Degrees of Freedom of Generic Block-Fading MIMO Channels without A Priori Channel State Information

Günther Koliander, *Student Member, IEEE,* Erwin Riegler, *Member, IEEE,*
Giuseppe Durisi, *Senior Member, IEEE,* and Franz Hlawatsch, *Fellow, IEEE*

***Abstract*—We study the high-SNR capacity of *generic* MIMO Rayleigh block-fading channels in the noncoherent setting where neither transmitter nor receiver has *a priori* channel state information but both are aware of the channel statistics. In contrast to the well-established constant block-fading model, we allow the fading to vary within each block with a temporal correlation that is "generic" (in the sense used in the interference-alignment literature). We show that the number of degrees of freedom of a generic MIMO Rayleigh block-fading channel with T transmit antennas and block length N is given by $T(1-1/N)$ provided that $T < N$ and the number of receive antennas is at least $T(N-1)/(N-T)$. A comparison with the constant block-fading channel (where the fading is constant within each block) shows that, for large block lengths, generic correlation increases the number of degrees of freedom by a factor of up to four.**

***Index Terms*—Block-fading channels, capacity pre-log, channel capacity, channel state information, degrees of freedom, MIMO, noncoherent communication, OFDM**

I. Introduction

The use of multiple antennas is a well-established method to increase data rates in wireless systems. A classic result in information theory states that the throughput achievable with multiple-input multiple-output (MIMO) wireless systems grows linearly in the number of antennas when perfect channel state information (CSI) is available at the receiver [1]. In practice, though, the MIMO data rates are limited by the need to acquire CSI [2]–[7]. A fundamental way to assess the rate penalty due to channel estimation (relative to the unrealistic case where perfect CSI is available) is to study capacity in the *noncoherent setting* where neither the transmitter nor the receiver has *a priori* CSI but both are aware of the channel statistics.

This paper was presented in part at the Allerton Conference on Communication, Control, and Computing, Monticello, IL, Oct. 2012 and at the IEEE International Symposium on Information Theory (ISIT), Istanbul, Turkey, July 2013.

This work was supported by the WWTF under grant ICT10-066 (NOWIRE) and by the Swedish Research Council under grant 2012-4571.

G. Koliander and F. Hlawatsch are with the Institute of Telecommunications, Vienna University of Technology, 1040 Vienna, Austria (e-mail: guenther.koliander@nt.tuwien.ac.at, franz.hlawatsch@nt.tuwien.ac.at).

E. Riegler is with the Department of Information Technology and Electrical Engineering, ETH Zurich, 8092 Zurich, Switzerland (e-mail: eriegler@nari.ee.ethz.ch).

G. Durisi is with the Department of Signals and Systems, Chalmers University of Technology, 41296 Gothenburg, Sweden (e-mail: durisi@chalmers.se).

The model most commonly used to capture channel variations for capacity analyses in the noncoherent MIMO setting is the Rayleigh-fading *constant block-fading channel model* [2], according to which the fading process takes on independent realizations across blocks of N channel uses ("block-memoryless" assumption), and within each block the fading coefficients stay constant. Thus, the N-dimensional vector describing the channel between antennas t and r (hereafter briefly termed "(t,r) channel") within a block is

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

Here, $\mathbf{1}_{N\times 1}$ denotes the N-dimensional all-one vector and $\mathsf{s}_{r,t}$, $r \in \{1,\ldots,R\}, t \in \{1,\ldots,T\}$, are independent $\mathcal{CN}(0,1)$ random variables; T and R denote the number of transmit and receive antennas, respectively. Unfortunately, even for this simple channel model, a closed-form expression for the capacity in the noncoherent setting is unavailable. However, an accurate characterization exists for high signal-to-noise ratio (SNR). Specifically, Zheng and Tse [3] proved that the number of degrees of freedom (i.e., the asymptotic ratio between capacity and the logarithm of the SNR as the SNR grows large, also referred to as *capacity pre-log*) for the constant block-fading model is given by

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

For the case $R+T \leq N$, they also provided a high-SNR capacity expansion that is accurate up to a $o(1)$ term (i.e., a term that vanishes as the SNR grows). This expansion was recently extended in [8] to the "large-MIMO" setting $R+T > N$.

A. Extending the Constant Block-fading Model

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 [9]. In such systems, the channel input-output relation is most conveniently described in the frequency domain: the vector of channel gains $\mathbf{h}_{r,t}$ is equal to the Fourier transform of the discrete-time impulse response $\mathbf{c}_{r,t}$ of the (t,r) channel. The constant block-fading model here corresponds to the situation where the impulse response of each (t,r) channel consists of a single tap, i.e., $\mathbf{c}_{r,t} = \sqrt{N}\mathsf{s}_{r,t}(1\ 0\cdots 0)^{\mathrm{T}}$, a situation for which the use of OFDM is unnecessary.

In this paper, we focus on a channel model that allows for impulse responses with multiple taps. Furthermore, we shall allow different (t, r) channels to have different correlation structures. One way to achieve these goals is to model the channel gains as

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

Here, the squared magnitude of the inverse Fourier transform of each deterministic vector $\boldsymbol{z}_{r,t}$ is equal to the power-delay profile of the corresponding (t, r) channel. To obtain an even more general system model, we assume that in each block the correlation is described by $Q \geq 1$ independent random variables according to

$$\mathsf{h}_{r,t} = \boldsymbol{Z}_{r,t}\mathsf{s}_{r,t} \tag{4}$$

where $\boldsymbol{Z}_{r,t} \in \mathbb{C}^{N\times Q}$ with $Q \leq N$ is a deterministic matrix and $\mathsf{s}_{r,t} \in \mathbb{C}^{Q}$ contains independent $\mathcal{CN}(0, 1)$ entries, which are also independent across $r \in \{1, \ldots, R\}$ and $t \in \{1, \ldots, T\}$. A similar system model, with the simplifying assumption that all matrices $\boldsymbol{Z}_{r,t}$ are equal, was analyzed in [4], where a lower bound on the number of degrees of freedom was derived. This lower bound is tight only for the single-antenna case [10]–[12].

B. Main Result

Building on our previous work in [13] and [14], we study the high-SNR capacity of MIMO block-fading channels modeled according to (4) and show that when the deterministic matrices $\boldsymbol{Z}_{r,t}$ are *generic*, the number of degrees of freedom can be larger than in the constant block-fading case as given in (2). Coarsely speaking, we can think of generic $\boldsymbol{Z}_{r,t}$ as being generated from an underlying joint probability density function.[1] We shall refer to (4) with generic $\boldsymbol{Z}_{r,t}$ as *generic block-fading model*. Our specific contribution is as follows: we show that for all matrices $\boldsymbol{Z}_{r,t}$ except for a set of Lebesgue measure zero, the number of degrees of freedom is given by

$$\chi_{\text{gen}} = T\left(1 - \frac{1}{N}\right) \tag{5}$$

provided that $T < N/Q$ and $R \geq T(N-1)/(N-TQ)$. We note that the set corresponding to the case where all matrices $\boldsymbol{Z}_{r,t}$ are *exactly* equal has Lebesgue measure zero, and thus we do not know whether (5) holds for equal $\boldsymbol{Z}_{r,t}$. Therefore, this specific case remains an open problem. We also provide an upper bound and a lower bound on χ_{gen} for the case $R < T(N-1)/(N-TQ)$.

C. Comparison with the Constant Block-fading Model

Let us compare the maximal values of χ_{const} and χ_{gen} for a fixed N, which are obtained for optimal choices of T and R. For the constant block-fading model (1) with block length N, it can be easily verified that the number of degrees of freedom χ_{const} given in (2) is maximized for $M = \lfloor N/2 \rfloor$. Setting $T = R = \lfloor N/2 \rfloor$ to obtain $M = \lfloor N/2 \rfloor$, we conclude that the maximal χ_{const} is given by

$$\chi_{\text{const,max}} = \left\lfloor \frac{N}{2} \right\rfloor \left(1 - \frac{\lfloor \frac{N}{2} \rfloor}{N}\right).$$

[1] We use the term "generic" in the same sense as it is used in the interference-alignment literature [15].

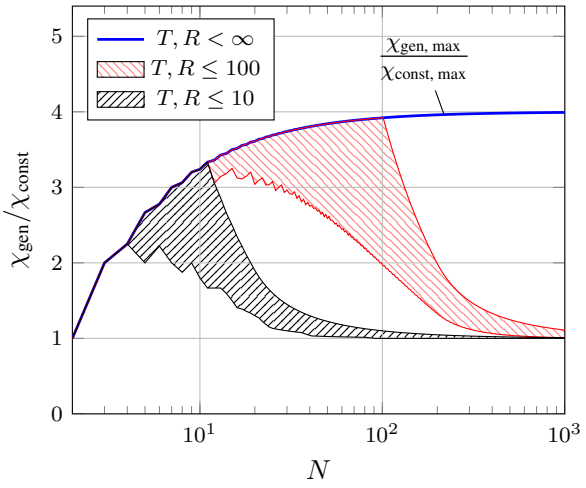

Fig. 1. Ratio between the maximal value of χ_{gen} (for the case $Q = 1$) and the maximal value of χ_{const} as a function of N, with and without a constraint on the maximal number of antennas. The shaded areas indicate the regions of $\chi_{\text{gen}}/\chi_{\text{const}}$ delimited by the upper bound (17) and lower bound (26) on χ_{gen}.

This can be easily shown to be upper-bounded by $N/4$. For the generic block-fading model with $Q = 1$ and $T < N$, it follows from (5) that the number of degrees of freedom is maximized for $T = N - 1$ and $R = (N-1)^2$, which results in

$$\chi_{\text{gen,max}} = \frac{(N-1)^2}{N}\,.$$

Fig. 1 shows the ratio between the maximal value of χ_{gen} (for $Q = 1$) and the maximal value of χ_{const} as a function of N. Because for the generic block-fading model the optimal number of receive antennas grows quadratically with N, which may yield an unreasonably large number of antennas for practically relevant values of N (e.g., 1000 symbols or more), in Fig. 1 we also show the ratio between the maximal values of χ_{gen} and χ_{const} under a constraint on the maximal number of antennas. For the case $R < T(N-1)/(N-T)$, which is relevant in the constrained setting, our upper and lower bounds on χ_{gen} (see (17) and (26) below) do not match. The degrees-of-freedom region delimited by the two bounds is represented in Fig. 1 by shaded areas. One can see from Fig. 1 that $\chi_{\text{gen, max}}$ is about four times $\chi_{\text{const, max}}$ when N grows large. However, when the maximal number of transmit and receive antennas is constrained, the ratio $\chi_{\text{gen}}/\chi_{\text{const}}$ converges to 1.

We emphasize that the only difference between the channel models (3) and (1) is that the generic (but deterministic) vectors $\boldsymbol{z}_{r,t}$ of (3) are replaced by the all-one vector in (1). It is important to note that the generic vectors $\boldsymbol{z}_{r,t}$ for which (5) holds include vectors that are arbitrarily close to the all-one vector. Hence, *arbitrarily small perturbations of the constant block-fading model may result in a significant increase in the number of degrees of freedom*. As we will demonstrate, the potential increase in the number of the degrees of freedom obtained when going from (1) to (3) is due to the fact that, under the generic block-fading model (3), the received signal vectors in the absence of noise span a subspace of higher dimension than under the constant block-fading model (1).

We conclude that the commonly used constant block-fading model results in largely pessimistic capacity estimates at high SNR.

D. Proof Techniques

To establish (5), we derive upper and lower bounds on capacity that match asymptotically (i.e., in terms of degrees of freedom). A similar approach was recently used in [11] to establish the degrees of freedom for the single-input multiple-output (SIMO) case. However, the proof techniques in [11] cannot be directly applied to the MIMO setting. A key step in [11] to obtain a tight lower bound on the number of degrees of freedom for the SIMO setting is to perform a change of variables using specific one-to-one mappings that relate the channel gains, the input signals, and the noiseless output signals. Unfortunately, the corresponding mappings for the MIMO case are not one-to-one, and hence the change-of-variable argument used in [11] cannot be applied. To overcome this problem, we invoke Bézout's theorem in algebraic geometry [16, Prop. B.2.7] and show that these mappings are at least finite-to-one almost everywhere. We also derive a bound on the change of differential entropy that occurs when a random variable undergoes a finite-to-one mapping. Finally, we use a property of subharmonic functions [17, Th. 2.6.2.1] to establish that a term appearing in this change of differential entropy is finite.

E. Notation

Sets are denoted by calligraphic letters (e.g., $\mathcal{I}$), and $|\mathcal{I}|$ denotes the cardinality of the set $\mathcal{I}$. The indicator function of a set $\mathcal{I}$ is denoted by $\mathbb{1}_{\mathcal{I}}$. Sets of sets are denoted by fraktur letters (e.g., $\mathfrak{M}$). The set of natural numbers (including zero) $\{0, 1, 2, \dots\}$ is denoted as $\mathbb{N}$. We use the notation $[M\!:\!N]$ to indicate the set $\{n \in \mathbb{N} : M \leq n \leq N\}$ for $M, N \in \mathbb{N}$. Boldface uppercase and lowercase letters denote matrices and vectors, respectively. Sans serif letters denote random quantities, e.g., A is a random matrix, x is a random vector, and s is a random scalar ($\boldsymbol{A}, \boldsymbol{x}$, and s denote the deterministic counterparts). The superscripts $^{\mathrm{T}}$ and $^{\mathrm{H}}$ stand for transposition and Hermitian transposition, respectively. The all-zero vector or matrix 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 $[\boldsymbol{x}]_i$. For an $M \times N$ matrix $\boldsymbol{A}$, we denote by $[\boldsymbol{A}]_{\mathcal{I}}^{\mathcal{J}}$, where $\mathcal{I} \subseteq [1\!:\!M]$ and $\mathcal{J} \subseteq [1\!:\!N]$, the $|\mathcal{I}| \times |\mathcal{J}|$ submatrix of $\boldsymbol{A}$ containing the entries $[\boldsymbol{A}]_i^j$ with $i \in \mathcal{I}$ and $j \in \mathcal{J}$; furthermore, we let $[\boldsymbol{A}]_{\mathcal{I}} \triangleq [\boldsymbol{A}]_{\mathcal{I}}^{[1:N]}$ and $[\boldsymbol{A}]^{\mathcal{J}} \triangleq [\boldsymbol{A}]_{[1:M]}^{\mathcal{J}}$. We denote by $[\boldsymbol{x}]_{\mathcal{I}} \in \mathbb{C}^{|\mathcal{I}|}$ the subvector of $\boldsymbol{x}$ containing the entries $[\boldsymbol{x}]_i$ with $i \in \mathcal{I}$. The diagonal matrix with the entries of $\boldsymbol{x}$ in its main diagonal is denoted by $\mathrm{diag}(\boldsymbol{x})$. We let $\mathrm{diag}(\boldsymbol{A}_1, \dots, \boldsymbol{A}_K)$ be the block-diagonal matrix having the matrices $\boldsymbol{A}_1, \dots, \boldsymbol{A}_K$ on the main block diagonal. By $|\boldsymbol{A}|$ we denote the modulus of the determinant of the square matrix $\boldsymbol{A}$. For $x \in \mathbb{R}$, we define $\lfloor x \rfloor \triangleq \max\{m \in \mathbb{Z} : m \leq x\}$ and $\lceil x \rceil \triangleq \min\{m \in \mathbb{Z} : m \geq x\}$. 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 written as $\boldsymbol{J}_\phi$. For a function ϕ with domain $\mathcal{D}$ and a subset $\widetilde{\mathcal{D}} \subseteq \mathcal{D}$, we denote by $\phi\big|_{\widetilde{\mathcal{D}}}$ the restriction of ϕ to the domain $\widetilde{\mathcal{D}}$. We use the Landau notation $f(\rho) = \mathcal{O}(g(\rho))$ to indicate that there exist constants $c_1, c_2 > 0$ such that $|f(\rho)| \leq c_1 |g(\rho)|$ for $\rho > c_2$. Similarly, we use $f(\rho) = o(g(\rho))$ to indicate that for every $\varepsilon > 0$ there exists a constant $c_3 > 0$ such that $|f(\rho)| \leq \varepsilon |g(\rho)|$ for $\rho > c_3$.

F. Organization of the Paper

The rest of this paper is organized as follows. The system model is formulated in Section II. In Section III, we present and discuss our main result on the number of degrees of freedom of the generic block-fading MIMO channel. An underlying upper bound is stated and proved in Section IV, and a corresponding lower bound is given in Section V. In Section VI and in four appendices, we provide a proof of the lower bound.

II. System Model

We consider a MIMO channel with T transmit and R receive antennas. The discrete-time fading process associated with each transmit-receive antenna pair conforms to a block-fading model, which results in the following channel input-output relations within a given block of N channel uses:

$$\mathsf{y}_r = \sqrt{\frac{\rho}{T}} \sum_{t \in [1:T]} \mathrm{diag}(\mathsf{h}_{r,t})\, \mathsf{x}_t + \mathsf{w}_r\,, \quad r \in [1\!:\!R]\,. \tag{6}$$

Here, $\mathsf{x}_t \in \mathbb{C}^N$ is the signal vector originating from the tth transmit antenna; $\mathsf{y}_r \in \mathbb{C}^N$ is the signal vector at the rth receive antenna; $\mathsf{h}_{r,t} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}_{r,t})$ is the vector of N channel coefficients between the tth transmit antenna and the rth receive antenna; $\mathsf{w}_r \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$ is the noise vector at the rth receive antenna; and $\rho \in \mathbb{R}^+$ is the SNR. The vectors $\mathsf{h}_{r,t}$ and w_r are assumed to be mutually independent and independent across $r \in [1:R]$ and $t \in [1:T]$, and to change in an independent fashion from block to block ("block-memoryless" assumption). The transmitted signal vectors x_t are assumed to be independent of the vectors $\mathsf{h}_{r,t}$ and w_r. We consider the noncoherent setting, where transmitter and receiver know the covariance matrix $\boldsymbol{\Sigma}_{r,t}$ of $\mathsf{h}_{r,t}$ but have no *a priori* knowledge of the realization of $\mathsf{h}_{r,t}$.

Because the covariance matrix $\boldsymbol{\Sigma}_{r,t}$ is positive-semidefinite, it can be factorized as

$$\boldsymbol{\Sigma}_{r,t} = \boldsymbol{Z}_{r,t} \boldsymbol{Z}_{r,t}^{\mathrm{H}}$$

with $\boldsymbol{Z}_{r,t} \in \mathbb{C}^{N \times Q}$ and $Q = \mathrm{rank}(\boldsymbol{\Sigma}_{r,t}) = \mathrm{rank}(\boldsymbol{Z}_{r,t})$. We can then rewrite the channel coefficient vectors $\mathsf{h}_{r,t}$ in terms of $\boldsymbol{Z}_{r,t}$ as in (4), i.e.,

$$\mathsf{h}_{r,t} = \boldsymbol{Z}_{r,t} \mathsf{s}_{r,t} \tag{7}$$

where $\mathsf{s}_{r,t} \in \mathbb{C}^Q$, $\mathsf{s}_{r,t} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_Q)$. Using (7), the R input-output relations (6) can be rewritten as

$$\mathsf{y}_r = \sqrt{\frac{\rho}{T}} \sum_{t \in [1:T]} \mathrm{diag}(\boldsymbol{Z}_{r,t} \mathsf{s}_{r,t})\, \mathsf{x}_t + \mathsf{w}_r\,, \quad r \in [1\!:\!R] \tag{8}$$

or in stacked form as

$$\mathsf{y} = \sqrt{\frac{\rho}{T}}\bar{\mathsf{y}} + \mathsf{w}, \quad \text{with } \bar{\mathsf{y}} \triangleq \mathsf{B}\mathsf{s} \tag{9}$$

where $\mathsf{y} \triangleq (\mathsf{y}_1^{\mathrm{T}} \cdots \mathsf{y}_R^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{C}^{RN}$, $\mathsf{w} \triangleq (\mathsf{w}_1^{\mathrm{T}} \cdots \mathsf{w}_R^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{C}^{RN}$, $\mathsf{s} \triangleq (\mathsf{s}_1^{\mathrm{T}} \cdots \mathsf{s}_R^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{C}^{RTQ}$ with $\mathsf{s}_r \triangleq (\mathsf{s}_{r,1}^{\mathrm{T}} \cdots \mathsf{s}_{r,T}^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{C}^{TQ}$, and

$$\mathsf{B} \triangleq \begin{pmatrix} \mathsf{B}_1 & & \\ & \ddots & \\ & & \mathsf{B}_R \end{pmatrix} \in \mathbb{C}^{RN \times RTQ},$$
$$\text{with } \mathsf{B}_r \triangleq (\mathsf{X}_1 \boldsymbol{Z}_{r,1} \cdots \mathsf{X}_T \boldsymbol{Z}_{r,T}) \in \mathbb{C}^{N \times TQ} \tag{10}$$

where $\mathsf{X}_t \triangleq \mathrm{diag}(\mathsf{x}_t) \in \mathbb{C}^{N \times N}$. For later use, we also define $\mathsf{x} \triangleq (\mathsf{x}_1^{\mathrm{T}} \cdots \mathsf{x}_T^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{C}^{TN}$ and

$$\boldsymbol{Z} \triangleq \begin{pmatrix} \boldsymbol{Z}_{1,1} & \cdots & \boldsymbol{Z}_{1,T} \\ \vdots & & \vdots \\ \boldsymbol{Z}_{R,1} & \cdots & \boldsymbol{Z}_{R,T} \end{pmatrix} \in \mathbb{C}^{RN \times TQ}.$$

The matrix $\boldsymbol{Z}$ contains all information about the correlation of the channel coefficients $\mathsf{h}_{r,t}$ (recall that $\boldsymbol{\Sigma}_{r,t} = \boldsymbol{Z}_{r,t}\boldsymbol{Z}_{r,t}^{\mathrm{H}}$). We will refer to $\boldsymbol{Z}$ as *coloring matrix* and use the phrase "for a generic coloring matrix $\boldsymbol{Z}$" to indicate that a property holds for *almost every* matrix $\boldsymbol{Z}$. Here, "almost every" is understood in the precise mathematical sense that the set of all matrices $\boldsymbol{Z}$ for which the property does *not* hold has Lebesgue measure zero.

In the special (nongeneric) case where $Q = 1$ and each $\boldsymbol{Z}_{r,t} \in \mathbb{C}^{N \times 1}$ is the all-one vector, (8) reduces to the input-output relation of the constant block-fading model given by (cf. (1))

$$\mathsf{y}_r = \sqrt{\frac{\rho}{T}} \sum_{t \in [1:T]} \mathsf{s}_{r,t}\, \mathsf{x}_t + \mathsf{w}_r\,, \quad r \in [1:R]\,. \tag{11}$$

III. Characterization of the Number of Degrees of Freedom

A. Main Result

Because of the block-memoryless assumption, the coding theorem in [18, Section 7.3] implies that the capacity of the channel (8) is given by

$$C(\rho) = \frac{1}{N} \sup I(\mathsf{x}; \mathsf{y})\,. \tag{12}$$

Here, $I(\cdot;\cdot)$ denotes mutual information [19, p. 251] and the supremum is taken over all probability distributions of x that satisfy the average-power constraint

$$\mathbb{E}[\|\mathsf{x}\|^2] \leq TN\,. \tag{13}$$

The number of degrees of freedom is defined as

$$\chi \triangleq \lim_{\rho \to \infty} \frac{C(\rho)}{\log \rho} \tag{14}$$

which corresponds to the expansion

$$C(\rho) = \chi \log \rho + o(\log \rho)\,. \tag{15}$$

Our main result is stated in the following theorem.

Theorem 1: Let $T < N/Q$ and $R \geq T(N-1)/(N-TQ)$. For a channel conforming to the generic block-fading model, i.e., the channel (8) with generic coloring matrix $\boldsymbol{Z}$, the number of degrees of freedom is given by

$$\chi_{\text{gen}} = T\left(1 - \frac{1}{N}\right). \tag{16}$$

Proof: In Section IV, we will show that χ_{gen} is upper-bounded by $T(1-1/N)$ for all choices of T, R, N, Q, and $\boldsymbol{Z}$. In Section V, we will show that this upper bound is achievable when $T < N/Q$, $R \geq T(N-1)/(N-TQ)$, and $\boldsymbol{Z}$ is generic (see Corollary 5). ∎

B. Degrees of Freedom Gain

As discussed in Section I, (16) implies that the maximal achieveable number of degrees of freedom in the generic block-fading model can be about four times as large as the number of degrees of freedom in the constant block-fading model (2). We will now provide some intuition regarding this gain. For concreteness, we consider the case $T = 2, R = 3, Q = 1, N = 4$. In this case, (2) and (16) give $\chi_{\text{const}} = 1$ and $\chi_{\text{gen}} = 3/2$, respectively.

The number of degrees of freedom characterizes the channel capacity in a regime where the noise can "effectively" be ignored. Thus, according to the intuitive argumentation in [12, Section III], the number of degrees of freedom should be equal to the number of entries of $\mathsf{x} \in \mathbb{C}^8$ that can be deduced from the corresponding received vector $\mathsf{y} \in \mathbb{C}^{12}$ in the absence of noise, divided by the block length $N = 4$.

In the constant block-fading model (11), 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 two of them contain all the information available about x. From two of the received vectors, we obtain $2 \cdot 4$ scalar equations in $8 + 4$ scalar 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 the remaining four entries of x. Hence, we can deduce four entries of x from $\bar{\mathsf{y}}$. We conclude that the number of degrees of freedom is $4/4 = 1$, which is in agreement with (2).

In the generic block-fading model (8), on the other hand, the received vectors without noise

$$\bar{\mathsf{y}}_r = \mathrm{diag}(\boldsymbol{Z}_{r,1}\mathsf{s}_{r,1})\mathsf{x}_1 + \mathrm{diag}(\boldsymbol{Z}_{r,2}\mathsf{s}_{r,2})\mathsf{x}_2, \quad r = 1, 2, 3$$

span a three-dimensional subspace almost surely. Hence, we obtain a system of $3 \cdot 4$ equations in $8 + 6$ variables $(\mathsf{x}, \mathsf{s}_{1,1}, \mathsf{s}_{1,2}, \mathsf{s}_{2,1}, \mathsf{s}_{2,2}, \mathsf{s}_{3,1}, \mathsf{s}_{3,2})$. Fixing two entries of x, we are able to recover the remaining six entries. Hence, the number of degrees of freedom is $6/4 = 3/2$, which is in agreement with (16).

This argument suggests that the reason why the generic block-fading model yields a larger number of degrees of freedom 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 number of degrees of freedom of the channel (8) holds for every T, R, Q, N, and $\boldsymbol{Z}$. The assumption of a generic coloring matrix $\boldsymbol{Z}$ is not required.

