Chalmers Publication Library

Copyright Notice

Capacity Pre-Log of Noncoherent SIMO Channels via Hironaka's Theorem

Veniamin I. Morgenshtern, Erwin Riegler,
Wei Yang, Giuseppe Durisi,
Shaowei Lin, Bernd Sturmfels, and Helmut Bölcskei

Abstract—We find the capacity pre-log of a temporally correlated Rayleigh block-fading single-input multiple-output (SIMO) channel in the noncoherent setting. It is well known that for block-length L and rank of the channel covariance matrix equal to Q, the capacity pre-log in the single-input single-output (SISO) case is given by $1 - Q/L$. Here, Q/L can be interpreted as the pre-log penalty incurred by channel uncertainty. Our main result reveals that, by adding only one receive antenna, this penalty can be reduced to $1/L$ and can, hence, be made to vanish for the block-length $L \to \infty$, even if Q/L remains constant as $L \to \infty$. Intuitively, even though the SISO channels between the transmit antenna and the two receive antennas are statistically independent, the transmit signal induces enough statistical dependence between the corresponding receive signals for the second receive antenna to be able to resolve the uncertainty associated with the first receive antenna's channel and thereby make the overall system appear coherent. The proof of our main theorem is based on a deep result from algebraic geometry known as Hironaka's Theorem on the Resolution of Singularities.

I. INTRODUCTION

It is well known that the capacity pre-log, i.e., the asymptotic ratio between capacity and the logarithm of signal-to-noise ratio (SNR), as SNR goes to infinity, of a single-input multiple-output (SIMO) fading channel in the *coherent setting* (i.e., when the receiver has perfect channel state information (CSI)) is equal to 1 and is, hence, the same as that of a single-input single-output (SISO) fading channel [4]. This result holds under very general assumptions on the channel statistics. Multiple antennas at the receiver only, hence, do not result in an increase of the capacity pre-log in the coherent setting [4]. In the *noncoherent setting*, where neither transmitter nor receiver have CSI, but both know the channel statistics, the effect of multiple antennas on the

The results in this paper appeared in part at the 2010 and 2011 IEEE International Symposia on Information Theory [1], [2], and at the International Symposium on Wireless Communication Systems (ISWCS) 2011, Aachen, Germany [3].

V. I. Morgenshtern is with the Department of Statistics, Stanford University, CA, USA, Email: vmorgen@stanford.edu

E. Riegler is with Vienna University of Technology, Vienna, Austria, Email: erwin.riegler@nt.tuwien.ac.at

W. Yang and G. Durisi are with the Department of Signals and Systems, Chalmers University of Technology, Gothenburg, Sweden, Email: {ywei, durisi}@chalmers.se

S. Lin is with the Institute for Infocomm Research, A*STAR, Singapore, Email: lins@i2r.a-star.edu.sg

B. Sturmfels is with the Department of Mathematics, University of California Berkeley, CA, USA, Email: bernd@math.berkeley.edu

H. Bölcskei is with the Dept. of IT & EE, ETH Zurich, Switzerland, Email: boelcskei@nari.ee.ethz.ch

capacity[1] pre-log is understood only for a specific simple channel model, namely, the Rayleigh *constant block-fading* model. In this model the channel is assumed to remain constant over a block (of L symbols) and to change in an independent fashion from block to block [5]. The corresponding SIMO capacity pre-log is again equal to the SISO capacity pre-log, but, differently from the coherent setting, is given by $1 - 1/L$ [6], [7].

An alternative approach to capturing channel variations in time is to assume that the fading process is *stationary*. In this case, the capacity pre-log is known only in the SISO [8] and the multiple-input single-output (MISO) [9, Thm. 4.15] cases. The capacity bounds for the SIMO stationary-fading channel available in the literature [9, Thm. 4.13] do not allow to determine whether the capacity pre-log in the SIMO case equals that in the SISO case. Resolving this question for stationary fading seems elusive at this point.

A widely used channel model that can be seen as lying in between the stationary-fading model considered in [8], [9], and the simpler constant block-fading model analyzed in [5], [7] is the *correlated block-fading* model, which assumes that the fading process is temporally correlated within blocks of length L and independent across blocks. The $L \times L$ channel covariance matrix of rank $Q \leq L$ is taken to be the same for each block. This channel model is relevant as it captures channel variations in time in an accurate yet simple fashion: the rank Q of the covariance matrix corresponds to the minimum number of channel coefficients per block that need to be known at the receiver to perfectly reconstruct all channel coefficients within the same block. Therefore, larger Q/L corresponds to faster channel variations.

The SISO capacity pre-log for correlated block-fading channels is given by $1 - Q/L$ [10]. In the SIMO and the multiple-input multiple-output (MIMO) cases the capacity pre-log is unknown. The main contribution of this paper is a full characterization of the capacity pre-log for SIMO correlated block-fading channels. Specifically, we prove that under a mild technical condition on the channel covariance matrix, the SIMO capacity pre-log, χ, of a channel with R receive antennas and independent identically distributed (i.i.d.) SISO subchannels is given by

$$\chi = \min[1 - 1/L, R(1 - Q/L)]. \tag{1}$$

This shows that even with $R = 2$ receive antennas a capacity pre-log of $1 - 1/L$ can be obtained in the SIMO case (provided

---

[1]In the remainder of the paper, we consider the noncoherent setting only. Consequently, we will refer to capacity in the noncoherent setting simply as capacity.

that $L \geq 2Q - 1$). This capacity pre-log is strictly larger than the capacity pre-log of the corresponding SISO channel (i.e., the capacity pre-log of one of the component channels), given by $1 - Q/L$. Here Q/L can be interpreted as pre-log penalty due to channel uncertainty. Our result reveals that, by adding at least one receive antenna, this penalty can be made to vanish in the large block-length limit, $L \to \infty$, even if the amount of channel uncertainty scales linearly in the block-length.

A conjecture for the correlated block-fading channel model stated in [10] for the MIMO case, when particularized to the SIMO case, implies that the capacity pre-log in the SIMO case would be the same as that in the SISO case. As a consequence of (1) this conjecture is disproved.

In terms of the technical aspects of our main result, we sandwich capacity between an upper and a lower bound that turn out to be asymptotically (in SNR) tight (in the sense of delivering the same capacity pre-log). The upper bound is established by proving that the capacity pre-log of a correlated block-fading channel with R receive antennas can be upper-bounded by the capacity pre-log of a constant block-fading channel with RQ receive antennas and the same SNR. The derivation of the capacity pre-log lower bound poses serious technical challenges. Specifically, after a change of variables argument applied to the integral expression for the differential entropy of the channel output signal, the main technical difficulty lies in showing that the expected logarithm of the Jacobian determinant corresponding to this change of variables is finite. As the Jacobian determinant takes on a very involved form, a per pedes approach appears infeasible. The problem is resolved by first distilling structural properties of the determinant through a suitable factorization and then introducing a powerful tool from algebraic geometry, namely [11, Th. 2.3], which is a consequence of Hironaka's Theorem on the Resolution of Singularities [12], [13]. Roughly speaking, this result allows to rewrite every real analytic function [14, Def. 1.1.5, Def. 2.2.1] locally as a product of a monomial and a nonvanishing real analytic function. This factorization is then used to show that the integral of the logarithm of the absolute value of a real analytic function over a compact set is finite, provided that the real analytic function is not identically zero. This method is quite general and may be of independent interest when one tries to show that integrals of certain functions with singularities are finite, in particular, functions involving logarithms. In information theory such integrals often occur when analyzing differential entropy.

Notation: Sets are denoted by calligraphic letters $\mathcal{A}, \mathcal{B}, \ldots$ Roman letters $A, B, \ldots$ and $a, b, \ldots$ designate deterministic matrices and vectors, respectively. Boldface letters $\mathbf{A}, \mathbf{B}, \ldots$ and $\mathbf{a}, \mathbf{b}, \ldots$ denote random matrices and random vectors, respectively. We let e_i be the vector (of appropriate dimension) that has the ith entry equal to one and all other entries equal to zero, and denote the $M \times M$ identity matrix as I_M. The element in the ith row and jth column of a deterministic matrix A is a_{ij} (italic letters), and the ith component of the deterministic vector u is u_i (italic letters); the element in the ith row and jth column of a random matrix $\mathbf{A}$ is $\mathbf{a}_{ij}$ (sans serif letters), and the ith component of the random vector $\mathbf{u}$ is $\mathbf{u}_i$ (sans serif letters). For a vector u, $\mathrm{diag}(u)$ stands for the diagonal matrix that has the entries of u on its main diagonal. The linear subspace spanned by the vectors $u_1, \ldots, u_n$ is denoted by $\mathrm{span}\{u_1, \ldots, u_n\}$. The superscripts $^\mathsf{T}$ and $^\mathsf{H}$ stand for transposition and Hermitian transposition, respectively. For two matrices A and B, we designate their Kronecker product as $A \otimes B$; to simplify notation, we use the convention that the ordinary matrix product precedes the Kronecker product, i.e., $AB \otimes C \triangleq (AB) \otimes C$. For a finite subset of the set of natural numbers, $\mathcal{I} \subset \mathbb{N}$, we write $|\mathcal{I}|$ for the cardinality of $\mathcal{I}$. For an $M \times N$ matrix A, and a set of indices $\mathcal{I} \subset [1:M]$, we use $A_\mathcal{I}$ to denote the $|\mathcal{I}| \times N$ submatrix of A containing the rows of A with indices in $\mathcal{I}$. For two matrices A and B of arbitrary size, $\mathrm{diag}(A, B)$ is the 2×2 block-diagonal matrix that has A in the upper left corner and B in the lower right corner. For N matrices $A_1, \ldots, A_N$, we let $\mathrm{diag}(A_1, \ldots, A_N) \triangleq \mathrm{diag}(\mathrm{diag}(A_1, \ldots, A_{N-1}), A_N)$. The ordered eigenvalues of the $N \times N$ matrix A are denoted by $\lambda_1(A) \geq \cdots \geq \lambda_N(A)$. For two functions $f(\cdot)$ and $g(\cdot)$, the notation $f(\cdot) = \mathcal{O}(g(\cdot))$ means that $\lim_{u \to \infty} |f(u)/g(u)|$ is bounded. For a function $f(\cdot)$, we say that $f(\cdot)$ is not identically zero and write $f(\cdot) \not\equiv 0$ if there exists at least one element u in the domain of $f(\cdot)$ such that $f(u) \neq 0$. We say that a function $f(\cdot)$ is nonvanishing on a subset $\mathcal{S}$ of its domain, if for all $u \in \mathcal{S}$, $f(u) \neq 0$. For two functions $f(\cdot)$ and $g(\cdot)$, $(f \circ g)(\cdot)$ denotes the composition $f(g(\cdot))$. For $x \in \mathbb{R}$, $\lceil x \rceil \triangleq \min\{m \in \mathbb{Z} \mid m \geq x\}$. We use $[n:m]$ to designate the set of natural numbers $\{n, n+1, \ldots, m\}$. Let $g : \mathbb{C}^M \to \mathbb{C}^N$, $u \mapsto g(u)$, be a vector-valued function; then $\partial g / \partial u$ denotes the $N \times M$ Jacobian matrix [15, Def. 3.8] of the function $g(\cdot)$, i.e., the matrix that contains the partial derivative $\partial g_i / \partial u_j$ in its ith row and jth column. The logarithm to the base 2 is written as $\log(\cdot)$. For sets $\mathcal{A}, \mathcal{B} \subseteq \mathbb{R}^M$, we define $\mathcal{A} \pm \mathcal{B} \triangleq \{a \pm b \mid a \in \mathcal{A}, b \in \mathcal{B}\}$. If $\mathcal{A} = \{a\}$, then $a \pm \mathcal{B} \triangleq \mathcal{A} \pm \mathcal{B}$. With $(-\epsilon, \epsilon) \triangleq \{u \in \mathbb{R} \mid |u| < \epsilon\}$, we denote by $\mathcal{C}(u, \epsilon) \triangleq u + (-\epsilon, \epsilon)^M \subset \mathbb{R}^M$ the open cube in $\mathbb{R}^M$ with side length 2ϵ centered at $u \in \mathbb{R}^M$. The set of natural numbers, including zero, is $\mathbb{N}_0$. For $u \in \mathbb{C}^M$ and $m \in \mathbb{N}_0^M$, we let $u^m \triangleq u_1^{m_1} \ldots u_M^{m_M}$. If $\mathcal{A}$ is a subset of the image of a map $f(\cdot)$ then $f^{-1}(\mathcal{A})$ denotes the inverse image of $\mathcal{A}$. The expectation operator is designated by $\mathbb{E}[\cdot]$. For random matrices $\mathbf{A}$ and $\mathbf{B}$, we write $\mathbf{A} \stackrel{d}{\sim} \mathbf{B}$ to indicate that $\mathbf{A}$ and $\mathbf{B}$ have the same distribution. Finally, $\mathcal{CN}(u, C)$ stands for the distribution of a jointly proper Gaussian (JPG) random vector with mean u and covariance matrix C.

II. System Model

We consider a SIMO channel with R receive antennas. The fading in each SISO component channel follows the correlated block-fading model described in the previous section. The input-output (IO) relation within any block of length L for the mth SISO component channel can be written as

$$\mathbf{y}_m = \sqrt{\rho}\, \mathrm{diag}(\mathbf{h}_m)\mathbf{x} + \mathbf{w}_m, \quad m \in [1:R], \qquad (2)$$

where $\mathbf{x} = [\mathbf{x}_1 \cdots \mathbf{x}_L]^\mathsf{T} \in \mathbb{C}^L$ is the signal vector transmitted in the given block, and the vectors $\mathbf{y}_m, \mathbf{w}_m \in \mathbb{C}^L$ are the corresponding received signal and additive noise, respectively, at the mth receive antenna. Finally, $\mathbf{h}_m \in \mathbb{C}^L$ contains the channel coefficients between the transmit antenna and the mth receive antenna. We assume that $\mathbf{h}_m \sim \mathcal{CN}(0, DD^\mathsf{H})$, for all

$m \in [1\!:\!R]$, where $\mathrm{D} \in \mathbb{C}^{L \times Q}$ (which is the same for all blocks and all component channels) has rank $Q \leq L$. The entries of the vectors $\mathbf{h}_m$ are taken to be of unit variance, which implies that the main diagonal entries of $\mathrm{DD^H}$ are equal to 1 and the average received power is constant across time slots. It will turn out convenient to write the channel coefficient vector in whitened form as $\mathbf{h}_m = \mathrm{D}\mathbf{s}_m$, where $\mathbf{s}_m \sim \mathcal{CN}(0, \mathrm{I}_Q)$. Further, we assume that $\mathbf{w}_m \sim \mathcal{CN}(0, \mathrm{I}_L)$. As the noise vector has unit variance components, ρ in (2) can be interpreted as the SNR. Finally, we assume that $\mathbf{s}_m$ and $\mathbf{w}_m$ are mutually independent, independent across m, and change in an independent fashion from block to block. Note that for $Q = 1$ the correlated block-fading model reduces to the constant block-fading model as used in [6], [7].

