the steps of the proof. For concreteness, we consider the special choice $T=2$, $R=3$, and $N=4$. For this choice, $T(1-1/N)=3/2$.

In the remainder of this paper, we choose the input distribution $\mathbf{x}\sim\mathcal{CN}(\mathbf{0},\mathbf{I}_8)$. Because of (6) and (7), we obtain

$$\chi \ge \frac{1}{4}\lim_{\rho\to\infty}\frac{I(\mathbf{x};\mathbf{y})}{\log(\rho)}\,. \tag{14}$$

Since

$$I(\mathbf{x};\mathbf{y}) = h(\mathbf{y}) - h(\mathbf{y}|\mathbf{x})\,, \tag{15}$$

we can lower-bound $I(\mathbf{x};\mathbf{y})$ by lower-bounding $h(\mathbf{y})$ and upper-bounding $h(\mathbf{y}|\mathbf{x})$. For later use, we note that the input-output relation (5) can be written as

$$\mathbf{y} = \sqrt{\frac{\rho}{2}}\,\bar{\mathbf{y}} + \mathbf{w}\,, \tag{16}$$

with

$$\bar{\mathbf{y}} \triangleq \underbrace{\begin{pmatrix} \boldsymbol{Z}_{1,1}\mathbf{x}_1 & & & \boldsymbol{Z}_{1,2}\mathbf{x}_2 & & \\ & \boldsymbol{Z}_{2,1}\mathbf{x}_1 & & & \boldsymbol{Z}_{2,2}\mathbf{x}_2 & \\ & & \boldsymbol{Z}_{3,1}\mathbf{x}_1 & & & \boldsymbol{Z}_{3,2}\mathbf{x}_2 \end{pmatrix}}_{\triangleq\,\mathbf{B}} \underbrace{\begin{pmatrix}\mathsf{s}_{1,1}\\ \mathsf{s}_{2,1}\\ \mathsf{s}_{3,1}\\ \mathsf{s}_{1,2}\\ \mathsf{s}_{2,2}\\ \mathsf{s}_{3,2}\end{pmatrix}}_{\triangleq\,\mathbf{s}}. \tag{17}$$

We will first upper-bound $h(\mathbf{y}|\mathbf{x})$. It follows from (16) that given $\mathbf{x}$, $\mathbf{y}$ is conditionally Gaussian with covariance matrix $(\rho/2)\mathbf{B}\mathbf{B}^{\mathrm{H}}+\mathbf{I}_{12}$. Hence, $h(\mathbf{y}|\mathbf{x}) = \mathbb{E}_{\mathbf{x}}\big[\log\big((\pi e)^{12}\,|(\rho/2)\mathbf{B}\mathbf{B}^{\mathrm{H}} + \mathbf{I}_{12}|\big)\big]$. By [10, Theorem 1.3.20], $|(\rho/2)\mathbf{B}\mathbf{B}^{\mathrm{H}} + \mathbf{I}_{12}| = |(\rho/2)\mathbf{B}^{\mathrm{H}}\mathbf{B}+\mathbf{I}_6|$. Furthermore, assuming $\rho>1$ (note that we are only interested in $\rho\to\infty$), we have $|(\rho/2)\mathbf{B}^{\mathrm{H}}\mathbf{B}+\mathbf{I}_6| \le \rho^6|(1/2)\mathbf{B}^{\mathrm{H}}\mathbf{B}+\mathbf{I}_6|$. Thus,

$$\begin{aligned}
h(\mathbf{y}|\mathbf{x}) &\le \mathbb{E}_{\mathbf{x}}\big[\log\big((\pi e)^{12}\rho^6\,|(1/2)\mathbf{B}^{\mathrm{H}}\mathbf{B}+\mathbf{I}_6|\big)\big] \\
&= 6\log(\rho) + \mathbb{E}_{\mathbf{x}}\big[\log|(1/2)\mathbf{B}^{\mathrm{H}}\mathbf{B}+\mathbf{I}_6|\big] + \mathcal{O}(1)\,.
\end{aligned}$$