Theorem 2: The number of degrees of freedom of the channel (8) satisfies

$$\chi_{\text{gen}} \leq T\left(1-\frac{1}{N}\right). \tag{17}$$

Proof: We will show that the number of degrees of freedom is upper-bounded by T times the number of degrees of freedom of a constant block-fading SIMO channel; the result then follows from (2). To this end, we will rewrite each output vector y_r as the sum of the output vectors of T SIMO systems with RQ receive antennas each. This will be achieved by splitting the additive noise variables appropriately.

From (8), the ith entry of the received vector y_r is given by

$$[\mathsf{y}_r]_i = \sqrt{\frac{\rho}{T}} \sum_{t\in[1:T]} \sum_{q\in[1:Q]} [\boldsymbol{Z}_{r,t}]_i^q [\mathsf{s}_{r,t}]_q [\mathsf{x}_t]_i + [\mathsf{w}_r]_i \tag{18}$$

for $r \in [1:R]$. We first decompose the noise variables according to

$$[\mathsf{w}_r]_i = \sum_{t\in[1:T]} \sum_{q\in[1:Q]} \frac{[\boldsymbol{Z}_{r,t}]_i^q}{\sqrt{KT}} [\tilde{\mathsf{w}}_{q,r,t}]_i + [\mathsf{w}'_r]_i \,. \tag{19}$$

Here, all $[\tilde{\mathsf{w}}_{q,r,t}]_i$ and $[\mathsf{w}'_r]_i$ are mutually independent and independent of all x_t and $\mathsf{s}_{r,t}$. Furthermore, $[\tilde{\mathsf{w}}_{q,r,t}]_i \sim \mathcal{CN}(0,1)$,

$$[\mathsf{w}'_r]_i \sim \mathcal{CN}\left(0, 1 - \sum_{t\in[1:T]} \sum_{q\in[1:Q]} \frac{|[\boldsymbol{Z}_{r,t}]_i^q|^2}{KT}\right),$$

and K is a finite constant satisfying[2]

$$K > \max_{r\in[1:R], i\in[1:N]} \sum_{t\in[1:T]} \sum_{q\in[1:Q]} |[\boldsymbol{Z}_{r,t}]_i^q|^2 \,.$$

We next define T "virtual" constant block-fading SIMO channels with RQ receive antennas each:

$$[\tilde{\mathsf{y}}_{q,r,t}]_i = \sqrt{K\rho}\,[\mathsf{s}_{r,t}]_q [\mathsf{x}_t]_i + [\tilde{\mathsf{w}}_{q,r,t}]_i \,, \quad i\in[1:N], r\in[1:R], q\in[1:Q] \tag{20}$$

for $t\in[1:T]$. Inserting (19) into (18) and using (20), it can be verified that (18) can be rewritten as

$$[\mathsf{y}_r]_i = \frac{1}{\sqrt{KT}} \sum_{t\in[1:T]} \sum_{q\in[1:Q]} [\boldsymbol{Z}_{r,t}]_i^q [\tilde{\mathsf{y}}_{q,r,t}]_i + [\mathsf{w}'_r]_i \,. \tag{21}$$

Let $\tilde{\mathsf{y}}_t \triangleq (\tilde{\mathsf{y}}_{1,1,t}^{\mathrm{T}} \cdots \tilde{\mathsf{y}}_{Q,R,t}^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{C}^{QRN}$. By (21), the random variable y depends on x only via the random variables $\{\tilde{\mathsf{y}}_t\}_{t\in[1:T]}$. Hence, the data-processing inequality [18, eq. (2.3.19)] yields

$$I(\mathsf{x};\mathsf{y}) \leq I(\mathsf{x};\tilde{\mathsf{y}}_1,\ldots,\tilde{\mathsf{y}}_T) \,. \tag{22}$$

The right-hand side of (22) can be upper-bounded as follows:

$$\begin{aligned} I(\mathsf{x};\tilde{\mathsf{y}}_1,\ldots,\tilde{\mathsf{y}}_T) &= h(\tilde{\mathsf{y}}_1,\ldots,\tilde{\mathsf{y}}_T) - h(\tilde{\mathsf{y}}_1,\ldots,\tilde{\mathsf{y}}_T|\mathsf{x}) \\ &\overset{(a)}{=} h(\tilde{\mathsf{y}}_1,\ldots,\tilde{\mathsf{y}}_T) - \sum_{t\in[1:T]} h(\tilde{\mathsf{y}}_t|\mathsf{x}_t) \\ &\overset{(b)}{\leq} \sum_{t\in[1:T]} \big[h(\tilde{\mathsf{y}}_t) - h(\tilde{\mathsf{y}}_t|\mathsf{x}_t)\big] \\ &= \sum_{t\in[1:T]} I(\mathsf{x}_t;\tilde{\mathsf{y}}_t) \,. \end{aligned} \tag{23}$$

Here, $h(\cdot)$ denotes differential entropy [19, Ch. 8], (a) holds because $\tilde{\mathsf{y}}_1,\ldots,\tilde{\mathsf{y}}_T$ are conditionally independent given x, and (b) follows from the chain rule for differential entropy [19, Th. 8.6.2] and because conditioning does not increase differential entropy. Since (by assumption) the input vector x satisfies the power constraint (13), we conclude that, trivially, also each subvector x_t satisfies the individual power constraint $\mathbb{E}[\|\mathsf{x}_t\|^2] \leq TN$. Thus, the SNR (i.e., the expected power of the noiseless received signal divided by the noise power) of each "virtual" constant block-fading SIMO channel (20) is given by

$$\begin{aligned} \frac{\mathbb{E}[\|\sqrt{K\rho}\,[\mathsf{s}_{r,t}]_q \mathsf{x}_t\|^2]}{\mathbb{E}[\|\tilde{\mathsf{w}}_{q,r,t}\|^2]} &= \frac{K\rho\,\mathbb{E}[|[\mathsf{s}_{r,t}]_q|^2]\,\mathbb{E}[\|\mathsf{x}_t\|^2]}{\mathbb{E}[\|\tilde{\mathsf{w}}_{q,r,t}\|^2]} \\ &\leq \frac{K\rho TN}{N} \\ &= TK\rho \,. \end{aligned}$$

By (2) and (15), the capacity of a constant block-fading SIMO channel of SNR $TK\rho$ is of the form[3] $(1-1/N)\log(TK\rho) + o(\log\rho)$. Since, by (12), the capacity is the supremum of the mutual information divided by the block length, we can upper-bound each mutual information $I(\mathsf{x}_t;\tilde{\mathsf{y}}_t)$, $t\in[1:T]$ by N times the capacity. This results in

$$\begin{aligned} I(\mathsf{x}_t;\tilde{\mathsf{y}}_t) &\leq N\left(\left(1-\frac{1}{N}\right)\log(TK\rho) + o(\log\rho)\right) \\ &= (N-1)\log(TK\rho) + o(\log\rho) \,. \end{aligned}$$

Hence, continuing (22) and (23), we obtain

$$\begin{aligned} I(\mathsf{x};\mathsf{y}) &\leq \sum_{t\in[1:T]} I(\mathsf{x}_t;\tilde{\mathsf{y}}_t) \\ &\leq T(N-1)\log(TK\rho) + o(\log\rho) \\ &\overset{(a)}{=} T(N-1)\log\rho + o(\log\rho) \end{aligned} \tag{24}$$

where (a) holds because $\log(TK\rho) = \log\rho + \log(TK)$. Thus, the mutual information $I(\mathsf{x};\mathsf{y})$ with x satisfying the power constraint (13) is upper-bounded by (24). Inserting (24) into (12) yields

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

from which (17) follows via (14). ∎

V. Lower Bound

We first derive a lower bound on χ_{gen} assuming that $\widetilde{T} \leq \min\{T,R\}$ transmit antennas are effectively used (i.e., $\mathsf{x}_{\widetilde{T}+1},\ldots,\mathsf{x}_T$ are set to zero). Then we maximize the lower

[2] This condition on K is required to ensure that the variance of all random variables $[\mathsf{w}'_r]_i$ is positive.

[3] Since the number of transmit antennas is one for a SIMO channel, we have $M=1$ in (2).

bound by identifying the optimal number $\widetilde{T}$ of transmit antennas to use.

Proposition 3: The number of degrees of freedom of the channel (8) for a generic coloring matrix $\boldsymbol{Z}$ is lower-bounded by

$$\chi_{\text{gen}} \geq \chi_{\text{low}}(\widetilde{T}) \triangleq \min\left\{\widetilde{T}\left(1-\frac{1}{N}\right), R\left(1-\frac{\widetilde{T}Q}{N}\right)\right\} \quad (25)$$

for all $\widetilde{T} \leq \min\{T, R\}$.

Proof: See Section VI. ∎

The minimum in (25) is given by $\chi_{\text{low}}(\widetilde{T}) = \widetilde{T}(1-1/N)$ when the number R of receive antennas is large enough (i.e., $R \geq \widetilde{T}(N-1)/(N-\widetilde{T}Q)$). In contrast, $\chi_{\text{low}}(\widetilde{T}) = R(1-\widetilde{T}Q/N)$ when the number of degrees of freedom is constrained by the limited number of receive antennas (i.e., $R < \widetilde{T}(N-1)/(N-\widetilde{T}Q)$).

The main result of this section is stated in the following theorem.

Theorem 4: The number of degrees of freedom of the channel (8) for a generic coloring matrix $\boldsymbol{Z}$ is lower-bounded by

$$\begin{aligned}\chi_{\text{gen}} \geq \chi^*_{\text{low}} &\triangleq \max_{\widetilde{T}\leq\min\{T,R\}} \chi_{\text{low}}(\widetilde{T}) \\ &= \begin{cases} T\left(1-\frac{1}{N}\right), & \text{if } T \leq T_{\text{opt}} \\ \eta, & \text{if } T > T_{\text{opt}} \end{cases} \end{aligned} \quad (26)$$

where

$$T_{\text{opt}} \triangleq \frac{RN}{N+RQ-1} \quad (27)$$

and

$$\eta \triangleq \max\left\{R\left(1-\frac{\lceil T_{\text{opt}}\rceil Q}{N}\right), \lfloor T_{\text{opt}}\rfloor\left(1-\frac{1}{N}\right)\right\}. \quad (28)$$

Proof: The idea behind the bound χ^*_{low} in (26) is to obtain the tightest (i.e., largest) of the lower bounds $\chi_{\text{low}}(\widetilde{T})$ in (25) for T transmit antennas by maximizing $\chi_{\text{low}}(\widetilde{T})$ with respect to the number of effectively used transmit antennas $\widetilde{T} \leq \min\{T, R\}$. According to (25), $\chi_{\text{low}}(\widetilde{T})$ is the minimum of two quantities where the first, $\widetilde{T}(1-1/N)$, is monotonically increasing in $\widetilde{T}$ and the second, $R(1-\widetilde{T}Q/N)$, is monotonically decreasing in $\widetilde{T}$. Hence, $\chi_{\text{low}}(\widetilde{T})$ attains its maximum at the intersection point T_{opt} defined in (27). If $T \leq T_{\text{opt}}$, we are for all $\widetilde{T} \leq \min\{T, R\}$ in the regime where $\chi_{\text{low}}(\widetilde{T})$ is monotonically increasing, and thus the best choice is to use $\widetilde{T} = T$ transmit antennas (note that because $T \leq T_{\text{opt}} \overset{(27)}{\leq} RN/N = R$, the choice $\widetilde{T} = T$ in Proposition 3 is possible). Thus, in this case we have $\chi^*_{\text{low}} = \chi_{\text{low}}(T) = T(1-1/N)$, which yields the first case in (26). If $T > T_{\text{opt}}$, we would like to use T_{opt} transmit antennas, but we have to take into account that T_{opt} may be noninteger. Thus, we take the maximum of the bounds $\chi_{\text{low}}(\widetilde{T})$ resulting from the closest integers, $\chi_{\text{low}}(\lfloor T_{\text{opt}}\rfloor)$ and $\chi_{\text{low}}(\lceil T_{\text{opt}}\rceil)$, which yields η in (28). This concludes the proof. ∎

Remark 1: For $N \geq 2$, the optimal number of transmit antennas T_{opt} is upper-bounded as follows:

$$T_{\text{opt}} < \frac{N}{Q}. \quad (29)$$

In fact, $T_{\text{opt}} = RN/(N+RQ-1) < RN/(RQ) = N/Q$.

Remark 2: For $N = Q \geq 2$, we have by (29) that $T_{\text{opt}} < 1$. Hence, $T > T_{\text{opt}}$ and thus, by (26) and (28), $\chi^*_{\text{low}} = \eta = \max\{R(1-Q/N), 0\} = 0$. Similarly, we obtain for $N = 1$ that $\chi_{\text{low}}(\widetilde{T}) \leq 0$ for all $\widetilde{T}$, which yields $\chi^*_{\text{low}} \leq 0$. Hence, our lower bound χ^*_{low} is trivial. In these scenarios, the capacity grows double-logarithmically in the SNR ρ [20], [21].

Remark 3: The lower bound χ^*_{low} in (26) can be equivalently expressed as

$$\chi^*_{\text{low}} = \min\left\{T\left(1-\frac{1}{N}\right), \eta\right\}.$$

Corollary 5: Let $N \geq 2$. For the lower bound χ^*_{low} in Theorem 4, the following properties hold:

(i) For $T \geq N/Q$, we have $T > T_{\text{opt}}$ and $\chi^*_{\text{low}} = \eta$.
(ii) For $T < N/Q$ and $R \geq T(N-1)/(N-TQ)$, we have $T \leq T_{\text{opt}}$ and $\chi^*_{\text{low}} = T(1-1/N)$.
(iii) For $T < N/Q$ and $R < T(N-1)/(N-TQ)$, we have $T > T_{\text{opt}}$ and $\chi^*_{\text{low}} = \eta$.
(iv) For fixed N and Q, χ^*_{low} attains its maximal value for $T = \lfloor (N-1)/Q\rfloor$ transmit antennas and $R = \lceil (N-1)^2/Q\rceil$ receive antennas; this maximal value of χ^*_{low} equals $\lfloor (N-1)/Q\rfloor(1-1/N)$.

Proof: By (29), the inequality $T \geq N/Q$ implies $T > T_{\text{opt}}$, from which Property (i) follows by (26). For $T < N/Q$, the following equivalence holds:

$$T \leq T_{\text{opt}} \overset{(27)}{=} \frac{RN}{N+RQ-1} \quad \Leftrightarrow \quad T\frac{N-1}{N-TQ} \leq R.$$

Thus, the conditions in Properties (ii) and (iii) imply $T \leq T_{\text{opt}}$ and $T > T_{\text{opt}}$, respectively, and the expressions of χ^*_{low} given in Properties (ii) and (iii) follow immediately from the case distinction in (26).

To prove Property (iv), we first show that $\chi^*_{\text{low}} \leq \lfloor (N-1)/Q\rfloor(1-1/N)$ for arbitrary T and R. Subsequently, we will show that this upper bound is achievable for the proposed number of antennas. We first note that for each $\widetilde{T} \leq N/Q$, the lower bound $\chi_{\text{low}}(\widetilde{T})$ in (25) is monotonically nondecreasing in R. Furthermore, for $\widetilde{T} > N/Q$, $\chi_{\text{low}}(\widetilde{T})$ is negative and can be ignored in the maximization process, i.e., we have $\chi^*_{\text{low}} = \max_{\widetilde{T}\leq\min\{T,R,N/Q\}} \chi_{\text{low}}(\widetilde{T})$. This implies that χ^*_{low} is—as a maximum of nondecreasing functions—also monotonically nondecreasing in R. Hence, to obtain an upper bound on χ^*_{low}, we can assume R arbitrarily large without loss of generality. We choose $R > (N-1)^2/Q$. Simple algebraic manipulations yield the equivalence

$$R > \frac{(N-1)^2}{Q} \quad \Leftrightarrow \quad T_{\text{opt}} = \frac{RN}{N+RQ-1} > \frac{N-1}{Q}. \quad (30)$$

This implies $\lceil T_{\text{opt}} \rceil Q > N-1$ and further, because both sides of this strict inequality are integers, that $\lceil T_{\text{opt}} \rceil Q \geq N$. Thus, the first argument of the maximum defining η in (28) satisfies

$$R\left(1 - \frac{\lceil T_{\text{opt}} \rceil Q}{N}\right) \leq R(1-1) = 0$$

and, hence, η reduces to $\eta = \lfloor T_{\text{opt}} \rfloor (1 - 1/N)$. By (26), we have that χ^*_{low} is either equal to $T(1-1/N)$ (for $T \leq T_{\text{opt}}$) or equal to $\eta = \lfloor T_{\text{opt}} \rfloor (1-1/N)$ (for $T > T_{\text{opt}}$). In both cases we have $\chi^*_{\text{low}} \leq \lfloor T_{\text{opt}} \rfloor (1-1/N)$. Since $\lfloor T_{\text{opt}} \rfloor \leq \lfloor (N-1)/Q \rfloor$ by[4] (29), this implies $\chi^*_{\text{low}} \leq \lfloor (N-1)/Q \rfloor (1-1/N)$.

It remains to be shown that this upper bound is achievable. For $R = \lceil (N-1)^2/Q \rceil \geq (N-1)^2/Q$, we obtain (see (30) with ">" replaced by "$\geq$") that $T_{\text{opt}} \geq (N-1)/Q$. Hence, for $T = \lfloor (N-1)/Q \rfloor \leq T_{\text{opt}}$, the lower bound (26) simplifies to $\chi^*_{\text{low}} = T(1-1/N) = \lfloor (N-1)/Q \rfloor (1-1/N)$. Thus, we have shown that χ^*_{low} is maximized for $T = \lfloor (N-1)/Q \rfloor$ and $R = \lceil (N-1)^2/Q \rceil$ and its maximum equals $\lfloor (N-1)/Q \rfloor (1-1/N)$. ∎

Remark 4: Property (ii) in Corollary 5 shows that for a fixed $T < N/Q$, we can achieve $\chi^*_{\text{low}} = T(1-1/N)$ by using a sufficiently large number of receive antennas R. This coincides with the upper bound presented in Section IV. Thus, in this regime, the number of degrees of freedom grows linearly in the number of transmit antennas.

VI. Proof of Proposition 3

In this section, we establish the lower bound (25). For $N \leq \widetilde{T}Q$, the inequality in (25) is trivially true, because in this case $R(1 - \widetilde{T}Q/N) \leq 0$ and hence $\chi_{\text{low}} \leq 0$. Therefore, we focus on the case

$$N > \widetilde{T}Q$$

which will thus be assumed in the remainder of this section. Furthermore, recall that we assumed in Proposition 3 that $\widetilde{T} \leq \min\{T, R\}$. Thus, setting $\mathbf{x}_{\widetilde{T}+1}, \ldots, \mathbf{x}_T$ to zero, we can replace T by $\widetilde{T}$ in the input-output relation (9) and the power constraint (13). Finally, we shall assume that

$$R \leq \left\lceil \frac{\widetilde{T}(N-1)}{N - \widetilde{T}Q} \right\rceil .$$

If more receive antennas are available, we simply turn them off. The following dimension counting argument provides some intuition on why the use of more than $\lceil \widetilde{T}(N-1)/(N - \widetilde{T}Q) \rceil$ receive antennas is not beneficial.

A. Dimension Counting

The noiseless received vector $\bar{\mathbf{y}} = \mathbf{B}\mathbf{s} \in \mathbb{C}^{RN}$ in (9) corresponds to RN polynomial equations. The unknown variables of these equations are the entries of the vectors $\mathbf{s}_{r,t} \in \mathbb{C}^Q$, $r \in [1:R]$, $t \in [1:\widetilde{T}]$ ($R\widetilde{T}Q$ unknown variables) and of the transmitted signal vectors $\mathbf{x}_t \in \mathbb{C}^N$, $t \in [1:\widetilde{T}]$ ($\widetilde{T}N$ unknown variables). Consider now a pair $(\mathbf{x}_t, \mathbf{s}_{r,t})$, consisting of a transmitted signal vector $\mathbf{x}_t$ and a fading vector $\mathbf{s}_{r,t}$ that is a solution of $\bar{\mathbf{y}} = \mathbf{B}\mathbf{s}$. Then the pair $(c_t \mathbf{x}_t, \mathbf{s}_{r,t}/c_t)$, where c_t is an arbitrary nonzero constant, is also a solution of $\bar{\mathbf{y}} = \mathbf{B}\mathbf{s}$. This implies that each $\mathbf{x}_t$ can be recovered from $\bar{\mathbf{y}}$ only up to a scaling factor. To resolve this ambiguity, we fix one entry in each $\mathbf{x}_t$. Hence, the total number of unknown variables becomes $R\widetilde{T}Q + \widetilde{T}N - \widetilde{T}$. As long as the number of equations is larger than or equal to the number of unknown variables, i.e., $RN \geq R\widetilde{T}Q + \widetilde{T}N - \widetilde{T}$, we are able to recover[5] the $N-1$ unknown entries of each $\mathbf{x}_t$. The above condition is equivalent to $R \geq \widetilde{T}(N-1)/(N - \widetilde{T}Q)$. Hence, it is reasonable to consider only the case $R \leq \lceil \widetilde{T}(N-1)/(N - \widetilde{T}Q) \rceil$, as the received vectors resulting from the use of additional receive antennas would not help us gain more information about the transmit vectors $\{\mathbf{x}_t\}_{t \in [1:\widetilde{T}]}$.

B. Bounding $I(\mathbf{x};\mathbf{y})$

By (12), the capacity $C(\rho)$ and, hence, χ_{gen} (cf. (14)) can be lower-bounded by evaluating $I(\mathbf{x};\mathbf{y})$ for any specific input distribution that satisfies the power constraint (13). In particular, in what follows, we will assume $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{\widetilde{T}N})$. Thus,

$$C(\rho) \geq \frac{1}{N} I(\mathbf{x};\mathbf{y}) . \tag{31}$$

As

$$I(\mathbf{x};\mathbf{y}) = h(\mathbf{y}) - h(\mathbf{y}|\mathbf{x}) \tag{32}$$

we can lower-bound $I(\mathbf{x};\mathbf{y})$ by upper-bounding $h(\mathbf{y}|\mathbf{x})$ and lower-bounding $h(\mathbf{y})$.

1) Upper Bound on $h(\mathbf{y}|\mathbf{x})$: It follows from (9) and (10) together with $\mathbf{s}_{r,t} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_Q)$ and $\mathbf{w}_r \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$ that $\mathbf{y}$ is conditionally Gaussian given $\mathbf{x}$, with conditional covariance matrix $(\rho/\widetilde{T})\mathbf{B}\mathbf{B}^{\text{H}} + \mathbf{I}_{RN}$ (note that $\mathbf{B} = \mathbf{B}(\mathbf{x})$). Hence,

$$h(\mathbf{y}|\mathbf{x}) = \mathbb{E}_{\mathbf{x}}\left[\log\left((\pi e)^{RN}\left|\frac{\rho}{\widetilde{T}}\mathbf{B}\mathbf{B}^{\text{H}} + \mathbf{I}_{RN}\right|\right)\right]$$

according to [22, Th. 2]. By [23, Th. 1.3.20], $\left|(\rho/\widetilde{T})\mathbf{B}\mathbf{B}^{\text{H}} + \mathbf{I}_{RN}\right| = \left|(\rho/\widetilde{T})\mathbf{B}^{\text{H}}\mathbf{B} + \mathbf{I}_{R\widetilde{T}Q}\right|$. Furthermore, assuming without loss of generality that $\rho > 1$ (note that we are only interested in the asymptotic regime $\rho \to \infty$), we have $\left|(\rho/\widetilde{T})\mathbf{B}^{\text{H}}\mathbf{B} + \mathbf{I}_{R\widetilde{T}Q}\right| \leq \left|\rho\big((1/\widetilde{T})\mathbf{B}^{\text{H}}\mathbf{B} + \mathbf{I}_{R\widetilde{T}Q}\big)\right| = \rho^{R\widetilde{T}Q}\left|(1/\widetilde{T})\mathbf{B}^{\text{H}}\mathbf{B} + \mathbf{I}_{R\widetilde{T}Q}\right|$. Thus,

$$\begin{aligned} h(\mathbf{y}|\mathbf{x}) &\leq \mathbb{E}_{\mathbf{x}}\left[\log\left((\pi e)^{RN}\rho^{R\widetilde{T}Q}\left|\frac{1}{\widetilde{T}}\mathbf{B}^{\text{H}}\mathbf{B} + \mathbf{I}_{R\widetilde{T}Q}\right|\right)\right] \\ &= R\widetilde{T}Q\log\rho + \mathbb{E}_{\mathbf{x}}\left[\log\left|\frac{1}{\widetilde{T}}\mathbf{B}^{\text{H}}\mathbf{B} + \mathbf{I}_{R\widetilde{T}Q}\right|\right] + \mathcal{O}(1). \end{aligned} \tag{33}$$

By using Jensen's inequality for the concave function $\log(\cdot)$, we obtain

$$\mathbb{E}_{\mathbf{x}}\left[\log\left|\frac{1}{\widetilde{T}}\mathbf{B}^{\text{H}}\mathbf{B} + \mathbf{I}_{R\widetilde{T}Q}\right|\right] \leq \log\mathbb{E}_{\mathbf{x}}\left[\left|\frac{1}{\widetilde{T}}\mathbf{B}^{\text{H}}\mathbf{B} + \mathbf{I}_{R\widetilde{T}Q}\right|\right] . \tag{34}$$

[4] By (29), $\lfloor T_{\text{opt}} \rfloor < N/Q$ and thus $Q\lfloor T_{\text{opt}} \rfloor < N$. Since both sides of this strict inequality are integers, we have $Q\lfloor T_{\text{opt}} \rfloor \leq N-1$ and hence $\lfloor T_{\text{opt}} \rfloor \leq (N-1)/Q$, which in turn implies $\lfloor T_{\text{opt}} \rfloor \leq \lfloor (N-1)/Q \rfloor$.

[5] Strictly speaking, this argument is true for linear equations. In our case, because we have polynomial rather than linear equations, we obtain in general a finite number of solutions for the variables $\mathbf{x}$ and not a unique solution, as will be discussed further in Section VI-C.