With $\mathbf{y} \triangleq [\mathbf{y}_1^\mathsf{T} \cdots \mathbf{y}_R^\mathsf{T}]^\mathsf{T}$, $\mathbf{s} \triangleq [\mathbf{s}_1^\mathsf{T} \cdots \mathbf{s}_R^\mathsf{T}]^\mathsf{T}$, $\mathbf{w} \triangleq [\mathbf{w}_1^\mathsf{T} \cdots \mathbf{w}_R^\mathsf{T}]^\mathsf{T}$, and $\mathbf{X} \triangleq \mathrm{diag}(\mathbf{x})$, we can write the IO relation (2) in the following—more compact—form

$$\mathbf{y} = \sqrt{\rho}\, (\mathrm{I}_R \otimes \mathbf{X}\mathrm{D})\, \mathbf{s} + \mathbf{w}. \tag{3}$$

The capacity of the channel (3) is defined as

$$C(\rho) \triangleq (1/L) \sup_{f_{\mathbf{x}}(\cdot)} I(\mathbf{x}; \mathbf{y}), \tag{4}$$

where the supremum is taken over all input distributions $f_{\mathbf{x}}(\cdot)$ that satisfy the average-power constraint

$$\mathbb{E}\big[\|\mathbf{x}\|^2\big] \leq L. \tag{5}$$

The capacity pre-log, the central quantity of interest in this paper, is defined as

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

III. Intuitive Analysis

We start with a simple "back-of-the-envelope" calculation that allows to develop some intuition on the main result in this paper, summarized in (1). The different steps in the intuitive analysis below will be seen to have rigorous counterparts in the formal proof of the capacity pre-log lower bound detailed in Section VI.

The capacity pre-log characterizes the channel capacity behavior in the regime where additive noise can "effectively" be ignored. To guess the capacity pre-log, it therefore appears prudent to consider the problem of identifying the transmit symbols x_i, $i \in [1\!:\!L]$, from the *noise-free* (and rescaled) observation

$$\hat{\mathbf{y}} \triangleq (\mathrm{I}_R \otimes \mathbf{X}\mathrm{D})\, \mathbf{s}. \tag{6}$$

Specifically, we shall ask the question: "How many symbols x_i can be identified *uniquely* from $\hat{\mathbf{y}}$ given that the vector of channel coefficients $\mathbf{s}$ is unknown but the statistics of the channel, i.e., the matrix D, are known?" The claim we make is that the capacity pre-log is given by the number of identifiable symbols divided by the block length L.

We start by noting that the unknown variables in (6) are $\mathbf{s}$ and $\mathbf{x}$, which means that we have a quadratic system of equations. It turns out, however, that the simple change of variables

$$\mathsf{z}_i \triangleq 1/\mathsf{x}_i, \ i \in [1\!:\!L], \tag{7}$$

(we make the technical assumption $|\mathsf{x}_i| > 0$, $i \in [1\!:\!L]$, in the remainder of this section) transforms (6) into a system of equations that is linear in $\mathbf{s}$ and z_i, $i \in [1\!:\!L]$. Since the transformation $\mathsf{z}_i \triangleq 1/\mathsf{x}_i$ is invertible for $|\mathsf{x}_i| > 0$, uniqueness of the solution of the linear system of equations in $\mathbf{s}$ and z_i, $i \in [1\!:\!L]$, is equivalent to uniqueness of the solution of the quadratic system of equations in $\mathbf{s}$ and x_i, $i \in [1\!:\!L]$.

For concreteness and simplicity of exposition, we first consider the case $L = 3$ and $R = Q = 2$ and assume that D satisfies the technical condition specified in Theorem 1, stated in Section IV. A direct computation reveals that upon change of variables according to (7), the quadratic system (6) can be rewritten as the following linear system of equations:

$$\begin{bmatrix} d_{11} & d_{12} & 0 & 0 & \hat{\mathsf{y}}_1 & 0 & 0 \\ d_{21} & d_{22} & 0 & 0 & 0 & \hat{\mathsf{y}}_2 & 0 \\ d_{31} & d_{32} & 0 & 0 & 0 & 0 & \hat{\mathsf{y}}_3 \\ 0 & 0 & d_{11} & d_{12} & \hat{\mathsf{y}}_4 & 0 & 0 \\ 0 & 0 & d_{21} & d_{22} & 0 & \hat{\mathsf{y}}_5 & 0 \\ 0 & 0 & d_{31} & d_{32} & 0 & 0 & \hat{\mathsf{y}}_6 \end{bmatrix} \begin{bmatrix} \mathsf{s}_1 \\ \mathsf{s}_2 \\ \mathsf{s}_3 \\ \mathsf{s}_4 \\ -\mathsf{z}_1 \\ -\mathsf{z}_2 \\ -\mathsf{z}_3 \end{bmatrix} = 0. \tag{8}$$

The solution of (8) can not be unique, as we have 6 equations in 7 unknowns. The $\mathsf{x}_i = 1/\mathsf{z}_i$, $i \in [1\!:\!3]$, can, therefore, not be determined uniquely from $\hat{\mathbf{y}}$. We can, however, make the solution of (8) to be unique if we devote one of the data symbols x_i to transmitting a pilot symbol (known to the receiver). Take, for concreteness, $\mathsf{x}_1 = 1$. Then (8) reduces to the following inhomogeneous system of 6 equations in 6 unknowns

$$\underbrace{\begin{bmatrix} d_{11} & d_{12} & 0 & 0 & 0 & 0 \\ d_{21} & d_{22} & 0 & 0 & \hat{\mathsf{y}}_2 & 0 \\ d_{31} & d_{32} & 0 & 0 & 0 & \hat{\mathsf{y}}_3 \\ 0 & 0 & d_{11} & d_{12} & 0 & 0 \\ 0 & 0 & d_{21} & d_{22} & \hat{\mathsf{y}}_5 & 0 \\ 0 & 0 & d_{31} & d_{32} & 0 & \hat{\mathsf{y}}_6 \end{bmatrix}}_{\triangleq \mathrm{B}} \begin{bmatrix} \mathsf{s}_1 \\ \mathsf{s}_2 \\ \mathsf{s}_3 \\ \mathsf{s}_4 \\ -\mathsf{z}_2 \\ -\mathsf{z}_3 \end{bmatrix} = \begin{bmatrix} \hat{\mathsf{y}}_1 \\ 0 \\ 0 \\ \hat{\mathsf{y}}_4 \\ 0 \\ 0 \end{bmatrix}. \tag{9}$$

This system of equations has a unique solution if $\det \mathrm{B} \neq 0$. We prove in Appendix C that under the technical condition on D specified in Theorem 1, stated in Section IV, we, indeed, have that $\det \mathrm{B} \neq 0$ for almost all[2] $\hat{\mathsf{y}}_2, \hat{\mathsf{y}}_3, \hat{\mathsf{y}}_5, \hat{\mathsf{y}}_6$. It, therefore, follows that for almost all $\hat{\mathbf{y}}$, the linear system of equations (9) has a unique solution. As explained above, this implies uniqueness of the solution of the original quadratic system of equations (6). We can therefore recover z_2 and z_3, and, hence, $\mathsf{x}_2 = 1/\mathsf{z}_2$ and $\mathsf{x}_3 = 1/\mathsf{z}_3$ from $\hat{\mathbf{y}}$. Summarizing our findings, we expect that the capacity pre-log of the channel (3), for the special case $L = 3$ and $R = Q = 2$, is equal to $2/3$, which is larger than the capacity pre-log of the corresponding SISO channel (i.e., one of the SISO component channels), given by $1 - Q/L = 1/3$ [10]. This answer, obtained through the back-of-the-envelope calculation above, coincides with the rigorous result in Theorem 1.

We next generalize what we learned in the example above to $L, R,$ and Q arbitrary, and start by noting that if $(\mathbf{X}, \mathbf{s})$ is a solution of $\hat{\mathbf{y}} = (\mathrm{I}_R \otimes \mathbf{X}\mathrm{D})\, \mathbf{s}$ for fixed $\hat{\mathbf{y}}$, then $(a\mathbf{X}, \mathbf{s}/a)$ with $a \in \mathbb{C}$ is also a solution of this system of equations. It is therefore immediately clear that at least one pilot symbol is needed to make this system of equations uniquely solvable.

---

[2]Except for a set of measure zero.

To guess the capacity pre-log for general parameters L, R, and Q, we first note that the homogeneous linear system of equations corresponding to that in (8), has RL equations for $RQ + L$ unknowns. As the example above indicates, we need to seek conditions under which this homogeneous linear system of equations can be converted into a linear system of equations that has a unique solution. Provided that D satisfies the technical condition specified in Theorem 1 below, this entails meeting the following two requirements: (i) at least one symbol is used as a pilot symbol to resolve the scaling ambiguity described in the previous paragraph; (ii) the number of unknowns in the system of equations corresponding to that in (8) must be smaller than or equal to the number of equations. To maximize the capacity pre-log we want to use the minimum number of pilot symbols that guarantees (i) and (ii). In order to identify this minimum, we have to distinguish two cases:

1) When $RL < RQ + L$ [in this case $\min[1 - 1/L, R(1 - Q/L)] = R(1 - Q/L)$] we will need at least $RQ + L - RL$ pilot symbols to satisfy requirement (ii). Since $RQ + L - RL \geq 1$, choosing exactly $RQ + L - RL$ pilot symbols will satisfy both requirements. The number of symbols left for communication will, therefore, be $L - (RQ + L - RL) = R(L - Q)$. Hence, we expect the capacity pre-log to be given by $R(1 - Q/L)$, which agrees with the result stated in (1).

2) When $RL \geq RQ + L$ [in this case $\min[1 - 1/L, R(1 - Q/L)] = 1 - 1/L$], we will need at least one pilot symbol to satisfy requirement (i). Since requirement (ii) is satisfied as a consequence of $RL \geq RQ + L$, it suffices to choose exactly one pilot symbol. The number of symbols left for communication will, therefore, be $L - 1$ and we hence expect the capacity pre-log to equal $1 - 1/L$, which again agrees with the result stated in (1). Note that the resulting inhomogeneous linear system of equations has RL equations in $RQ + L - 1$ unknowns. As there are more equations than unknowns, $RL - RQ - L + 1$ equations are redundant and can be eliminated.

The proof of our main result, stated in the next section, will provide rigorous justification for the casual arguments put forward in this section.

IV. THE CAPACITY PRE-LOG

The main result of this paper is the following theorem.

Theorem 1. *Suppose that* D *satisfies the following Property (A): Every Q rows of* D *are linearly independent. Then, the capacity pre-log of the SIMO channel* (3) *is given by*

$$\chi = \min[1 - 1/L, R(1 - Q/L)]. \tag{10}$$

Remark 1. We will prove Theorem 1 by showing, in Section V, that the capacity pre-log of the SIMO channel (3) can be upper-bounded as

$$\chi \leq \min[1 - 1/L, R(1 - Q/L)] \tag{11}$$

and by establishing, in Section VI, the lower bound

$$\chi \geq \min[1 - 1/L, R(1 - Q/L)]. \tag{12}$$

While the upper bound (11) can be shown to hold even if D does not satisfy Property (A), this property is crucial to establish the lower bound (12).

Remark 2. The lower bound (12) continues to hold if Property (A) is replaced by the following milder condition on D.

Property (A'): There exists a subset of indices $\mathcal{K} \subseteq [1:L]$ with cardinality

$$|\mathcal{K}| \triangleq \min(\lceil (RQ - 1)/(R - 1)\rceil, L)$$

such that every Q rows of $\mathrm{D}_{\mathcal{K}}$ are linearly independent.

We decided, however, to state our main result under the stronger Property (A) as both Property (A) and Property (A') are very mild and the proof of the lower bound (12) under Property (A') is significantly more cumbersome and does not contain any new conceptual aspects. A sketch of the proof of the stronger result (i.e., under Property (A')) can be found in [2].

We proceed to discussing the significance of Theorem 1.

A. Eliminating the prediction penalty

According to (10) the capacity pre-log of the SIMO channel (3) with $R = 2$ receive antennas is given by $\chi = 1 - 1/L$, provided that Property (A) holds, and $L \geq 2Q - 1$. Comparing to the capacity pre-log $\chi_{\mathrm{SISO}} = 1 - Q/L$ in the SISO case[3] [10] (this result also follows from (10) with $R = 1$), we see that—under a mild condition on the channel covariance matrix D—adding only one receive antenna yields a reduction of the channel uncertainty-induced pre-log penalty from Q/L to $1/L$. How significant is this reduction? Recall that Q is the number of uncertain channel parameters within each given block of length L. Hence, the ratio between the rank of the covariance matrix and the block-length, Q/L, is a measure that can be seen as quantifying the amount of channel uncertainty relative to the number of degrees of freedom for communication. It often makes sense to consider $L \to \infty$ with the amount of channel uncertainty Q/L held constant. For concreteness, consider $L, Q \to \infty$ with $L = 2Q - 1$ so that $Q/L \to 1/2$. The capacity pre-log penalty due to channel uncertainty in the SISO case is then given by $1/2$. Theorem 1 reveals that, by adding a second receive antenna, this penalty can be reduced to $1/L$ and, hence, be made to vanish in the limit $L \to \infty$. Intuitively, even though the SISO channels between the transmit antenna and the two receive antennas are statistically independent, the transmit signal induces enough statistical dependence between the corresponding receive signals for the second receive antenna to be able to resolve the channel uncertainty associated with the first receive antenna's channel and thereby make the overall system appear coherent.