Finally, using $\mathbb{E}_{\mathbf{x}}\big[\log|(1/2)\mathbf{B}^{\mathrm{H}}\mathbf{B}+\mathbf{I}_6|\big] \le \log\mathbb{E}_{\mathbf{x}}\big[|(1/2)\mathbf{B}^{\mathrm{H}}\mathbf{B} + \mathbf{I}_6|\big] = \mathcal{O}(1)$ [9, Theorem 17.1.1], we obtain

$$h(\mathbf{y}|\mathbf{x}) \leq 6\log(\rho) + \mathcal{O}(1)\,. \tag{18}$$

Next, we will lower-bound $h(\mathbf{y})$. Using (16), we obtain

$$\begin{aligned} h(\mathbf{y}) &\geq h\left(\sqrt{\frac{\rho}{2}}\,\bar{\mathbf{y}} + \mathbf{w}\,\middle|\,\mathbf{w}\right) = h\left(\sqrt{\frac{\rho}{2}}\,\bar{\mathbf{y}}\right) \\ &= 12\log(\rho) + h(\bar{\mathbf{y}}) + \mathcal{O}(1)\,. \end{aligned}$$

In Section VI, we will show that $h(\bar{\mathbf{y}}) > -\infty$. Hence, $h(\mathbf{y}) \geq 12\log(\rho) + \mathcal{O}(1)$ (note that $h(\bar{\mathbf{y}})$ does not depend on ρ). Inserting this bound and (18) into (15), we conclude that $I(\mathbf{x};\mathbf{y}) \geq 6\log(\rho) + \mathcal{O}(1)$. With (14), this implies $\chi \geq 3/2 = T(1-1/N)$.

VI. Proof that $h(\bar{\mathbf{y}}) > -\infty$

According to (17), $\bar{\mathbf{y}}$ is a function of $\mathbf{s}$ and $\mathbf{x}$. We will relate $h(\bar{\mathbf{y}})$ to $h(\mathbf{s},\mathbf{x})$. To equalize the dimensions—note that $\bar{\mathbf{y}} \in \mathbb{C}^{12}$ and $(\mathbf{s}^{\mathrm{T}}\,\mathbf{x}^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{C}^{14}$—we condition on $[\mathbf{x}_1]_1$ and $[\mathbf{x}_2]_2$, which results in $h(\bar{\mathbf{y}}) \geq h(\bar{\mathbf{y}}\,|\,[\mathbf{x}_1]_1, [\mathbf{x}_2]_2)$. For easier notation, we set $\mathbf{x}_{\mathcal{P}} \triangleq ([\mathbf{x}_1]_1\ [\mathbf{x}_2]_2)^{\mathrm{T}}$ and $\mathbf{x}_{\mathcal{D}} \triangleq ([\mathbf{x}_1]_2\ [\mathbf{x}_1]_3\ [\mathbf{x}_1]_4\ [\mathbf{x}_2]_1\ [\mathbf{x}_2]_3\ [\mathbf{x}_2]_4)^{\mathrm{T}}$. One can think of $\mathbf{x}_{\mathcal{P}}$ as pilot symbols and of $\mathbf{x}_{\mathcal{D}}$ as data symbols. The above inequality then becomes

$$h(\bar{\mathbf{y}}) \geq h(\bar{\mathbf{y}}|\mathbf{x}_{\mathcal{P}})\,. \tag{19}$$

We conclude the proof by showing that $h(\bar{\mathbf{y}}\,|\,\mathbf{x}_{\mathcal{P}}) > -\infty$. This will be done in the following five steps: (i) Relate $(\boldsymbol{s}, \boldsymbol{x}_{\mathcal{D}})$ to $\bar{\boldsymbol{y}}$ via polynomial mappings $\phi_{\boldsymbol{x}_{\mathcal{P}}}$. (ii) Show that the Jacobian matrices $\boldsymbol{J}_{\phi_{\boldsymbol{x}_{\mathcal{P}}}}(\boldsymbol{s}, \boldsymbol{x}_{\mathcal{D}})$ are nonsingular almost everywhere (a.e.) for almost all (a.a.) $\boldsymbol{x}_{\mathcal{P}}$. (iii) Show that the mappings $\phi_{\boldsymbol{x}_{\mathcal{P}}}$ are finite-to-one a.e. for a.a. $\boldsymbol{x}_{\mathcal{P}}$. (iv) Apply a novel result on the change in differential entropy under a finite-to-one mapping to $h(\bar{\mathbf{y}}\,|\,\mathbf{x}_{\mathcal{P}})$. (v) Bound the terms resulting from this change in differential entropy.