The right-hand side in (34) is independent of ρ and the determinant $\left|(1/\widetilde{T})\mathbf{B}^{\mathrm{H}}\mathbf{B}+\mathbf{I}_{R\widetilde{T}Q}\right|$ is some polynomial in the entries of x and x^{H} (cf. (10)). Since $\mathsf{x}\sim\mathcal{CN}(\mathbf{0},\mathbf{I}_{\widetilde{T}N})$, all moments of x, and, hence, the expectation $\mathbb{E}_{\mathsf{x}}\big[\left|(1/\widetilde{T})\mathbf{B}^{\mathrm{H}}\mathbf{B}+\mathbf{I}_{R\widetilde{T}Q}\right|\big]$, are finite. Therefore, the right-hand side in (34) is a finite constant with respect to ρ. Hence, (33) together with (34) implies

$$h(\mathsf{y}\,|\,\mathsf{x}) \leq R\widetilde{T}Q\log\rho+\mathcal{O}(1)\,. \tag{35}$$

2) Lower Bound on $h(\mathsf{y})$: The dimension counting argument provided in Section VI-A suggests that $R\leq\lceil T(N-1)/(N-\widetilde{T}Q)\rceil$ receive antennas are sufficient to identify all unknown input parameters. By comparing more carefully the number of equations RN and the number of variables $R\widetilde{T}Q+\widetilde{T}N-\widetilde{T}$, we see that we can get rid of

$$\ell\triangleq\max\{0,RN-(R\widetilde{T}Q+\widetilde{T}N-\widetilde{T})\} \tag{36}$$

equations. Since we assumed that $R\leq\lceil\widetilde{T}(N-1)/(N-\widetilde{T}Q)\rceil$ and $N>\widetilde{T}Q$, we have that

$$\begin{aligned}
\ell&=\max\{0,R(N-\widetilde{T}Q)-\widetilde{T}(N-1)\}\\
&\leq\max\left\{0,\left\lceil\frac{\widetilde{T}(N-1)}{N-\widetilde{T}Q}\right\rceil(N-\widetilde{T}Q)-\widetilde{T}(N-1)\right\}\\
&=\max\Bigg\{0,\underbrace{\left(\left\lceil\frac{\widetilde{T}(N-1)}{N-\widetilde{T}Q}\right\rceil-\frac{\widetilde{T}(N-1)}{N-\widetilde{T}Q}\right)}_{\geq 0}\underbrace{(N-\widetilde{T}Q)}_{>0}\Bigg\}\\
&=\underbrace{\left(\left\lceil\frac{\widetilde{T}(N-1)}{N-\widetilde{T}Q}\right\rceil-\frac{\widetilde{T}(N-1)}{N-\widetilde{T}Q}\right)}_{<1}(N-\widetilde{T}Q)\\
&<N-\widetilde{T}Q\,.
\end{aligned} \tag{37}$$

Thus, we can make the number of equations equal to the number of unknown variables by removing at most $N-\widetilde{T}Q-1$ equations. To do so, it is convenient to separate the RN received variables into a "useful" part, which we denote by $[\mathsf{y}]_{\mathcal{I}}$ with[6]

$$\mathcal{I}\triangleq[1:RN-\ell] \tag{38}$$

and a "redundant" part $[\mathsf{y}]_{\mathcal{J}}$ with $\mathcal{J}\triangleq[1:RN]\setminus\mathcal{I}=[RN-\ell+1:RN]$. Note that in the case $\ell=0$, i.e., when the number of equations does not exceed the number of unknown variables, we have $[\mathsf{y}]_{\mathcal{I}}=\mathsf{y}$ and the redundant part $[\mathsf{y}]_{\mathcal{J}}$ is empty.

We can now lower-bound $h(\mathsf{y})$ as follows:

$$\begin{aligned}
h(\mathsf{y})&=h([\mathsf{y}]_{\mathcal{I}},[\mathsf{y}]_{\mathcal{J}})\\
&\overset{(a)}{=}h([\mathsf{y}]_{\mathcal{I}})+h\big([\mathsf{y}]_{\mathcal{J}}\,\big|\,[\mathsf{y}]_{\mathcal{I}}\big)\\
&\overset{(b)}{\geq}h\left(\sqrt{\frac{\rho}{\widetilde{T}}}[\bar{\mathsf{y}}]_{\mathcal{I}}+[\mathsf{w}]_{\mathcal{I}}\,\middle|\,[\mathsf{w}]_{\mathcal{I}}\right)+h\big([\mathsf{y}]_{\mathcal{J}}\,\big|\,\mathsf{s},\mathsf{x},[\mathsf{y}]_{\mathcal{I}}\big)\\
&\overset{(c)}{=}h\left(\sqrt{\frac{\rho}{\widetilde{T}}}[\bar{\mathsf{y}}]_{\mathcal{I}}\right)+\mathcal{O}(1)\\
&\overset{(d)}{=}\log\left(\sqrt{\frac{\rho}{\widetilde{T}}}\right)^{2(RN-\ell)}+h([\bar{\mathsf{y}}]_{\mathcal{I}})+\mathcal{O}(1)\\
&=(RN-\ell)\log\rho+h([\bar{\mathsf{y}}]_{\mathcal{I}})+\mathcal{O}(1)\,.
\end{aligned} \tag{39}$$

Here, (a) follows from the chain rule for differential entropy, in (b) we used (9) and the fact that conditioning reduces differential entropy, (c) holds since $h\big([\mathsf{y}]_{\mathcal{J}}\,\big|\,\mathsf{s},\mathsf{x},[\mathsf{y}]_{\mathcal{I}}\big)=h\big([\mathsf{w}]_{\mathcal{J}}\big)$ is a finite constant, and (d) holds by the transformation property of differential entropy [19, eq. (8.71)]. Using (35) and (39) in (32), we obtain

$$\begin{aligned}
I(\mathsf{x};\mathsf{y})&\geq(RN-\ell-R\widetilde{T}Q)\log\rho+h([\bar{\mathsf{y}}]_{\mathcal{I}})+\mathcal{O}(1)\\
&\overset{(36)}{=}\Big(RN-\max\{0,RN-(R\widetilde{T}Q+\widetilde{T}N-\widetilde{T})\}\\
&\qquad\qquad-R\widetilde{T}Q\Big)\log\rho+h([\bar{\mathsf{y}}]_{\mathcal{I}})+\mathcal{O}(1)\\
&=\min\{RN-R\widetilde{T}Q,\widetilde{T}N-\widetilde{T}\}\log\rho\\
&\qquad\qquad+h([\bar{\mathsf{y}}]_{\mathcal{I}})+\mathcal{O}(1)\,.
\end{aligned} \tag{40}$$

The degrees of freedom lower bound (25) follows by inserting (40) into (31):

$$\begin{aligned}
C(\rho)&\geq\frac{1}{N}I(\mathsf{x};\mathsf{y})\\
&\geq\min\left\{R\left(1-\frac{\widetilde{T}Q}{N}\right),\widetilde{T}\left(1-\frac{1}{N}\right)\right\}\log\rho\\
&\qquad+\frac{1}{N}h([\bar{\mathsf{y}}]_{\mathcal{I}})+\mathcal{O}(1)
\end{aligned}$$

whence, by (14) and because $h([\bar{\mathsf{y}}]_{\mathcal{I}})$ does not depend on ρ,

$$\begin{aligned}
\chi_{\text{gen}}&\geq\lim_{\rho\to\infty}\frac{\min\left\{R\left(1-\frac{\widetilde{T}Q}{N}\right),\widetilde{T}\left(1-\frac{1}{N}\right)\right\}\log\rho+\mathcal{O}(1)}{\log\rho}\\
&=\min\left\{R\left(1-\frac{\widetilde{T}Q}{N}\right),\widetilde{T}\left(1-\frac{1}{N}\right)\right\}
\end{aligned}$$

provided that $h([\bar{\mathsf{y}}]_{\mathcal{I}})>-\infty$. To conclude the proof, we will next show that $h([\bar{\mathsf{y}}]_{\mathcal{I}})>-\infty$ for a generic coloring matrix $\boldsymbol{Z}$. This is the most technical part of the proof.

C. Proof that $h([\bar{\mathsf{y}}]_{\mathcal{I}})>-\infty$

As $[\bar{\mathsf{y}}]_{\mathcal{I}}$ is a function of s and x (see (9) and (10)), the idea behind our proof is to relate $h([\bar{\mathsf{y}}]_{\mathcal{I}})$, which we are not able to calculate directly, to $h(\mathsf{s},\mathsf{x})$, which can be calculated trivially. The underlying intuition is that the image of a random variable of finite differential entropy, such as (s,x), under a "well-behaved" mapping, such as $(\mathsf{s},\mathsf{x})\mapsto[\bar{\mathsf{y}}]_{\mathcal{I}}$, cannot have an infinite differential entropy. At the heart of the proof is the bounding of differential entropy under finite-to-one mappings, to be established in Lemma 8 below.

We first need to characterize the mapping between (s,x) and $[\bar{\mathsf{y}}]_{\mathcal{I}}$. To equalize the dimensions—note that $[\bar{\mathsf{y}}]_{\mathcal{I}}\in\mathbb{C}^{|\mathcal{I}|}$ and $(\mathsf{s}^{\mathrm{T}}\,\mathsf{x}^{\mathrm{T}})^{\mathrm{T}}\in\mathbb{C}^{R\widetilde{T}Q+\widetilde{T}N}$—we condition on $R\widetilde{T}Q+\widetilde{T}N-|\mathcal{I}|$ entries of x, which we denote by $[\mathsf{x}]_{\mathcal{P}}$ (hence, $|\mathcal{P}|=R\widetilde{T}Q+\widetilde{T}N-|\mathcal{I}|$). This results in

$$h([\bar{\mathsf{y}}]_{\mathcal{I}})\geq h([\bar{\mathsf{y}}]_{\mathcal{I}}\,|\,[\mathsf{x}]_{\mathcal{P}})\,. \tag{41}$$

We shall denote by $[\mathsf{x}]_{\mathcal{D}}$ the remaining entries of x, i.e., $\mathcal{D}\triangleq[1:\widetilde{T}N]\setminus\mathcal{P}$. Note that $|\mathcal{D}|+|\mathcal{P}|=\widetilde{T}N$ and thus

$$|\mathcal{I}|=R\widetilde{T}Q+|\mathcal{D}|\,. \tag{42}$$

[6]It is convenient to choose $\mathcal{I}$ this way, but other choices may also be possible.

One can think of $[\mathbf{x}]_{\mathcal{P}}$ as pilot symbols and of $[\mathbf{x}]_{\mathcal{D}}$ as data symbols. The set $\mathcal{P}$ will be defined in Appendix A.B. At this point, we are only concerned with its size, which is equal to

$$|\mathcal{P}| = R\widetilde{T}Q + \widetilde{T}N - |\mathcal{I}|\,. \tag{43}$$

Because of (41), it suffices to show that

$$h([\bar{\mathbf{y}}]_{\mathcal{I}}\,|\,[\mathbf{x}]_{\mathcal{P}}) > -\infty\,.$$

This will be done by relating $h([\bar{\mathbf{y}}]_{\mathcal{I}}\,|\,[\mathbf{x}]_{\mathcal{P}})$ to $h(\mathbf{s}, [\mathbf{x}]_{\mathcal{D}})$. Before doing so, we have to understand the connection between the variables $[\bar{\mathbf{y}}]_{\mathcal{I}}$ and $(\mathbf{s}, [\mathbf{x}]_{\mathcal{D}})$. This leads us to the following program:

(i) Define the polynomial mapping $\phi_{[\boldsymbol{x}]_{\mathcal{P}}}$ relating $(\mathbf{s}, [\mathbf{x}]_{\mathcal{D}})$ and $[\bar{\mathbf{y}}]_{\mathcal{I}}$.

(ii) Prove that $\phi_{[\boldsymbol{x}]_{\mathcal{P}}}$ satisfies the following two properties:
   a) Its Jacobian matrix is nonsingular almost everywhere (a.e.) for almost all (a.a.) $[\boldsymbol{x}]_{\mathcal{P}}$.
   b) It is finite-to-one[7] a.e. for a.a. $[\boldsymbol{x}]_{\mathcal{P}}$.

(iii) Apply a novel result on the change in differential entropy that occurs when a random variable undergoes a finite-to-one mapping to relate $h([\bar{\mathbf{y}}]_{\mathcal{I}}\,|\,[\mathbf{x}]_{\mathcal{P}})$ to $h(\mathbf{s}, [\mathbf{x}]_{\mathcal{D}})$.

(iv) Bound the terms resulting from this change in differential entropy.

Step (i):

We consider the $[\boldsymbol{x}]_{\mathcal{P}}$-parametrized mapping

$$\phi_{[\boldsymbol{x}]_{\mathcal{P}}} \colon \mathbb{C}^{R\widetilde{T}Q+|\mathcal{D}|} \to \mathbb{C}^{|\mathcal{I}|};\ (\boldsymbol{s}, [\boldsymbol{x}]_{\mathcal{D}}) \mapsto\ [\bar{\boldsymbol{y}}]_{\mathcal{I}} \tag{44}$$

in which $\bar{\boldsymbol{y}}$ is defined in (9) and (10), i.e.,

$$\bar{\boldsymbol{y}} = \boldsymbol{B}\boldsymbol{s}, \quad \text{with } \boldsymbol{B} = \begin{pmatrix} \boldsymbol{B}_1 & & \\ & \ddots & \\ & & \boldsymbol{B}_R \end{pmatrix} \tag{45}$$

where

$$\boldsymbol{B}_r = (\boldsymbol{X}_1 \boldsymbol{Z}_{r,1} \cdots \boldsymbol{X}_{\widetilde{T}} \boldsymbol{Z}_{r,\widetilde{T}}), \quad \text{with } \boldsymbol{X}_t = \operatorname{diag}(\boldsymbol{x}_t). \tag{46}$$

We see from (45) and (46) that the components of the vector-valued mapping $\phi_{[\boldsymbol{x}]_{\mathcal{P}}}$ are multivariate polynomials of degree 2 in the entries of $\boldsymbol{s}$ and $[\boldsymbol{x}]_{\mathcal{P}}$. The Jacobian matrix $\boldsymbol{J}_{\phi_{[\boldsymbol{x}]_{\mathcal{P}}}}$ of $\phi_{[\boldsymbol{x}]_{\mathcal{P}}}$ is equal to

$$\boldsymbol{J}_{\phi_{[\boldsymbol{x}]_{\mathcal{P}}}}(\boldsymbol{s}, [\boldsymbol{x}]_{\mathcal{D}}) \;=\; \left[\left(\boldsymbol{B}\ [\boldsymbol{A}]^{\mathcal{D}}\right)\right]_{\mathcal{I}} \in \mathbb{C}^{|\mathcal{I}|\times|\mathcal{I}|},$$
$$\text{with } \boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{1,1} & \cdots & \boldsymbol{A}_{1,\widetilde{T}} \\ \vdots & & \vdots \\ \boldsymbol{A}_{R,1} & \cdots & \boldsymbol{A}_{R,\widetilde{T}} \end{pmatrix} \in \mathbb{C}^{RN\times\widetilde{T}N} \tag{47}$$

where

$$\boldsymbol{A}_{r,t} \triangleq \operatorname{diag}(\boldsymbol{a}_{r,t}), \quad t \in [1\!:\!\widetilde{T}],\ r \in [1\!:\!R]\,,$$
$$\text{with } \boldsymbol{a}_{r,t} \triangleq \boldsymbol{Z}_{r,t}\boldsymbol{s}_{r,t} \tag{48}$$

and where in (47) we used that $|\mathcal{I}| = R\widetilde{T}Q + |\mathcal{D}|$ (see (42)). Note that we did not take derivatives with respect to $[\boldsymbol{x}]_{\mathcal{P}}$, since the entries of $[\boldsymbol{x}]_{\mathcal{P}}$ are treated as fixed parameters.

[7] A mapping is called finite-to-one if every element in the codomain has a preimage of finite cardinality.

Step (ii-a):

We have 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 coloring matrix $\boldsymbol{Z}$. 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 $\boldsymbol{Z}$, $\boldsymbol{s}$, $[\boldsymbol{x}]_{\mathcal{D}}$, and $[\boldsymbol{x}]_{\mathcal{P}}$). We will show that $p(\boldsymbol{Z}, \boldsymbol{s}, \boldsymbol{x})$ does not vanish at a specific point $(\tilde{\boldsymbol{Z}}, \tilde{\boldsymbol{s}}, \tilde{\boldsymbol{x}})$, i.e., $p(\tilde{\boldsymbol{Z}}, \tilde{\boldsymbol{s}}, \tilde{\boldsymbol{x}}) \neq 0$. This implies that $p(\boldsymbol{Z}, \tilde{\boldsymbol{s}}, \tilde{\boldsymbol{x}})$ (as a function of $\boldsymbol{Z}$, for fixed $\tilde{\boldsymbol{s}}$ and $\tilde{\boldsymbol{x}}$) is not identically zero. Since a polynomial vanishes either identically or on a set of measure zero[24, Cor. 10], we conclude that $p(\boldsymbol{Z}, \tilde{\boldsymbol{s}}, \tilde{\boldsymbol{x}}) \neq 0$ for $\boldsymbol{Z} \in \mathcal{Z}$, where $\mathcal{Z}$ is a set with a complement of measure zero. Using the same argument, we find that, for a fixed $\boldsymbol{Z} \in \mathcal{Z}$, the function $p(\boldsymbol{Z}, \boldsymbol{s}, \boldsymbol{x})$ does not vanish a.e. (as a function of $(\boldsymbol{s}, \boldsymbol{x})$). Hence, $|\boldsymbol{J}_{\phi_{[\boldsymbol{x}]_{\mathcal{P}}}}(\boldsymbol{s}, [\boldsymbol{x}]_{\mathcal{D}})| \neq 0$ for a.a. $(\boldsymbol{s}, [\boldsymbol{x}]_{\mathcal{D}}, [\boldsymbol{x}]_{\mathcal{P}})$ and all $\boldsymbol{Z} \in \mathcal{Z}$. In other words, for a generic coloring matrix $\boldsymbol{Z}$, the matrix $\boldsymbol{J}_{\phi_{[\boldsymbol{x}]_{\mathcal{P}}}}$ is nonsingular a.e. for a.a. $[\boldsymbol{x}]_{\mathcal{P}}$.

It remains to find the point $(\tilde{\boldsymbol{Z}}, \tilde{\boldsymbol{s}}, \tilde{\boldsymbol{x}})$, i.e., a specific point $(\tilde{\boldsymbol{Z}}, \tilde{\boldsymbol{s}}, \tilde{\boldsymbol{x}})$ such that $p(\tilde{\boldsymbol{Z}}, \tilde{\boldsymbol{s}}, \tilde{\boldsymbol{x}}) \neq 0$. This, in turn, requires to find a specific set $\mathcal{P}$. This is done in the proof of the following lemma.

Lemma 6: Let $R \geq \widetilde{T}$, $N > \widetilde{T}Q$, and $R \leq \lceil \widetilde{T}(N-1)/(N-\widetilde{T}Q) \rceil$. Then there exists a triple $(\boldsymbol{Z}, \boldsymbol{s}, \boldsymbol{x})$ and a choice of $\mathcal{P}$ for which the determinant of the Jacobian matrix $\boldsymbol{J}_{\phi_{[\boldsymbol{x}]_{\mathcal{P}}}}$ in (47) is nonzero.

Proof: See Appendix A. ∎

Step (ii-b):

We will invoke Bézout's theorem [16, Prop. B.2.7] to show that the mapping $\phi_{[\boldsymbol{x}]_{\mathcal{P}}}$ is finite-to-one a.e. for a.a. $[\boldsymbol{x}]_{\mathcal{P}}$. In what follows, note that for a given $[\bar{\boldsymbol{y}}]_{\mathcal{I}}$ in the codomain of $\phi_{[\boldsymbol{x}]_{\mathcal{P}}}$, the quantity $\phi^{-1}_{[\boldsymbol{x}]_{\mathcal{P}}}([\bar{\boldsymbol{y}}]_{\mathcal{I}})$ is the preimage $\phi^{-1}_{[\boldsymbol{x}]_{\mathcal{P}}}([\bar{\boldsymbol{y}}]_{\mathcal{I}}) = \{(\boldsymbol{s}, [\boldsymbol{x}]_{\mathcal{D}}) : \phi_{[\boldsymbol{x}]_{\mathcal{P}}}(\boldsymbol{s}, [\boldsymbol{x}]_{\mathcal{D}}) = [\bar{\boldsymbol{y}}]_{\mathcal{I}}\}$ and not the function value of the inverse function (which does not even exist in most cases). Furthermore, for a given $[\boldsymbol{x}]_{\mathcal{P}}$, we denote by $\widetilde{\mathcal{M}} \subseteq \mathbb{C}^{|\mathcal{I}|}$ the set of all $(\boldsymbol{s}, [\boldsymbol{x}]_{\mathcal{D}})$ for which $\boldsymbol{J}_{\phi_{[\boldsymbol{x}]_{\mathcal{P}}}}(\boldsymbol{s}, [\boldsymbol{x}]_{\mathcal{D}})$ is nonsingular, i.e.,

$$\widetilde{\mathcal{M}} \triangleq \left\{(\boldsymbol{s}, [\boldsymbol{x}]_{\mathcal{D}}) \in \mathbb{C}^{R\widetilde{T}Q+|\mathcal{D}|} \colon |\boldsymbol{J}_{\phi_{[\boldsymbol{x}]_{\mathcal{P}}}}(\boldsymbol{s}, [\boldsymbol{x}]_{\mathcal{D}})| \neq 0\right\}.$$

Lemma 7: For a given $[\boldsymbol{x}]_{\mathcal{P}}$, let $\widetilde{\mathcal{M}}$ be defined as above. Then for all $[\bar{\boldsymbol{y}}]_{\mathcal{I}} \in \phi_{[\boldsymbol{x}]_{\mathcal{P}}}(\widetilde{\mathcal{M}})$,

$$\left|\phi^{-1}_{[\boldsymbol{x}]_{\mathcal{P}}}([\bar{\boldsymbol{y}}]_{\mathcal{I}}) \cap \widetilde{\mathcal{M}}\right| \;\leq\; \widetilde{m} \;\triangleq\; 2^{R\widetilde{T}Q+|\mathcal{D}|}\,.$$

Proof: Let $[\bar{\boldsymbol{y}}]_{\mathcal{I}} \in \phi_{[\boldsymbol{x}]_{\mathcal{P}}}(\widetilde{\mathcal{M}})$. The set $\phi^{-1}_{[\boldsymbol{x}]_{\mathcal{P}}}([\bar{\boldsymbol{y}}]_{\mathcal{I}})$ contains all points $(\boldsymbol{s}, [\boldsymbol{x}]_{\mathcal{D}})$ such that $\phi_{[\boldsymbol{x}]_{\mathcal{P}}}(\boldsymbol{s}, [\boldsymbol{x}]_{\mathcal{D}}) = [\bar{\boldsymbol{y}}]_{\mathcal{I}}$. Thus, these points are the zeros of the vector-valued mapping

$$(\boldsymbol{s}, [\boldsymbol{x}]_{\mathcal{D}}) \mapsto \phi_{[\boldsymbol{x}]_{\mathcal{P}}}(\boldsymbol{s}, [\boldsymbol{x}]_{\mathcal{D}}) - [\bar{\boldsymbol{y}}]_{\mathcal{I}}\,. \tag{49}$$

It follows from (44)–(46) that each component of the vector-valued mapping (49) is a polynomial of degree 2. Hence, the zeros of the mapping (49) are the common zeros of $|\mathcal{I}| = R\widetilde{T}Q + |\mathcal{D}|$ polynomials of degree 2. By a weak version of Bézout's theorem [16, Prop. B.2.7], the number of isolated zeros (i.e., with no other zeros in some neighborhood) cannot exceed $\widetilde{m} = 2^{R\widetilde{T}Q+|\mathcal{D}|}$. Since $\boldsymbol{J}_{\phi_{[\boldsymbol{x}]_{\mathcal{P}}}}$ is nonsingular on $\widetilde{\mathcal{M}}$,

the function $\phi_{[\boldsymbol{x}]_{\mathcal{P}}}$ restricted to $\widetilde{\mathcal{M}}$ is locally one-to-one [25, Th. 9.24] and, hence, each zero of $\phi_{[\boldsymbol{x}]_{\mathcal{P}}}-[\bar{\boldsymbol{y}}]_{\mathcal{I}}$ on $\widetilde{\mathcal{M}}$ has to be an isolated zero. Therefore, the number of points $(\boldsymbol{s},[\boldsymbol{x}]_{\mathcal{D}})\in\widetilde{\mathcal{M}}$ such that $\phi_{[\boldsymbol{x}]_{\mathcal{P}}}(\boldsymbol{s},[\boldsymbol{x}]_{\mathcal{D}})=[\bar{\boldsymbol{y}}]_{\mathcal{I}}$ cannot exceed $\tilde{m}$. ∎

By Lemma 7, the function $\phi_{[\boldsymbol{x}]_{\mathcal{P}}}$ for a given $[\boldsymbol{x}]_{\mathcal{P}}$ is finite-to-one on the set $\widetilde{\mathcal{M}}$. Because by Step (ii-a) the matrix $\boldsymbol{J}_{\phi_{[\boldsymbol{x}]_{\mathcal{P}}}}(\boldsymbol{s},[\boldsymbol{x}]_{\mathcal{D}})$ is nonsingular a.e. for a.a. $[\boldsymbol{x}]_{\mathcal{P}}$, and because $\widetilde{\mathcal{M}}\subseteq\mathbb{C}^{|\mathcal{I}|}$ is the set of all $(\boldsymbol{s},[\boldsymbol{x}]_{\mathcal{D}})$ for which $\boldsymbol{J}_{\phi_{[\boldsymbol{x}]_{\mathcal{P}}}}(\boldsymbol{s},[\boldsymbol{x}]_{\mathcal{D}})$ is nonsingular, we conclude that $\phi_{[\boldsymbol{x}]_{\mathcal{P}}}$ is finite-to-one a.e. for a.a. $[\boldsymbol{x}]_{\mathcal{P}}$.

Step (iii):

We will use the following novel result bounding the change in differential entropy that occurs when a random variable undergoes a finite-to-one mapping.

Lemma 8: Let $\mathbf{u}\in\mathbb{C}^n$ be a random vector with probability density function $f_{\mathbf{u}}$. Consider a continuously differentiable mapping $\kappa\colon\mathbb{C}^n\to\mathbb{C}^n$ with Jacobian matrix $\boldsymbol{J}_\kappa$. Assume that $\boldsymbol{J}_\kappa$ is nonsingular a.e. and let $\mathcal{M}\triangleq\{\boldsymbol{u}\in\mathbb{C}^n\colon|\boldsymbol{J}_\kappa(\boldsymbol{u})|\neq 0\}$ (thus, $\mathbb{C}^n\setminus\mathcal{M}$ has Lebesgue measure zero). Furthermore, let $\mathbf{v}\triangleq\kappa(\mathbf{u})$, and assume that for all $\boldsymbol{v}\in\mathbb{C}^n$, the cardinality of the set $\kappa^{-1}(\boldsymbol{v})\cap\mathcal{M}$ satisfies $|\kappa^{-1}(\boldsymbol{v})\cap\mathcal{M}|\leq m<\infty$, for some $m\in\mathbb{N}$ (i.e., $\kappa\big|_{\mathcal{M}}$ is finite-to-one). Then:

(I) There exist disjoint measurable sets $\{\mathcal{U}_k\}_{k\in[1:m]}$ such that $\kappa\big|_{\mathcal{U}_k}$ is one-to-one for each $k\in[1:m]$ and $\bigcup_{k\in[1:m]}\mathcal{U}_k$ covers almost all of $\mathcal{M}$.