B. Number of receive antennas

Note that for $Q < L$, we can rewrite (10) as

$$\chi = \min[1 - 1/L, R(1 - Q/L)]$$
$$= \begin{cases} 1 - 1/L, & \text{if } R \geq \lceil \frac{L-1}{L-Q} \rceil \\ R(1 - Q/L), & \text{else.} \end{cases} \tag{13}$$

As illustrated in Fig. 1, it follows from (13) that for fixed L and Q

---

[3]Note that the results in [10] are stated for general channel covariance matrix D.
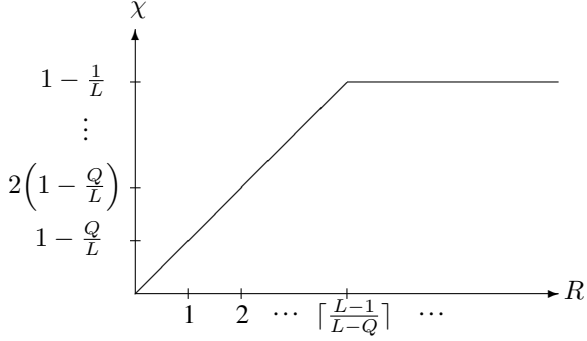
Fig. 1. The capacity pre-log of the SIMO channel (3).

with $Q < L$ the capacity pre-log of the SIMO channel (3) grows linearly with R as long as R is smaller than the critical value $\lceil (L-1)/(L-Q) \rceil$. Once R reaches this critical value, further increasing the number of receive antennas does not increase the capacity pre-log.

C. Property (A) is mild

Property (A) is not very restrictive and is satisfied by many practically relevant channel covariance matrices D. For example, removing an arbitrary set of $L - Q$ columns from an $L \times L$ discrete Fourier transform (DFT) matrix results in a matrix that satisfies Property (A) when L is prime [16]. (Weighted) DFT covariance matrices arise naturally in so-called basis-expansion models for time-selective channels [10].

Property (A) can furthermore be shown to be satisfied by "generic" matrices D. Specifically, if the entries of D are chosen randomly and independently from a continuous distribution [17, Sec. 2-3, Def. (2)] (i.e., a distribution with a well-defined probability density function (PDF)), then the resulting matrix D will satisfy Property (A) with probability one. The proof of this statement follows from a union bound argument together with the fact that N independent N-dimensional vectors drawn independently from a continuous distribution are linearly independent with probability one.

V. PROOF OF THE UPPER BOUND (11)

The proof of (11) consists of two parts. First, in Section V-A, we prove that $\chi \le R(1 - Q/L)$. This will be accomplished by generalizing—to the SIMO case—the approach developed in [10, Prop. 4] for establishing an upper bound on the SISO capacity pre-log. Second, in Section V-B, we prove that $\chi \le 1 - 1/L$ by showing that the capacity of a SIMO channel with R receive antennas and channel covariance matrix of rank Q can be upper-bounded by the capacity of a SIMO channel with RQ receive antennas, the same SNR, and a rank-1 covariance matrix. The desired result, $\chi \le 1 - 1/L$, then follows by application of [7, Eq. (27)], [18, Eq. (7)] as detailed below.

A. First part: $\chi \le R(1 - Q/L)$

To simplify notation, we first rewrite (3) as

$$\mathbf{Y} = \sqrt{\rho}\,\text{diag}(\mathbf{x})\mathbf{D}\mathbf{S} + \mathbf{W}, \tag{14}$$

where $\mathbf{Y} \triangleq [\mathbf{y}_1 \cdots \mathbf{y}_R]$, $\mathbf{H} \triangleq [\mathbf{h}_1 \cdots \mathbf{h}_R]$, $\mathbf{W} \triangleq [\mathbf{w}_1 \cdots \mathbf{w}_R]$, and $\mathbf{S} \triangleq [\mathbf{s}_1 \cdots \mathbf{s}_R]$.

Recall that D has rank Q. Without loss of generality, we assume, in what follows, that the first Q rows of D are linearly independent. This can always be ensured by reordering the scalar IO relations in (2). With $\mathcal{Q} \triangleq [1:Q]$ and $\mathcal{L} \triangleq [Q+1:L]$ we can write

$$
\begin{aligned}
I(\mathbf{Y}; \mathbf{x}) &= I(\mathbf{Y}_\mathcal{Q}, \mathbf{Y}_\mathcal{L}; \mathbf{x}) \\
&\overset{(a)}{=} I(\mathbf{Y}_\mathcal{Q}; \mathbf{x}) + I(\mathbf{Y}_\mathcal{L}; \mathbf{x} \,|\, \mathbf{Y}_\mathcal{Q}) \\
&\overset{(b)}{=} I(\mathbf{Y}_\mathcal{Q}; \mathbf{x}_\mathcal{Q}) + \underbrace{I(\mathbf{Y}_\mathcal{Q}; \mathbf{x}_\mathcal{L} \,|\, \mathbf{x}_\mathcal{Q})}_{0} + I(\mathbf{Y}_\mathcal{L}; \mathbf{x} \,|\, \mathbf{Y}_\mathcal{Q}) \\
&\overset{(c)}{=} I(\mathbf{Y}_\mathcal{Q}; \mathbf{x}_\mathcal{Q}) + I(\mathbf{Y}_\mathcal{L}; \mathbf{x} \,|\, \mathbf{Y}_\mathcal{Q}),
\end{aligned} \tag{15}
$$

where (a) and (b) follow by the chain rule for mutual information and in (c) we used that $\mathbf{Y}_\mathcal{Q}$ and $\mathbf{x}_\mathcal{L}$ are independent conditional on $\mathbf{x}_\mathcal{Q}$. Next, we upper-bound each term in (15) separately.

From [19, Thm. 4.2] we can conclude that the assumption of the first Q rows of D being linearly independent implies that the first term on the RHS of (15) grows at most double-logarithmically with SNR and hence does not contribute to the capacity pre-log. For the reader's convenience, we repeat the corresponding brief calculation from [19, Thm. 4.2] in Appendix A and show that:

$$I(\mathbf{Y}_\mathcal{Q}; \mathbf{x}_\mathcal{Q}) \le Q \log\log(\rho) + \mathcal{O}(1). \tag{16}$$

Here and in what follows, $\mathcal{O}(1)$ refers to the limit $\rho \to \infty$.

For the second term in (15) we can write

$$
\begin{aligned}
I(\mathbf{Y}_\mathcal{L}; \mathbf{x} \,|\, \mathbf{Y}_\mathcal{Q}) &= h(\mathbf{Y}_\mathcal{L} \,|\, \mathbf{Y}_\mathcal{Q}) - h(\mathbf{Y}_\mathcal{L} \,|\, \mathbf{x}, \mathbf{Y}_\mathcal{Q}) \\
&\overset{(a)}{\le} h(\mathbf{Y}_\mathcal{L}) - h(\mathbf{Y}_\mathcal{L} \,|\, \mathbf{x}, \mathbf{Y}_\mathcal{Q}, \mathbf{s}) \\
&= h(\mathbf{Y}_\mathcal{L}) - h(\mathbf{W}_\mathcal{L}) \\
&\overset{(b)}{\le} \sum_{l=Q+1}^{L} \sum_{r=1}^{R} (h(\mathsf{y}_{lr}) - h(\mathsf{w}_{lr})) \\
&\overset{(c)}{\le} \sum_{l=Q+1}^{L} \sum_{r=1}^{R} \log\left(1 + \rho\,\mathbb{E}\left[|\mathsf{h}_{lr}|^2\right] \mathbb{E}\left[|\mathsf{x}_l|^2\right]\right) \\
&\overset{(d)}{\le} \sum_{l=Q+1}^{L} \sum_{r=1}^{R} \log\left(1 + L\rho\,\mathbb{E}\left[|\mathsf{h}_{lr}|^2\right]\right) \\
&\overset{(e)}{=} R(L - Q)\log(\rho) + \mathcal{O}(1),
\end{aligned} \tag{17}
$$

where in (a) we used the fact that conditioning reduces entropy; (b) follows from the chain rule for differential entropy and the fact that conditioning reduces entropy; (c) follows because Gaussian random variables are differential-entropy-maximizers for fixed variance and because h_{lr} and x_l are independent; (d) is a consequence of the power constraint (5); and (e) follows because $\mathbb{E}\left[|\mathsf{h}_{lr}|^2\right] = 1$.

Combining (15), (16), and (17) yields

$$C(\rho) \le R(1 - Q/L)\log(\rho) + (Q/L)\log\log(\rho) + \mathcal{O}(1). \tag{18}$$

Since $\lim_{\rho \to \infty} \log\log(\rho)/\log(\rho) = 0$, this completes the proof of the bound $\chi \le R(1 - Q/L)$.

It follows from (18) that for $Q = L$, the capacity pre-log is zero and $C(\rho)$ can grow no faster than double-logarithmically in ρ.

Recall that $1 - Q/L$ is the capacity pre-log of the correlated block-fading SISO channel [10]. As the proof of the upper bound $\chi \leq R(1 - Q/L)$ reveals, the capacity pre-log of the SIMO channel (3) can not be larger than R times the capacity pre-log of the corresponding SISO channel (i.e., the capacity pre-log of one of the SISO component channels). The upper bound $R(1 - Q/L)$ may seem crude, but, surprisingly, it matches the lower bound for $R < \lceil (L-1)/(L-Q) \rceil$.

B. Second part: $\chi \leq 1 - 1/L$

The proof of $\chi \leq 1 - 1/L$ will be accomplished in two steps. In the first step, we show that the capacity of a SIMO channel with R receive antennas and rank-Q channel covariance matrix is upper-bounded by the capacity of a SIMO channel with RQ receive antennas, the same SNR, and rank-1 covariance matrix. In the second step, we exploit the fact that the channel (14) with rank-1 covariance matrix (under the assumption that the rows of D have unit norm) is a constant block-fading channel for which the capacity pre-log was shown in [7] to equal $1 - 1/L$. We now implement the proof program just outlined.

Let $d_1, \ldots, d_Q \in \mathbb{C}^L$ denote the columns of the $L \times Q$ matrix D so that $D = [d_1 \cdots d_Q]$. Let $\bar{\mathbf{s}}_1, \ldots, \bar{\mathbf{s}}_Q \in \mathbb{C}^R$ denote the transposed rows of the $Q \times R$ matrix $\mathbf{S}$ so that $\mathbf{S}^\mathsf{T} = [\bar{\mathbf{s}}_1 \cdots \bar{\mathbf{s}}_Q]$. We can rewrite the IO relation (14) in the following form that is more convenient for the ensuing analysis:

$$\mathbf{Y} = \sqrt{\rho} \sum_{q=1}^{Q} \text{diag}(d_q) \mathbf{x} \bar{\mathbf{s}}_q^\mathsf{T} + \mathbf{W}.$$

Let $\mathbf{W}_1, \ldots, \mathbf{W}_Q$ be independent random matrices of dimension $L \times R$, each with i.i.d. $\mathcal{CN}(0, 1)$ entries. As, by assumption, the rows of D have unit norm, we have that

$$\mathbf{W} \overset{d}{\sim} \sum_{q=1}^{Q} \text{diag}(d_q) \mathbf{W}_q.$$

Hence, we can rewrite $\mathbf{Y}$ as

$$\mathbf{Y} \overset{d}{\sim} \sum_{q=1}^{Q} \text{diag}(d_q) \mathbf{Y}_q, \qquad (19)$$

where

$$\mathbf{Y}_q \triangleq \sqrt{\rho} \mathbf{x} \bar{\mathbf{s}}_q^\mathsf{T} + \mathbf{W}_q, \quad q \in [1\!:\!Q]. \qquad (20)$$

Note now that each $\mathbf{Y}_q$ is the output of a SIMO channel with R receive antennas, rank-1 channel covariance matrix, and SNR ρ. Realizing that, by (19) and (20), $\mathbf{x} \to \{\mathbf{Y}_1, \ldots, \mathbf{Y}_Q\} \to \mathbf{Y}$ forms a Markov chain, we conclude, by the data-processing inequality [20, Sec. 2.8], that

$$I(\mathbf{Y}; \mathbf{x}) \leq I(\mathbf{Y}_1, \ldots, \mathbf{Y}_Q; \mathbf{x}).$$

The claim now follows by noting that the $L \times (RQ)$ matrix obtained by stacking the matrices $\mathbf{Y}_q$ next to each other can be interpreted as the output of a SIMO channel with RQ receive antennas, rank-1 covariance matrix, independent fading across

receive antennas, and SNR ρ. The proof is completed by upper-bounding the capacity of this channel by means of the following lemma.

Lemma 2. *The capacity of the SIMO channel* (14) *with R receive antennas, $Q = 1$, and $L \geq 2$ can be upper-bounded according to*

$$C(\rho) \leq (1 - 1/L) \log \rho + \mathcal{O}(1), \quad \rho \to \infty.$$

This result follows from [7, Eq. (27)]. A simpler and more detailed proof can be found in [18, Eq. (7)].

VI. PROOF OF THE LOWER BOUND (12)

To help the reader navigate through the proof of the lower bound (12), we start by explaining the architecture of the proof.

A. Architecture of the proof

The proof consists of the following steps, each of which corresponds to a subsection in this section:

Step 1: Choose an input distribution; we will see that i.i.d. $\mathcal{CN}(0, 1)$ input symbols allow us to establish the capacity pre-log lower bound (12).

Step 2: Decompose the mutual information between the input and the output of the channel according to $I(\mathbf{x}; \mathbf{y}) = h(\mathbf{y}) - h(\mathbf{y} \,|\, \mathbf{x})$.

Step 3: Using standard information-theoretic bounds show that $h(\mathbf{y} \,|\, \mathbf{x})$ is upper-bounded by $RQ \log(\rho) + \mathcal{O}(1)$.

Step 4: Split $h(\mathbf{y})$ into three terms: a term that depends on SNR, a differential entropy term that depends on the noiseless channel output $\hat{\mathbf{y}}$ only, and a differential entropy term that depends on the noise vector $\mathbf{w}$ only. Conclude that the last of these three terms is a finite constant[4].

Step 5: Conclude that the SNR-dependent term obtained in Step 4 scales (in SNR) as $\min[RQ + L - 1, RL] \log(\rho)$. Together with the decomposition from Step 2 and the result from Step 3 this gives the desired lower bound (12) provided that the $\hat{\mathbf{y}}$-dependent differential entropy obtained in Step 4 can be lower-bounded by a finite constant.