Step (i): We consider the $\boldsymbol{x}_{\mathcal{P}}$-parametrized mappings

$$\phi_{\boldsymbol{x}_{\mathcal{P}}} : (\boldsymbol{s}, \boldsymbol{x}_{\mathcal{D}}) \mapsto \bar{\boldsymbol{y}} = \begin{pmatrix} s_{1,1}\boldsymbol{Z}_{1,1}\boldsymbol{x}_1 + s_{1,2}\boldsymbol{Z}_{1,2}\boldsymbol{x}_2 \\ s_{2,1}\boldsymbol{Z}_{2,1}\boldsymbol{x}_1 + s_{2,2}\boldsymbol{Z}_{2,2}\boldsymbol{x}_2 \\ s_{3,1}\boldsymbol{Z}_{3,1}\boldsymbol{x}_1 + s_{3,2}\boldsymbol{Z}_{3,2}\boldsymbol{x}_2 \end{pmatrix}, \tag{20}$$

which map $\mathbb{C}^{12}$ to itself. The Jacobian matrix of $\phi_{\boldsymbol{x}_{\mathcal{P}}}$ is

$$\boldsymbol{J}_{\phi_{\boldsymbol{x}_{\mathcal{P}}}} = \begin{pmatrix} & \boldsymbol{A}_{1,1} & \boldsymbol{A}_{1,2} \\ \boldsymbol{B} & \boldsymbol{A}_{2,1} & \boldsymbol{A}_{2,2} \\ & \boldsymbol{A}_{3,1} & \boldsymbol{A}_{3,2} \end{pmatrix},$$

where $\boldsymbol{B}$ was defined in (17) and

$$\boldsymbol{A}_{r,1} \triangleq \begin{pmatrix} 0 & & \\ s_{r,1}[\boldsymbol{z}_{r,1}]_2 & & \\ & s_{r,1}[\boldsymbol{z}_{r,1}]_3 & \\ & & s_{r,1}[\boldsymbol{z}_{r,1}]_4 \end{pmatrix},$$

$$\boldsymbol{A}_{r,2} \triangleq \begin{pmatrix} s_{r,2}[\boldsymbol{z}_{r,2}]_1 & & \\ 0 & & \\ & s_{r,2}[\boldsymbol{z}_{r,2}]_3 & \\ & & s_{r,2}[\boldsymbol{z}_{r,2}]_4 \end{pmatrix}.$$

Note that we did not take derivatives with respect to $[\boldsymbol{x}_1]_1$ and $[\boldsymbol{x}_2]_2$, since these variables are treated as fixed parameters.

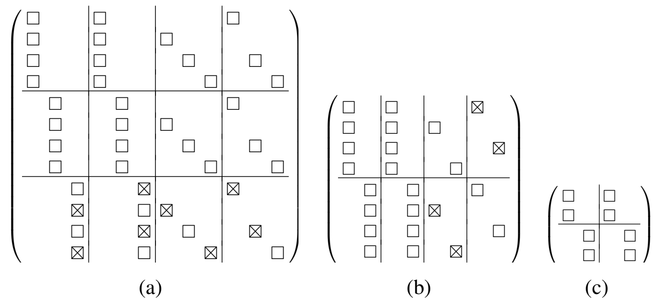

Fig. 1. Three matrices considered in Step (ii). ☐ indicates a potentially nonzero entry; ⊠ indicates a potentially nonzero entry that is set to zero. All the other entries are zero.

Step (ii): To show that $\boldsymbol{J}_{\phi_{\boldsymbol{x}_{\mathcal{P}}}}$ is nonsingular (i.e., $|\boldsymbol{J}_{\phi_{\boldsymbol{x}_{\mathcal{P}}}}| \neq 0$) a.e. for a.a. $\boldsymbol{x}_{\mathcal{P}}$ and a generic $\boldsymbol{Z}$, we use the approach of [7, Appendix C]. The determinant of $\boldsymbol{J}_{\phi_{\boldsymbol{x}_{\mathcal{P}}}}$ is a polynomial $p(\boldsymbol{Z}, \boldsymbol{s}, \boldsymbol{x})$ (i.e., a polynomial in all the entries of $\boldsymbol{Z}$, $\boldsymbol{s}$, and $\boldsymbol{x}$), which we will show to be nonzero at a specific point $(\tilde{\boldsymbol{Z}}, \tilde{\boldsymbol{s}}, \tilde{\boldsymbol{x}})$. Fixing $\tilde{\boldsymbol{s}}$ and $\tilde{\boldsymbol{x}}$, we can then conclude that $p(\boldsymbol{Z}, \tilde{\boldsymbol{s}}, \tilde{\boldsymbol{x}})$ (as a function of $\boldsymbol{Z}$) does not vanish identically. Since a polynomial vanishes either identically or on a set of measure zero, we conclude that $p(\boldsymbol{Z}, \tilde{\boldsymbol{s}}, \tilde{\boldsymbol{x}}) \neq 0$ for a generic $\boldsymbol{Z}$. Using the same argument, we conclude that, for a generic fixed $\boldsymbol{Z}$, $p(\boldsymbol{Z}, \boldsymbol{s}, \boldsymbol{x}) \neq 0$ a.e. (as a function of $(\boldsymbol{s}, \boldsymbol{x})$). Hence, $|\boldsymbol{J}_{\phi_{\boldsymbol{x}_{\mathcal{P}}}}| \neq 0$ a.e. for a.a. $\boldsymbol{x}_{\mathcal{P}}$ and a generic $\boldsymbol{Z}$.