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

$$h(\mathbf{v})\geq h(\mathbf{u})+\int_{\mathbb{C}^n}f_{\mathbf{u}}(\boldsymbol{u})\log(|\boldsymbol{J}_\kappa(\boldsymbol{u})|^2)\,d\boldsymbol{u}-H(\mathsf{k}) \tag{50}$$

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

Proof: See Appendix B. ∎

Since by Step (ii-b) the mapping $\phi_{[\boldsymbol{x}]_{\mathcal{P}}}\big|_{\widetilde{\mathcal{M}}}$ is finite-to-one for a.a. $[\boldsymbol{x}]_{\mathcal{P}}$, we can use Lemma 8 with $\mathbf{u}=(\mathbf{s},[\mathbf{x}]_{\mathcal{D}})$, $\kappa=\phi_{[\boldsymbol{x}]_{\mathcal{P}}}$, $n=R\widetilde{T}Q+|\mathcal{D}|$, $m=\tilde{m}$, and $\mathcal{M}=\widetilde{\mathcal{M}}$ and obtain

$$\begin{aligned}h\big(\phi_{[\boldsymbol{x}]_{\mathcal{P}}}(\mathbf{s},[\mathbf{x}]_{\mathcal{D}})\big)&\geq h(\mathbf{s},[\mathbf{x}]_{\mathcal{D}})+\int_{\mathbb{C}^{R\tilde{T}Q+|\mathcal{D}|}}f_{\mathbf{s},[\mathbf{x}]_{\mathcal{D}}}(\boldsymbol{s},[\boldsymbol{x}]_{\mathcal{D}})\\&\quad\times\log\big(|\boldsymbol{J}_{\phi_{[\boldsymbol{x}]_{\mathcal{P}}}}(\boldsymbol{s},[\boldsymbol{x}]_{\mathcal{D}})|^2\big)\,d(\boldsymbol{s},[\boldsymbol{x}]_{\mathcal{D}})-H(\mathsf{k}_{[\boldsymbol{x}]_{\mathcal{P}}})\end{aligned} \tag{51}$$

where $\mathsf{k}_{[\boldsymbol{x}]_{\mathcal{P}}}$ corresponds to the random variable k from Lemma 8 (since $\kappa=\phi_{[\boldsymbol{x}]_{\mathcal{P}}}$, we have a different k for each $[\boldsymbol{x}]_{\mathcal{P}}$). Because of $[\bar{\mathbf{y}}]_{\mathcal{I}}=\phi_{[\mathbf{x}]_{\mathcal{P}}}(\mathbf{s},[\mathbf{x}]_{\mathcal{D}})$, we have $h\big([\bar{\mathbf{y}}]_{\mathcal{I}}\big|[\mathbf{x}]_{\mathcal{P}}=[\boldsymbol{x}]_{\mathcal{P}}\big)=h\big(\phi_{[\boldsymbol{x}]_{\mathcal{P}}}(\mathbf{s},[\mathbf{x}]_{\mathcal{D}})\big)$. Thus, (51) entails

$$\begin{aligned}&h\big([\bar{\mathbf{y}}]_{\mathcal{I}}\big|[\mathbf{x}]_{\mathcal{P}}\big)\\&=\mathbb{E}_{[\mathbf{x}]_{\mathcal{P}}}\big[h\big(\phi_{[\mathbf{x}]_{\mathcal{P}}}(\mathbf{s},[\mathbf{x}]_{\mathcal{D}})\big)\big]\\&\geq h(\mathbf{s},[\mathbf{x}]_{\mathcal{D}})+\mathbb{E}_{[\mathbf{x}]_{\mathcal{P}}}\bigg[\int_{\mathbb{C}^{R\tilde{T}Q+|\mathcal{D}|}}f_{\mathbf{s},[\mathbf{x}]_{\mathcal{D}}}(\boldsymbol{s},[\boldsymbol{x}]_{\mathcal{D}})\\&\quad\times\log\big(\big|\boldsymbol{J}_{\phi_{[\mathbf{x}]_{\mathcal{P}}}}(\boldsymbol{s},[\boldsymbol{x}]_{\mathcal{D}})\big|^2\big)\,d(\boldsymbol{s},[\boldsymbol{x}]_{\mathcal{D}})-H\big(\mathsf{k}_{[\mathbf{x}]_{\mathcal{P}}}\big)\bigg].\end{aligned} \tag{52}$$

Step (iv):

We show now that the right-hand side of (52) is lower-bounded by a finite constant. The differential entropy $h(\mathbf{s},[\mathbf{x}]_{\mathcal{D}})$ is the differential entropy of a standard multivariate Gaussian random vector and thus a finite constant. The entropy $H\big(\mathsf{k}_{[\boldsymbol{x}]_{\mathcal{P}}}\big)$ for a.a. $[\boldsymbol{x}]_{\mathcal{P}}$ does not exceed $\log(\tilde{m})$, where $\tilde{m}=2^{R\widetilde{T}Q+|\mathcal{D}|}$. Hence, it remains to lower-bound

$$\begin{aligned}&\mathbb{E}_{[\mathbf{x}]_{\mathcal{P}}}\bigg[\int_{\mathbb{C}^{R\tilde{T}Q+|\mathcal{D}|}}f_{\mathbf{s},[\mathbf{x}]_{\mathcal{D}}}(\boldsymbol{s},[\boldsymbol{x}]_{\mathcal{D}})\\&\qquad\times\log\big(|\boldsymbol{J}_{\phi_{[\mathbf{x}]_{\mathcal{P}}}}(\boldsymbol{s},[\boldsymbol{x}]_{\mathcal{D}})|^2\big)\,d(\boldsymbol{s},[\boldsymbol{x}]_{\mathcal{D}})\bigg]\\&=\int_{\mathbb{C}^{|\mathcal{P}|}}\int_{\mathbb{C}^{R\tilde{T}Q+|\mathcal{D}|}}f_{[\mathbf{x}]_{\mathcal{P}}}([\boldsymbol{x}]_{\mathcal{P}})\,f_{\mathbf{s},[\mathbf{x}]_{\mathcal{D}}}(\boldsymbol{s},[\boldsymbol{x}]_{\mathcal{D}})\\&\qquad\times\log\big(|\boldsymbol{J}_{\phi_{[\boldsymbol{x}]_{\mathcal{P}}}}(\boldsymbol{s},[\boldsymbol{x}]_{\mathcal{D}})|^2\big)\,d(\boldsymbol{s},[\boldsymbol{x}]_{\mathcal{D}})\,d[\boldsymbol{x}]_{\mathcal{P}}\\&\overset{(a)}{=}\int_{\mathbb{C}^{R\tilde{T}Q+\tilde{T}N}}f_{\mathbf{s},\mathbf{x}}(\boldsymbol{s},\boldsymbol{x})\log\big(|\boldsymbol{J}_{\phi_{[\boldsymbol{x}]_{\mathcal{P}}}}(\boldsymbol{s},[\boldsymbol{x}]_{\mathcal{D}})|^2\big)\,d(\boldsymbol{s},\boldsymbol{x})\end{aligned} \tag{53}$$

where (a) holds because $(\mathbf{s},[\mathbf{x}]_{\mathcal{D}})$ and $[\mathbf{x}]_{\mathcal{P}}$ are independent. A similar problem was recently solved in [12] using Hironaka's theorem on the resolution of singularities. Here, we take a much simpler approach, which relies on the fact that $\det\big(\boldsymbol{J}_{\phi_{[\boldsymbol{x}]_{\mathcal{P}}}}\big)$ in (53) is an analytic function [26, Ch. 10] that does not vanish identically, and on a property of subharmonic functions[8] (see [17, Th. 2.6.2.1]).

Lemma 9: Let f be an analytic function on $\mathbb{C}^n$ that is not identically zero. Then

$$I_1\triangleq\int_{\mathbb{C}^n}\exp(-\|\boldsymbol{\xi}\|^2)\log(|f(\boldsymbol{\xi})|)\,d\boldsymbol{\xi}>-\infty\,. \tag{54}$$

Proof: See Appendix C. ∎

The function $f_{\mathbf{s},\mathbf{x}}$ is the probability density function of a standard multivariate Gaussian random vector. Furthermore, since the function $\det(\boldsymbol{J}_{\phi_{[\boldsymbol{x}]_{\mathcal{P}}}}(\boldsymbol{s},[\boldsymbol{x}]_{\mathcal{D}}))$ is a complex polynomial that is nonzero a.e. (see Step (ii-a)), it is an analytic function that is not identically zero. Hence, by Lemma 9, the integral in (53) is finite. Thus, with (52), we obtain $h\big([\bar{\mathbf{y}}]_{\mathcal{I}}\big|[\mathbf{x}]_{\mathcal{P}}\big)>-\infty$ and, because of (41), that $h([\bar{\mathbf{y}}]_{\mathcal{I}})>-\infty$. This concludes the proof.

VII. Conclusion

We characterized the number of degrees of freedom for generic block-fading MIMO channels in the noncoherent setting. Although the generic block-fading model seems to be just a minor variation of the classically used constant block-fading model, our result shows that the assumption of generic correlation may strongly affect the number of degrees of freedom. In fact, we showed that the (potentially small) perturbation in the channel model that results from making the coloring matrix $\boldsymbol{Z}$ generic may yield a significant increase in the number of degrees of freedom. This suggests once more (see also [20], [27]) that care must be exercised in using this asymptotic quantity as a performance measure.

The highest gain in terms of the number of degrees of freedom is obtained for a sufficiently large number of receive

[8] See [17, Ch. 2.6] for a definition of subharmonic functions.

$$\begin{pmatrix}
[\boldsymbol{Z}_{1,1}]_1 & [\boldsymbol{Z}_{1,2}]_1 & & & & & 0 & & & & [\boldsymbol{Z}_{1,2}]_1 s_{1,2} & & & \\
[\boldsymbol{Z}_{1,1}]_2 & [\boldsymbol{Z}_{1,2}]_2 & & & & & [\boldsymbol{Z}_{1,1}]_2 s_{1,1} & & & & & 0 & & \\
[\boldsymbol{Z}_{1,1}]_3 & [\boldsymbol{Z}_{1,2}]_3 & & & & & & [\boldsymbol{Z}_{1,1}]_3 s_{1,1} & & & & & [\boldsymbol{Z}_{1,2}]_3 s_{1,2} & \\
[\boldsymbol{Z}_{1,1}]_4 & [\boldsymbol{Z}_{1,2}]_4 & & & & & & & [\boldsymbol{Z}_{1,1}]_4 s_{1,1} & & & & & [\boldsymbol{Z}_{1,2}]_4 s_{1,2} \\
 & & [\boldsymbol{Z}_{2,1}]_1 & [\boldsymbol{Z}_{2,2}]_1 & & & 0 & & & & [\boldsymbol{Z}_{2,2}]_1 s_{2,2} & & & \\
 & & [\boldsymbol{Z}_{2,1}]_2 & [\boldsymbol{Z}_{2,2}]_2 & & & [\boldsymbol{Z}_{2,1}]_2 s_{2,1} & & & & & 0 & & \\
 & & [\boldsymbol{Z}_{2,1}]_3 & [\boldsymbol{Z}_{2,2}]_3 & & & & [\boldsymbol{Z}_{2,1}]_3 s_{2,1} & & & & & [\boldsymbol{Z}_{2,2}]_3 s_{2,2} & \\
 & & [\boldsymbol{Z}_{2,1}]_4 & [\boldsymbol{Z}_{2,2}]_4 & & & & & [\boldsymbol{Z}_{2,1}]_4 s_{2,1} & & & & & [\boldsymbol{Z}_{2,2}]_4 s_{2,2} \\
 & & & & [\boldsymbol{Z}_{3,1}]_1 & [\boldsymbol{Z}_{3,2}]_1 & 0 & & & & [\boldsymbol{Z}_{3,2}]_1 s_{3,2} & & & \\
 & & & & [\boldsymbol{Z}_{3,1}]_2 & [\boldsymbol{Z}_{3,2}]_2 & [\boldsymbol{Z}_{3,1}]_2 s_{3,1} & & & & & 0 & & \\
 & & & & [\boldsymbol{Z}_{3,1}]_3 & [\boldsymbol{Z}_{3,2}]_3 & & [\boldsymbol{Z}_{3,1}]_3 s_{3,1} & & & & & [\boldsymbol{Z}_{3,2}]_3 s_{3,2} & \\
 & & & & [\boldsymbol{Z}_{3,1}]_4 & [\boldsymbol{Z}_{3,2}]_4 & & & [\boldsymbol{Z}_{3,1}]_4 s_{3,1} & & & & & [\boldsymbol{Z}_{3,2}]_4 s_{3,2}
\end{pmatrix} \tag{55}$$

antennas. In this case, the number of degrees of freedom is equal to T times the number of degrees of freedom in the SIMO case, as long as the number T of transmit antennas satisfies $T < N/Q$. This may be of interest for the uplink of massive-MIMO systems [28].

From a practical point of view, the generic block-fading model is of particular interest for CP-OFDM systems. These systems cannot be described appropriately by the constant block-fading model, which corresponds to an impulse response of each (t, r) channel that consists of a single tap. By contrast, the generic block-fading model allows for impulse responses with multiple taps.

For CP-OFDM systems with colocated antennas, it may appear questionable to assume that all coloring matrices $\boldsymbol{Z}_{r,t}$ are different—an assumption that is needed for our result to hold (although MIMO channel matrices with nonidentical distributions arise, e.g., when pattern diversity is used [29]). The case where all matrices $\boldsymbol{Z}_{r,t}$ are *exactly* equal is still an open problem. However, it should be noted that any nonzero perturbation of the model with exactly equal $\boldsymbol{Z}_{r,t}$—be it arbitrarily small—yields the generic model considered in this paper. One may then argue that the assumption of exactly equal $\boldsymbol{Z}_{r,t}$ is an idealization that may be convenient in theoretical analyses but will not be satisfied in practical systems. An important conclusion to be drawn from our analysis is the fact that, as far as the number of degrees of freedom is concerned, the model with exactly equal $\boldsymbol{Z}_{r,t}$ is highly nonrobust, since arbitrarily small perturbations yield a potentially large change in the number of degrees of freedom.

The proof of Proposition 3 in Section VI does not provide a characterization of the class of coloring matrices $\boldsymbol{Z}$ for which Theorem 4 does not hold. However, the only part of the proof where a generic $\boldsymbol{Z}$ is needed is in the statement that $\left|\boldsymbol{J}_{\phi_{[\boldsymbol{x}]_{\mathcal{P}}}}(\boldsymbol{s}, [\boldsymbol{x}]_{\mathcal{D}})\right| \neq 0$ a.e. (see Step (ii-a) in Section VI-C). If a specific $\boldsymbol{Z}$ is given, one can search for two vectors $\boldsymbol{s}$ and $\boldsymbol{x}$ and a set $\mathcal{P}$ for which $\left|\boldsymbol{J}_{\phi_{[\boldsymbol{x}]_{\mathcal{P}}}}(\boldsymbol{s}, [\boldsymbol{x}]_{\mathcal{D}})\right| \neq 0$. If the search is successful, then Theorem 4 holds for this $\boldsymbol{Z}$. Note that the converse is not necessarily true: if $\left|\boldsymbol{J}_{\phi_{[\boldsymbol{x}]_{\mathcal{P}}}}(\boldsymbol{s}, [\boldsymbol{x}]_{\mathcal{D}})\right|$ vanishes for all choices of $\boldsymbol{s}$, $\boldsymbol{x}$, and $\mathcal{P}$, one cannot conclude that Theorem 4 does not hold.

An open problem is a characterization of the capacity of generic block-fading MIMO channels beyond the number of degrees of freedom. Such a characterization would help understand whether the sensitivity of the number of degrees of freedom discussed above is an indication of a similar sensitivity of the capacity that occurs already at moderate SNR, or merely an asymptotic peculiarity. Furthermore, a capacity characterization that is nonasymptotic in the SNR can be analyzed for asymptotic block length, which would enable a capacity analysis of, e.g., stationary channel models.

Appendix A
Proof of Lemma 6

Since the proof of Lemma 6 is quite technical, we shall first (in Section A.A) illustrate its key steps by focusing on the special case $\widetilde{T} = 2, R = 3, N = 4$, and $Q = 1$. The proof for arbitrary $\widetilde{T}, R, N$, and Q will be provided in Section A.B.

A. *Special Case* $\widetilde{T} = 2, R = 3, N = 4$, $Q = 1$

By (36), we have $\ell = 0$ and, thus, $\mathcal{I} = [1\!:\!12]$. Furthermore, by (43), we have $|\mathcal{P}| = 2$. We choose $\mathcal{P} = \{1, 6\}$. Hence, recalling that $\boldsymbol{x} = (\boldsymbol{x}_1^{\mathrm{T}}\ \boldsymbol{x}_2^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{C}^8$, we have

$$[\boldsymbol{x}]_{\mathcal{P}} = ([\boldsymbol{x}_1]_1\ [\boldsymbol{x}_2]_2)^{\mathrm{T}}$$

and

$$[\boldsymbol{x}]_{\mathcal{D}} = ([\boldsymbol{x}_1]_2\ [\boldsymbol{x}_1]_3\ [\boldsymbol{x}_1]_4\ [\boldsymbol{x}_2]_1\ [\boldsymbol{x}_2]_3\ [\boldsymbol{x}_2]_4)^{\mathrm{T}}.$$

We also choose $\boldsymbol{x}$ as the all-one vector. For these choices, the Jacobian $\boldsymbol{J}_{\phi_{[\boldsymbol{x}]_{\mathcal{P}}}}$ in (47) is equal to (55) at the top of this page. We have to find $\boldsymbol{Z}$ and $\boldsymbol{s}$ such that the determinant of this matrix is nonzero. Setting $[\boldsymbol{Z}_{3,2}]_1 = [\boldsymbol{Z}_{3,1}]_2 = [\boldsymbol{Z}_{3,2}]_3 = [\boldsymbol{Z}_{3,1}]_4 = 0$, the entries highlighted in gray in (55) become zero. Furthermore, choosing nonzero $[\boldsymbol{Z}_{3,1}]_1$, $[\boldsymbol{Z}_{3,1}]_3$, $[\boldsymbol{Z}_{3,2}]_2$, $[\boldsymbol{Z}_{3,2}]_4$, $s_{3,1}$, and $s_{3,2}$ and operating a Laplace expansion on the last four rows in (55), it is seen that the determinant of the matrix in (55) is nonzero if the determinant of the matrix in (56) at the top of the next page is nonzero. This is the Jacobian matrix corresponding to the case $\widetilde{T} = 2, R = 2, N = 4$, and $Q = 1$. In other words, by performing the matrix manipulations just described, we reduced the case $R = 3$ to the case $R = 2$. A similar idea will be used in the proof for the general case provided in Section A.B, where we will reduce R inductively until $R = \widetilde{T}$. Setting $s_{1,2} = s_{2,1} = 0$, the entries highlighted in gray in (56) become zero. By choosing nonzero $[\boldsymbol{Z}_{1,1}]_2$, $[\boldsymbol{Z}_{1,1}]_4$, $[\boldsymbol{Z}_{2,2}]_1$, $[\boldsymbol{Z}_{2,2}]_3$, $s_{1,1}$, and $s_{2,2}$ and operating a Laplace expansion on the last four columns, it is seen that

$$\begin{pmatrix} [\boldsymbol{Z}_{1,1}]_1 & [\boldsymbol{Z}_{1,2}]_1 & 0 & [\boldsymbol{Z}_{1,2}]_1 s_{1,2} & \\ [\boldsymbol{Z}_{1,1}]_2 & [\boldsymbol{Z}_{1,2}]_2 & [\boldsymbol{Z}_{1,1}]_2 s_{1,1} & & 0 \\ [\boldsymbol{Z}_{1,1}]_3 & [\boldsymbol{Z}_{1,2}]_3 & & 0 & [\boldsymbol{Z}_{1,2}]_3 s_{1,2} \\ [\boldsymbol{Z}_{1,1}]_4 & [\boldsymbol{Z}_{1,2}]_4 & & [\boldsymbol{Z}_{1,1}]_4 s_{1,1} & 0 \\ & [\boldsymbol{Z}_{2,1}]_1 & [\boldsymbol{Z}_{2,2}]_1 & 0 & [\boldsymbol{Z}_{2,2}]_1 s_{2,2} \\ & [\boldsymbol{Z}_{2,1}]_2 & [\boldsymbol{Z}_{2,2}]_2 & [\boldsymbol{Z}_{2,1}]_2 s_{2,1} & 0 \\ & [\boldsymbol{Z}_{2,1}]_3 & [\boldsymbol{Z}_{2,2}]_3 & 0 & [\boldsymbol{Z}_{2,2}]_3 s_{2,2} \\ & [\boldsymbol{Z}_{2,1}]_4 & [\boldsymbol{Z}_{2,2}]_4 & [\boldsymbol{Z}_{2,1}]_4 s_{2,1} & 0 \end{pmatrix} \tag{56}$$

it is sufficient to show that the determinant of the following matrix is nonzero:

$$\begin{pmatrix} [\boldsymbol{Z}_{1,1}]_1 & [\boldsymbol{Z}_{1,2}]_1 & & \\ [\boldsymbol{Z}_{1,1}]_3 & [\boldsymbol{Z}_{1,2}]_3 & & \\ & & [\boldsymbol{Z}_{2,1}]_2 & [\boldsymbol{Z}_{2,2}]_2 \\ & & [\boldsymbol{Z}_{2,1}]_4 & [\boldsymbol{Z}_{2,2}]_4 \end{pmatrix}.$$

This can be achieved, e.g., by setting all off-diagonal entries (i.e., $[\boldsymbol{Z}_{1,1}]_3$, $[\boldsymbol{Z}_{1,2}]_1$, $[\boldsymbol{Z}_{2,1}]_4$, and $[\boldsymbol{Z}_{2,2}]_2$) to zero and choosing all diagonal entries (i.e., $[\boldsymbol{Z}_{1,1}]_1$, $[\boldsymbol{Z}_{1,2}]_3$, $[\boldsymbol{Z}_{2,1}]_2$, and $[\boldsymbol{Z}_{2,2}]_4$) nonzero.

B. Proof for the General Case

We have to find $\boldsymbol{Z}$, $\boldsymbol{s}$, $\boldsymbol{x}$, and $\mathcal{P}$ such that

$$\left|\boldsymbol{J}_{\phi_{[\boldsymbol{x}]_{\mathcal{P}}}}(\boldsymbol{s}, [\boldsymbol{x}]_{\mathcal{D}})\right| \neq 0\,.$$

1) Construction of $\mathcal{P}$: We start by constructing the set $\mathcal{P}$. Recall that $\mathcal{P}$ specifies the indices of the pilot symbols in the vector $\boldsymbol{x} = (\boldsymbol{x}_1^{\mathrm{T}} \cdots \boldsymbol{x}_{\widetilde{T}}^{\mathrm{T}})^{\mathrm{T}}$. It will turn out convenient to use the expression

$$\mathcal{P} = \{i + (t-1)N : i \in \mathcal{P}_t, t \in [1\!:\!\widetilde{T}]\} \tag{57}$$

where $\mathcal{P}_t \subseteq [1\!:\!N]$ specifies the indices of the pilot symbols in the vector $\boldsymbol{x}_t$, $t \in [1\!:\!\widetilde{T}]$. The sets $\{\mathcal{P}_t\}_{t\in[1:\widetilde{T}]}$ have to satisfy

$$\begin{aligned} \sum_{t\in[1:\widetilde{T}]} |\mathcal{P}_t| &= |\mathcal{P}| \\ &\overset{(43)}{=} R\widetilde{T}Q + \widetilde{T}N - |\mathcal{I}| \\ &\overset{(38)}{=} R\widetilde{T}Q + \widetilde{T}N - RN + \ell \\ &\overset{(36)}{=} R\widetilde{T}Q + \widetilde{T}N - RN \\ &\qquad + \max\{0, RN - (R\widetilde{T}Q + \widetilde{T}N - \widetilde{T})\} \\ &= \max\{\widetilde{T}, R\widetilde{T}Q - (R-\widetilde{T})N\} \\ &\triangleq \vartheta_R\,. \end{aligned} \tag{58}$$

(We use the subscript R in ϑ_R because the dependence on R will be important later.) To provide intuition about our choice of the sets $\mathcal{P}_t$, we use a card game metaphor. Consider a deck of $\widetilde{T}N$ cards showing numbers from 1 to N sorted as follows: $1, 2, \ldots, N, \ldots, 1, 2, \ldots, N$ (i.e., the sequence $1, 2, \ldots, N$ repeated $\widetilde{T}$ times). The idea is to choose the ϑ_R positions of the pilot symbols by assigning the indices $i \in [1\!:\!N]$ to the sets $\mathcal{P}_t$ in the same way as the first ϑ_R cards are distributed to $\widetilde{T}$ players (in Fig. 2, we give an example of the algorithm for $\vartheta_R = 14$, $N = 6$, and $\widetilde{T} = 4$): The first card shows 1 and goes to $\mathcal{P}_1$, i.e., $1 \in \mathcal{P}_1$, and in the same way we proceed with $2 \in \mathcal{P}_2, \ldots, \widetilde{T} \in \mathcal{P}_{\widetilde{T}}$ (this corresponds to the 1st to 4th card in Fig. 2). When we run out of sets (players), we start with the first set (player) again: $\widetilde{T}+1 \in \mathcal{P}_1$, $\widetilde{T}+2 \in \mathcal{P}_2$, etc. After the card showing index N (recall that $\mathcal{P}_t \subseteq [1\!:\!N]$), the next card starts with index 1 again (in Fig. 2, the 6th card shows $N = 6$ and goes to $\mathcal{P}_2$ and the 7th card shows 1 and goes to $\mathcal{P}_3$). This scheme works as long as we avoid assigning an index to a set $\mathcal{P}_t$ to which that index was already assigned in a previous round. (In Fig. 2, this would happen after the 12th card. The 13th card shows 1 and the algorithm would set $1 \in \mathcal{P}_1$, which was already assigned to $\mathcal{P}_1$ in the first round.) To avoid this issue, we introduce an offset and skip one set (resulting in the 13th card going to $\mathcal{P}_2$ in Fig. 2) and proceed as before. The algorithm stops when ϑ_R indices (cards) have been assigned to the sets (players) $\mathcal{P}_t$.

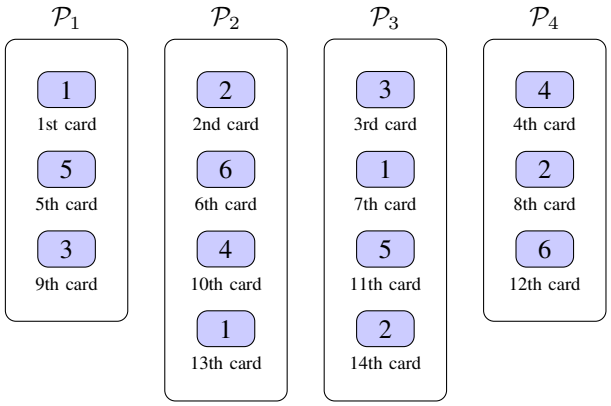

Fig. 2. Construction of the sets $\mathcal{P}_t$ for $\widetilde{T} = 4$, $N = 6$, and $\vartheta_R = 14$.