Step 6: To show that the $\hat{\mathbf{y}}$-dependent differential entropy obtained in Step 4 can be lower-bounded by a finite constant, apply the change of variables $\hat{\mathbf{y}} \to (\mathbf{x}, \mathbf{s})$ to rewrite the differential entropy as a sum of the differential entropy of $(\mathbf{x}, \mathbf{s})$ and the expected (w.r.t. $\mathbf{x}$ and $\mathbf{s}$) logarithm of the Jacobian determinant corresponding to the transformation $\hat{\mathbf{y}} \to (\mathbf{x}, \mathbf{s})$. Conclude that the differential entropy of $(\mathbf{x}, \mathbf{s})$ is a finite constant. It remains to show that the expected logarithm of the Jacobian determinant is lower-bounded by a finite constant as well.

Step 7: Factor out the $\mathbf{x}$-dependent terms from the expected logarithm of the Jacobian determinant and conclude that these terms are finite constants. It remains to show that the expected logarithm of the $\mathbf{s}$-dependent factor in the Jacobian determinant is lower-bounded by a

---

[4]Here, and in what follows, whenever we say "finite constant", we mean SNR-independent and finite.

finite constant as well. This poses the greatest technical difficulties in the proof of the lower bound (12) and is addressed in the remaining steps.

Step 8: Based on a deep result from algebraic geometry, known as Hironaka's Theorem on the Resolution of Singularities, conclude that the expected logarithm of the $\mathbf{s}$-dependent factor in the Jacobian determinant is lower-bounded by a finite constant, provided that this factor is nonzero for at least one element in its domain.

Step 9: Prove by explicit construction that there exists at least one $\mathbf{s}$, for which the $\mathbf{s}$-dependent factor in the Jacobian determinant is nonzero.

We next implement the proof program outlined above.

B. Step 1: Choice of input distribution

First note that for $Q = L$ the lower bound in (12) is reduced to $\chi \geq 0$ and is hence trivially satisfied. In the remainder of the paper we shall therefore assume that $Q < L$.

We shall furthermore work under the assumption

$$R \leq \left\lceil \frac{L-1}{L-Q} \right\rceil, \tag{21}$$

which trivially leads to a capacity pre-log lower bound as capacity is a nondecreasing function of R (one can always switch off receive antennas).

A capacity lower bound is trivially obtained by evaluating the mutual information in (4) for an appropriate input distribution. Specifically, we take i.i.d. $\mathsf{x}_i \sim \mathcal{CN}(0,1)$, $i \in [1:L]$. This implies that $h(\mathsf{x}_i) > -\infty, i \in [1:L]$, and, hence [19, Lem. 6.7],

$$\mathbb{E}[\log(|\mathsf{x}_i|)] > -\infty, \quad i \in [1:L]. \tag{22}$$

We point out that every input vector with i.i.d., zero mean, unit variance entries x_i that satisfy $h(\mathsf{x}_i) > -\infty, i \in [1:L]$, would allow us to prove (12). The choice $\mathsf{x}_i \sim \mathcal{CN}(0,1)$ is made for concreteness and convenience.

C. Step 2: Mutual information decomposition

Decompose

$$I(\mathbf{x}; \mathbf{y}) = h(\mathbf{y}) - h(\mathbf{y} \mid \mathbf{x}) \tag{23}$$

and separately bound the two differential entropy terms for the input distribution chosen in Step 1.

D. Step 3: Analysis of $h(\mathbf{y} \mid \mathbf{x})$

As $\mathbf{y}$ conditioned on $\mathbf{x}$ is JPG, the conditional differential entropy $h(\mathbf{y} \mid \mathbf{x})$ can be upper-bounded in a straightforward

manner as follows:

$$
\begin{aligned}
h(\mathbf{y} \mid \mathbf{x}) &= RL \log(\pi e) \\
&+ \mathbb{E}_{\mathbf{x}}\big[\log\det\big(\mathrm{I}_{RL} + \rho\,(\mathrm{I}_R \otimes \mathbf{X}\mathrm{D})\,\mathbb{E}_{\mathbf{s}}\big[\mathbf{ss}^{\mathsf{H}}\big]\,(\mathrm{I}_R \otimes \mathrm{D}^{\mathsf{H}}\mathbf{X}^{\mathsf{H}})\big)\big] \\
&= RL\log(\pi e) + R\,\mathbb{E}_{\mathbf{x}}\big[\log\det\big(\mathrm{I}_L + \rho\big(\mathbf{X}\mathrm{D}\mathrm{D}^{\mathsf{H}}\mathbf{X}^{\mathsf{H}}\big)\big)\big] \\
&= RL\log(\pi e) + R\,\mathbb{E}_{\mathbf{x}}\big[\log\det\big(\mathrm{I}_Q + \rho\big(\mathrm{D}^{\mathsf{H}}\mathbf{X}^{\mathsf{H}}\mathbf{X}\mathrm{D}\big)\big)\big] \\
&\overset{(a)}{\leq} RL\log(\pi e) + R\log\det\big(\mathrm{I}_Q + \rho\big(\mathrm{D}^{\mathsf{H}}\,\mathbb{E}_{\mathbf{x}}[\mathbf{X}^{\mathsf{H}}\mathbf{X}]\,\mathrm{D}\big)\big) \\
&= RL\log(\pi e) + R\sum_{i=1}^{Q}\log\big(1 + \rho\lambda_i\big(\mathrm{D}^{\mathsf{H}}\mathrm{D}\big)\big) \\
&\overset{(b)}{\leq} RQ\log(\rho) + \mathcal{O}(1). \tag{24}
\end{aligned}
$$

Here, (a) follows from Jensen's inequality, and (b) holds because D has rank Q and, therefore, $\lambda_i\big(\mathrm{D}^{\mathsf{H}}\mathrm{D}\big) > 0$ for all $i \in [1:Q]$.

E. Step 4: Splitting $h(\mathbf{y})$ into three terms

Finding an asymptotically (in SNR) tight lower bound on $h(\mathbf{y})$ is the main technical challenge of the proof of Theorem 1. The back-of-the-envelope calculation presented in Section III suggests that the problem can be approached by splitting $h(\mathbf{y})$ into a term that depends on the *noiseless* channel output $\hat{\mathbf{y}} = (\mathrm{I}_R \otimes \mathbf{X}\mathrm{D})\,\mathbf{s}$ only and a term that depends on noise $\mathbf{w}$ only. This can be realized as follows.

Consider a set of indices $\mathcal{I} \subseteq [1:LR]$ (we shall later discuss how to choose $\mathcal{I}$) and define the following projection matrices

$$
\begin{aligned}
\mathrm{P} &\triangleq (\mathrm{I}_{LR})_{\mathcal{I}} \\
\mathrm{Q} &\triangleq (\mathrm{I}_{LR})_{[1:LR]\setminus\mathcal{I}}.
\end{aligned}
$$

We can lower-bound $h(\mathbf{y})$ according to

$$
\begin{aligned}
h(\mathbf{y}) &= h(\mathrm{P}\mathbf{y}, \mathrm{Q}\mathbf{y}) \\
&\overset{(a)}{=} h(\mathrm{P}\mathbf{y}) + h(\mathrm{Q}\mathbf{y} \mid \mathrm{P}\mathbf{y}) \\
&\overset{(b)}{\geq} h(\sqrt{\rho}\,\mathrm{P}\hat{\mathbf{y}} + \mathrm{P}\mathbf{w} \mid \mathrm{P}\mathbf{w}) + h(\mathrm{Q}\hat{\mathbf{y}} + \mathrm{Q}\mathbf{w} \mid \mathrm{Q}\hat{\mathbf{y}}, \mathrm{P}\mathbf{y}) \\
&\overset{(c)}{=} h(\sqrt{\rho}\,\mathrm{P}\hat{\mathbf{y}}) + h(\mathrm{Q}\mathbf{w} \mid \mathrm{P}\mathbf{y}) \\
&\overset{(d)}{=} h(\sqrt{\rho}\,\mathrm{P}\hat{\mathbf{y}}) + h(\mathrm{Q}\mathbf{w}) \\
&\overset{(e)}{=} |\mathcal{I}|\log(\rho) + h(\mathrm{P}\hat{\mathbf{y}}) + c. \tag{25}
\end{aligned}
$$

Here, (a) follows by the chain rule for differential entropy; (b) follows from (3), (6), and because conditioning reduces entropy; (c) follows because differential entropy is invariant under translations and because $\mathbf{w}$ and $\hat{\mathbf{y}}$ are independent; (d) follows because $\mathrm{Q}\mathbf{w}$ and $\mathrm{P}\mathbf{y}$ are independent; and in (e) we used the fact that $\mathrm{P}\hat{\mathbf{y}}$ is a $|\mathcal{I}|$-dimensional vector and $h(\mathrm{Q}\mathbf{w}) = c$, where c here and in what follows denotes a constant that is independent of ρ and can take a different value at each appearance.

Through this chain of inequalities, we disposed of noise $\mathbf{w}$ and isolated SNR-dependence into a separate term. This corresponds to considering the noise-free IO relation (6) in the back-of-the-envelope calculation. Note further that we also rid ourselves of the components of $\hat{\mathbf{y}}$ indexed by $[1:LR] \setminus \mathcal{I}$; this corresponds to eliminating unnecessary equations in the back-of-the-envelope calculation. The specific choice of the set $\mathcal{I}$ is crucial and will be discussed next.

F. Step 5: Analysis of the SNR-dependent term in (25)

If $h(P\hat{y}) > -\infty$, we can substitute (25) and (24) into (23) which then yields a capacity lower bound of the form

$$C(\rho) \geq \frac{|\mathcal{I}| - RQ}{L} \log(\rho) + \mathcal{O}(1). \qquad (26)$$

This bound needs to be tightened by choosing the set $\mathcal{I}$ such that $|\mathcal{I}|$ is as large as possible while guaranteeing $h(P\hat{y}) > -\infty$. Comparing the lower bound (26) to the upper bound (11) we see that the bounds match if

$$|\mathcal{I}| = \min[RQ + L - 1, RL]. \qquad (27)$$

Condition (27) dictates that for $RL \leq RQ + L - 1$ we must set $\mathcal{I} = [1:RL]$, which yields $P\hat{y} = \hat{y}$. When $RL > RQ + L - 1$ the set $\mathcal{I}$ must be a proper subset of $[1:RL]$. Specifically, we shall choose $\mathcal{I}$ as follows. Set

$$\tilde{R} = \begin{cases} R(L-Q) - (L-1), & \text{if } RL > RQ + L - 1 \\ 0, & \text{if } RL \leq RQ + L - 1, \end{cases} \qquad (28)$$

let

$$\mathcal{I}_r = \begin{cases} [(r-1)L + 1 : rL - 1], & 1 \leq r \leq \tilde{R} \\ [(r-1)L + 1 : rL], & \tilde{R} + 1 \leq r \leq R, \end{cases}$$

and define $\mathcal{I} \triangleq \bigcup_{r=1}^{R} \mathcal{I}_r$.

This choice can be verified to satisfy (27). Obviously, this is not the only choice for $\mathcal{I}$ that satisfies (27). The specific set $\mathcal{I}$ chosen here will be seen to guarantee $h(P\hat{y}) > -\infty$ and at the same time simplify the calculations in Section VI-I.

Substituting (27) into (26), we obtain the desired result (12), provided that $h(P\hat{y}) > -\infty$. Establishing that $h(P\hat{y}) > -\infty$ is, as already mentioned, the major technical difficulty in the proof of Theorem 1 and will be addressed next.

G. Step 6: Analysis of $h(P\hat{y})$ through change of variables

It is difficult to analyze $h(P\hat{y})$ directly since $\hat{y} = (I_R \otimes XD) s$ depends on the pair of variables (s, x) in a nonlinear fashion. We have seen, in Section III, that (6) has a unique solution in (s, x), provided that the appropriate number of pilot symbols is used. This suggests that there must be a one-to-one correspondence between $P\hat{y}$ and the pair (s, x). The existence of such a one-to-one correspondence allows us to locally linearize the equation $\hat{y} = (I_R \otimes XD) s$ and to relate $h(P\hat{y})$ to $h(s, x) = h(s) + h(x)$. This idea is key to bringing $h(P\hat{y})$ into a form that eventually allows us to conclude that $h(P\hat{y}) > -\infty$.

Formally, it is possible to relate the differential entropies of two random vectors of the same dimension that are related by a *deterministic one-to-one* function (in the sense of [21, p.7]) according to the following lemma.

Lemma 3 (Transformation of differential entropy). *Assume that* $g : \mathbb{C}^N \to \mathbb{C}^N$ *is a continuous vector-valued function that is one-to-one and differentiable almost everywhere (a.e.) on* $\mathbb{C}^N$. *Let* $u \in \mathbb{C}^N$ *be a continuous [17, Sec. 2-3, Def. (2)] random vector (i.e., it has a well-defined PDF) and let* $v = g(u)$. *Then*

$$h(v) = h(u) + 2\, \mathbb{E}_u[\log|\det(\partial g / \partial u)|],$$

where $\partial g / \partial u$ is the Jacobian of the function $g(\cdot)$.

The proof follows from the change-of-variables theorem for integrals [21, Thm. 7.26] and is given in Appendix B for completeness since the version of the theorem for complex-valued functions does not seem to be well documented in the literature.