It remains to find the point $(\tilde{\boldsymbol{Z}}, \tilde{\boldsymbol{s}}, \tilde{\boldsymbol{x}})$. The matrix $\boldsymbol{J}_{\phi_{\boldsymbol{x}_{\mathcal{P}}}}$ has the form sketched in Fig. 1(a). Setting $[\tilde{\boldsymbol{z}}_{3,2}]_3 = [\tilde{\boldsymbol{z}}_{3,1}]_4 = [\tilde{\boldsymbol{z}}_{3,2}]_1 = [\tilde{\boldsymbol{z}}_{3,1}]_2 = 0$, the entries marked by ⊠ become zero. Choosing $[\tilde{\boldsymbol{z}}_{3,1}]_1$, $[\tilde{\boldsymbol{z}}_{3,1}]_3$, $[\tilde{\boldsymbol{z}}_{3,2}]_2$, $[\tilde{\boldsymbol{z}}_{3,2}]_4$, $\tilde{s}_{3,1}$, $\tilde{s}_{3,2}$, $[\tilde{\boldsymbol{x}}_1]_1$, and $[\tilde{\boldsymbol{x}}_2]_2$ nonzero and operating a Laplace expansion on the last four rows in Fig. 1(a), we see that the matrix in Fig. 1(a) is nonsingular if the matrix in Fig. 1(b) is nonsingular. Setting $\tilde{s}_{1,2} = \tilde{s}_{2,1} = 0$, the entries marked by ⊠ in Fig. 1(b) become zero. By choosing $[\tilde{\boldsymbol{z}}_{1,1}]_2$, $[\tilde{\boldsymbol{z}}_{1,1}]_4$, $[\tilde{\boldsymbol{z}}_{2,2}]_1$, $[\tilde{\boldsymbol{z}}_{2,2}]_3$, $\tilde{s}_{1,1}$, and $\tilde{s}_{2,2}$ nonzero and operating a Laplace expansion on the last four columns, it remains to show nonsingularity of the matrix in Fig. 1(c). This can be achieved by suitably choosing $[\tilde{\boldsymbol{z}}_{1,1}]_1$, $[\tilde{\boldsymbol{z}}_{1,1}]_3$, $[\tilde{\boldsymbol{z}}_{1,2}]_1$, $[\tilde{\boldsymbol{z}}_{1,2}]_3$, $[\tilde{\boldsymbol{z}}_{2,1}]_2$, $[\tilde{\boldsymbol{z}}_{2,1}]_4$, $[\tilde{\boldsymbol{z}}_{2,2}]_2$, and $[\tilde{\boldsymbol{z}}_{2,2}]_4$.