We now present a mathematical formulation of the algorithm we just outlined. Let the function $\boldsymbol{\beta}\colon [1 : \widetilde{T}N] \to [1 : \widetilde{T}] \times [1\!:\!N]$ be defined as

$$\boldsymbol{\beta}(j) = \begin{pmatrix} \beta_1(j) \\ \beta_2(j) \end{pmatrix} \triangleq \begin{pmatrix} \left(j + \left\lfloor \frac{j-1}{\mathrm{lcm}(\widetilde{T},N)} \right\rfloor\right) \bmod^* \widetilde{T} \\ j \bmod^* N \end{pmatrix}, \quad j \in [1\!:\!\widetilde{T}N]\,. \tag{59}$$

Here $\mathrm{lcm}(\cdot,\cdot)$ denotes the least common multiple and

$$a \bmod^* b \triangleq a - b\left\lfloor \frac{a-1}{b} \right\rfloor$$

denotes the residuum of a divided by b in $[1 : b]$ (and not in $[0 : b-1]$ as commonly done). We use the function $\boldsymbol{\beta}$ to assign up to $\widetilde{T}N$ elements (note that $\vartheta_R \leq \widetilde{T}N$) to the sets $\mathcal{P}_t$ as follows: for $j \in [1 : \vartheta_R]$, the function $\beta_1(j)$ specifies $t \in [1\!:\!\widetilde{T}]$ (equivalently, one of the sets $\mathcal{P}_t$, $t \in [1\!:\!\widetilde{T}]$), and the function

$\beta_2(j)$ specifies the index $i \in [1:N]$ that is assigned to $\mathcal{P}_t$ (again invoking our card game metaphor, the jth card shows the index $\beta_2(j)$ and is assigned to player $\mathcal{P}_{\beta_1(j)}$). Using $\beta_1(j)$ and $\beta_2(j)$, we can compactly describe each set $\mathcal{P}_t$ as follows:[9]

$$\mathcal{P}_t \triangleq \beta_2\big(\beta_1^{-1}(t) \cap [1:\vartheta_R]\big), \quad t \in [1:\widetilde{T}]. \tag{60}$$

Here, the set $\beta_1^{-1}(t)$ consists of all values $j \in [1:\widetilde{T}N]$ that correspond to an assignment of an index i to the set $\mathcal{P}_t$. Since we only want to assign a total of ϑ_R indices, we take the intersection with $[1:\vartheta_R]$. For each $j \in \beta_1^{-1}(t) \cap [1:\vartheta_R]$, the function β_2 now chooses an index $i \in [1:N]$, and we obtain the definition (60).

The sets $\mathcal{P}_t$ in (60) satisfy the properties listed in the following lemma.

Lemma 10: Suppose that $R \geq \widetilde{T}$, $N > \widetilde{T}Q$, and $R \leq \lceil \widetilde{T}(N-1)/(N-\widetilde{T}Q) \rceil$. Let the sets $\{\mathcal{P}_t\}_{t\in[1:\widetilde{T}]}$ be defined as in (60). Then the following properties hold:

(i) $\sum_{t\in[1:\widetilde{T}]} |\mathcal{P}_t| = \vartheta_R$;

(ii) $|\mathcal{P}_t| \leq \widetilde{T}Q$;

If $R > \widetilde{T}$, let $\{\widetilde{\mathcal{P}}_t\}_{t\in[1:\widetilde{T}]}$ be the corresponding sets for the case of $R-1$ receive antennas, i.e.,

$$\widetilde{\mathcal{P}}_t \triangleq \beta_2\big(\beta_1^{-1}(t) \cap [1:\vartheta_{R-1}]\big)\,. \tag{61}$$

Furthermore, we set

$$\mathcal{L}_t \triangleq \widetilde{\mathcal{P}}_t \setminus \mathcal{P}_t \tag{62}$$

and

$$\widetilde{\mathcal{L}} \triangleq \bigcup_{t\in[1:\widetilde{T}]} \mathcal{L}_t\,. \tag{63}$$

Then the following properties hold:

(iii) $\mathcal{L}_t \cap \mathcal{L}_{t'} = \emptyset$ for $t \neq t'$;

(iv) $\mathcal{L}_t \subseteq [1:N-\ell]$, where ℓ is defined in (36);

(v) There exist sets $\mathcal{G}_t \subseteq [1:N-\ell]$, $t \in [1:\widetilde{T}]$ satisfying

a) $|\mathcal{G}_t| = Q$,

b) $\mathcal{G}_t \cap \mathcal{G}_{t'} = \emptyset$ for $t \neq t'$,

c) $\mathcal{G}_t \cap \mathcal{P}_t \neq \emptyset$,

d) $\bigcup_{t\in[1:\widetilde{T}]} \mathcal{G}_t = \mathcal{G} \triangleq [1:N-\ell] \setminus \widetilde{\mathcal{L}}$.

Proof: See Appendix D. ∎

Remark 5: Property (i) states that the sets $\mathcal{P}_t$ have the correct size (see (58)). Properties (iii), (iv), and (v) state that we can partition the set $[1:N-\ell]$ into $2\widetilde{T}$ disjoint sets $\mathcal{L}_t$ and $\mathcal{G}_{t'}$, $t,t' \in [1:\widetilde{T}]$, i.e., $\mathcal{G}_t \cap \mathcal{G}_{t'} = \emptyset$ for $t \neq t'$ (see (v-b)), $\mathcal{L}_t \cap \mathcal{L}_{t'} = \emptyset$ for $t \neq t'$ (see (iii)), and $\mathcal{G}_t \cap \mathcal{L}_{t'} = \emptyset$ for $t,t' \in [1:\widetilde{T}]$ (see (v-d)). Furthermore, in each $\mathcal{G}_t$ there is a point $g_t \in \mathcal{P}_t$ (see (v-c)).

[9]For a set $\mathcal{A} \subseteq [1:\widetilde{T}N]$, we use the notation $\beta_2(\mathcal{A})$ to denote the image of the set $\mathcal{A}$ under the function β_2, i.e., $\beta_2(\mathcal{A}) = \{\beta_2(j) : j \in \mathcal{A}\}$.

2) Construction of $\boldsymbol{Z}$, $\boldsymbol{s}$, and $\boldsymbol{x}$: For the choice of $\{\mathcal{P}_t\}_{t\in[1:\widetilde{T}]}$ described above, it now remains to find a triple $(\boldsymbol{Z},\boldsymbol{s},\boldsymbol{x})$ for which $p(\boldsymbol{Z},\boldsymbol{s},\boldsymbol{x}) = \det\big(\boldsymbol{J}_{\phi_{[\boldsymbol{x}]_{\mathcal{P}}}}(\boldsymbol{s},[\boldsymbol{x}]_{\mathcal{D}})\big)$ is nonzero. This will be done by an induction argument over $R \geq \widetilde{T}$. For this purpose, it is convenient to define the sets

$$\mathcal{D}_t \triangleq [1:N] \setminus \mathcal{P}_t\,. \tag{64}$$

Note that by (57) and because $\mathcal{D} = [1:\widetilde{T}N] \setminus \mathcal{P}$, we have that

$$\begin{aligned}\mathcal{D} &= [1:\widetilde{T}N] \setminus \mathcal{P} \\ &= [1:\widetilde{T}N] \setminus \{i+(t-1)N : i \in \mathcal{P}_t, t \in [1:\widetilde{T}]\} \\ &= \{i+(t-1)N : i \in \mathcal{D}_t, t \in [1:\widetilde{T}]\}\end{aligned} \tag{65}$$

i.e., $\mathcal{D}_t \subseteq [1:N]$ specifies the positions of the data symbols in the vector $\boldsymbol{x}_t$, $t \in [1:\widetilde{T}]$. Furthermore, we will make repeated use of the next result, which follows from [23, Sec. 0.8.5].

Lemma 11: Let $\boldsymbol{M} \in \mathbb{C}^{n\times n}$, and let $\mathcal{E},\mathcal{F} \subseteq [1:n]$ with $|\mathcal{E}| = |\mathcal{F}|$. If $[\boldsymbol{M}]^{\mathcal{F}}_{[1:n]\setminus\mathcal{E}} = \boldsymbol{0}$ or $[\boldsymbol{M}]^{[1:n]\setminus\mathcal{F}}_{\mathcal{E}} = \boldsymbol{0}$, and if $[\boldsymbol{M}]^{\mathcal{F}}_{\mathcal{E}}$ is nonsingular, then $\det(\boldsymbol{M}) \neq 0$ if and only if $\det\Big([\boldsymbol{M}]^{[1:n]\setminus\mathcal{F}}_{[1:n]\setminus\mathcal{E}}\Big) \neq 0$.

Remark 6: Lemma 11 is just an abstract way to describe a situation where given a matrix $\boldsymbol{M}$, one is able to perform row and column interchanges that yield a new matrix of the form $\left(\begin{smallmatrix}\boldsymbol{A} & \boldsymbol{B}\\ \boldsymbol{0} & \boldsymbol{C}\end{smallmatrix}\right)$, where $\boldsymbol{A}$ and $\boldsymbol{C}$ are square matrices. In this case, a basic result in linear algebra states that the determinant of $\boldsymbol{M}$ equals the product of the determinants of $\boldsymbol{A}$ and $\boldsymbol{C}$, and hence, assuming that $\boldsymbol{C}$ is nonsingular, $\det(\boldsymbol{M}) \neq 0$ if and only if $\det(\boldsymbol{A}) \neq 0$.

We will now present the inductive construction of $\boldsymbol{Z}$, $\boldsymbol{s}$, and $\boldsymbol{x}$.

Induction hypothesis: For $\widetilde{T} \leq R \leq \lceil \widetilde{T}(N-1)/(N-\widetilde{T}Q) \rceil$, $\widetilde{T}Q < N$ (as assumed throughout the proof), and $\{\mathcal{P}_t\}_{t\in[1:\widetilde{T}]}$ as in (60), there exists a triple $(\boldsymbol{Z},\boldsymbol{s},\boldsymbol{x})$ with $\boldsymbol{x} = (1\cdots 1)^{\mathrm{T}}$ such that $p(\boldsymbol{Z},\boldsymbol{s},\boldsymbol{x}) = \det\big(\boldsymbol{J}_{\phi_{[\boldsymbol{x}]_{\mathcal{P}}}}(\boldsymbol{s},[\boldsymbol{x}]_{\mathcal{D}})\big)$ is nonzero.

Base case (proof for $R = \widetilde{T}$): When $R = \widetilde{T}$, (58) reduces to $\sum_{t\in[1:\widetilde{T}]} |\mathcal{P}_t| = \widetilde{T}^2Q$. Using Property (ii) in Lemma 10, this implies that $|\mathcal{P}_t| = \widetilde{T}Q$. Furthermore, $\ell = 0$ (see (36)), resulting in $\mathcal{I} = [1:RN]$. To establish the desired result, we first choose $\boldsymbol{s}_{r,t} = \boldsymbol{0}$ for $r \neq t$. With this choice, the matrix $\boldsymbol{J}_{\phi_{[\boldsymbol{x}]_{\mathcal{P}}}}(\boldsymbol{s},[\boldsymbol{x}]_{\mathcal{D}})$ in (47) looks as follows:

$$\begin{aligned}\boldsymbol{J}_{\phi_{[\boldsymbol{x}]_{\mathcal{P}}}}(\boldsymbol{s},[\boldsymbol{x}]_{\mathcal{D}}) &= \left[\left(\boldsymbol{B}\left[\begin{pmatrix}\boldsymbol{A}_{1,1} & \cdots & \boldsymbol{A}_{1,\widetilde{T}}\\ \vdots & & \vdots \\ \boldsymbol{A}_{\widetilde{T},1} & \cdots & \boldsymbol{A}_{\widetilde{T},\widetilde{T}}\end{pmatrix}\right]^{\mathcal{D}}\right)\right]_{\mathcal{I}} \\ &= \begin{pmatrix}\boldsymbol{B}_1 & & [\boldsymbol{A}_{1,1}]^{\mathcal{D}_1} & & \\ & \ddots & & \ddots & \\ & & \boldsymbol{B}_{\widetilde{T}} & & [\boldsymbol{A}_{\widetilde{T},\widetilde{T}}]^{\mathcal{D}_{\widetilde{T}}}\end{pmatrix} \\ &\in \mathbb{C}^{(\widetilde{T}^2Q+|\mathcal{D}|)\times(\widetilde{T}^2Q+|\mathcal{D}|)}\end{aligned} \tag{66}$$

where we used the sets $\{\mathcal{D}_t\}_{t\in[1:\widetilde{T}]}$ given in (64), and where (cf. (46))

$$\boldsymbol{B}_r = (\boldsymbol{Z}_{r,1}\cdots\boldsymbol{Z}_{r,\widetilde{T}}), \quad r \in [1:\widetilde{T}]$$

and (cf. (48))

$$\boldsymbol{A}_{t,t} = \mathrm{diag}(\boldsymbol{a}_{t,t}),\ t \in [1:\widetilde{T}],\ \text{with } \boldsymbol{a}_{t,t} \triangleq \boldsymbol{Z}_{t,t}\boldsymbol{s}_{t,t}\,. \tag{67}$$

We choose[10] $[\boldsymbol{Z}_{r,t}]_{\mathcal{P}_r} \in \mathbb{C}^{\widetilde{T}Q\times Q}$ such that the square matrices $[\boldsymbol{B}_r]_{\mathcal{P}_r} = \left[\begin{pmatrix}\boldsymbol{Z}_{r,1} & \cdots & \boldsymbol{Z}_{r,\widetilde{T}}\end{pmatrix}\right]_{\mathcal{P}_r} \in \mathbb{C}^{\widetilde{T}Q\times \widetilde{T}Q}$ are nonsingular. Furthermore, we have that $[\boldsymbol{A}_{t,t}]_{\mathcal{P}_t}^{\mathcal{D}_t} = \boldsymbol{0}$ (by (67), $\boldsymbol{A}_{t,t}$ is a diagonal matrix, and because $\mathcal{P}_t \cap \mathcal{D}_t \overset{(64)}{=} \emptyset$, the matrix $[\boldsymbol{A}_{t,t}]_{\mathcal{P}_t}^{\mathcal{D}_t}$ contains only off-diagonal entries). We will use Lemma 11 with $\boldsymbol{M} = \boldsymbol{J}_{\phi_{[\boldsymbol{x}]_{\mathcal{P}}}}(\boldsymbol{s},[\boldsymbol{x}]_{\mathcal{D}})$ given by (66), $n = \widetilde{T}^2Q + |\mathcal{D}|$, $\mathcal{E} = \mathcal{P}$ (i.e., the rows where $[\boldsymbol{A}_{t,t}]^{\mathcal{D}_t}$ is zero), and $\mathcal{F} = [1:\widetilde{T}^2Q]$ (i.e., the columns of all $\boldsymbol{B}_r$, $r \in [1:\widetilde{T}]$). This choice yields $[\boldsymbol{M}]_{\mathcal{E}}^{\mathcal{F}} = \operatorname{diag}\left([\boldsymbol{B}_1]_{\mathcal{P}_1},\ldots,[\boldsymbol{B}_{\widetilde{T}}]_{\mathcal{P}_{\widetilde{T}}}\right)$, which is nonsingular because it is a block-diagonal matrix where each block on the diagonal, $[\boldsymbol{B}_r]_{\mathcal{P}_r}$, was chosen nonsingular. Furthermore, we have that $[\boldsymbol{M}]_{\mathcal{E}}^{[1:n]\setminus\mathcal{F}} = \operatorname{diag}\left([\boldsymbol{A}_{1,1}]_{\mathcal{P}_1}^{\mathcal{D}_1},\ldots,[\boldsymbol{A}_{\widetilde{T},\widetilde{T}}]_{\mathcal{P}_{\widetilde{T}}}^{\mathcal{D}_{\widetilde{T}}}\right) = \boldsymbol{0}$. Thus, the requirements of Lemma 11 are met and, hence, $\det(\boldsymbol{M}) = \det\left(\boldsymbol{J}_{\phi_{[\boldsymbol{x}]_{\mathcal{P}}}}(\boldsymbol{s},[\boldsymbol{x}]_{\mathcal{D}})\right) \neq 0$ if and only if the determinant of the following matrix is nonzero:

$$[\boldsymbol{M}]_{[1:n]\setminus\mathcal{E}}^{[1:n]\setminus\mathcal{F}} = \begin{pmatrix}[\boldsymbol{A}_{1,1}]_{\mathcal{D}_1}^{\mathcal{D}_1} & & \\ & \ddots & \\ & & [\boldsymbol{A}_{\widetilde{T},\widetilde{T}}]_{\mathcal{D}_{\widetilde{T}}}^{\mathcal{D}_{\widetilde{T}}}\end{pmatrix}. \tag{68}$$

Because of (67), we have $[\boldsymbol{A}_{t,t}]_{\mathcal{D}_t}^{\mathcal{D}_t} = [\operatorname{diag}(\boldsymbol{a}_{t,t})]_{\mathcal{D}_t}^{\mathcal{D}_t}$. Hence, the matrix in (68) is a diagonal matrix and can be chosen to have nonzero diagonal entries by choosing $[\boldsymbol{Z}_{t,t}]_{\mathcal{D}_t}$ and $\boldsymbol{s}_{t,t}$ such that $[\boldsymbol{a}_{t,t}]_i = [\boldsymbol{Z}_{t,t}]_{\{i\}}\boldsymbol{s}_{t,t} \neq 0$ for all $i \in \mathcal{D}_t$ (again see (67)). Thus, its determinant is nonzero and, in turn, $\det(\boldsymbol{M}) \neq 0$.

Inductive step (transition from $R-1$ to R): Assuming that $\boldsymbol{Z}_{r,t}$ and $\boldsymbol{s}_{r,t}$ for $t \in [1:\widetilde{T}]$, $r \in [1:R-1]$ have already been chosen such that the determinant of $\boldsymbol{J}_{\phi_{[\boldsymbol{x}]_{\mathcal{P}}}}(\boldsymbol{s},[\boldsymbol{x}]_{\mathcal{D}})$ is nonzero in the $R-1$ setting, we want to show that there exist $\boldsymbol{Z}_{R,t}$ and $\boldsymbol{s}_{R,t}$, $t \in [1:\widetilde{T}]$ for which the determinant of the matrix $\boldsymbol{J}_{\phi_{[\boldsymbol{x}]_{\mathcal{P}}}}(\boldsymbol{s},[\boldsymbol{x}]_{\mathcal{D}})$ in (47) is nonzero. To facilitate the exposition, we rewrite the matrices involved in a more convenient form. For the case of R receive antennas, denoted by the superscript $[R]$, we rewrite the Jacobian matrix $\boldsymbol{J}_{\phi_{[\boldsymbol{x}]_{\mathcal{P}}}}(\boldsymbol{s},[\boldsymbol{x}]_{\mathcal{D}})$ in (47) as

$$\boldsymbol{J}_{\phi_{[\boldsymbol{x}]_{\mathcal{P}}}}(\boldsymbol{s},[\boldsymbol{x}]_{\mathcal{D}})^{[R]} = \left([\boldsymbol{B}]_{\mathcal{I}}\ [\boldsymbol{A}]_{\mathcal{I}}^{\mathcal{D}}\right) \in \mathbb{C}^{(RN-\ell)\times(R\widetilde{T}Q+|\mathcal{D}|)} \tag{69}$$

with

$$[\boldsymbol{B}]_{\mathcal{I}} = \begin{pmatrix}\boldsymbol{B}_1 & & & \\ & \ddots & & \\ & & \boldsymbol{B}_{R-1} & \\ & & & [\boldsymbol{B}_R]_{[1:N-\ell]}\end{pmatrix}$$

and

$$\begin{aligned}[\boldsymbol{A}]_{\mathcal{I}}^{\mathcal{D}} &= \left[\begin{pmatrix}\boldsymbol{A}_{1,1} & \cdots & \boldsymbol{A}_{1,\widetilde{T}} \\ \vdots & & \vdots \\ \boldsymbol{A}_{R,1} & \cdots & \boldsymbol{A}_{R,\widetilde{T}}\end{pmatrix}\right]_{\mathcal{I}}^{\mathcal{D}} \\ &= \begin{pmatrix}[\boldsymbol{A}_{1,1}]^{\mathcal{D}_1} & \cdots & [\boldsymbol{A}_{1,\widetilde{T}}]^{\mathcal{D}_{\widetilde{T}}} \\ \vdots & & \vdots \\ [\boldsymbol{A}_{R-1,1}]^{\mathcal{D}_1} & \cdots & [\boldsymbol{A}_{R-1,\widetilde{T}}]^{\mathcal{D}_{\widetilde{T}}} \\ [\boldsymbol{A}_{R,1}]_{[1:N-\ell]}^{\mathcal{D}_1} & \cdots & [\boldsymbol{A}_{R,\widetilde{T}}]_{[1:N-\ell]}^{\mathcal{D}_{\widetilde{T}}}\end{pmatrix}\end{aligned}$$

[10]Note that so far we used the index t for the sets $\mathcal{P}_t$. Now we consider the matrix $[\boldsymbol{B}_r]_{\mathcal{P}_t}$ for $t = r$. Thus, it is convenient to use only the index r.

where we used (65) and (38). For the $R-1$ case, the Jacobian matrix is given by

$$\begin{aligned}&\boldsymbol{J}_{\phi_{[\boldsymbol{x}]_{\mathcal{P}}}}(\boldsymbol{s},[\boldsymbol{x}]_{\mathcal{D}})^{[R-1]} \\ &\quad= \begin{pmatrix}\boldsymbol{B}_1 & & & [\boldsymbol{A}_{1,1}]^{\widetilde{\mathcal{D}}_1} & \cdots & [\boldsymbol{A}_{1,\widetilde{T}}]^{\widetilde{\mathcal{D}}_{\widetilde{T}}} \\ & \ddots & & \vdots & & \vdots \\ & & \boldsymbol{B}_{R-1} & [\boldsymbol{A}_{R-1,1}]^{\widetilde{\mathcal{D}}_1} & \cdots & [\boldsymbol{A}_{R-1,\widetilde{T}}]^{\widetilde{\mathcal{D}}_{\widetilde{T}}}\end{pmatrix} \\ &\qquad\in \mathbb{C}^{(R-1)N\times\left((R-1)\widetilde{T}Q+\sum_{t\in[1:\widetilde{T}]}|\widetilde{\mathcal{D}}_t|\right)}\end{aligned} \tag{70}$$

where

$$\widetilde{\mathcal{D}}_t \triangleq [1:N] \setminus \widetilde{\mathcal{P}}_t \tag{71}$$

with the sets $\widetilde{\mathcal{P}}_t$ introduced in Lemma 10. Note that in (70), we do not need to truncate the matrix when selecting the rows in the set $\mathcal{I}$ as required by (47). This follows because $\ell = 0$ for $R-1 \leq \widetilde{T}(N-1)/(N-\widetilde{T}Q)$ (which holds because $R \leq \lceil \widetilde{T}(N-1)/(N-\widetilde{T}Q)\rceil$) and, hence, $\mathcal{I} = [1:(R-1)N]$.

Let $\mathcal{G}$, $\mathcal{G}_t$, and $\mathcal{L}_t$ be defined as in Lemma 10. Set $[\boldsymbol{Z}_{R,t}]_{\mathcal{G}\setminus\mathcal{G}_t} = \boldsymbol{0}$ for all $t \in [1:\widetilde{T}]$, and choose $[\boldsymbol{Z}_{R,t}]_{\mathcal{G}_t} \in \mathbb{C}^{Q\times Q}$ nonsingular for all $t \in [1:\widetilde{T}]$. With these choices, and recalling that we set $\boldsymbol{x} = (1\cdots 1)^{\mathrm{T}}$ in the induction hypothesis (whence $\boldsymbol{X}_t = \mathbf{I}_N$), it follows from (46) that $[\boldsymbol{B}_R]_{\mathcal{G}} = \left([\boldsymbol{Z}_{R,1}]_{\mathcal{G}}\ \cdots\ [\boldsymbol{Z}_{R,\widetilde{T}}]_{\mathcal{G}}\right)$ is nonsingular. Next, for each $t \in [1:\widetilde{T}]$, select an index g_t in the set $\mathcal{G}_t \cap \mathcal{P}_t$ (note that this set is non-empty due to Property (v-c) in Lemma 10). Furthermore, choose $\boldsymbol{s}_{R,t}$ to be orthogonal to the rows of $[\boldsymbol{Z}_{R,t}]_{\mathcal{G}_t\setminus\{g_t\}} \in \mathbb{C}^{(Q-1)\times Q}$ and to satisfy $[\boldsymbol{Z}_{R,t}]_{\{g_t\}}\boldsymbol{s}_{R,t} \neq 0$ (note that since $\boldsymbol{s}_{r,t} \in \mathbb{C}^Q$, it is always possible to choose $\boldsymbol{s}_{r,t}$ such that it is orthogonal to $Q-1$ vectors of a set of Q linearly independent vectors and not orthogonal to the last one). Recalling (48), we have

$$[\boldsymbol{A}_{R,t}]_{\mathcal{G}} = \left[\operatorname{diag}(\boldsymbol{a}_{R,t})\right]_{\mathcal{G}}, \quad t \in [1:\widetilde{T}]$$

where $[\boldsymbol{a}_{R,t}]_i = [\boldsymbol{Z}_{R,t}]_{\{i\}}\boldsymbol{s}_{R,t} = 0$ for $i \in \mathcal{G}\setminus\mathcal{G}_t$ by our choice $[\boldsymbol{Z}_{R,t}]_{\mathcal{G}\setminus\mathcal{G}_t} = \boldsymbol{0}$, and for $i \in \mathcal{G}_t\setminus\{g_t\}$ because we chose $\boldsymbol{s}_{R,t}$ to be orthogonal to the rows of $[\boldsymbol{Z}_{R,t}]_{\mathcal{G}_t\setminus\{g_t\}}$. Thus, $[\boldsymbol{A}_{R,t}]_{\mathcal{G}}$ has only one nonzero entry $[\boldsymbol{a}_{R,t}]_{g_t}$, which is in the g_tth column. But since $g_t \in \mathcal{P}_t$ and $\mathcal{P}_t \cap \mathcal{D}_t = \emptyset$, taking only the columns indexed by $\mathcal{D}_t$ results in $[\boldsymbol{A}_{R,t}]_{\mathcal{G}}^{\mathcal{D}_t} = \boldsymbol{0}$. We will use Lemma 11 with $\boldsymbol{M} = \boldsymbol{J}_{\phi_{[\boldsymbol{x}]_{\mathcal{P}}}}(\boldsymbol{s},[\boldsymbol{x}]_{\mathcal{D}}^{[R]})$ given in (69), $n = R\widetilde{T}Q + |\mathcal{D}|$, $\mathcal{E} = \{i + (R-1)N : i \in \mathcal{G}\}$ (i.e., the rows of $[\boldsymbol{B}_R]_{[1:N-\ell]}$ specified by $\mathcal{G}$), and $\mathcal{F} = [(R-1)\widetilde{T}Q + 1 : R\widetilde{T}Q]$ (i.e., the columns of $[\boldsymbol{B}_R]_{[1:N-\ell]}$). This choice yields

$$[\boldsymbol{M}]_{\mathcal{E}}^{\mathcal{F}} = [\boldsymbol{B}_R]_{\mathcal{G}} = \left([\boldsymbol{Z}_{R,1}]_{\mathcal{G}}\ \cdots\ [\boldsymbol{Z}_{R,\widetilde{T}}]_{\mathcal{G}}\right)$$

which is nonsingular as noted above. Furthermore, we have

$$[\boldsymbol{M}]_{\mathcal{E}}^{[1:n]\setminus\mathcal{F}} = \left(\boldsymbol{0}\ \ [\boldsymbol{A}_{R,1}]_{\mathcal{G}}^{\mathcal{D}_1}\ \cdots\ [\boldsymbol{A}_{R,\widetilde{T}}]_{\mathcal{G}}^{\mathcal{D}_{\widetilde{T}}}\right) = \boldsymbol{0}\,.$$

Hence, the requirements of Lemma 11 are satisfied. We obtain that the determinant of $\boldsymbol{M} = \boldsymbol{J}_{\phi_{[\boldsymbol{x}]_{\mathcal{P}}}}(\boldsymbol{s},[\boldsymbol{x}]_{\mathcal{D}})^{[R]}$ in (69) is nonzero if and only if the determinant of the following matrix is nonzero:

$$\boldsymbol{K} \triangleq [\boldsymbol{M}]_{[1:n]\setminus\mathcal{E}}^{[1:n]\setminus\mathcal{F}}$$

$$= \begin{pmatrix} \boldsymbol{B}_1 & & [\boldsymbol{A}_{1,1}]^{\mathcal{D}_1} & \dots & [\boldsymbol{A}_{1,\widetilde{T}}]^{\mathcal{D}_{\widetilde{T}}} \\ & \ddots & \vdots & & \vdots \\ & \boldsymbol{B}_{R-1} & [\boldsymbol{A}_{R-1,1}]^{\mathcal{D}_1} & \dots & [\boldsymbol{A}_{R-1,\widetilde{T}}]^{\mathcal{D}_{\widetilde{T}}} \\ \boldsymbol{0} & & [\boldsymbol{A}_{R,1}]^{\mathcal{D}_1}_{\widetilde{\mathcal{L}}} & \dots & [\boldsymbol{A}_{R,\widetilde{T}}]^{\mathcal{D}_{\widetilde{T}}}_{\widetilde{\mathcal{L}}} \end{pmatrix}. \tag{72}$$

Here, we used Property (v-d) in Lemma 10, i.e., that $[1 : N - \ell] \setminus \mathcal{G} = \widetilde{\mathcal{L}}$.