Note that $P\hat{y} \in \mathbb{C}^{|\mathcal{I}|}$ with $|\mathcal{I}|$ given in (27) and $[s^\mathsf{T} \ x^\mathsf{T}]^\mathsf{T} \in \mathbb{C}^{RQ+L}$. Since $|\mathcal{I}| < RQ + L$ (see (27)), the vectors $P\hat{y}$ and $[s^\mathsf{T} \ x^\mathsf{T}]^\mathsf{T}$ are of different dimensions and Lemma 3 can therefore not be applied directly to relate $h(P\hat{y})$ to $h(s, x)$. This problem can be resolved by conditioning on a subset $\mathcal{P} \subset [1:L]$ (specified below) of components of x according to

$$h(P\hat{y}) \geq h(P\hat{y} \,|\, x_\mathcal{P}). \qquad (29)$$

The components $x_\mathcal{P}$ correspond to the pilot symbols in the back-of-the-envelope calculation. The set $\mathcal{P}$ is chosen such that (i) the set of remaining components in x, $\mathcal{J} = [1:L] \setminus \mathcal{P}$, is of appropriate size ensuring that $P\hat{y}$ and $[s^\mathsf{T} \ x_\mathcal{J}^\mathsf{T}]^\mathsf{T}$ are of the same dimension, and (ii) $P\hat{y}$ and $[s^\mathsf{T} \ x_\mathcal{J}^\mathsf{T}]^\mathsf{T}$ are related by a deterministic bijection so that Lemma 3 can be applied to relate $h(P\hat{y} \,|\, x_\mathcal{P})$ to $h(s, x_\mathcal{J} \,|\, x_\mathcal{P})$. Specifically, set

$$\alpha = \max[1, RQ + L - RL], \qquad (30)$$

let $\mathcal{P} \triangleq [1:\alpha]$, which implies $\mathcal{J} = [\alpha + 1 : L]$. Observe that $P\hat{y}$ (conditioned on $x_\mathcal{P}$) depends only on $[s^\mathsf{T} \ x_\mathcal{J}^\mathsf{T}]^\mathsf{T}$, and due to our choice of $\mathcal{J}$ (it is actually the choice of $|\mathcal{J}|$ that is important here), the vectors $P\hat{y}$ and $[s^\mathsf{T} \ x_\mathcal{J}^\mathsf{T}]^\mathsf{T}$ are of the same dimension. Furthermore, these two vectors are related through a deterministic bijection: Consider the vector-valued function $g_{x_\mathcal{P}} : \mathbb{C}^{|\mathcal{I}|} \to \mathbb{C}^{|\mathcal{I}|}$

$$g_{x_\mathcal{P}}(s, x_\mathcal{J}) = P(I_R \otimes XD)s. \qquad (31)$$

Here, and whenever we refer to the function $g_{x_\mathcal{P}}(\cdot)$ in the following, we use the convention that the parameter vector $x_\mathcal{P} \in \mathbb{C}^{|\mathcal{P}|}$ and the variable vector $x_\mathcal{J} \in \mathbb{C}^{|\mathcal{J}|}$ are stacked into the vector $x \triangleq [x_\mathcal{P}^\mathsf{T} \ x_\mathcal{J}^\mathsf{T}]^\mathsf{T}$ and we set $X \triangleq \text{diag}(x)$.

Lemma 4. *If $x_\mathcal{P}$ has nonzero components only, i.e., $x_i \neq 0$ for all $i \in \mathcal{P}$, then the function $g_{x_\mathcal{P}}(\cdot)$ is one-to-one a.e. on $\mathbb{C}^{|\mathcal{I}|}$.*

The proof of Lemma 4 is given in Appendix C and is based on the results obtained later in this section. We therefore invite the reader to first study the remainder of Section V and to return to Appendix C afterwards.

Recall that $P\hat{y} = P(I_R \otimes XD)s$ and hence $P\hat{y} = g_{x_\mathcal{P}}(s, x_\mathcal{J})$. Therefore, it follows from Lemma 4 that as long as $x_\mathcal{P} = x_\mathcal{P}$ is fixed and satisfies $x_i \neq 0$, for all $i \in \mathcal{P}$, $P\hat{y}$ and $[s^\mathsf{T} \ x_\mathcal{J}^\mathsf{T}]^\mathsf{T}$ are related through the bijection $g_{x_\mathcal{P}}(\cdot)$ as claimed.

Comments: A few comments on Lemma 4 are in order. For $L = 3$ and $R = Q = 2$ as in the simple example in Section III, we see from (27) that $\mathcal{I} = [1:RL]$ so that $P = I_{RL}$ and $P\hat{y} = \hat{y}$. Further, for this example, it follows from (30) that $\alpha = 1$ and hence $\mathcal{P} = \{1\}$ and $\mathcal{J} = \{2, 3\}$. Therefore, Lemma 4 simply says that (6) has a unique solution for fixed $x_1 \neq 0$. As already mentioned, conditioning w.r.t. $x_\mathcal{P} = x_1$ in (29) in order to make the relation between $P\hat{y}$ and $[s^\mathsf{T} \ x_\mathcal{J}^\mathsf{T}]^\mathsf{T}$ be one-to-one corresponds to transmitting a pilot symbol, as was done in the back-of-the-envelope calculation by setting $x_1 = 1$.

We can now use Lemma 3 to relate $h(\mathrm{P}\hat{\mathbf{y}} \mid \mathbf{x}_\mathcal{P})$ to $h(\mathbf{s}, \mathbf{x}_\mathcal{J})$ as follows. Let $f_{\mathbf{x}_\mathcal{P}}(\cdot)$ denote the PDF of $\mathbf{x}_\mathcal{P}$. Then, we can write

$$h(\mathrm{P}\hat{\mathbf{y}} \mid \mathbf{x}_\mathcal{P}) = \int f_{\mathbf{x}_\mathcal{P}}(\mathrm{x}_\mathcal{P}) h(\mathrm{P}\hat{\mathbf{y}} \mid \mathbf{x}_\mathcal{P} = \mathrm{x}_\mathcal{P}) \, d\mathrm{x}_\mathcal{P}. \quad (32)$$

Let

$$\mathrm{J}(\mathrm{s}, \mathrm{x}) \triangleq \frac{\partial \mathrm{g}_{\mathrm{x}_\mathcal{P}}}{\partial(\mathrm{s}, \mathrm{x}_\mathcal{J})} \quad (33)$$

be the Jacobian of the mapping in (31) (where we again use the convention $\mathrm{x} = [\mathrm{x}_\mathcal{P}^\mathsf{T} \, \mathrm{x}_\mathcal{J}^\mathsf{T}]^\mathsf{T}$). Applying Lemma 3 to $h(\mathrm{P}\hat{\mathbf{y}} \mid \mathbf{x}_\mathcal{P} = \mathrm{x}_\mathcal{P})$, we get for all $\mathrm{x}_\mathcal{P}$ with $x_i \neq 0$, $i \in \mathcal{P}$, that

$$h(\mathrm{P}\hat{\mathbf{y}} \mid \mathbf{x}_\mathcal{P} = \mathrm{x}_\mathcal{P}) = h(\mathbf{s}, \mathbf{x}_\mathcal{J} \mid \mathbf{x}_\mathcal{P} = \mathrm{x}_\mathcal{P})$$
$$+ 2\,\mathbb{E}_{\mathbf{s},\mathbf{x}_\mathcal{J}}\Big[\log(|\det \mathrm{J}(\mathbf{s},\mathbf{x})|) \,\Big|\, \mathbf{x}_\mathcal{P} = \mathrm{x}_\mathcal{P}\Big]. \quad (34)$$

Substituting (34) into (32), we finally obtain

$$h(\mathrm{P}\hat{\mathbf{y}} \mid \mathbf{x}_\mathcal{P})$$
$$\overset{(a)}{=} \int f_{\mathbf{x}_\mathcal{P}}(\mathrm{x}_\mathcal{P}) h(\mathbf{s}, \mathbf{x}_\mathcal{J} \mid \mathbf{x}_\mathcal{P} = \mathrm{x}_\mathcal{P}) \, d\mathrm{x}_\mathcal{P}$$
$$+ 2\int f_{\mathbf{x}_\mathcal{P}}(\mathrm{x}_\mathcal{P}) \mathbb{E}_{\mathbf{s},\mathbf{x}_\mathcal{J}}\Big[\log(|\det \mathrm{J}(\mathbf{s},\mathbf{x})|) \,\Big|\, \mathbf{x}_\mathcal{P} = \mathrm{x}_\mathcal{P}\Big] d\mathrm{x}_\mathcal{P}$$
$$= h(\mathbf{s}, \mathbf{x}_\mathcal{J} \mid \mathbf{x}_\mathcal{P}) + 2\,\mathbb{E}_{\mathbf{s},\mathbf{x}}[\log(|\det(\mathrm{J}(\mathbf{s},\mathbf{x}))|)]. \quad (35)$$

Here, in (a), to be able to use (34), we exclude the set $\{\mathrm{x}_\mathcal{P} | x_i = 0 \text{ for at least one } i \in \mathcal{P}\}$ from the domain of integration. This is legitimate since that set has measure zero.

The first term on the right-hand side (RHS) of (35) satisfies

$$h(\mathbf{s}, \mathbf{x}_\mathcal{J} \mid \mathbf{x}_\mathcal{P}) \overset{(a)}{=} h(\mathbf{s} \mid \mathbf{x}_\mathcal{P}) + h(\mathbf{x}_\mathcal{J} \mid \mathbf{s}, \mathbf{x}_\mathcal{P})$$
$$\overset{(b)}{=} h(\mathbf{s}) + h(\mathbf{x}_\mathcal{J}) \overset{(c)}{=} c, \quad (36)$$

where (a) follows by the chain rule for differential entropy; in (b) we used that $\mathbf{x}$ is independent of $\mathbf{s}$, and $\mathbf{x}_\mathcal{P}$ is independent of $\mathbf{x}_\mathcal{J}$ because the x_i, $i \in [1\!:\!L]$, are i.i.d. and $\mathcal{J} \cap \mathcal{P} = \emptyset$; and (c) follows because the x_i, $i \in \mathcal{J}$, and the s_i, $i \in [1\!:\!RQ]$, are i.i.d. and have finite differential entropy, by assumption.

Combining (36), (35), and (29), we obtain

$$h(\mathrm{P}\hat{\mathbf{y}}) \geq c + 2\,\mathbb{E}_{\mathbf{s},\mathbf{x}}[\log(|\det(\mathrm{J}(\mathbf{s},\mathbf{x}))|)].$$

To show that $h(\mathrm{P}\hat{\mathbf{y}}) > -\infty$, it therefore remains to prove that

$$\mathbb{E}_{\mathbf{s},\mathbf{x}}[\log(|\det(\mathrm{J}(\mathbf{s},\mathbf{x}))|)] > -\infty. \quad (37)$$

This requires an in-depth analysis of the structure of $|\det(\mathrm{J}(\cdot))|$, which will be carried out in the next section.

H. Step 7: Factorization of $\det(\mathrm{J}(\cdot))$ *and analysis of* $\mathbf{x}$-*dependent terms*

The following lemma shows that the determinant of the Jacobian in (33) can be factorized into a product of simpler terms.

Lemma 5. *The determinant of the Jacobian in* (33) *factorizes as*

$$\det(\mathrm{J}(\mathbf{s},\mathbf{x})) = \det(\mathrm{J}_1(\mathbf{x}))\det(\mathrm{J}_2(\mathbf{s}))\det(\mathrm{J}_3(\mathbf{x}_\mathcal{J})),$$

where

$$\mathrm{J}_1(\mathrm{x}) \triangleq \mathrm{P}(\mathrm{I}_R \otimes \mathrm{X})\mathrm{P}^\mathsf{T}$$
$$\mathrm{J}_2(\mathrm{s}) \triangleq \mathrm{P}[\mathrm{I}_R \otimes \mathrm{D} \mid a_{\alpha+1} \mid \cdots \mid a_L] \quad (38)$$
$$\mathrm{J}_3(\mathrm{x}_\mathcal{J}) \triangleq \mathrm{diag}(\mathrm{I}_{RQ}, (\mathrm{diag}(\mathrm{x}_\mathcal{J}))^{-1})$$

with

$$a_i \triangleq (\mathrm{I}_R \otimes \mathrm{diag}(e_i)\mathrm{D})\mathrm{s}, \quad i \in \mathcal{J} = [1\!:\!L]. \quad (39)$$

Proof: First note that $\mathrm{g}_{\mathrm{x}_\mathcal{P}}(\mathrm{s}, \mathrm{x}_\mathcal{J})$ in (31) can be written as

$$\mathrm{g}_{\mathrm{x}_\mathcal{P}}(\mathrm{s}, \mathrm{x}_\mathcal{J}) = \sum_{j \in [1:L]} x_j(\mathrm{I}_R \otimes \mathrm{diag}(e_j)\mathrm{D})\mathrm{s}$$

and, therefore,

$$\frac{\partial \mathrm{g}_{\mathrm{x}_\mathcal{P}}}{\partial x_i} = \frac{\partial}{\partial x_i}\Big(\sum_{j \in [1:L]} x_j(\mathrm{I}_R \otimes \mathrm{diag}(e_j)\mathrm{D})\mathrm{s}\Big)$$
$$= a_i, \quad i \in \mathcal{J}.$$

With

$$\frac{\partial \mathrm{g}_{\mathrm{x}_\mathcal{P}}}{\partial \mathrm{s}} = \mathrm{I}_R \otimes \mathrm{XD}$$

we can now rewrite the Jacobian in (33) as

$$\mathrm{J}(\mathrm{s},\mathrm{x}) = \mathrm{P}[\mathrm{I}_R \otimes \mathrm{XD} \mid a_{\alpha+1} \mid \cdots \mid a_L]$$
$$= (\mathrm{P}(\mathrm{I}_R \otimes \mathrm{X})\mathrm{P}^\mathsf{T})\,\mathrm{J}_2(\mathrm{s})\,\mathrm{diag}(\mathrm{I}_{RQ}, (\mathrm{diag}(\mathrm{x}_\mathcal{J}))^{-1}), \quad (40)$$

which concludes the proof. ∎

Using Lemma 5, we can rewrite the second term on the RHS of (35) according to

$$\mathbb{E}_{\mathbf{s},\mathbf{x}}[\log(|\det(\mathrm{J}(\mathbf{s},\mathbf{x}))|)] = \mathbb{E}[\log(|\det(\mathrm{J}_1(\mathbf{x}))|)]$$
$$+ \mathbb{E}[\log(|\det(\mathrm{J}_2(\mathbf{s}))|)]$$
$$+ \mathbb{E}[\log(|\det(\mathrm{J}_3(\mathbf{x}_\mathcal{J}))|)]. \quad (41)$$

The first and the third term in (41) can be expanded as

$$\mathbb{E}[\log(|\det(\mathrm{J}_1(\mathbf{x}))|)] = \tilde{R}\sum_{j=1}^{L-1}\mathbb{E}[\log(|x_j|)]$$
$$+ (R - \tilde{R})\sum_{j=1}^{L}\mathbb{E}[\log(|x_j|)] \quad (42)$$

$$\mathbb{E}[\log(|\det(\mathrm{J}_3(\mathbf{x}_\mathcal{J}))|)] = -\sum_{j \in \mathcal{J}}\mathbb{E}[\log(|x_j|)]. \quad (43)$$

Using (22), (5), and Jensen's inequality, we have

$$-\infty < \mathbb{E}[\log(|x_j|)] \leq \log(\mathbb{E}[|x_j|]) < \infty,$$

which immediately implies that the terms on the left-hand side (LHS) of (42) and (43) are finite. It remains to show that $\mathbb{E}\big[\log(|\det(\mathrm{J}_2(\mathbf{s}))|)\big] > -\infty$.