Step (iii): By Bézout's theorem [11, Proposition B.2.7], d multivariate polynomials of degree k can have at most k^d isolated common zeros. Since the equation $\phi_{\boldsymbol{x}_{\mathcal{P}}}(\boldsymbol{s}, \boldsymbol{x}_{\mathcal{D}}) = \bar{\boldsymbol{y}}$ can be reformulated as the system of polynomial equations $\phi_{\boldsymbol{x}_{\mathcal{P}}}(\boldsymbol{s}, \boldsymbol{x}_{\mathcal{D}}) - \bar{\boldsymbol{y}} = \boldsymbol{0} \in \mathbb{C}^{12}$, where each of the 12 polynomials is of degree two (see (20)), the points $(\boldsymbol{s}, \boldsymbol{x}_{\mathcal{D}})$ that are mapped by $\phi_{\boldsymbol{x}_{\mathcal{P}}}$ to the same $\bar{\boldsymbol{y}}$ are the common zeros of 12 polynomials of degree two. Nonisolated common zeros of these polynomials can only exist in the set where $\boldsymbol{J}_{\phi_{\boldsymbol{x}_{\mathcal{P}}}}$ is singular. Hence, the set $\mathcal{M} \triangleq \{(\boldsymbol{s}, \boldsymbol{x}_{\mathcal{D}}) : |\boldsymbol{J}_{\phi_{\boldsymbol{x}_{\mathcal{P}}}}| \neq 0\}$ contains only isolated common zeros, whose number is upper-bounded by Bézout's theorem by 2^{12}. It follows that the number of points $(\boldsymbol{s}, \boldsymbol{x}_{\mathcal{D}}) \in \mathcal{M}$ that are mapped by $\phi_{\boldsymbol{x}_{\mathcal{P}}}$ to the same $\bar{\boldsymbol{y}}$ is upper-bounded by 2^{12}, i.e., $\phi_{\boldsymbol{x}_{\mathcal{P}}}\big|_{\mathcal{M}}$ is finite-to-one for a.a. $\boldsymbol{x}_{\mathcal{P}}$. Because by Step (ii) the complement of the set $\mathcal{M}$ has Lebesgue measure zero for a.a. $\boldsymbol{x}_{\mathcal{P}}$, the mapping $\phi_{\boldsymbol{x}_{\mathcal{P}}}$ is finite-to-one a.e. for a.a. $\boldsymbol{x}_{\mathcal{P}}$.

Step (iv): We will use the following novel result bounding the change in differential entropy under a finite-to-one mapping. A proof is provided in the appendix.

Lemma 1: Let $\mathsf{u} \in \mathbb{C}^n$ be a random vector with continuous probability density function f_{u}. Consider a continuously differentiable mapping $\vartheta\colon \mathbb{C}^n \to \mathbb{C}^n$ with Jacobian matrix $\boldsymbol{J}_\vartheta$. Let $\mathsf{v} \triangleq \vartheta(\mathsf{u})$, and assume that the cardinality of the set $\vartheta^{-1}(\{\boldsymbol{v}\})$ satisfies $|\vartheta^{-1}(\{\boldsymbol{v}\})| \le m < \infty$ a.e., for some $m \in \mathbb{N}$ (i.e., ϑ is finite-to-one a.e.). Then:

(I) There exist disjoint measurable sets $\{\mathcal{U}_k\}_{k\in[1:m]}$ such that $\vartheta\big|_{\mathcal{U}_k}$ is one-to-one for each $k \in [1:m]$ and $\bigcup_{k\in[1:m]} \mathcal{U}_k = \mathbb{C}^n \setminus \mathcal{N}$, where $\mathcal{N}$ is a set of Lebesgue measure zero.

(II) For any such sets $\{\mathcal{U}_k\}_{k\in[1:m]}$,

$$h(\mathsf{v}) \ge h(\mathsf{u}) + \int_{\mathbb{C}^n} f_{\mathsf{u}}(\boldsymbol{u}) \log(|\boldsymbol{J}_\vartheta(\boldsymbol{u})|^2)\, d\boldsymbol{u} - H(\mathsf{k})\,, \quad (21)$$

where k is the discrete random variable that takes on the value k when $\mathsf{u} \in \mathcal{U}_k$ and H denotes entropy.

Since by Step (iii) the mappings $\phi_{\boldsymbol{x}_\mathcal{P}}$ are finite-to-one a.e. for a.a. $\boldsymbol{x}_\mathcal{P}$, we can use Lemma 1 with $\mathsf{u} = (\mathsf{s}, \mathsf{x}_\mathcal{D})$ and $\vartheta = \phi_{\boldsymbol{x}_\mathcal{P}}$. We thus obtain

$$h(\bar{\mathsf{y}}|\mathsf{x}_\mathcal{P}) \ge h(\mathsf{s}, \mathsf{x}_\mathcal{D}) + \mathbb{E}_{\mathsf{x}_\mathcal{P}}\bigg[\int_{\mathbb{C}^{12}} f_{\mathsf{s},\mathsf{x}_\mathcal{D}}(\boldsymbol{s}, \boldsymbol{x}_\mathcal{D}) \times \log\big(|\boldsymbol{J}_{\phi_{\mathsf{x}_\mathcal{P}}}(\boldsymbol{s}, \boldsymbol{x}_\mathcal{D})|^2\big)\, d(\boldsymbol{s}, \boldsymbol{x}_\mathcal{D}) - H(\mathsf{k})\bigg].$$