So far, we specified only the rows $[\boldsymbol{Z}_{R,t}]_{\mathcal{G}}$. Because $\mathcal{G} \cap \widetilde{\mathcal{L}} = \emptyset$ by Property (v-d) in Lemma 10, we can still freely choose the remaining rows $[\boldsymbol{Z}_{R,t}]_{\widetilde{\mathcal{L}}}$. We first choose the rows indexed by $\mathcal{L}_t$ such that $[\boldsymbol{Z}_{R,t}]_{\mathcal{L}_t} \boldsymbol{s}_{R,t}$ does not have zero entries (e.g., $[\boldsymbol{Z}_{R,t}]_{\{i\}} = \boldsymbol{s}^{\mathrm{H}}_{R,t}$ for $i \in \mathcal{L}_t$, resulting in $[\boldsymbol{Z}_{R,t}]_{\{i\}} \boldsymbol{s}_{R,t} = \|\boldsymbol{s}_{R,t}\|^2 \neq 0$). Next, we choose the remaining rows, indexed by $\widetilde{\mathcal{L}} \setminus \mathcal{L}_t$, to be zero, i.e., $[\boldsymbol{Z}_{R,t}]_{\widetilde{\mathcal{L}} \setminus \mathcal{L}_t} = \boldsymbol{0}$. With these choices and using (48), we obtain $[\boldsymbol{A}_{R,t}]^{\mathcal{D}_t}_{\widetilde{\mathcal{L}} \setminus \mathcal{L}_t} = \boldsymbol{0}$ and $\det\big([\boldsymbol{A}_{R,t}]^{\mathcal{L}_t}_{\mathcal{L}_t}\big) \neq 0$.

We will next use another application of Lemma 11 with $\boldsymbol{M} = \boldsymbol{K}$ given in (72), $n = (R-1)\widetilde{T}Q + |\mathcal{D}|$,

$$\mathcal{E} = [(R-1)N + 1 : (R-1)\widetilde{T}Q + |\mathcal{D}|]$$

(i.e., all rows of $\boldsymbol{K}$ below $\boldsymbol{B}_{R-1}$), and

$$\mathcal{F} = \bigcup_{t \in [1:\widetilde{T}]} \left\{ i + (R-1)\widetilde{T}Q + \sum_{t' \in [1:t-1]} |\mathcal{D}_{t'}| : i \in \mathcal{L}_t \right\}$$

(i.e., the columns of $[\boldsymbol{A}_{R,t}]^{\mathcal{D}_t}_{\mathcal{L}_t}$ for all $t \in [1 : \widetilde{T}]$). This choice results in

$$[\boldsymbol{M}]^{\mathcal{F}}_{\mathcal{E}} = \operatorname{diag}\big([\boldsymbol{A}_{R,1}]^{\mathcal{L}_1}_{\mathcal{L}_1}, \dots, [\boldsymbol{A}_{R,\widetilde{T}}]^{\mathcal{L}_{\widetilde{T}}}_{\mathcal{L}_{\widetilde{T}}}\big)$$

which is nonsingular because $\det\big([\boldsymbol{A}_{R,t}]^{\mathcal{L}_t}_{\mathcal{L}_t}\big) \neq 0$. Furthermore, we have

$$[\boldsymbol{M}]^{[1:n]\setminus\mathcal{F}}_{\mathcal{E}} = \Big(\boldsymbol{0} \;\; [\boldsymbol{A}_{R,1}]^{\mathcal{D}_1}_{\widetilde{\mathcal{L}} \setminus \mathcal{L}_1} \;\; \cdots \;\; [\boldsymbol{A}_{R,\widetilde{T}}]^{\mathcal{D}_{\widetilde{T}}}_{\widetilde{\mathcal{L}} \setminus \mathcal{L}_{\widetilde{T}}}\Big) = \boldsymbol{0}\,.$$

Thus, the requirements of Lemma 11 are satisfied, and we obtain that the determinant of $\boldsymbol{K}$ in (72) is nonzero if and only if the determinant of the following matrix is nonzero:

$$[\boldsymbol{M}]^{[1:n]\setminus\mathcal{F}}_{[1:n]\setminus\mathcal{E}} = \begin{pmatrix} \boldsymbol{B}_1 & & [\boldsymbol{A}_{1,1}]^{\mathcal{D}_1\setminus\mathcal{L}_1} & \dots & [\boldsymbol{A}_{1,\widetilde{T}}]^{\mathcal{D}_{\widetilde{T}}\setminus\mathcal{L}_{\widetilde{T}}} \\ & \ddots & \vdots & & \vdots \\ & \boldsymbol{B}_{R-1} & [\boldsymbol{A}_{R-1,1}]^{\mathcal{D}_1\setminus\mathcal{L}_1} & \dots & [\boldsymbol{A}_{R-1,\widetilde{T}}]^{\mathcal{D}_{\widetilde{T}}\setminus\mathcal{L}_{\widetilde{T}}} \end{pmatrix}. \tag{73}$$

By the definitions $\mathcal{L}_t = \widetilde{\mathcal{P}}_t \setminus \mathcal{P}_t$, $\mathcal{D}_t = [1 : N] \setminus \mathcal{P}_t$, and $\widetilde{\mathcal{D}}_t = [1 : N] \setminus \widetilde{\mathcal{P}}_t$ (see (62), (64), and (71)), we obtain

$$\mathcal{D}_t \setminus \mathcal{L}_t = ([1 : N] \setminus \mathcal{P}_t) \setminus (\widetilde{\mathcal{P}}_t \setminus \mathcal{P}_t) \overset{(a)}{=} [1 : N] \setminus \widetilde{\mathcal{P}}_t = \widetilde{\mathcal{D}}_t$$

for all $t \in [1 : \widetilde{T}]$, where (a) holds because $\mathcal{P}_t \subseteq \widetilde{\mathcal{P}}_t$. Thus, $[\boldsymbol{M}]^{[1:n]\setminus\mathcal{F}}_{[1:n]\setminus\mathcal{E}}$ in (73) is equal to $\boldsymbol{J}_{\phi_{[\boldsymbol{x}]_{\mathcal{P}}}}(\boldsymbol{s}, [\boldsymbol{x}]_{\mathcal{D}})^{[R-1]}$ in (70). Altogether, we obtain that the determinant of $\boldsymbol{J}_{\phi_{[\boldsymbol{x}]_{\mathcal{P}}}}(\boldsymbol{s}, [\boldsymbol{x}]_{\mathcal{D}})^{[R]}$ in (69) is nonzero if and only if the determinant of $[\boldsymbol{M}]^{[1:n]\setminus\mathcal{F}}_{[1:n]\setminus\mathcal{E}} = \boldsymbol{J}_{\phi_{[\boldsymbol{x}]_{\mathcal{P}}}}(\boldsymbol{s}, [\boldsymbol{x}]^{[R-1]}_{\mathcal{D}})$ in (70) is nonzero. But the determinant of $\boldsymbol{J}_{\phi_{[\boldsymbol{x}]_{\mathcal{P}}}}(\boldsymbol{s}, [\boldsymbol{x}]_{\mathcal{D}})^{[R-1]}$ is nonzero by the induction hypothesis.

Appendix B
Proof of Lemma 8

A. Proof of Part (I)

To prove part (I) of Lemma 8, i.e., that almost all of $\mathcal{M}$ can be covered by the union of disjoint measurable subsets $\mathcal{U}_k$, we will use the following lemma, which is an application of the result reported in [30, Cor. 3.2.4].

Lemma 12: Let $\mathcal{A} \subseteq \mathbb{C}^n$ be a Lebesgue measurable set and $\kappa\colon \mathbb{C}^n \to \mathbb{C}^n$ a continuously differentiable mapping (e.g., the mapping in Lemma 8). Then there exists a Lebesgue measurable set $\mathcal{B} \subseteq \mathcal{A} \cap \{\boldsymbol{u} \in \mathbb{C}^n : |\boldsymbol{J}_\kappa(\boldsymbol{u})| \neq 0\}$ such that $\kappa\big|_{\mathcal{B}}$ is one-to-one and $\kappa(\mathcal{A}) \setminus \kappa(\mathcal{B}) = \mathcal{N}$, where $\mathcal{N}$ is a set of Lebesgue measure zero.

We will use Lemma 12 repeatedly to construct the disjoint sets $\{\mathcal{U}_j\}_{j \in [1:m]}$.

Lemma 13: Let κ and $\mathcal{M}$ be as in Lemma 8, i.e., $\kappa\colon \mathbb{C}^n \to \mathbb{C}^n$ is a continuously differentiable mapping with Jacobian matrix $\boldsymbol{J}_\kappa$ such that $\boldsymbol{J}_\kappa(\boldsymbol{u})$ is nonsingular a.e. and $\mathcal{M} \triangleq \{\boldsymbol{u} \in \mathbb{C}^n : |\boldsymbol{J}_\kappa(\boldsymbol{u})| \neq 0\}$. Again as in Lemma 8, assume that for all $\boldsymbol{v} \in \mathbb{C}^n$, the cardinality of the set $\kappa^{-1}(\boldsymbol{v}) \cap \mathcal{M}$ satisfies $|\kappa^{-1}(\boldsymbol{v}) \cap \mathcal{M}| \leq m < \infty$, for some $m \in \mathbb{N}$ (i.e., $\kappa\big|_{\mathcal{M}}$ is finite-to-one). Then, for $k \in [1 : m]$, there exist disjoint Lebesgue measurable sets $\{\mathcal{U}_j\}_{j \in [1:k]}$ with $\mathcal{U}_j \subseteq \mathcal{M}$ such that $\kappa\big|_{\mathcal{U}_j}$ is one-to-one for $j \in [1 : k]$. Furthermore, there exists a set $\mathcal{N}_k$ of Lebesgue measure zero such that

$$\left|\kappa^{-1}(\boldsymbol{v}) \cap \left(\mathcal{M} \setminus \bigcup_{j \in [1:k]} \mathcal{U}_j\right)\right| \leq m - k\,,$$
$$\text{for all } \boldsymbol{v} \in \kappa\left(\mathcal{M} \setminus \bigcup_{j \in [1:k-1]} \mathcal{U}_j\right) \setminus \mathcal{N}_k\,. \tag{74}$$

Proof: We prove Lemma 13 by induction over k.

Base case (proof for $k = 1$): By Lemma 12 with $\mathcal{A} = \mathcal{M}$, we obtain a set $\mathcal{B} \subseteq \mathcal{M}$ (recall that $\mathcal{M} = \{\boldsymbol{u} \in \mathbb{C}^n\colon |\boldsymbol{J}_\kappa(\boldsymbol{u})| \neq 0\}$ and thus $\mathcal{M} \cap \{\boldsymbol{u} \in \mathbb{C}^n\colon |\boldsymbol{J}_\kappa(\boldsymbol{u})| \neq 0\} = \mathcal{M}$) such that $\kappa\big|_{\mathcal{B}}$ is one-to-one. Furthermore, $\kappa(\mathcal{M}) \setminus \kappa(\mathcal{B}) = \mathcal{N}_1$ for a set $\mathcal{N}_1$ of Lebesgue measure zero. Because $\kappa(\mathcal{B}) \subseteq \kappa(\mathcal{M})$, this implies $\kappa(\mathcal{M}) \setminus \mathcal{N}_1 = \kappa(\mathcal{B})$. Thus, for each $\boldsymbol{v} \in \kappa(\mathcal{M}) \setminus \mathcal{N}_1$, there exists $\boldsymbol{u} \in \mathcal{B}$ such that $\kappa(\boldsymbol{u}) = \boldsymbol{v}$. Equivalently, $\kappa^{-1}(\boldsymbol{v}) \cap \mathcal{B} \neq \emptyset$. Hence, for $\boldsymbol{v} \in \kappa(\mathcal{M}) \setminus \mathcal{N}_1$,

$$\begin{aligned} |\kappa^{-1}(\boldsymbol{v}) \cap (\mathcal{M} \setminus \mathcal{B})| &= |(\kappa^{-1}(\boldsymbol{v}) \cap \mathcal{M}) \setminus (\kappa^{-1}(\boldsymbol{v}) \cap \mathcal{B})| \\ &\overset{(a)}{=} |\kappa^{-1}(\boldsymbol{v}) \cap \mathcal{M}| - |\kappa^{-1}(\boldsymbol{v}) \cap \mathcal{B}| \\ &\overset{(b)}{=} |\kappa^{-1}(\boldsymbol{v}) \cap \mathcal{M}| - 1 \\ &\overset{(c)}{\leq} m - 1 \end{aligned} \tag{75}$$

where (a) holds because $\kappa^{-1}(\boldsymbol{v}) \cap \mathcal{B} \subseteq \kappa^{-1}(\boldsymbol{v}) \cap \mathcal{M}$, (b) holds because $\kappa^{-1}(\boldsymbol{v}) \cap \mathcal{B}$ is nonempty and contains at most one element since $\kappa\big|_{\mathcal{B}}$ is one-to-one, and (c) holds because we assumed that $|\kappa^{-1}(\boldsymbol{v}) \cap \mathcal{M}| \leq m$. We set $\mathcal{U}_1 \triangleq \mathcal{B}$ and, by (75), the property (74) is satisfied for $k = 1$. Furthermore, $\kappa\big|_{\mathcal{U}_1} = \kappa\big|_{\mathcal{B}}$ is one-to-one, which concludes the proof for the base case.

Inductive step (transition from k to $k+1$): Suppose we already constructed the k disjoint measureable sets $\{\mathcal{U}_j\}_{j\in[1:k]}$ and the set $\mathcal{N}_k$ satisfying (74). To simplify notation, define

$$\mathcal{U}^{[k]} \triangleq \bigcup_{j\in[1:k]} \mathcal{U}_j\,.$$

Note that (74) can now be written as

$$\big|\kappa^{-1}(\boldsymbol{v}) \cap \big(\mathcal{M}\setminus\mathcal{U}^{[k]}\big)\big| \le m-k\,, \quad \text{for all } \boldsymbol{v}\in\kappa\big(\mathcal{M}\setminus\mathcal{U}^{[k-1]}\big)\setminus\mathcal{N}_k\,. \tag{76}$$

By Lemma 12 with $\mathcal{A}=\mathcal{M}\setminus\mathcal{U}^{[k]}$, we obtain a set $\mathcal{B}$ such that $\kappa\big|_{\mathcal{B}}$ is one-to-one and

$$\mathcal{B}\subseteq\mathcal{M}\setminus\mathcal{U}^{[k]}\,. \tag{77}$$

Furthermore, $\kappa\big(\mathcal{M}\setminus\mathcal{U}^{[k]}\big)\setminus\kappa(\mathcal{B}) = \widetilde{\mathcal{N}}_{k+1}$ for a set $\widetilde{\mathcal{N}}_{k+1}$ of Lebesgue measure zero. Because $\kappa(\mathcal{B})\subseteq\kappa\big(\mathcal{M}\setminus\mathcal{U}^{[k]}\big)$, this implies $\kappa\big(\mathcal{M}\setminus\mathcal{U}^{[k]}\big)\setminus\widetilde{\mathcal{N}}_{k+1} = \kappa(\mathcal{B})$. Hence, for $\boldsymbol{v}\in\kappa\big(\mathcal{M}\setminus\mathcal{U}^{[k]}\big)\setminus\widetilde{\mathcal{N}}_{k+1}$, there exists $\boldsymbol{u}\in\mathcal{B}$ such that $\kappa(\boldsymbol{u})=\boldsymbol{v}$, or equivalently, $\kappa^{-1}(\boldsymbol{v})\cap\mathcal{B}\neq\emptyset$. Thus, similarly to (75), we obtain for $\boldsymbol{v}\in\kappa\big(\mathcal{M}\setminus\mathcal{U}^{[k]}\big)\setminus\big(\widetilde{\mathcal{N}}_{k+1}\cup\mathcal{N}_k\big)$

$$\begin{aligned}
&\big|\kappa^{-1}(\boldsymbol{v})\cap\big(\big(\mathcal{M}\setminus\mathcal{U}^{[k]}\big)\setminus\mathcal{B}\big)\big| \\
&\quad= \big|\big(\kappa^{-1}(\boldsymbol{v})\cap\big(\mathcal{M}\setminus\mathcal{U}^{[k]}\big)\big)\setminus\big(\kappa^{-1}(\boldsymbol{v})\cap\mathcal{B}\big)\big| \\
&\quad\overset{(a)}{=} \big|\kappa^{-1}(\boldsymbol{v})\cap\big(\mathcal{M}\setminus\mathcal{U}^{[k]}\big)\big| - \big|\kappa^{-1}(\boldsymbol{v})\cap\mathcal{B}\big| \\
&\quad\overset{(b)}{=} \big|\kappa^{-1}(\boldsymbol{v})\cap\big(\mathcal{M}\setminus\mathcal{U}^{[k]}\big)\big| - 1 \\
&\quad\overset{(c)}{\le} m-k-1
\end{aligned} \tag{78}$$

where (a) holds because $\kappa^{-1}(\boldsymbol{v})\cap\mathcal{B}\subseteq\kappa^{-1}(\boldsymbol{v})\cap\big(\mathcal{M}\setminus\mathcal{U}^{[k]}\big)$, (b) holds because $\kappa^{-1}(\boldsymbol{v})\cap\mathcal{B}$ is nonempty and contains at most one element since $\kappa\big|_{\mathcal{B}}$ is one-to-one, and (c) holds because of our induction hypothesis (76). Setting $\mathcal{U}_{k+1}\triangleq\mathcal{B}$, the left-hand side in (78) is equal to $\big|\kappa^{-1}(\boldsymbol{v})\cap\big(\big(\mathcal{M}\setminus\mathcal{U}^{[k]}\big)\setminus\mathcal{U}_{k+1}\big)\big| = \big|\kappa^{-1}(\boldsymbol{v})\cap\big(\mathcal{M}\setminus\mathcal{U}^{[k+1]}\big)\big|$, so that (78) becomes $\big|\kappa^{-1}(\boldsymbol{v})\cap\big(\mathcal{M}\setminus\mathcal{U}^{[k+1]}\big)\big|\le m-k-1$ for all $\boldsymbol{v}\in\kappa\big(\mathcal{M}\setminus\mathcal{U}^{[k]}\big)\setminus\big(\widetilde{\mathcal{N}}_{k+1}\cup\mathcal{N}_k\big)$. This is exactly the property (76) with k replaced by $k+1$ and $\mathcal{N}_k$ replaced by $\mathcal{N}_{k+1}\triangleq\widetilde{\mathcal{N}}_{k+1}\cup\mathcal{N}_k$. Furthermore, we have by (77) that $\mathcal{U}_{k+1}=\mathcal{B}\subseteq\mathcal{M}\setminus\mathcal{U}^{[k]}$ and thus $\mathcal{U}_{k+1}\cap\mathcal{U}_j=\emptyset$ for $j\in[1:k]$. Finally, $\kappa\big|_{\mathcal{U}_{k+1}} = \kappa\big|_{\mathcal{B}}$ is one-to-one, which concludes the proof. ∎

The sets $\{\mathcal{U}_j\}_{j\in[1:m]}$ constructed in Lemma 13 are disjoint and $\kappa\big|_{\mathcal{U}_j}$ is one-to-one for all $j\in[1:m]$. It remains to be shown that $\mathcal{U}^{[m]}=\bigcup_{j\in[1:m]}\mathcal{U}_j$ covers almost all of $\mathcal{M}$. To this end, we first show that $\kappa\big(\mathcal{M}\setminus\mathcal{U}^{[m]}\big)\setminus\mathcal{N}_m$ is empty. Assume by contradiction that $\boldsymbol{v}\in\kappa\big(\mathcal{M}\setminus\mathcal{U}^{[m]}\big)\setminus\mathcal{N}_m$. By (74) with $k=m$, we have that for all $\boldsymbol{v}\in\kappa\big(\mathcal{M}\setminus\mathcal{U}^{[m-1]}\big)\setminus\mathcal{N}_m$

$$\big|\kappa^{-1}(\boldsymbol{v})\cap\big(\mathcal{M}\setminus\mathcal{U}^{[m]}\big)\big|\le m-m=0$$

i.e., there exists no $\boldsymbol{u}\in\mathcal{M}\setminus\mathcal{U}^{[m]}$ such that $\kappa(\boldsymbol{u})=\boldsymbol{v}$. This is a contradiction to the assumption $\boldsymbol{v}\in\kappa\big(\mathcal{M}\setminus\mathcal{U}^{[m]}\big)\setminus\mathcal{N}_m$, and thus we conclude that there is no $\boldsymbol{v}\in\kappa\big(\mathcal{M}\setminus\mathcal{U}^{[m]}\big)\setminus\mathcal{N}_m$, i.e., $\kappa\big(\mathcal{M}\setminus\mathcal{U}^{[m]}\big)\setminus\mathcal{N}_m=\emptyset$. Hence, we have

$$\kappa\big(\mathcal{M}\setminus\mathcal{U}^{[m]}\big)\subseteq\mathcal{N}_m\,. \tag{79}$$

We next use the integral transformation reported in [30, Th. 3.2.3] to obtain

$$\begin{aligned}
\int_{\mathcal{M}\setminus\mathcal{U}^{[m]}} |\boldsymbol{J}_\kappa(\boldsymbol{u})|^2\,d\boldsymbol{u} &\le m\int_{\kappa(\mathcal{M}\setminus\mathcal{U}^{[m]})} d\boldsymbol{v} \\
&\overset{(79)}{\le} m\int_{\mathcal{N}_m} d\boldsymbol{v} \\
&= 0\,.
\end{aligned}$$

Because the function $|\boldsymbol{J}_\kappa(\boldsymbol{u})|$ is positive on $\mathcal{M}$, it follows that the Lebesgue measure of the set $\mathcal{M}\setminus\mathcal{U}^{[m]}$ has to be zero, i.e., $\mathcal{U}^{[m]}$ covers almost all of $\mathcal{M}$. This concludes the proof of part (I).