I. Step 8: Proving $\mathbb{E}\left[\log\left(\left|\det(J_2(s))\right|\right)\right] > -\infty$ *through resolution of singularities*

This is the most technical part of the proof of Theorem 1. We need to show that

$$
\begin{aligned}
&\mathbb{E}[\log(|\det(J_2(\mathbf{s}))|)] \\
&= \frac{1}{\pi^{RQ}} \int_{\mathbb{C}^{RQ}} \exp(-\|\mathbf{s}\|^2) \log(|\det(J_2(s))|) ds > -\infty. \quad (44)
\end{aligned}
$$

Since $J_2(\cdot)$ is a large matrix with little structure to exploit, a direct evaluation of the integral in (44) seems daunting. Note, however, that by (38), (39), and [22, 4.2.1(2)] it follows that $\det(J_2(s))$ is a homogeneous polynomial in $s_1, \ldots, s_{RQ}$; in other words $\det(J_2(\cdot))$ is a well-behaved function of its arguments. It turns out that this mild property is sufficient to prove the inequality in (44). The proof, however, requires powerful tools, which will be described next.

Lemma 6. *Let* $p(u)$, $u \in \mathbb{C}^N$, *be a homogeneous polynomial in* $u_1, \ldots, u_N$. *Then,* $p(\cdot) \not\equiv 0$ *implies that*

$$
\int_{\mathbb{C}^N} \exp(-\|u\|^2) \log(|p(u)|) du > -\infty.
$$

Lemma 6 is proved in Appendix D using the following general result, which is a consequence of Hironaka's Theorem on the Resolution of Singularities [11, Theorem 2.3].

Theorem 7. *Let* $f(\cdot) \not\equiv 0$ *be a real analytic function[5] [14, Def. 2.2.1] on an open set* $\Omega \subset \mathbb{R}^K$. *Then*

$$
\int_\Delta |\log(|f(u)|)| du < \infty \quad (45)
$$

for all compact sets $\Delta \subset \Omega$.

For a formal proof of Theorem 7 see Appendix E. Here, we explain intuitively why this result holds. The only reason why the integral in (45) could diverge, is because $|f(\cdot)|$ may take on the value zero and $\log(0) = -\infty$. Since $f(\cdot)$ is a real analytic function and since $f(\cdot) \not\equiv 0$, the zero set $f^{-1}(\{0\})$ has measure zero. To prove (45), it remains to examine the detailed behavior of $f(\cdot)$ around the zero set $f^{-1}(\{0\})$. The integral of $|\log(|f(\cdot)|)|$ over a small enough neighborhood around each smooth (i.e. nonsingular) point in the zero set is bounded, but it is difficult to determine what happens near the singularities. Hironaka's Theorem on the Resolution of Singularities "untangles" the singularities so that we can understand their structure. More formally, Hironaka's Theorem states that in a small neighborhood around every point in $f^{-1}(\{0\})$, the real analytic function $f(\cdot)$ behaves like a product of a monomial of finite degree and a *nonvanishing* real analytic function. The integral of the logarithm of the absolute value of this product over a small enough neighborhood around each point in $f^{-1}(\{0\})$ is then easily bounded and turns out to be finite. The union of the neighborhoods of the points in $f^{-1}(\{0\})$ forms an open cover for $f^{-1}(\{0\})$. Since Δ is a compact set, it is possible to find a finite subcover for $f^{-1}(\{0\})$. Summing up the integrals over the elements of this subcover, each of which is finite as explained

above, allows us to deduce that the integral in (45) must be finite as well.

On account of Lemma 6, to show (44) it suffices to verify that $\det(J_2(\cdot)) \not\equiv 0$. This is indeed the case as demonstrated next.

J. Step 9: Identifying an s *for which* $\det(J_2(s)) \neq 0$

Lemma 8. *Property (A) in Theorem 1 implies that* $\det(J_2(\cdot)) \not\equiv 0$.

Proof: The proof is effected by showing that Property (A) implies the existence of a vector $s \in \mathbb{C}^{RQ}$ such that $\det(J_2(s)) \neq 0$. To this end, we first note that $J_2(s)$ in (38) can be written as $J_2(s) = [P\,(I_R \otimes D) \quad A]$ with

$$
A \triangleq \begin{bmatrix} A_1 \\ \vdots \\ A_R \end{bmatrix} \quad (46)
$$

and

$$
A_i \triangleq \begin{pmatrix} 0_\alpha & \cdots & 0_\alpha & 0_\alpha \\ \bar{d}_{\alpha+1}^\mathsf{T} s_i & \cdots & 0 & 0 \\ \vdots & \ddots & 0 & 0 \\ 0 & \cdots & \bar{d}_{L-1}^\mathsf{T} s_i & 0 \end{pmatrix}, \quad i \in [1:\tilde{R}],
$$

$$
A_i \triangleq \begin{pmatrix} 0_\alpha & \cdots & 0_\alpha & 0_\alpha \\ \bar{d}_{\alpha+1}^\mathsf{T} s_i & \cdots & 0 & 0 \\ \vdots & \ddots & 0 & 0 \\ 0 & \cdots & \bar{d}_{L-1}^\mathsf{T} s_i & 0 \\ 0 & \cdots & 0 & \bar{d}_L^\mathsf{T} s_i \end{pmatrix}, \quad i \in [\tilde{R}+1:R].
$$

Here, α was defined in (30); 0_α denotes an all-zero vector of dimension α; $\bar{d}_1, \ldots, \bar{d}_L \in \mathbb{C}^Q$ are the transposed rows of the $L \times Q$ matrix D so that $D^\mathsf{T} = [\bar{d}_1 \cdots \bar{d}_L]$; and the $s_i \in \mathbb{C}^Q$, $i \in [1:R]$, are defined through $s \triangleq [s_1^\mathsf{T} \cdots s_R^\mathsf{T}]^\mathsf{T}$. The calculations below are somewhat tedious but the idea is simple. Thanks to Property (A) in Theorem 1, it is possible to find vectors $s_i \in \mathbb{C}^Q$, $i \in [1:R]$, such that each column of the matrix A defined in (46) has exactly one nonzero element. For this choice of $s_i \in \mathbb{C}^Q$, $i \in [1:R]$, we can then conveniently factorize $|J_2(s)|$ using the Laplace formula [23, p. 7]; the resulting factors are easily seen to all be nonzero. We next detail the program just outlined.

Take an $i \in [1:R]$ and consider a set $\mathcal{K}_i$ satisfying

$$
\mathcal{K}_i \subseteq \begin{cases} [\alpha+1:L-1], & \text{if } i \in [1:\tilde{R}], \\ [\alpha+1:L], & \text{if } i \in [\tilde{R}+1:R], \end{cases} \quad (47)
$$

with

$$
|\mathcal{K}_i| = \begin{cases} Q-1, & \text{if } RL > RQ+L-1 \\ (R-1)(L-Q), & \text{if } RL \leq RQ+L-1. \end{cases} \quad (48)
$$

The freedom in choice of the set $\mathcal{K}_i$ will be used later to ensure that each column of the matrix A has exactly one nonzero element. We shall next show that the vector $s_i \in \mathbb{C}^Q$ can be chosen such that the entries of A_i given by $\bar{d}_j^\mathsf{T} s_i$, $j \in \mathcal{K}_i$, equal zero and the entries $\bar{d}_j^\mathsf{T} s_i$, $j \notin \mathcal{K}_i$, are nonzero. Since, by (48), $|\mathcal{K}_i| \leq Q-1$, Property (A) in Theorem 1 guarantees that the

---

[5]Let Ω be an open subset of $\mathbb{R}^K$. A function $f(\cdot) : \Omega \to \mathbb{R}$ is real analytic if for every $x_0 \in \Omega$, $f(\cdot)$ can be represented by a convergent power series in some neighborhood of x_0.
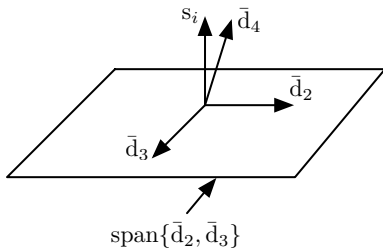
Fig. 2. Choice of the vector s_i for $L = 4$, $Q = 3$, $\alpha = 1$, $\mathcal{K}_i = \{2, 3\}$, $\mathcal{K}_i^c = \{4\}$.

vectors $\{\bar{d}_j\}_{j \in \mathcal{K}_i}$ are linearly independent. Furthermore, the vectors $\bar{d}_j$, $j \in \mathcal{K}_i^c$, with

$$\mathcal{K}_i^c \triangleq \begin{cases} [\alpha + 1 : L - 1] \setminus \mathcal{K}_i, & \text{if } i \in [1 : \tilde{R}], \\ [\alpha + 1 : L] \setminus \mathcal{K}_i, & \text{if } i \in [\tilde{R} + 1 : R], \end{cases} \quad (49)$$

do not belong to $\text{span}\{\bar{d}_j\}_{j \in \mathcal{K}_i}$. Hence, we can find a vector $s_i \in \mathbb{C}^Q$ such that

(a) $\bar{d}_j^\mathsf{T} s_i = 0$ for all $j \in \mathcal{K}_i$;

(b) $\bar{d}_j^\mathsf{T} s_i \neq 0$ for all $j \in \mathcal{K}_i^c$.

Geometrically, this simply means that s_i must be chosen such that it is orthogonal to $\text{span}\{\bar{d}_j\}_{j \in \mathcal{K}_i}$ (which is a subspace of $\mathbb{C}^Q$ of dimension less than or equal to $Q - 1$) and, in addition, is not orthogonal to every vector in the set $\{\bar{d}_j\}_{j \in \mathcal{K}_i^c}$ (see Fig. 2). Note that if Property (A) in Theorem 1 were not satisfied, we could have a vector $\bar{d}_{j'}$, $j' \in \mathcal{K}_i^c$, that belongs to the $\text{span}\{\bar{d}_j\}_{j \in \mathcal{K}_i}$; in this case there would not exist a vector s_i that satisfies (a) and (b) simultaneously. Based on (30), (48), and (49), we can see that if the vector s_i is chosen such that conditions (a) and (b) above are satisfied, the number of nonzero elements, $|\mathcal{K}_i^c|$, in the matrix A_i is [see (28)]

$$|\mathcal{K}_i^c| = \begin{cases} L - Q - 1, & \text{if } i \in [1 : \tilde{R}], \\ L - Q, & \text{if } i \in [\tilde{R} + 1 : R]. \end{cases}$$

Hence, applying the procedure described above to every $i \in [1 : R]$ and choosing the corresponding vector s_i such that (a) and (b) are satisfied, we obtain a matrix A [see (46)] with total number of nonzero elements equal to the number of columns in A and given by

$$\sum_{i \in [1 : R]} |\mathcal{K}_i^c| = L - \alpha.$$

Now, recall that we have full freedom in our choice of $\mathcal{K}_i$, $i \in [1 : R]$, as long as (47) and (48) are satisfied; this implies that we have control over the locations of the nonzero elements of A. Hence, by appropriate choice of the sets $\mathcal{K}_i$, $i \in [1 : R]$, we can ensure that each column of A contains precisely one nonzero element.

Applying the Laplace formula [23, p. 7] iteratively, we then get

$$|\det(J_2(s))| = c \prod_{i \in [1 : R]} |\det(D_{\mathcal{K}_i \cup [1 : \alpha]})|, \quad (50)$$

where c is a positive constant. Finally, since for every $i \in [1 : R]$, $D_{\mathcal{K}_i \cup [1 : \alpha]}$ is a $Q \times Q$ submatrix of D [see (30) and (48)], it

follows from Property (A) in Theorem 1 that $D_{\mathcal{K}_i \cup [1 : \alpha]}$ has linearly independent rows and hence $|\det(D_{\mathcal{K}_i \cup [1 : \alpha]})| > 0$, for all $i \in [1 : R]$, which by (50) concludes the proof. ∎

The proof of Theorem 1 is now completed as follows. Combining Lemmas 6 and 8, we conclude that (44) holds. Substituting (44) into (41) and using (42) and (43), we conclude that (37) holds. Therefore, by (29), (35), and (36), it follows that $h(P\hat{y}) > -\infty$.

VII. Conclusions and Future Work

We characterized the capacity pre-log of a temporally correlated block-fading SIMO channel in the noncoherent setting under a mild assumption on the channel covariance matrix. The most striking implication of this result is that the pre-log penalty in the SISO case due to channel uncertainty can be made to vanish in the large block length regime by adding only one receive antenna.

It would be interesting to generalize the results in this paper to the MIMO case. Preliminary work in this direction was reported in [24], which establishes a lower bound on the capacity pre-log of a temporally correlated block-fading MIMO channel. This lower bound is not accompanied by a matching upper bound so that the problem of determining the capacity pre-log in the MIMO case remains open. It is also interesting to note that [24] avoids the use of Hironaka's theorem through an alternative proof technique based on properties of subharmonic functions.

Further interesting open questions include the generalization of the results in this paper to the stationary case and the development of coding schemes that achieve the SIMO capacity pre-log.