Step (v): The differential entropy $h(\mathsf{s}, \mathsf{x}_\mathcal{D})$ is a finite constant, and the entropy $H(\mathsf{k})$ can be upper-bounded by the entropy of a uniformly distributed discrete random variable. Hence, it remains to bound

$$\mathbb{E}_{\mathsf{x}_\mathcal{P}}\bigg[\int_{\mathbb{C}^{12}} f_{\mathsf{s},\mathsf{x}_\mathcal{D}}(\boldsymbol{s}, \boldsymbol{x}_\mathcal{D}) \log\big(|\boldsymbol{J}_{\phi_{\mathsf{x}_\mathcal{P}}}(\boldsymbol{s}, \boldsymbol{x}_\mathcal{D})|^2\big)\, d(\boldsymbol{s}, \boldsymbol{x}_\mathcal{D})\bigg] = \int_{\mathbb{C}^{14}} f_{\mathsf{s},\mathsf{x}}(\boldsymbol{s}, \boldsymbol{x}) \log\big(|\boldsymbol{J}_{\phi_{\boldsymbol{x}_\mathcal{P}}}(\boldsymbol{s}, \boldsymbol{x}_\mathcal{D})|^2\big)\, d(\boldsymbol{s}, \boldsymbol{x})\,. \quad (22)$$

In [7, Appendix C], it is shown that for an analytic function $g\colon \mathbb{C}^n \to \mathbb{C}$ that is not identically zero,

$$\int_{\mathbb{C}^n} \exp(-\|\boldsymbol{\xi}\|^2) \log(|g(\boldsymbol{\xi})|)\, d\boldsymbol{\xi} > -\infty\,.$$

Since $f_{\mathsf{s},\mathsf{x}}$ is the probability density function of a standard multivariate Gaussian random vector and $\det\big(\boldsymbol{J}_{\phi_{\boldsymbol{x}_\mathcal{P}}}(\boldsymbol{s}, \boldsymbol{x}_\mathcal{D})\big)$ is a complex polynomial that is not identically zero as shown in Step (ii), it follows that the integral in (22) is finite. Hence, $h(\bar{\mathsf{y}}|\mathsf{x}_\mathcal{P}) > -\infty$. With (19), this concludes the proof that $h(\bar{\mathsf{y}}) > -\infty$.

Appendix: Proof of Lemma 1

Part (I), the separation of $\mathbb{C}^n$ into measurable subsets $\mathcal{U}_k$, can be shown using Zorn's Lemma (for details see [7, Lemma 8]). To establish part (II), i.e., the bound (21), we first note that

$$h(\mathsf{v}) \ge h(\mathsf{v}|\mathsf{k}) = \sum_{k\in[1:m]} h(\mathsf{v}|\mathsf{k}=k)\, p_k\,, \quad (23)$$

where $p_k \triangleq \Pr[\mathsf{u} \in \mathcal{U}_k] = \int_{\mathcal{U}_k} f_{\mathsf{u}}(\boldsymbol{u})\, d\boldsymbol{u}$. We assume without loss of generality that $p_k \neq 0$ for $k \in [1:m]$ (if $p_k = 0$ for some k, we simply omit the corresponding term in (23)). Since $\vartheta\big|_{\mathcal{U}_k}$ is one-to-one, $h(\mathsf{v}|\mathsf{k}=k)$ can be transformed using the transformation rule for one-to-one mappings [12, Lemma 3]:

$$h(\mathsf{v}|\mathsf{k}=k) = h(\mathsf{u}|\mathsf{k}=k) + \int_{\mathbb{C}^n} f_{\mathsf{u}|\mathsf{k}=k}(\boldsymbol{u}) \log(|\boldsymbol{J}_\vartheta(\boldsymbol{u})|^2)\, d\boldsymbol{u}\,. \quad (24)$$