B. Proof of Part (II)

To establish part (II), i.e., the bound (50), we first note that

$$h(\mathbf{v}) \ge h(\mathbf{v}|\mathsf{k}) = \sum_{k\in[1:m]} h(\mathbf{v}|\mathsf{k}{=}k)\,p_k \tag{80}$$

where k is the discrete random variable that takes on the value k when $\mathbf{u}\in\mathcal{U}_k$, and $p_k\triangleq\Pr\{\mathbf{u}\in\mathcal{U}_k\}=\int_{\mathcal{U}_k} f_{\mathbf{u}}(\boldsymbol{u})\,d\boldsymbol{u}$. We assume without loss of generality[11] that $p_k\neq 0$, $k\in[1:m]$. Since $\kappa\big|_{\mathcal{U}_k}$ is one-to-one, we can use the transformation rule for one-to-one mappings [12, Lemma 3] to relate $h(\mathbf{v}|\mathsf{k}{=}k)$ to $h(\mathbf{u}|\mathsf{k}{=}k)$:

$$h(\mathbf{v}|\mathsf{k}{=}k) = h(\mathbf{u}|\mathsf{k}{=}k) + \int_{\mathbb{C}^n} f_{\mathbf{u}|\mathsf{k}=k}(\boldsymbol{u})\log(|\boldsymbol{J}_\kappa(\boldsymbol{u})|^2)\,d\boldsymbol{u}\,. \tag{81}$$

The conditional probability density function of $\mathbf{u}$ given $\mathsf{k}{=}k$ is $f_{\mathbf{u}|\mathsf{k}=k}(\boldsymbol{u}) = \mathbb{1}_{\mathcal{U}_k}(\boldsymbol{u})\,f_{\mathbf{u}}(\boldsymbol{u})/p_k$. Thus, $h(\mathbf{u}|\mathsf{k}=k) = -\int_{\mathcal{U}_k}\big(f_{\mathbf{u}}(\boldsymbol{u})/p_k\big)\log\big(f_{\mathbf{u}}(\boldsymbol{u})/p_k\big)\,d\boldsymbol{u}$, and (81) becomes

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

Inserting this expression into (80), and recalling that the sets $\mathcal{U}_k$ are disjoint, that $\mathcal{U}^{[m]}=\bigcup_{k\in[1:m]}\mathcal{U}_k$ covers almost all of $\mathcal{M}$, and that $\mathbb{C}^n\setminus\mathcal{M}$ has Lebesgue measure zero, we obtain

$$\begin{aligned}
h(\mathbf{v}) &\ge \sum_{k\in[1:m]}\bigg[-\int_{\mathcal{U}_k} f_{\mathbf{u}}(\boldsymbol{u})\log\big(f_{\mathbf{u}}(\boldsymbol{u})\big)\,d\boldsymbol{u} \\
&\qquad + \int_{\mathcal{U}_k} f_{\mathbf{u}}(\boldsymbol{u})\log(|\boldsymbol{J}_\kappa(\boldsymbol{u})|^2)\,d\boldsymbol{u} + p_k\log(p_k)\bigg] \\
&= -\int_{\mathcal{U}^{[m]}} f_{\mathbf{u}}(\boldsymbol{u})\log\big(f_{\mathbf{u}}(\boldsymbol{u})\big)\,d\boldsymbol{u} \\
&\qquad + \int_{\mathcal{U}^{[m]}} f_{\mathbf{u}}(\boldsymbol{u})\log(|\boldsymbol{J}_\kappa(\boldsymbol{u})|^2)\,d\boldsymbol{u} + \underbrace{\sum_{k\in[1:m]} p_k\log(p_k)}_{-H(\mathsf{k})}
\end{aligned}$$

[11] If $p_k=0$ for some k, we simply omit the corresponding term in (80).

$$= -\int_{\mathbb{C}^n} f_{\mathbf{u}}(\boldsymbol{u}) \log\big(f_{\mathbf{u}}(\boldsymbol{u})\big)\, d\boldsymbol{u} + \int_{\mathbb{C}^n} f_{\mathbf{u}}(\boldsymbol{u}) \log(|\boldsymbol{J}_\kappa(\boldsymbol{u})|^2)\, d\boldsymbol{u} - H(\mathsf{k})$$
$$= h(\mathbf{u}) + \int_{\mathbb{C}^n} f_{\mathbf{u}}(\boldsymbol{u}) \log(|\boldsymbol{J}_\kappa(\boldsymbol{u})|^2)\, d\boldsymbol{u} - H(\mathsf{k})\,.$$

Appendix C
Proof of Lemma 9

Since f is not identically zero, there exists a $\boldsymbol{\xi}_0 \in \mathbb{C}^n$ such that $f(\boldsymbol{\xi}_0) \neq 0$. The function $g(\boldsymbol{\xi}) \triangleq f(\boldsymbol{\xi} + \boldsymbol{\xi}_0)$ is an analytic function that satisfies $g(\mathbf{0}) \neq 0$. By performing the change of variables $\boldsymbol{\xi} \mapsto \boldsymbol{\xi} + \boldsymbol{\xi}_0$, we can rewrite I_1 in (54) in the following more convenient form:

$$I_1 = \int_{\mathbb{C}^n} \exp(-\|\boldsymbol{\xi} + \boldsymbol{\xi}_0\|^2) \log(|g(\boldsymbol{\xi})|)\, d\boldsymbol{\xi}\,.$$

We have

$$\begin{aligned}\|\boldsymbol{\xi} + \boldsymbol{\xi}_0\|^2 &\leq (\|\boldsymbol{\xi}\| + \|\boldsymbol{\xi}_0\|)^2 \\ &= \|\boldsymbol{\xi}\|^2 + 2\|\boldsymbol{\xi}\|\|\boldsymbol{\xi}_0\| + \|\boldsymbol{\xi}_0\|^2 \\ &\leq \|\boldsymbol{\xi}\|^2 + 2\max\{\|\boldsymbol{\xi}\|^2, \|\boldsymbol{\xi}_0\|^2\} + \|\boldsymbol{\xi}_0\|^2 \\ &\leq 3\|\boldsymbol{\xi}\|^2 + 3\|\boldsymbol{\xi}_0\|^2\,. \end{aligned} \tag{82}$$

Using (82), we lower-bound I_1 as follows:

$$I_1 \geq c \int_{\mathbb{C}^n} \exp(-3\|\boldsymbol{\xi}\|^2) \log(|g(\boldsymbol{\xi})|)\, d\boldsymbol{\xi} \triangleq I_2 \tag{83}$$

where $c \triangleq \exp(-3\|\boldsymbol{\xi}_0\|^2)$. We next define the mapping $\varphi\colon \mathbb{R}^{2n} \to \mathbb{C}^n;\ \boldsymbol{x} \mapsto \big([\boldsymbol{x}]_{[1:n]} + i[\boldsymbol{x}]_{[n+1:2n]}\big)$, and rewrite I_2 in (83) as

$$I_2 = c \int_{\mathbb{R}^{2n}} \exp(-3\|\boldsymbol{x}\|^2)\, u(\boldsymbol{x})\, d\boldsymbol{x} \tag{84}$$

with $u(\boldsymbol{x}) \triangleq \log(|g(\varphi(\boldsymbol{x}))|)$. Since $g(\mathbf{0}) \neq 0$, we have that $u(\mathbf{0}) > -\infty$. By [17, Example 2.6.1.3], $u(\boldsymbol{x})$ is a *subharmonic function*. We shall use the following property of subharmonic functions, which is a special case of the more general result reported in [17, Th. 2.6.2.1].

Lemma 14: Let u be a subharmonic function on $\mathcal{W} \subseteq \mathbb{R}^{2n}$. If $\{\boldsymbol{x} \in \mathbb{R}^{2n}\colon \|\boldsymbol{x}\| \leq r\} \subseteq \mathcal{W}$ for some $r > 0$, then

$$u(\mathbf{0}) \leq \frac{1}{\sigma_{2n}\, r^{2n-1}} \int_{\mathcal{S}_r} u(\boldsymbol{x})\, ds(\boldsymbol{x})$$

where $\mathcal{S}_r \triangleq \{\boldsymbol{x} \in \mathbb{R}^{2n}\colon \|\boldsymbol{x}\| = r\}$, the constant σ_{2n} denotes the area of the unit sphere in $\mathbb{R}^{2n}$, and ds denotes integration with respect to the $(2n-1)$-dimensional Hausdorff measure (cf. [30, Sec. 2.10.2]).

Using a well-known measure-theoretic result (see, e.g., [30, Th. 3.2.12]), we have for $u(\boldsymbol{x}) = \log(|g(\varphi(\boldsymbol{x}))|)$

$$\int_{\mathbb{R}^{2n}} \exp(-3\|\boldsymbol{x}\|^2)\, u(\boldsymbol{x})\, d\boldsymbol{x} = \int_0^\infty \left[\int_{\mathcal{S}_r} u(\boldsymbol{x})\, ds(\boldsymbol{x})\right] \exp(-3r^2)\, dr\,. \tag{85}$$

Inserting (85) in (84), we obtain

$$I_2 = c \int_0^\infty \left[\int_{\mathcal{S}_r} u(\boldsymbol{x})\, ds(\boldsymbol{x})\right] \exp(-3r^2)\, dr$$
$$\overset{(a)}{\geq} c\, \sigma_{2n}\, u(\mathbf{0}) \int_0^\infty \exp(-3r^2)\, r^{2n-1}\, dr$$
$$\overset{(b)}{>} -\infty\,.$$

Here, (a) is due to Lemma 14 and (b) holds because $u(\mathbf{0}) > -\infty$ and $0 < \int_0^\infty \exp(-3r^2)\, r^{2n-1}\, dr < \infty$. Using (83), we conclude that $I_1 > -\infty$.

Appendix D
Proof of Lemma 10

A. Bijectivity of $\boldsymbol{\beta}$

In order to prove Lemma 10, we will use the following property of the function $\boldsymbol{\beta}$ in (59).

Lemma 15: The function $\boldsymbol{\beta}$ defined in (59) is bijective.

Proof: To facilitate the exposition, we introduce the notation

$$L \triangleq \operatorname{lcm}(\widetilde{T}, N)\,.$$

Recall that $\boldsymbol{\beta}(j) = \big(\beta_1(j)\ \beta_2(j)\big)^{\mathrm{T}}$ with $\beta_1(j) = \big(j + \lfloor (j-1)/L \rfloor\big) \bmod^* \widetilde{T} \in [1:\widetilde{T}]$ and $\beta_2(j) = j \bmod^* N \in [1:N]$, for $j \in [1:\widetilde{T}N]$. We start by proving that $\boldsymbol{\beta}$ is one-to-one. Assume that there exist $j_1, j_2 \in [1:\widetilde{T}N]$ with $j_1 \leq j_2$ such that $\boldsymbol{\beta}(j_1) = \boldsymbol{\beta}(j_2)$. From $\beta_2(j_1) = \beta_2(j_2)$, it follows that $j_1 \bmod^* N = j_2 \bmod^* N$ and, hence,[12] $j_2 = j_1 + nN$ for some $n \in [0\!:\!\widetilde{T} - 1]$. Similarly, $\beta_1(j_1) = \beta_1(j_2)$ implies that

$$j_1 + \left\lfloor \frac{j_1 - 1}{L} \right\rfloor = j_2 + \left\lfloor \frac{j_2 - 1}{L} \right\rfloor - m\widetilde{T}$$

for some $m \in \mathbb{N}$, and thus

$$j_1 + \left\lfloor \frac{j_1 - 1}{L} \right\rfloor = j_1 + nN + \left\lfloor \frac{j_1 + nN - 1}{L} \right\rfloor - m\widetilde{T}$$

or, equivalently,

$$m\widetilde{T} - nN = \left\lfloor \frac{j_1 + nN - 1}{L} \right\rfloor - \left\lfloor \frac{j_1 - 1}{L} \right\rfloor. \tag{86}$$

We can write $j_1 = kL + \tilde{\jmath}_1$ with some $k \in \mathbb{N}$ and $\tilde{\jmath}_1 \in [1:L]$ and simplify (86) as follows:

$$\begin{aligned} m\widetilde{T} - nN &= \left\lfloor \frac{kL + \tilde{\jmath}_1 + nN - 1}{L} \right\rfloor - \left\lfloor \frac{kL + \tilde{\jmath}_1 - 1}{L} \right\rfloor \\ &= k + \left\lfloor \frac{\tilde{\jmath}_1 + nN - 1}{L} \right\rfloor - k - \left\lfloor \frac{\tilde{\jmath}_1 - 1}{L} \right\rfloor \\ &\overset{(a)}{=} \left\lfloor \frac{\tilde{\jmath}_1 + nN - 1}{L} \right\rfloor. \end{aligned} \tag{87}$$

Here, (a) holds because $\tilde{\jmath}_1 - 1 < L$ and thus $\lfloor (\tilde{\jmath}_1 - 1)/L \rfloor = 0$. We will next show that the right-hand side of (87) is zero, by establishing the following chain of inequalities:

$$0 \leq \left\lfloor \frac{\tilde{\jmath}_1 + nN - 1}{L} \right\rfloor \overset{(a)}{\leq} \left\lfloor \frac{j_1 + nN - 1}{L} \right\rfloor$$

[12] Recall that we defined $a \bmod^* b \triangleq a - b\lfloor (a-1)/b \rfloor$ to be the residuum of a divided by b in $[1:b]$ (and not in $[0:b-1]$ as commonly done).

$$\overset{(b)}{\leq} \left\lfloor \frac{\widetilde{T}N-1}{L} \right\rfloor$$
$$\overset{(c)}{=} \left\lfloor \gcd(\widetilde{T},N) - \frac{1}{L} \right\rfloor$$
$$= \gcd(\widetilde{T},N) - 1\,. \tag{88}$$

Here, (a) holds because $\tilde{j}_1 \leq j_1$, (b) holds because $j_1+nN = j_2 \leq \widetilde{T}N$, and (c) holds because $\widetilde{T}N = \gcd(\widetilde{T},N)L$ [31, Th. 52] (here, $\gcd(\cdot,\cdot)$ denotes the greatest common divisor). Note now that $\gcd(\widetilde{T},N)$ divides the left-hand side of (87) and, hence, also the right-hand side. But by (88), the right-hand side of (87) is an element of $[0\,{:}\,\gcd(\widetilde{T},N)-1]$. Hence, it must be zero, and thus (87) becomes

$$m\widetilde{T} - nN = \left\lfloor \frac{\tilde{j}_1 + nN - 1}{L} \right\rfloor = 0\,. \tag{89}$$

Therefore, $\tilde{j}_1+nN-1 < L$. Since $nN \leq \tilde{j}_1+nN-1$, we obtain $nN < L$. Furthermore, by (89), we have that $m\widetilde{T} = nN$. Thus, nN is a common multiple of $\widetilde{T}$ and N that is less than the least (positive) common multiple. Therefore, $n = 0$ and, hence, $j_1 = j_1 + nN = j_2$. We have thus shown that $\boldsymbol{\beta}(j_1) = \boldsymbol{\beta}(j_2)$ implies $j_1 = j_2$, which means that $\boldsymbol{\beta}$ is one-to-one. Since the domain of $\boldsymbol{\beta}$, $[1\,{:}\,\widetilde{T}N]$, and its codomain, $[1\,{:}\,\widetilde{T}] \times [1\,{:}\,N]$, are finite and of the same cardinality (namely, $\widetilde{T}N$), we conclude that $\boldsymbol{\beta}$ is also bijective. ∎

We will now prove the individual properties stated in Lemma 10.

B. Proof of Property (i)

We first show that $\beta_2\big|_{\beta_1^{-1}(t)}$ is one-to-one, i.e., if $\beta_2(j_1) = \beta_2(j_2)$ for $j_1, j_2 \in \beta_1^{-1}(t)$ then $j_1 = j_2$. To this end, let $j_1, j_2 \in \beta_1^{-1}(t)$ (i.e., $\beta_1(j_1) = \beta_1(j_2) = t$) and assume that $\beta_2(j_1) = \beta_2(j_2) = i$. Then $\boldsymbol{\beta}(j_1) = \boldsymbol{\beta}(j_2) = (t\ i)^{\mathrm{T}}$. Since $\boldsymbol{\beta}$ is one-to-one by Lemma 15, we conclude that $j_1 = j_2$. Hence, $\beta_2\big|_{\beta_1^{-1}(t)}$ is one-to-one. Furthermore, since $\beta_1^{-1}(t) \cap [1\,{:}\,\vartheta_R] \subseteq \beta_1^{-1}(t)$, we have (cf. (60))

$$|\mathcal{P}_t| = \big|\beta_2\big(\beta_1^{-1}(t) \cap [1\,{:}\,\vartheta_R]\big)\big| = \big|\beta_1^{-1}(t) \cap [1\,{:}\,\vartheta_R]\big| \tag{90}$$

for $t \in [1\,{:}\,\widetilde{T}]$. To conclude the proof, we will use the following basic lemma.

Lemma 16: The sets $\{\beta_1^{-1}(t)\}_{t\in[1:\widetilde{T}]}$ form a partition of the domain $[1\,{:}\,\widetilde{T}N]$ of β_1, i.e.,

$$\beta_1^{-1}(t) \cap \beta_1^{-1}(t') = \emptyset, \quad \text{for } t, t' \in [1\,{:}\,\widetilde{T}] \text{ with } t \neq t' \tag{91}$$

and

$$\bigcup_{t\in[1:\widetilde{T}]} \beta_1^{-1}(t) = [1\,{:}\,\widetilde{T}N]\,. \tag{92}$$

Proof: This lemma follows from the definition of a function, i.e., the fact that β_1 maps every element in the domain to exactly one element in the codomain. ∎

By Lemma 16, we obtain

$$\sum_{t\in[1:\widetilde{T}]} |\mathcal{P}_t| \overset{(90)}{=} \sum_{t\in[1:\widetilde{T}]} \big|\beta_1^{-1}(t) \cap [1\,{:}\,\vartheta_R]\big|$$
$$\overset{(91)}{=} \left| \left(\bigcup_{t\in[1:\widetilde{T}]} \beta_1^{-1}(t) \right) \cap [1\,{:}\,\vartheta_R] \right|$$
$$\overset{(92)}{=} \big|[1\,{:}\,\widetilde{T}N] \cap [1\,{:}\,\vartheta_R]\big|$$
$$= \min\{\widetilde{T}N, \vartheta_R\}\,. \tag{93}$$

Since $N > \widetilde{T}Q$, we have that $\vartheta_R = \max\{\widetilde{T}, R\widetilde{T}Q - (R-\widetilde{T})N\} = \max\{\widetilde{T}, \widetilde{T}N - R(N-\widetilde{T}Q)\} < \widetilde{T}N$. Combining this with (93), we conclude that

$$\sum_{t\in[1:\widetilde{T}]} |\mathcal{P}_t| \;=\; \vartheta_R\,.$$

C. Proof of Property (ii)

We will make use of the following lemma.

Lemma 17: Let $p, q \in \mathbb{N}$ with $p < q$. Then

$$\big|\{j \in [p+1\,{:}\,q] : (j+a) \bmod^* b = c\}\big| \leq \left\lceil \frac{q-p}{b} \right\rceil$$

for all $a, b, c \in \mathbb{N}$ with $b \geq 2$, $c \geq 1$, and $c \leq b$.

Proof: We prove Lemma 17 by contradiction. Assume

$$\big|\{j \in [p+1\,{:}\,q] : (j+a) \bmod^* b = c\}\big| > \left\lceil \frac{q-p}{b} \right\rceil \triangleq d\,.$$

Thus, the set $\{j \in [p+1\,{:}\,q] : (j+a) \bmod^* b = c\}$ contains at least $d+1$ elements $\{j_i\}_{i\in[1:d+1]}$, i.e., there exist at least $d+1$ distinct elements $j_i \in [p+1\,{:}\,q]$ satisfying $(j_i+a) \bmod^* b = c$. Hence, there exist distinct $k_i \in \mathbb{N}$, $i \in [1\,{:}\,d+1]$ such that

$$j_i + a = c + k_i b \in [p+1\,{:}\,q]\,. \tag{94}$$

Assume, without loss of generality, that $k_i < k_{i+1}$ for $i \in [1\,{:}\,d]$. Because $k_i \in \mathbb{N}$, we obtain $k_i \leq k_{i+1} - 1$ and thus, iteratively, $k_1 \leq k_2 - 1 \leq k_3 - 2 \leq \cdots$, and finally

$$k_1 \leq k_{d+1} - d\,. \tag{95}$$

Hence,

$$j_{d+1} - j_1 \overset{(94)}{=} k_{d+1}b - k_1 b$$
$$= (k_{d+1} - k_1)b$$
$$\overset{(95)}{\geq} d\,b$$
$$= \left\lceil \frac{q-p}{b} \right\rceil b$$
$$\geq q - p$$

which contradicts $j_1, j_{d+1} \in [p+1\,{:}\,q]$. ∎

To prove Property (ii), we first establish an upper bound on ϑ_R. We have that

$$R\widetilde{T}Q - (R-\widetilde{T})N = (R-\widetilde{T})\widetilde{T}Q - (R-\widetilde{T})N + \widetilde{T}^2Q$$
$$= \underbrace{(R-\widetilde{T})}_{\geq 0}\underbrace{(\widetilde{T}Q - N)}_{<0} + \widetilde{T}^2Q$$
$$\leq \widetilde{T}^2Q$$

and, hence,

$$\vartheta_R = \max\{\widetilde{T}, R\widetilde{T}Q - (R-\widetilde{T})N\} \leq \widetilde{T}^2Q\,. \tag{96}$$

To bound the size of the sets $\mathcal{P}_t$, we use (90) and the definition of β_1 to conclude that

$$\begin{aligned}|\mathcal{P}_t| &= \big|\{j \in [1\!:\!\vartheta_R] : \beta_1(j) = t\}\big| \\ &= \left|\left\{j \in [1\!:\!\vartheta_R] : \left(j + \left\lfloor \frac{j-1}{L} \right\rfloor\right) \bmod^* \widetilde{T} = t\right\}\right|. \quad (97)\end{aligned}$$

Choose $m \in \mathbb{N}$ such that $(m-1)L < \vartheta_R \leq mL$. We can partition the set $[1\!:\!\vartheta_R]$ as follows:

$$\begin{aligned}[1\!:\!\vartheta_R] = \Bigg(\bigcup_{n\in[0:m-2]} \big[nL+1 : (n+1)L\big]\Bigg) \\ \cup \big[(m-1)L+1 : \vartheta_R\big]. \quad (98)\end{aligned}$$

Note that the intervals $\big[nL+1 : (n+1)L\big]$, $n \in [0\!:\!m-2]$ and $\big[(m-1)L+1 : \vartheta_R\big]$ in (98) are disjoint and satisfy

$$\left\lfloor \frac{j-1}{L} \right\rfloor = \begin{cases} n, & \text{for } j \in \big[nL+1 : (n+1)L\big] \\ m-1, & \text{for } j \in \big[(m-1)L+1 : \vartheta_R\big]. \end{cases} \quad (99)$$

Thus, using (98) and (99) in (97), we obtain

$$\begin{aligned}|\mathcal{P}_t| = \sum_{n\in[0:m-2]} \big|\big\{j \in \big[nL+1 : (n+1)L\big] : \\ (j+n) \bmod^* \widetilde{T} = t\big\}\big| \\ + \big|\big\{j \in \big[(m-1)L+1 : \vartheta_R\big] : \\ (j+m-1) \bmod^* \widetilde{T} = t\big\}\big|. \quad (100)\end{aligned}$$

By Lemma 17, we have

$$\begin{aligned}\big|\big\{j \in \big[nL+1 : (n+1)L\big] : (j+n) \bmod^* \widetilde{T} = t\big\}\big| \\ \leq \left\lceil \frac{L}{\widetilde{T}} \right\rceil = \frac{L}{\widetilde{T}} \quad (101)\end{aligned}$$

and

$$\begin{aligned}\big|\big\{j \in \big[(m-1)L+1 : \vartheta_R\big] : (j+m-1) \bmod^* \widetilde{T} = t\big\}\big| \\ \leq \left\lceil \frac{\vartheta_R - (m-1)L}{\widetilde{T}} \right\rceil. \quad (102)\end{aligned}$$

Thus, inserting (101) and (102) into (100), we obtain

$$\begin{aligned}|\mathcal{P}_t| &\leq (m-1)\frac{L}{\widetilde{T}} + \left\lceil \frac{\vartheta_R - (m-1)L}{\widetilde{T}} \right\rceil \\ &\overset{(a)}{=} (m-1)\frac{L}{\widetilde{T}} + \left\lceil \frac{\vartheta_R}{\widetilde{T}} \right\rceil - (m-1)\frac{L}{\widetilde{T}} \\ &= \left\lceil \frac{\vartheta_R}{\widetilde{T}} \right\rceil \\ &\overset{(96)}{\leq} \left\lceil \frac{\widetilde{T}^2 Q}{\widetilde{T}} \right\rceil \\ &= \widetilde{T}Q\end{aligned}$$

where (a) holds because $L/\widetilde{T} \in \mathbb{N}$ (recall that $L = \operatorname{lcm}(\widetilde{T}, N)$).

D. Proof of Property (iii)

To prove Properties (iii)–(v), we calculate the difference $\vartheta_{R-1} - \vartheta_R$. Because we assumed that $R \leq \lceil \widetilde{T}(N-1)/(N-\widetilde{T}Q)\rceil$, we have $R-1 < \widetilde{T}(N-1)/(N-\widetilde{T}Q)$. This is easily verified to be equivalent to $(R-1)\widetilde{T}Q - (R-1-\widetilde{T})N > \widetilde{T}$. Hence, using (58),

$$\begin{aligned}\vartheta_{R-1} &= \max\{\widetilde{T}, (R-1)\widetilde{T}Q - (R-1-\widetilde{T})N\} \\ &= (R-1)\widetilde{T}Q - (R-1-\widetilde{T})N. \quad (103)\end{aligned}$$

Thus, we have

$$\begin{aligned}\vartheta_{R-1} - \vartheta_R \\ &= (R-1)\widetilde{T}Q - (R-1-\widetilde{T})N \\ &\qquad - \max\{\widetilde{T}, R\widetilde{T}Q - (R-\widetilde{T})N\} \\ &= R\widetilde{T}Q - (R-\widetilde{T})N + N - \widetilde{T}Q \\ &\qquad - \max\{\widetilde{T}, R\widetilde{T}Q - (R-\widetilde{T})N\} \\ &= N - \widetilde{T}Q - \max\big\{\widetilde{T} - \big(R\widetilde{T}Q - (R-\widetilde{T})N\big), 0\big\} \\ &= N - \widetilde{T}Q - \ell \quad (104)\end{aligned}$$

where ℓ was defined in (36). Furthermore, by (37), $\ell < N - \widetilde{T}Q$ and thus (104) implies

$$\vartheta_{R-1} - \vartheta_R > 0. \quad (105)$$

We are now ready to prove Property (iii). From the definitions $\mathcal{P}_t \triangleq \beta_2\big(\beta_1^{-1}(t) \cap [1\!:\!\vartheta_R]\big)$ in (60) and $\widetilde{\mathcal{P}}_t \triangleq \beta_2\big(\beta_1^{-1}(t) \cap [1\!:\!\vartheta_{R-1}]\big)$ in (61), it follows that $\mathcal{L}_t = \widetilde{\mathcal{P}}_t \setminus \mathcal{P}_t$ (recall (62)) can be written as

$$\begin{aligned}\mathcal{L}_t &= \beta_2\big(\beta_1^{-1}(t) \cap [1\!:\!\vartheta_{R-1}]\big) \setminus \beta_2\big(\beta_1^{-1}(t) \cap [1\!:\!\vartheta_R]\big) \\ &\overset{(a)}{=} \beta_2\big((\beta_1^{-1}(t) \cap [1\!:\!\vartheta_{R-1}]) \setminus (\beta_1^{-1}(t) \cap [1\!:\!\vartheta_R])\big) \\ &= \beta_2\big(\beta_1^{-1}(t) \cap [\vartheta_R + 1\!:\!\vartheta_{R-1}]\big) \quad (106)\end{aligned}$$

where (a) holds because $\beta_2\big|_{\beta_1^{-1}(t)}$ is one-to-one (see Section D.B). Since $\beta_2(j) = j \bmod^* N$, the function β_2 is one-to-one on every set consisting of up to N consecutive integers. In particular, (105) and (104) imply that $\big|[\vartheta_R + 1\!:\!\vartheta_{R-1}]\big| = \vartheta_{R-1} - \vartheta_R = N - \widetilde{T}Q - \ell$ and hence $\beta_2\big|_{[\vartheta_R+1:\vartheta_{R-1}]}$ is one-to-one. Because by Lemma 16 the sets $\beta_1^{-1}(t)$, $t \in [1\!:\!\widetilde{T}]$ are pairwise disjoint, we conclude that the sets $\beta_1^{-1}(t) \cap [\vartheta_R + 1\!:\!\vartheta_{R-1}]$, $t \in [1\!:\!\widetilde{T}]$ are pairwise disjoint too. Hence, by (106) and because $\beta_2\big|_{[\vartheta_R+1:\vartheta_{R-1}]}$ is one-to-one, the sets $\mathcal{L}_t$ are pairwise disjoint.