Appendix A
Proof of (16)

The following calculation repeats the steps in [19, Thm. 4.2] and is provided for the reader's convenience:

$$I(\mathbf{Y}_\mathcal{Q}; \mathbf{x}_\mathcal{Q}) = \sum_{q=1}^Q I(\mathbf{Y}_{\{q\}}; \mathbf{x}_\mathcal{Q} \mid \mathbf{Y}_{[1 : q-1]})$$

$$= \sum_{q=1}^Q \left(I(\mathbf{Y}_{\{q\}}; \mathbf{Y}_{[1 : q-1]}, \mathbf{x}_\mathcal{Q}) - I(\mathbf{Y}_{\{q\}}; \mathbf{Y}_{[1 : q-1]}) \right)$$

$$\leq \sum_{q=1}^Q I(\mathbf{Y}_{\{q\}}; \mathbf{Y}_{[1 : q-1]}, \mathbf{x}_\mathcal{Q})$$

$$= \sum_{q=1}^Q \left(I(\mathbf{Y}_{\{q\}}; \mathbf{Y}_{[1 : q-1]}, \mathbf{H}_{[1 : q-1]}, \mathbf{x}_\mathcal{Q}) \right.$$

$$\left. - I(\mathbf{Y}_{\{q\}}; \mathbf{H}_{[1 : q-1]} \mid \mathbf{Y}_{[1 : q-1]}, \mathbf{x}_\mathcal{Q}) \right)$$

$$\leq \sum_{q=1}^Q I(\mathbf{Y}_{\{q\}}; \mathbf{Y}_{[1 : q-1]}, \mathbf{H}_{[1 : q-1]}, \mathbf{x}_\mathcal{Q})$$

$$\overset{(a)}{=} \sum_{q=1}^Q I(\mathbf{Y}_{\{q\}}; \mathbf{H}_{[1 : q-1]}, \mathbf{x}_q)$$

$$= \sum_{q=1}^Q \left(I(\mathbf{Y}_{\{q\}}; \mathbf{H}_{[1 : q-1]} \mid \mathbf{x}_q) + I(\mathbf{Y}_{\{q\}}; \mathbf{x}_q) \right)$$

$$\overset{(b)}{\leq} \sum_{q=1}^{Q} I\big(\mathbf{Y}_{\{q\}}; \mathbf{H}_{[1:q-1]} \,|\, \mathsf{x}_q\big) + Q \log\log(\rho) + \mathcal{O}(1)$$

$$\overset{(c)}{=} \sum_{q=1}^{Q} I\big(\mathbf{Y}_{\{q\}}, \mathsf{x}_q; \mathbf{H}_{[1:q-1]}\big) + Q \log\log(\rho) + \mathcal{O}(1)$$

$$\leq \sum_{q=1}^{Q} I\big(\mathbf{Y}_{\{q\}}, \mathsf{x}_q, \mathbf{H}_{\{q\}}; \mathbf{H}_{[1:q-1]}\big) + Q \log\log(\rho) + \mathcal{O}(1)$$

$$\overset{(d)}{=} \sum_{q=1}^{Q} I\big(\mathbf{H}_{\{q\}}; \mathbf{H}_{[1:q-1]}\big) + Q \log\log(\rho) + \mathcal{O}(1)$$

$$= Q h\big(\mathbf{H}_{\{1\}}\big) - h(\mathbf{H}_{\mathcal{Q}}) + Q \log\log(\rho) + \mathcal{O}(1)$$

$$= Q \sum_{i=1}^{R} h(\mathsf{h}_{i1}) - R \log\det\big(\mathrm{D}_{\mathcal{Q}} \mathrm{D}_{\mathcal{Q}}^{\mathsf{H}}\big) + Q \log\log(\rho) + \mathcal{O}(1)$$

$$\overset{(e)}{=} Q \log\log(\rho) + \mathcal{O}(1)$$

where (a) follows because $\mathbf{Y}_{\{q\}}$ is conditionally independent of $\mathbf{x}_{[1:q-1]}$ and of $\mathbf{Y}_{[1:q-1]}$ given x_q and $\mathbf{H}_{[1:q-1]}$; (b) follows from [19, Th. 4.2]; (c) follows because x_q is independent of $\mathbf{H}_{[1:q-1]}$; (d) follows because $\mathbf{Y}_q$ and x_q are conditionally independent of $\mathbf{H}_{[1:q-1]}$ given $\mathbf{H}_q$; and (e) follows because the matrix $\mathrm{D}_{\mathcal{Q}}$ is full-rank and $h(\mathsf{h}_{i1}) = c$, $i \in [1:R]$.

APPENDIX B
PROOF OF LEMMA 3

The lemma is based on the change of variables theorem for integrals, which we restate for the reader's convenience.

Theorem 9. *[21, Thm. 7.26], [15, p. 31, Thm. 7.2] Assume that* $\mathrm{g} : \mathcal{U} \subset \mathbb{C}^N \to \mathbb{C}^N$ *is a continuous vector-valued function that is one-to-one and differentiable a.e. on $\mathcal{U}$. Let $\mathcal{V} = \mathrm{g}(\mathcal{U})$. Then,*

$$\int_{\mathcal{V}} f(\mathrm{v}) d\mathrm{v} = \int_{\mathcal{U}} f(\mathrm{g(u)}) |\det(\partial \mathrm{g}/\partial \mathrm{u})|^2 d\mathrm{u}$$

for every measurable $f : \mathbb{C}^N \to [0, \infty]$.

To prove Lemma 3, we let $f_{\mathbf{v}}(\cdot)$ and $f_{\mathbf{u}}(\cdot)$ denote the PDFs of random vectors $\mathbf{v}$ and $\mathbf{u}$, respectively. Then, according to [25, (7-8)] and [15, p.31, Thm. 7.2]

$$f_{\mathbf{v}}(\mathrm{g(u)}) = \frac{f_{\mathbf{u}}(\mathrm{u})}{|\det(\partial \mathrm{g}/\partial \mathrm{u})|^2}. \tag{51}$$

Next, let $\mathcal{U}$ and $\mathcal{V}$ denote the support of $f_{\mathbf{u}}(\cdot)$ and $f_{\mathbf{v}}(\cdot)$, respectively. Then, $\mathcal{V} = \mathrm{g}(\mathcal{U})$ and, on account of Theorem 9, we have

$$h(\mathbf{v}) = -\int_{\mathcal{V}} f_{\mathbf{v}}(\mathrm{v}) \log(f_{\mathbf{v}}(\mathrm{v})) \, d\mathrm{v}$$

$$= -\int_{\mathcal{U}} f_{\mathbf{v}}(\mathrm{g(u)}) \log(f_{\mathbf{v}}(\mathrm{g(u)})) \, |\det(\partial \mathrm{g}/\partial \mathrm{u})|^2 d\mathrm{u}$$

$$\overset{(a)}{=} -\int_{\mathcal{U}} \frac{f_{\mathbf{u}}(\mathrm{u})}{|\det(\partial \mathrm{g}/\partial \mathrm{u})|^2} \times$$

$$\times \log\left(\frac{f_{\mathbf{u}}(\mathrm{u})}{|\det(\partial \mathrm{g}/\partial \mathrm{u})|^2}\right) |\det(\partial \mathrm{g}/\partial \mathrm{u})|^2 d\mathrm{u}$$

$$= -\int_{\mathcal{U}} f_{\mathbf{u}}(\mathrm{u}) \log(f_{\mathbf{u}}(\mathrm{u})) \, d\mathrm{u}$$

$$+ 2 \int_{\mathcal{U}} f_{\mathbf{u}}(\mathrm{u}) \log(|\det(\partial \mathrm{g}/\partial \mathrm{u})|) \, d\mathrm{u}$$

$$= h(\mathbf{u}) + 2 \, \mathbb{E}_{\mathbf{u}}[\log |\det(\partial \mathrm{g}/\partial \mathbf{u})|]$$

where in (a) we used (51). This concludes the proof.

APPENDIX C
PROOF OF LEMMA 4

We need to show that the function $\mathrm{g}_{\mathsf{x}_{\mathcal{P}}}(\mathrm{s}, \mathrm{x}_{\mathcal{J}})$ is one-to-one *almost everywhere*. It is therefore legitimate to exclude sets of measure zero from its domain. In particular, we consider the restriction of the function $\mathrm{g}_{\mathsf{x}_{\mathcal{P}}}(\mathrm{s}, \mathrm{x}_{\mathcal{J}})$ to the set of pairs $(\mathrm{s}, \mathrm{x}_{\mathcal{J}})$ that satisfy

  (i) $|x_i| > 0$ for all $i \in \mathcal{J}$;
  (ii) $\det \mathrm{J}_2(\mathrm{s}) \neq 0$ with $\mathrm{J}_2(\cdot)$ defined in (38).

Condition (i) excludes those $\mathrm{x}_{\mathcal{J}}$ from the domain of $\mathrm{g}_{\mathsf{x}_{\mathcal{P}}}(\cdot)$ that have at least one component equal to zero; since the $x_i, i \in \mathcal{J}$, take on values in a continuum, the excluded set has measure zero. Condition (ii) excludes those s from the domain of $\mathrm{g}_{\mathsf{x}_{\mathcal{P}}}(\cdot)$ that have $\det(\mathrm{J}_2(\mathrm{s})) = 0$. Remember that we proved in Section VI-I (see (44)) that $\mathbb{E}\big[\log(|\det(\mathrm{J}_2(\mathbf{s}))|)\big] > -\infty$, which implies $\det(\mathrm{J}_2(\cdot)) \neq 0$ a.e. Therefore, the set excluded in (ii) must be a set of measure zero. We conclude that the set of pairs $(\mathrm{s}, \mathrm{x}_{\mathcal{J}})$ that violates at least one of the conditions (i) and (ii) is a set of measure zero.

To show that the resulting restriction of the function $\mathrm{g}_{\mathsf{x}_{\mathcal{P}}}(\cdot)$ [which, with slight abuse of notation we still call $\mathrm{g}_{\mathsf{x}_{\mathcal{P}}}(\cdot)$] is one-to-one, we take two pairs $(\tilde{\mathrm{s}}, \tilde{\mathrm{x}}_{\mathcal{J}})$ and $(\mathrm{s}, \mathrm{x}_{\mathcal{J}})$ from the domain of $\mathrm{g}_{\mathsf{x}_{\mathcal{P}}}(\cdot)$ and show that if $\mathrm{g}_{\mathsf{x}_{\mathcal{P}}}(\tilde{\mathrm{s}}, \tilde{\mathrm{x}}_{\mathcal{J}}) = \mathrm{g}_{\mathsf{x}_{\mathcal{P}}}(\mathrm{s}, \mathrm{x}_{\mathcal{J}})$, then necessarily $(\tilde{\mathrm{s}}, \tilde{\mathrm{x}}_{\mathcal{J}}) = (\mathrm{s}, \mathrm{x}_{\mathcal{J}})$.

Indeed, assume that both $(\tilde{\mathrm{s}}, \tilde{\mathrm{x}}_{\mathcal{J}})$ and $(\mathrm{s}, \mathrm{x}_{\mathcal{J}})$ belong to the domain of $\mathrm{g}_{\mathsf{x}_{\mathcal{P}}}(\cdot)$, i.e., both pairs satisfy conditions (i) and (ii) above. Suppose that $\mathrm{g}_{\mathsf{x}_{\mathcal{P}}}(\tilde{\mathrm{s}}, \tilde{\mathrm{x}}_{\mathcal{J}}) = \mathrm{g}_{\mathsf{x}_{\mathcal{P}}}(\mathrm{s}, \mathrm{x}_{\mathcal{J}})$, or, equivalently,

$$\mathrm{P}(\mathrm{I}_R \otimes \tilde{\mathrm{X}} \mathrm{D})\tilde{\mathrm{s}} = \mathrm{P}(\mathrm{I}_R \otimes \mathrm{X}\mathrm{D})\mathrm{s} \tag{52}$$

where $\mathrm{x} = [\mathrm{x}_{\mathcal{P}}^{\mathsf{T}} \, \mathrm{x}_{\mathcal{J}}^{\mathsf{T}}]^{\mathsf{T}}$, $\mathrm{X} = \mathrm{diag}(\mathrm{x})$, $\tilde{\mathrm{x}} = [\mathrm{x}_{\mathcal{P}}^{\mathsf{T}} \, \tilde{\mathrm{x}}_{\mathcal{J}}^{\mathsf{T}}]^{\mathsf{T}}$, and $\tilde{\mathrm{X}} = \mathrm{diag}(\tilde{\mathrm{x}})$. We next consider (52) as an equation parametrized by $(\mathrm{s}, \mathrm{x}_{\mathcal{J}})$ in the variables $(\tilde{\mathrm{s}}, \tilde{\mathrm{x}}_{\mathcal{J}})$ and show that this equation has a *unique* solution. Since $(\tilde{\mathrm{s}}, \tilde{\mathrm{x}}_{\mathcal{J}}) = (\mathrm{s}, \mathrm{x}_{\mathcal{J}})$ (trivially) satisfies (52), uniqueness then implies that $(\tilde{\mathrm{s}}, \tilde{\mathrm{x}}_{\mathcal{J}}) = (\mathrm{s}, \mathrm{x}_{\mathcal{J}})$.

To prove that (52) has a unique solution, we follow the approach described in Section III and convert (52) into a linear system of equations through a change of variables. In particular, thanks to constraint (i), we can left-multiply both sides of (52) by $\mathrm{P}[\mathrm{I}_R \otimes \mathrm{X}]^{-1}\mathrm{P}^{\mathsf{T}}\mathrm{P}[\mathrm{I}_R \otimes \tilde{\mathrm{X}}]^{-1}\mathrm{P}^{\mathsf{T}}$ to transform (52) into the equivalent equation

$$\mathrm{P}\big(\mathrm{I}_R \otimes \mathrm{X}^{-1}\mathrm{D}\big)\tilde{\mathrm{s}} = \mathrm{P}\big(\mathrm{I}_R \otimes \tilde{\mathrm{X}}^{-1}\mathrm{D}\big)\mathrm{s}. \tag{53}$$

Next, perform the substitutions $z_i = 1/x_i$, $\tilde{z}_i = 1/\tilde{x}_i$, $i \in [1:L]$, define $\mathrm{z} \triangleq [z_1 \ldots z_L]^{\mathsf{T}}$, and set $\mathrm{Z} \triangleq \mathrm{diag}(\mathrm{z})$ so that (53) can be written as

$$\mathrm{P}\left(\mathrm{I}_R \otimes \mathrm{Z}\mathrm{D}\right)\tilde{\mathrm{s}} = \sum_{i=1}^{L} \tilde{z}_i \mathrm{P}\mathrm{a}_i, \tag{54}$$

where $\mathrm{a}_i = (\mathrm{I}_R \otimes \mathrm{diag}(\mathrm{e}_i)\mathrm{D})\mathrm{s}$, $i \in [1 : L]$, as defined in (39). Finally, moving the terms containing the unknowns $\tilde{z}_i$, $i \in \mathcal{J}$, to the LHS of (54) while keeping the terms containing the fixed

parameters $\tilde{z}_i$, $i \in \mathcal{P}$, on the RHS, we transform (54) into the equivalent equation

$$\mathrm{P}\left(\mathrm{I}_R \otimes \mathrm{ZD}\right)\tilde{\mathrm{s}} - \sum_{i \in \mathcal{J}} \tilde{z}_i \mathrm{Pa}_i = \sum_{i \in \mathcal{P}} \tilde{z}_i \mathrm{Pa}_i. \qquad (55)$$