The conditional probability density function of u given $\mathsf{k}=k$ is $f_{\mathsf{u}|\mathsf{k}=k}(\boldsymbol{u}) = \mathbb{1}_{\mathcal{U}_k}(\boldsymbol{u}) f_{\mathsf{u}}(\boldsymbol{u})/p_k$. Thus, $h(\mathsf{u}|\mathsf{k}=k) = -\int_{\mathcal{U}_k} \big(f_{\mathsf{u}}(\boldsymbol{u})/p_k\big) \log\big(f_{\mathsf{u}}(\boldsymbol{u})/p_k\big)\, d\boldsymbol{u}$, and (24) becomes

$$\begin{aligned} h(\mathsf{v}|\mathsf{k}=k) &= \frac{1}{p_k}\bigg[-\int_{\mathcal{U}_k} f_{\mathsf{u}}(\boldsymbol{u}) \log\bigg(\frac{f_{\mathsf{u}}(\boldsymbol{u})}{p_k}\bigg) d\boldsymbol{u} \\ &\qquad + \int_{\mathcal{U}_k} f_{\mathsf{u}}(\boldsymbol{u}) \log(|\boldsymbol{J}_\vartheta(\boldsymbol{u})|^2)\, d\boldsymbol{u}\bigg] \\ &= \frac{1}{p_k}\bigg[-\int_{\mathcal{U}_k} f_{\mathsf{u}}(\boldsymbol{u}) \log\big(f_{\mathsf{u}}(\boldsymbol{u})\big)\, d\boldsymbol{u} \\ &\qquad + \int_{\mathcal{U}_k} f_{\mathsf{u}}(\boldsymbol{u}) \log(|\boldsymbol{J}_\vartheta(\boldsymbol{u})|^2)\, d\boldsymbol{u}\bigg] + \log(p_k)\,. \end{aligned}$$

Inserting this expression into (23) and recalling that the sets $\mathcal{U}_k$ are disjoint and $\bigcup_{k\in[1:m]} \mathcal{U}_k = \mathbb{C}^n \setminus \mathcal{N}$, we obtain

$$\begin{aligned} h(\mathsf{v}) &\ge -\int_{\mathbb{C}^n} f_{\mathsf{u}}(\boldsymbol{u}) \log\big(f_{\mathsf{u}}(\boldsymbol{u})\big)\, d\boldsymbol{u} \\ &\qquad + \int_{\mathbb{C}^n} f_{\mathsf{u}}(\boldsymbol{u}) \log(|\boldsymbol{J}_\vartheta(\boldsymbol{u})|^2)\, d\boldsymbol{u} + \sum_{k\in[1:m]} p_k \log(p_k) \\ &= h(\mathsf{u}) + \int_{\mathbb{C}^n} f_{\mathsf{u}}(\boldsymbol{u}) \log(|\boldsymbol{J}_\vartheta(\boldsymbol{u})|^2)\, d\boldsymbol{u} - H(\mathsf{k})\,, \end{aligned}$$

which is (21).

References

[1] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing," *arXiv:1209.1402v1 [cs.IT]*, Sep. 2012.

[2] 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.

[3] L. Zheng and D. 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.

[4] W. Yang, G. Durisi, and E. Riegler, "On the capacity of large-MIMO block-fading channels," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 117–132, Feb. 2013.

[5] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications*. Cambridge, UK: Cambridge Univ. Press, 2005.

[6] S. A. Jafar, *Interference Alignment: A New Look at Signal Dimensions in a Communication Network*, ser. Foundations and Trends in Communications and Information Theory. now publisher, 2011, vol. 7, no. 1.

[7] G. Koliander, E. Riegler, G. Durisi, V. I. Morgenshtern, and F. Hlawatsch, "A lower bound on the noncoherent capacity pre-log for the MIMO channel with temporally correlated fading," in *Proc. Allerton Conference*, Monticello, IL, Sep. 2012, pp. 1198–1205.

[8] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY: Wiley, 1968.

[9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY: Wiley, 2006.

[10] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, UK: Cambridge Univ. Press, 1985.

[11] A. R. P. van den Essen, *Polynomial Automorphisms and the Jacobian Conjecture*. Basel, Switzerland: Birkhäuser, 2000.

[12] V. I. Morgenshtern, E. Riegler, W. Yang, G. Durisi, S. Lin, B. Sturmfels, and H. Bölcskei, "Capacity pre-log of noncoherent SIMO channels via Hironaka's theorem," *IEEE Trans. Inf. Theory*, 2013, arXiv:1204.2775v1 [cs.IT].