E. Proof of Property (iv)

By (106), we have

$$\mathcal{L}_t = \beta_2\big(\beta_1^{-1}(t) \cap [\vartheta_R + 1\!:\!\vartheta_{R-1}]\big) \subseteq \beta_2([1\!:\!\vartheta_{R-1}]). \quad (107)$$

Hence, it remains to prove that

$$\beta_2([1\!:\!\vartheta_{R-1}]) \subseteq [1\!:\!N-\ell]. \quad (108)$$

Recall that we assumed $R \leq \lceil \widetilde{T}(N-1)/(N-\widetilde{T}Q)\rceil$. If $R < \lceil \widetilde{T}(N-1)/(N-\widetilde{T}Q)\rceil$, then $R < \widetilde{T}(N-1)/(N-\widetilde{T}Q)$ (because $R \in \mathbb{N}$), which implies $RN - (R\widetilde{T}Q + \widetilde{T}N - \widetilde{T}) < 0$; hence, it follows from the definition of ℓ in (36) that $\ell = 0$.

In this case, it follows from the definition of β_2 in (59), i.e., $\beta_2(j) = j \bmod^* N$ for $j \in [1\!:\!\widetilde{T}N]$, that (108) is trivially true. For the complementary case $R = \lceil \widetilde{T}(N-1)/(N-\widetilde{T}Q) \rceil$, we note that $RN - (R\widetilde{T}Q + \widetilde{T}N - \widetilde{T}) \geq 0$ and hence, using the definition of ℓ in (36),

$$\begin{aligned} N - \ell &= N - (RN - R\widetilde{T}Q - \widetilde{T}N + \widetilde{T}) \\ &= R\widetilde{T}Q - (R - 1 - \widetilde{T})N - \widetilde{T} \\ &\geq (R-1)\widetilde{T}Q - (R - 1 - \widetilde{T})N \\ &\overset{(103)}{=} \vartheta_{R-1}\,. \end{aligned}$$

Thus, $[1 : \vartheta_{R-1}] \subseteq [1 : N - \ell]$ and, further, $\beta_2([1 : \vartheta_{R-1}]) \subseteq \beta_2([1 : N - \ell]) = [1 : N - \ell]$, i.e., (108) is again true. Combining (107) and (108) concludes the proof that $\mathcal{L}_t \subseteq [1\!:\!N - \ell]$.

F. Proof of Property (v)

We have

$$\begin{aligned} \widetilde{\mathcal{L}} &\overset{(63)}{=} \bigcup_{t\in[1:\widetilde{T}]} \mathcal{L}_t \\ &\overset{(106)}{=} \bigcup_{t\in[1:\widetilde{T}]} \beta_2\big(\beta_1^{-1}(t) \cap [\vartheta_R + 1\!:\!\vartheta_{R-1}]\big) \\ &\overset{(a)}{=} \beta_2\Bigg(\bigcup_{t\in[1:\widetilde{T}]} \Big(\beta_1^{-1}(t) \cap [\vartheta_R + 1\!:\!\vartheta_{R-1}]\Big)\Bigg) \\ &= \beta_2\Bigg(\Bigg(\bigcup_{t\in[1:\widetilde{T}]} \beta_1^{-1}(t)\Bigg) \cap [\vartheta_R + 1\!:\!\vartheta_{R-1}]\Bigg) \\ &\overset{(92)}{=} \beta_2([\vartheta_R + 1\!:\!\vartheta_{R-1}]) \end{aligned} \tag{109}$$

where (a) holds because β_2 is one-to-one on every set consisting of up to N consecutive integers. Thus, $\big|\widetilde{\mathcal{L}}\big| = \big|\beta_2([\vartheta_R + 1\!:\!\vartheta_{R-1}])\big| = \vartheta_{R-1} - \vartheta_R \overset{(104)}{=} N - \widetilde{T}Q - \ell$. Furthermore, Property (iv) implies that the set $\widetilde{\mathcal{L}}$ is a subset of $[1\!:\!N - \ell]$, and hence we obtain for the size of $\mathcal{G} = [1\!:\!N - \ell] \setminus \widetilde{\mathcal{L}}$

$$\begin{aligned} |\mathcal{G}| &= \big|[1\!:\!N - \ell] \setminus \widetilde{\mathcal{L}}\big| \\ &= N - \ell - (N - \widetilde{T}Q - \ell) \\ &= \widetilde{T}Q\,. \end{aligned}$$

Thus, we can partition $\mathcal{G}$ as $\mathcal{G} = \bigcup_{t\in[1:\widetilde{T}]} \mathcal{G}_t$, with disjoint $\mathcal{G}_t$ of size Q each. We have thus shown the existence of sets $\mathcal{G}_t$ satisfying (v-a), (v-b), and (v-d).

It remains to show (v-c), i.e., that we can choose $\{\mathcal{G}_t\}_{t\in[1:\widetilde{T}]}$ such that each $\mathcal{G}_t$ has a nonempty intersection with $\mathcal{P}_t$. Because β_2 is one-to-one on sets of up to N consecutive integers and

$$\begin{aligned} \vartheta_{R-1} - (\vartheta_R - \widetilde{T}) &\overset{(104)}{=} N - \widetilde{T}Q - \ell + \widetilde{T} \\ &= N - \ell - \widetilde{T}(Q-1) \\ &\leq N - \ell \end{aligned}$$

we obtain that $\beta_2\big|_{[\vartheta_R - \widetilde{T} + 1:\vartheta_{R-1}]}$ is one-to-one. Thus,

$$\begin{aligned} &\beta_2([\vartheta_R - \widetilde{T} + 1\!:\!\vartheta_R]) \cap \beta_2([\vartheta_R + 1\!:\!\vartheta_{R-1}]) \\ &\quad= \beta_2([\vartheta_R - \widetilde{T} + 1\!:\!\vartheta_R] \cap [\vartheta_R + 1\!:\!\vartheta_{R-1}]) \\ &\quad= \beta_2(\emptyset) \\ &\quad= \emptyset\,. \end{aligned} \tag{110}$$

Inserting (109) into (110), we obtain

$$\beta_2([\vartheta_R - \widetilde{T} + 1\!:\!\vartheta_R]) \cap \widetilde{\mathcal{L}} = \emptyset\,. \tag{111}$$

By the fact that $[\vartheta_R - \widetilde{T} + 1\!:\!\vartheta_R] \subseteq [1\!:\!\vartheta_{R-1}]$ and (108), we have that $\beta_2([\vartheta_R - \widetilde{T} + 1\!:\!\vartheta_R]) \subseteq \beta_2([1\!:\!\vartheta_{R-1}]) \subseteq [1\!:\!N - \ell]$. Hence, (111) implies that

$$\beta_2([\vartheta_R - \widetilde{T} + 1\!:\!\vartheta_R]) \subseteq [1\!:\!N - \ell] \setminus \widetilde{\mathcal{L}} = \mathcal{G}\,.$$

Thus, we identified $\widetilde{T}$ elements $\beta_2(\vartheta_R - \widetilde{T} + 1), \beta_2(\vartheta_R - \widetilde{T} + 2), \ldots, \beta_2(\vartheta_R)$ in the set $\mathcal{G}$, which will now be used to construct the sets $\mathcal{G}_t$. We will show that we can assign a different index $t \in [1\!:\!\widetilde{T}]$ to each of these $\widetilde{T}$ elements such that the element with index t belongs to $\mathcal{P}_t$, i.e.,

$$\begin{aligned} \beta_2([\vartheta_R - \widetilde{T} + 1\!:\!\vartheta_R]) = \{g_1, \ldots, g_{\widetilde{T}}\}, \\ \text{with } g_t \in \mathcal{P}_t, t \in [1\!:\!\widetilde{T}]\,. \end{aligned} \tag{112}$$

The desired sets $\mathcal{G}_t$ are then obtained by assigning g_t to $\mathcal{G}_t$, for $t \in [1\!:\!\widetilde{T}]$. Thus, recalling that $|\mathcal{G}_t| = Q$, $\mathcal{G}_t$ consists of $g_t \in \mathcal{P}_t$ and $Q - 1$ additional elements taken from the set $\mathcal{G} \setminus \beta_2([\vartheta_R - \widetilde{T} + 1\!:\!\vartheta_R])$.

In order to prove (112), we distinguish two cases.

Case $nL \notin [\vartheta_R - \widetilde{T} + 1\!:\!\vartheta_R - 1]$ for All $n \in \mathbb{N}$

In this case, there exists $m \in \mathbb{N}$ such that $mL \leq \vartheta_R - \widetilde{T}$ and $(m+1)L \geq \vartheta_R$. Thus, for all $j \in [\vartheta_R - \widetilde{T} + 1\!:\!\vartheta_R]$, we have

$$\left\lfloor \frac{j-1}{L} \right\rfloor \geq \left\lfloor \frac{\vartheta_R - \widetilde{T}}{L} \right\rfloor \geq \left\lfloor \frac{mL}{L} \right\rfloor = m \tag{113}$$

and

$$\left\lfloor \frac{j-1}{L} \right\rfloor < \left\lfloor \frac{\vartheta_R}{L} \right\rfloor \leq \left\lfloor \frac{(m+1)L}{L} \right\rfloor = m+1\,. \tag{114}$$

Combining (113) and (114), we obtain that the offset in (59) satisfies $\lfloor (j-1)/L \rfloor = m$ for all $j \in [\vartheta_R - \widetilde{T} + 1\!:\!\vartheta_R]$. Thus, we have $\beta_1\big|_{[\vartheta_R - \widetilde{T} + 1:\vartheta_R]}(j) = (j+m) \bmod^* \widetilde{T}$, which implies that $\beta_1([\vartheta_R - \widetilde{T} + 1\!:\!\vartheta_R]) = [1\!:\!\widetilde{T}]$. Hence, we can write

$$\begin{aligned} [\vartheta_R - \widetilde{T} + 1\!:\!\vartheta_R] = \{\tilde{\jmath}_1, \ldots, \tilde{\jmath}_{\widetilde{T}}\}, \\ \text{where } \tilde{\jmath}_t \in \beta_1^{-1}(t) \text{ for } t \in [1\!:\!\widetilde{T}]\,. \end{aligned}$$

We then obtain

$$\beta_2([\vartheta_R - \widetilde{T} + 1\!:\!\vartheta_R]) = \{\beta_2(\tilde{\jmath}_1), \ldots, \beta_2(\tilde{\jmath}_{\widetilde{T}})\}$$

and assign the indices $t \in [1 : \widetilde{T}]$ according to $g_t = \beta_2(\tilde{\jmath}_t)$. By construction, we have both $g_t = \beta_2(\tilde{\jmath}_t) \in \beta_2(\beta_1^{-1}(t))$ and $g_t = \beta_2(\tilde{\jmath}_t) \in \beta_2([\vartheta_R - \widetilde{T} + 1 : \vartheta_R]) \subseteq \beta_2([1 : \vartheta_R])$, so that we also have

$$g_t \in \beta_2(\beta_1^{-1}(t) \cap [1\!:\!\vartheta_R]) = \mathcal{P}_t$$

(recall (60)). Thus, our choice of the g_t satisfies (112).

Case $nL \in [\vartheta_R - \widetilde{T} + 1 : \vartheta_R - 1]$ *for Some* $n \in \mathbb{N}$

We first note that

$$\begin{aligned}
&\beta_2([\vartheta_R - \widetilde{T} + 1 : \vartheta_R]) \\
&\quad = \beta_2([\vartheta_R - \widetilde{T} + 1 : nL]) \cup \beta_2([nL + 1 : \vartheta_R]) \\
&\quad \overset{(a)}{=} \beta_2([\vartheta_R - \widetilde{T} + 1 : nL]) \cup \beta_2([nL - L + 1 : \vartheta_R - L]) \\
&\quad = \beta_2([\vartheta_R - \widetilde{T} + 1 : nL]) \cup \beta_2([(n-1)L + 1 : \vartheta_R - L])
\end{aligned} \tag{115}$$

where (a) holds because (recall that $L = \operatorname{lcm}(\widetilde{T}, N)$ is a multiple of N)

$$\beta_2(j) = j \bmod^* N = (j - L) \bmod^* N = \beta_2(j - L)$$

for $j > L$. We will next calculate the offset $\lfloor (j-1)/L \rfloor$ in (59) for j belonging to either of the intervals in the arguments in (115), i.e., $j \in [\vartheta_R - \widetilde{T} + 1 : nL]$ or $j \in [(n-1)L + 1 : \vartheta_R - L]$. Note that

$$nL \in [\vartheta_R - \widetilde{T} + 1 : \vartheta_R - 1] \tag{116}$$

and

$$L \geq \widetilde{T}\,. \tag{117}$$

Thus, we have

$$(n-1)L = nL - L \overset{(116)}{<} \vartheta_R - L \overset{(117)}{\leq} \vartheta_R - \widetilde{T} \tag{118}$$

and

$$\vartheta_R \overset{(116)}{<} nL + \widetilde{T} \overset{(117)}{\leq} (n+1)L\,. \tag{119}$$

For $j \in [\vartheta_R - \widetilde{T} + 1 : nL]$, we obtain that $j - 1 \geq \vartheta_R - \widetilde{T} \overset{(118)}{>} (n-1)L$ and $j - 1 \leq nL - 1$. Hence, $n - 1 < (j-1)/L < n$ and further

$$\left\lfloor \frac{j-1}{L} \right\rfloor = n - 1, \quad \text{for } j \in [\vartheta_R - \widetilde{T} + 1 : nL]\,. \tag{120}$$

Similarly, for $j \in [(n-1)L + 1 : \vartheta_R - L]$, we obtain $j - 1 \leq \vartheta_R - L - 1 \overset{(119)}{<} (n+1)L - L - 1 = nL - 1$ and $j - 1 \geq (n-1)L$. Thus, $n - 1 \leq (j-1)/L < n$ and further

$$\left\lfloor \frac{j-1}{L} \right\rfloor = n - 1, \quad \text{for } j \in [(n-1)L + 1 : \vartheta_R - L]\,. \tag{121}$$

Combining (120) and (121), we conclude that the offset in (59) satisfies

$$\begin{aligned}
\left\lfloor \frac{j-1}{L} \right\rfloor = n - 1\,, \text{ for } & j \in [\vartheta_R - \widetilde{T} + 1 : nL] \\
& \cup [(n-1)L + 1 : \vartheta_R - L]\,.
\end{aligned} \tag{122}$$

Let us next consider β_1 on the sets $[\vartheta_R - \widetilde{T} + 1 : nL]$ and $[(n-1)L + 1 : \vartheta_R - L]$. We obtain

$$\begin{aligned}
&\beta_1([\vartheta_R - \widetilde{T} + 1 : nL]) \\
&\quad = \big\{k = \beta_1(j) = \big(j + \lfloor (j-1)/L \rfloor\big) \bmod^* \widetilde{T} : \\
&\qquad\qquad j \in [\vartheta_R - \widetilde{T} + 1 : nL]\big\} \\
&\quad \overset{(122)}{=} \big\{k = \beta_1(j) = (j + n - 1) \bmod^* \widetilde{T} : \\
&\qquad\qquad j \in [\vartheta_R - \widetilde{T} + 1 : nL]\big\} \\
&\quad = \big\{k = j \bmod^* \widetilde{T} : j \in [\vartheta_R - \widetilde{T} + n : nL + n - 1]\big\} \\
&\quad \overset{(a)}{=} \big\{k \in [1 : \widetilde{T}] : \exists\, m \in \mathbb{N} \text{ such that} \\
&\qquad\qquad k + m\widetilde{T} \in [\vartheta_R - \widetilde{T} + n : nL + n - 1]\big\}
\end{aligned} \tag{123}$$

where (a) holds because $k = j \bmod^* \widetilde{T}$ is equivalent to $j = k + m\widetilde{T}$ for some $m \in \mathbb{N}$. Similarly,

$$\begin{aligned}
&\beta_1([(n-1)L + 1 : \vartheta_R - L]) \\
&\quad = \big\{k = \beta_1(j) = \big(j + \lfloor (j-1)/L \rfloor\big) \bmod^* \widetilde{T} : \\
&\qquad\qquad j \in [(n-1)L + 1 : \vartheta_R - L]\big\} \\
&\quad \overset{(122)}{=} \big\{k = \beta_1(j) = (j + n - 1) \bmod^* \widetilde{T} : \\
&\qquad\qquad j \in [(n-1)L + 1 : \vartheta_R - L]\big\} \\
&\quad = \big\{k = j \bmod^* \widetilde{T} : \\
&\qquad\qquad j \in [(n-1)L + n : \vartheta_R - L + n - 1]\big\} \\
&\quad = \big\{k \in [1 : \widetilde{T}] : \exists\, m \in \mathbb{N} \text{ such that} \\
&\qquad\qquad k + m\widetilde{T} \in [(n-1)L + n : \vartheta_R - L + n - 1]\big\} \\
&\quad \overset{(a)}{=} \big\{k \in [1 : \widetilde{T}] : \exists\, m \in \mathbb{N} \text{ such that} \\
&\qquad\qquad k + m\widetilde{T} \in [nL + n : \vartheta_R + n - 1]\big\}
\end{aligned} \tag{124}$$

where (a) holds because a shift of the interval by L (which is a multiple of $\widetilde{T}$) can be compensated by choosing a different $m \in \mathbb{N}$. Combining (123) and (124), we obtain

$$\begin{aligned}
&\beta_1\big([\vartheta_R - \widetilde{T} + 1 : nL] \cup [(n-1)L + 1 : \vartheta_R - L]\big) \\
&\quad = \big\{k \in [1 : \widetilde{T}] : \exists\, m \in \mathbb{N} \text{ such that} \\
&\qquad\qquad k + m\widetilde{T} \in [\vartheta_R - \widetilde{T} + n : nL + n - 1] \\
&\qquad\qquad\qquad \cup [nL + n : \vartheta_R + n - 1]\big\} \\
&\quad = \big\{k \in [1 : \widetilde{T}] : \exists\, m \in \mathbb{N} \text{ such that} \\
&\qquad\qquad k + m\widetilde{T} \in [\vartheta_R - \widetilde{T} + n : \vartheta_R + n - 1]\big\} \\
&\quad \overset{(a)}{=} [1 : \widetilde{T}]
\end{aligned} \tag{125}$$

where (a) holds because $[\vartheta_R - \widetilde{T} + n : \vartheta_R + n - 1]$ is an interval of length $\widetilde{T}$ and thus for every $k \in [1 : \widetilde{T}]$ we can find an $m \in \mathbb{N}$ such that $k + m\widetilde{T} \in [\vartheta_R - \widetilde{T} + n : \vartheta_R + n - 1]$. Similarly to the previous case, (125) allows us to write

$$[\vartheta_R - \widetilde{T} + 1 : nL] \cup [(n-1)L + 1 : \vartheta_R - L] = \{\tilde{\jmath}_1, \ldots, \tilde{\jmath}_{\widetilde{T}}\}$$

where $\tilde{\jmath}_t \in \beta_1^{-1}(t)$ for $t \in [1 : \widetilde{T}]$. By (115), we then obtain

$$\begin{aligned}
&\beta_2([\vartheta_R - \widetilde{T} + 1 : \vartheta_R]) \\
&\quad = \beta_2([\vartheta_R - \widetilde{T} + 1 : nL] \cup [(n-1)L + 1 : \vartheta_R - L]) \\
&\quad = \{\beta_2(\tilde{\jmath}_1), \ldots, \beta_2(\tilde{\jmath}_{\widetilde{T}})\}\,.
\end{aligned}$$

By the same arguments as in the previous case, we find that assigning $g_t = \beta_2(\tilde{\jmath}_t)$ satisfies (112).

Acknowledgment

The authors would like to thank Dr. Shaowei Lin for pointing them to the weak version of Bézout's theorem. Furthermore, they would like to thank the associate editor and the anonymous reviewers, whose insightful comments helped them improve the presentation of their results.

REFERENCES

[1] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecomm.*, vol. 10, no. 6, pp. 585–595, Nov. 1999.

[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] Y. Liang and V. V. Veeravalli, "Capacity of noncoherent time-selective Rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3095–3110, Dec. 2004.

[5] U. G. Schuster, G. Durisi, H. Bölcskei, and H. V. Poor, "Capacity bounds for peak-constrained multiantenna wideband channels," *IEEE Trans. Commun.*, vol. 57, no. 9, pp. 2686–2696, Sep. 2009.

[6] S. M. Moser, "The fading number of multiple-input multiple-output fading channels with memory," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2716–2755, Jun. 2009.

[7] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing—the large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.

[8] 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.

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

[10] V. I. Morgenshtern, G. Durisi, and H. Bölcskei, "The SIMO pre-log can be larger than the SISO pre-log," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT 2010)*, Austin, TX, June 2010, pp. 320–324.

[11] E. Riegler, V. I. Morgenshtern, G. Durisi, S. Lin, B. Sturmfels, and H. Bölcskei, "Noncoherent SIMO pre-log via resolution of singularities," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT 2011)*, St. Petersburg, Russia, Aug. 2011, pp. 2020–2024.

[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*, vol. 59, no. 7, pp. 4213–4229, Jul. 2013.

[13] 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 Conf. Commun. Control Comput.*, Monticello, IL, Sep. 2012, pp. 1198–1205.

[14] G. Koliander, E. Riegler, G. Durisi, and F. Hlawatsch, "Generic correlation increases noncoherent MIMO capacity," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT 2013)*, Istanbul, Turkey, Jul. 2013, pp. 2084–2088.

[15] 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.

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

[17] V. Azarin, *Growth Theory of Subharmonic Functions*. Basel, Switzerland: Birkhäuser, 2009.

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

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

[20] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2426–2467, Oct. 2003.

[21] G. Durisi and H. Bölcskei, "High-SNR capacity of wireless communication channels in the noncoherent setting: A primer," *Int. J. Electron. Commun. (AEÜ)*, vol. 65, no. 8, pp. 707–712, Aug. 2011.

[22] F. Neeser and J. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1293–1302, Jul. 1993.

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

[24] R. C. Gunning and H. Rossi, *Analytic Functions of Several Complex Variables*. Englewood Cliffs, NJ: Prentice-Hall, 1965.

[25] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed. New York, NY: McGraw-Hill, 1976.

[26] ——, *Real and Complex Analysis*, 3rd ed. New York, NY: McGraw-Hill, 1987.

[27] G. Durisi, V. I. Morgenshtern, and H. Bölcskei, "On the sensitivity of continuous-time noncoherent fading channel capacity," *IEEE Trans. Inf. Theory*, vol. 58, no. 10, pp. 6372–6391, Oct. 2012.

[28] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Processing Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[29] A. M. 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, July 2005.

[30] H. Federer, *Geometric Measure Theory*. New York, NY: Springer, 1969.

[31] G. Hardy and E. Wright, *An Introduction to the Theory of Numbers*, 4th ed. Oxford, UK: Oxford Univ. Press, 1975.

Günther Koliander (S'13) received the Master degree in Technical Mathematics (with distinction) from Vienna University of Technology, Austria, in 2011. Since 2011 he has been with the Institute of Telecommunications, Vienna University of Technology, Austria, where he is currently working towards his PhD. He twice held visiting researcher positions at Chalmers University of Technology, Gothenburg, Sweden. His research interests are in the areas of noncoherent communications and information theory.

Erwin Riegler (M'07) received the Dipl-Ing. degree in Technical Physics (with distinction) in 2001 and the Dr. techn. degree in Technical Physics (with distinction) in 2004 from Vienna University of Technology. From 2005 to 2006, he was a post-doctoral researcher at the Institute for Analysis and Scientific Computing, Vienna University of Technology. From 2007 to 2010, he was a senior researcher at the Telecommunications Research Center Vienna (FTW). From 2010 to 2014, he was a post-doctoral researcher at the Institute of Telecommunications, Vienna University of Technology. Since 2014, he has been a senior researcher with the Communication Theory Group at ETH Zurich, Switzerland.

Dr. Riegler was a visiting researcher at the Max Planck Institute for Mathematics in the Sciences in Leipzig, Germany (Sep. 2004 to Feb. 2005), the Communication Theory Group at ETH Zurich, Switzerland (Sep. 2010 to Feb. 2011 and June 2012 to Nov. 2012), the Department of Electrical and Computer Engineering at The Ohio State University in Columbus, Ohio (Mar. 2012), and the Department of Signals and Systems at Chalmers University of Technology in Gothenburg, Sweden (Nov. 2013). He is a co-author of a paper that won a Student Paper Award at the 2012 International Symposium on Information Theory.

His research interests include noncoherent communications, machine learning, interference management, large system analysis, and transceiver design.

Giuseppe Durisi (S'02–M'06–SM'12) received the Laurea degree summa cum laude and the Doctor degree both from Politecnico di Torino, Italy, in 2001 and 2006, respectively. From 2002 to 2006, he was with Istituto Superiore Mario Boella, Torino, Italy. From 2006 to 2010 he was a postdoctoral researcher at ETH Zurich, Switzerland. Since 2010 he has been with Chalmers University of Technology, Gothenburg, Sweden, where he is now an associate professor. He held visiting researcher positions at IMST, Germany, University of Pisa, Italy, ETH Zurich, Switzerland, and Vienna University of Technology, Austria.

Dr. Durisi is a senior member of the IEEE. He is the recipient of the 2013 IEEE ComSoc Best Young Researcher Award for the Europe, Middle East, and Africa Region, and is co-author of a paper that won a Student Paper Award at the 2012 International Symposium on Information Theory, and of a paper that won the 2013 IEEE Sweden VT-COM-IT joint chapter best student conference paper award. He served as TPC member in several IEEE conferences, and is currently publications editor of the IEEE TRANSACTIONS ON INFORMATION THEORY. His research interests are in the areas of communication and information theory.

Franz Hlawatsch (S'85–M'88–SM'00–F'12) received the Diplom-Ingenieur, Dr. techn., and Univ.-Dozent (habilitation) degrees in electrical engineering/signal processing from Vienna University of Technology, Vienna, Austria in 1983, 1988, and 1996, respectively. Since 1983, he has been with the Institute of Telecommunications, Vienna University of Technology, where he is currently an Associate Professor. During 1991–1992, as a recipient of an Erwin Schrödinger Fellowship, he spent a sabbatical year with the Department of Electrical Engineering, University of Rhode Island, Kingston, RI, USA. In 1999, 2000, and 2001, he held one-month Visiting Professor positions with INP/ENSEEIHT, Toulouse, France and IRCCyN, Nantes, France. He (co)authored a book, three review papers that appeared in the IEEE SIGNAL PROCESSING MAGAZINE, about 200 refereed scientific papers and book chapters, and three patents. He coedited three books. His research interests include wireless communications, sensor networks, and statistical and compressive signal processing.

Prof. Hlawatsch was Technical Program Co-Chair of EUSIPCO 2004 and served on the technical committees of numerous IEEE conferences. He was an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2003 to 2007 and for the IEEE TRANSACTIONS ON INFORMATION THEORY from 2008 to 2011. From 2004 to 2009, he was a member of the IEEE SPCOM Technical Committee. He is coauthor of papers that won an IEEE Signal Processing Society Young Author Best Paper Award and a Best Student Paper Award at IEEE ICASSP 2011.