Defining $\tilde{z}_{\mathcal{J}} \triangleq [\tilde{z}_{\alpha+1} \ldots \tilde{z}_L]^{\mathsf{T}}$ and using the expression for $\mathrm{J}(\cdot)$ in (40), we can write (55) as

$$\mathrm{J}(\mathrm{s}, \mathrm{z})\begin{bmatrix}\tilde{\mathrm{s}} \\ -\tilde{z}_{\mathcal{J}}\end{bmatrix} = \sum_{i \in \mathcal{P}} \tilde{z}_i \mathrm{Pa}_i. \qquad (56)$$

The solution of (56) is unique if and only if $\det \mathrm{J}(\mathrm{s}, \mathrm{z}) \neq 0$. We use Lemma 5 to factorize $\det \mathrm{J}(\mathrm{s}, \mathrm{z})$ according to

$$\det(\mathrm{J}(\mathrm{s}, \mathrm{z})) = \det(\mathrm{J}_1(\mathrm{z}))\det(\mathrm{J}_2(\mathrm{s}))\det(\mathrm{J}_3(\mathrm{z}_{\mathcal{J}})). \qquad (57)$$

The first and the third term on the RHS of (57) can be written as follows

$$\det(\mathrm{J}_1(\mathrm{z})) = \left(\prod_{j=1}^{L-1} z_j\right)^{\tilde{R}}\left(\prod_{j=1}^{L} z_j\right)^{(R-\tilde{R})}$$

$$= \left(\prod_{j=1}^{L-1} \frac{1}{x_j}\right)^{\tilde{R}}\left(\prod_{j=1}^{L} \frac{1}{x_j}\right)^{(R-\tilde{R})}$$

$$\det(\mathrm{J}_3(\mathrm{z}_{\mathcal{J}})) = \prod_{j \in \mathcal{J}} \frac{1}{z_j} = \prod_{j \in \mathcal{J}} x_j$$

and are nonzero due to constraint (i) stated at the beginning of this Appendix; $\det(\mathrm{J}_2(\mathrm{s})) \neq 0$ due to constraint (ii). Hence $\det \mathrm{J}(\mathrm{s}, \mathrm{z}) \neq 0$ and the solution of (56) in the variables $(\tilde{\mathrm{s}}, \tilde{z}_{\mathcal{J}})$ is unique. Therefore, the solution of (52) [parametrized by $(\mathrm{s}, \mathrm{x}_{\mathcal{J}})$] in the variables $(\tilde{\mathrm{s}}, \tilde{\mathrm{x}}_{\mathcal{J}})$ is unique. This completes the proof.

We conclude this section by closing an issue that was left open in the back-of-the-envelope calculation in Section III. Specifically, we will show that the matrix B in (9) is full-rank. For $L = 3$ and $R = Q = 2$, the matrix B in (9) is related to $\mathrm{J}(\cdot)$ in (40) according to $\mathrm{B} = (\mathrm{I}_2 \otimes \mathbf{X})\mathrm{J}(\mathbf{s}, \mathbf{z})$ with $\mathbf{z} = [\mathbf{z}_1 \ldots \mathbf{z}_L]^{\mathsf{T}}$. Hence $\det(\mathrm{B}) = \det(\mathrm{I}_2 \otimes \mathbf{X})\det(\mathrm{J}(\mathbf{s}, \mathbf{z}))$. Since we assumed in Section III that $|\mathrm{x}_i| > 0$, $i \in [1:L]$, we have $\det(\mathrm{I}_2 \otimes \mathbf{X}) \neq 0$. Together with $\det(\mathrm{J}(\mathbf{s}, \mathbf{z})) \neq 0$, a.e., as shown above, we can conclude that, indeed, $\det(\mathrm{B}) \neq 0$, a.e., as claimed in Section III.

APPENDIX D
PROOF OF LEMMA 6

Instead of working with

$$I \triangleq \int_{\mathbb{C}^N} \exp(-\|\mathrm{u}\|^2)\log(|p(\mathrm{u})|)d\mathrm{u} \qquad (58)$$

it will turn out convenient to consider $|I|$ and to show that $|I| < \infty$, which trivially implies $I > -\infty$. As already mentioned, the proof of $|I| < \infty$ is based on Theorem 7. In order to be able to apply Theorem 7 we will need to transform the integration domain in (58) into a compact set in $\mathbb{R}^{2N}$, transform the complex-valued polynomial $p(\cdot)$ into a real-valued function, and get rid of the term $\exp(-\|\mathrm{u}\|^2)$. All this will be accomplished as follows. First, we bound $|I|$ by a sum of two integrals over the set $\mathbb{C}^N$, then, we apply a change of variables to transform these two integrals into three new integrals. The first two of these three

integrals are over the set $[0, \infty]$, which is still not compact, but the resulting integrals are simple enough to be bounded directly. The third integral is over a compact set and can, thus, be bounded using Theorem 7. We now implement the program just outlined.

Let K denote the degree of the homogeneous polynomial $p(\cdot)$. Then, by homogeneity of $p(\cdot)$,

$$p(\mathrm{u}) = p\left(\|\mathrm{u}\|\frac{\mathrm{u}}{\|\mathrm{u}\|}\right) = \|\mathrm{u}\|^K p\left(\frac{\mathrm{u}}{\|\mathrm{u}\|}\right)$$

and, therefore,

$$I = \int_{\mathbb{C}^N} \exp(-\|\mathrm{u}\|^2)\log(|p(\mathrm{u})|)d\mathrm{u}$$

$$= \underbrace{K\int_{\mathbb{C}^N} \exp(-\|\mathrm{u}\|^2)\log(\|\mathrm{u}\|)d\mathrm{u}}_{I_1}$$

$$+ \underbrace{\int_{\mathbb{C}^N} \exp(-\|\mathrm{u}\|^2)\log(|p(\mathrm{u}/\|\mathrm{u}\|)|)d\mathrm{u}}_{I_2}.$$

We next change variables in I_1 and I_2 by first transforming the domain of integration from $\mathbb{C}^N$ to $\mathbb{R}^{2N}$ and then using polar coordinates [26, p. 55]. Specifically, we introduce the function $\mathrm{u} : \mathbb{R}^{2N} \to \mathbb{C}^N$ that acts according to

$$\mathrm{u}(v) \triangleq [v_1 + iv_2 \; \cdots \; v_{2N-1} + iv_{2N}]^{\mathsf{T}}, \qquad (59)$$

and the function $\mathrm{v} : \mathbb{R}_+ \times \Delta \to \mathbb{R}^{2N}$ with $\Delta \triangleq [0, \pi]^{2N-2} \times [0, 2\pi]$ defined through

$$\mathrm{v}(r, \mathrm{t}) \triangleq r\mathrm{f}(\mathrm{t}) \qquad (60)$$

with

$$\mathrm{f}(\mathrm{t}) \triangleq \begin{bmatrix} \sin(t_1)\sin(t_2)\ldots\sin(t_{2N-2})\sin(t_{2N-1}) \\ \sin(t_1)\sin(t_2)\ldots\sin(t_{2N-2})\cos(t_{2N-1}) \\ \sin(t_1)\sin(t_2)\ldots\cos(t_{2N-2}) \\ \vdots \\ \sin(t_1)\cos(t_2) \\ \cos(t_1) \end{bmatrix}. \qquad (61)$$

It follows from (59)–(61) that

$$\|\mathrm{u}(\mathrm{v}(r, \mathrm{t}))\| = \|\mathrm{v}(r, \mathrm{t})\| = r$$

and therefore

$$\frac{\mathrm{u}(\mathrm{v}(r, \mathrm{t}))}{\|\mathrm{u}(\mathrm{v}(r, \mathrm{t}))\|} = \frac{\mathrm{u}(r\mathrm{f}(\mathrm{t}))}{r} = \mathrm{u}(\mathrm{f}(\mathrm{t})).$$

The determinant of the Jacobian of the function $\mathrm{v}(\cdot)$ is well-known and is given by [26, p. 55]

$$\det\frac{\partial \mathrm{v}}{\partial(r, \mathrm{t})} = r^{2N-1}\underbrace{\sin(t_1)^{2N-2}\sin(t_2)^{2N-3}\ldots\sin(t_{2N-2})}_{\mathrm{g}(\mathrm{t})}.$$

Changing variables in I_1 and I_2 according to $\mathrm{u} \to \mathrm{v} \to (r, \mathrm{t})$, we obtain

$$I_1 = K\int_{r, \mathrm{t}} \exp(-r^2)\log(r)r^{2N-1}\mathrm{g}(\mathrm{t})dr d\mathrm{t}$$

$$I_2 = \int_{r, \mathrm{t}} \exp(-r^2)\log(|p(\mathrm{u}(\mathrm{f}(\mathrm{t})))|)r^{2N-1}\mathrm{g}(\mathrm{t})dr d\mathrm{t}.$$

By the triangle inequality we have

$$|I| \leq |I_1| + |I_2|.$$

Using $\big|\mathrm{g}(\mathrm{t})\big| < 1$, we get

$$|I_1| \leq K\, 2\pi^{2M-1} \int_0^\infty \exp(-r^2)|\log(r)|r^{2N-1} dr < \infty$$

$$|I_2| \leq \int_0^\infty \exp(-r^2) r^{2N-1} dr \times \int_\Delta \big|\log(|p(\mathrm{u}(\mathrm{f}(\mathrm{t})))|)\big| d\mathrm{t}$$

$$\leq c \int_\Delta \big|\log(|p(\mathrm{u}(\mathrm{f}(\mathrm{t})))|^2)\big| d\mathrm{t}. \tag{62}$$

We hereby disposed of the integrals over unbounded domains and are left only with an integral over the compact set Δ. Note also that by absorbing a factor $1/2$ into c we introduced a square in (62), which will turn out useful later. In order to prove that $\big|I\big| < \infty$ it now remains to show that

$$I_3 \triangleq \int_\Delta \big|\log(|p(\mathrm{u}(\mathrm{f}(\mathrm{t})))|^2)\big| d\mathrm{t} < \infty. \tag{63}$$

Note that $\big|p(\mathrm{u}(\mathrm{f}(\cdot)))\big|^2 : \Delta \to \mathbb{R}_+$ is a real analytic function by [14, Prop. 2.2.2], because it is a composition of the polynomial $\big|p(\mathrm{u}(\cdot))\big|^2 : \mathbb{R}^{2N} \to \mathbb{R}_+$ and the function $\mathrm{f}(\cdot) : \Delta \to \mathbb{R}^{2N}$ that has real analytic components (trigonometric functions are real analytic on $\mathbb{R}$). Furthermore, by assumption, $p(\cdot) \not\equiv 0$ and hence $\big|p(\mathrm{u}(\mathrm{f}(\cdot)))\big|^2 \not\equiv 0$. Finally, Δ is a compact set. The inequality (63) now follows by application of Theorem 7. This concludes the proof.

APPENDIX E
PROOF OF THEOREM 7 VIA RESOLUTION OF SINGULARITIES

In order to prove Theorem 7 note that $\int_{\Delta \subset \mathbb{R}^M} \big|\log(\big|f(\mathrm{u})\big|)\big| d\mathrm{u}$ would clearly be finite if the function $f(\cdot)$ were bounded away from zero on the set Δ. Unfortunately, this is not the case. However, because $f(\cdot)$ is real analytic and $f(\cdot) \not\equiv 0$, it can take on the value zero only on a set of measure zero [14, Cor. 1.2.6]. Establishing whether the integral $\int_{\Delta \subset \mathbb{R}^M} \big|\log(\big|f(\mathrm{u})\big|)\big| d\mathrm{u}$ is finite, hence requires a fine analysis of the behavior of $\big|\log(\big|f(\cdot)\big|)\big|$ in the neighborhood of the zero-measure set $f^{-1}(\{0\})$. This can be accomplished using Hironaka's Theorem on the Resolution of Singularities, which allows one to write $f(\cdot)$ as a product of a monomial and a *nonvanishing* real analytic function in the neighborhood of each point u where $f(\mathrm{u}) = 0$. The logarithm of this product can then easily be bounded and shown to be finite. As the tools used in the following are non-standard, at least in the information theory literature, we review the main ingredients in some detail. Formally, Hironaka's Theorem states the following:

Theorem 10. *[11, Theorem 2.3] Let $f(\cdot) \not\equiv 0$ be a real analytic function [14, Def. 1.1.5] from a neighborhood of the origin 0, denoted $\Omega \subseteq \mathbb{R}^K$, to $\mathbb{R}$, which satisfies $f(0) = 0$. Then, there exists a triple $(\mathcal{W}, \mathcal{M}, \psi(\cdot))$ such that*
(a) $\mathcal{W} \subset \Omega$ is an open set in $\mathbb{R}^K$ with $0 \in \mathcal{W}$,
(b) $\mathcal{M}$ is a K-dimensional real analytic manifold [11, Def. 2.10] with coordinate charts $\{\mathcal{M}_\mathrm{p}, \varphi_\mathrm{p} : \mathcal{C}(0, \epsilon_\mathrm{p}) \to \mathcal{M}_\mathrm{p}\}$ for each point $\mathrm{p} \in \mathcal{M}$, where $\varphi_\mathrm{p}(\cdot)$ is an isomorphism[6] between $\mathcal{C}(0, \epsilon_\mathrm{p})$ and $\mathcal{M}_\mathrm{p}$ with $\varphi_\mathrm{p}(0) = \mathrm{p}$.

[6] Let $\mathcal{U}$ and $\mathcal{V}$ be two real analytic manifolds. A real analytic map $f : \mathcal{U} \to \mathcal{V}$ is called an isomorphism between $\tilde{\mathcal{U}} \subset \mathcal{U}$ and $\tilde{\mathcal{V}} \subset \mathcal{V}$ if it is one-to-one and an onto map from $\tilde{\mathcal{U}}$ to $\tilde{\mathcal{V}}$ whose inverse on $\tilde{\mathcal{V}}$ is also a real analytic map.

(c) $\psi : \mathcal{M} \to \mathcal{W}$ is a real analytic map,
that satisfies the following conditions:
(i) The map $\psi(\cdot)$ is proper, i.e., the inverse image of every compact set under $\psi(\cdot)$ is compact.
(ii) The map $\psi(\cdot)$ is an isomorphism [11, Def. 2.5] between $\mathcal{M} \setminus (f \circ \psi)^{-1}(\{0\})$ and $\mathcal{W} \setminus f^{-1}(\{0\})$.
(iii) For every point $\mathrm{p} \in \mathcal{M} \cap ((f \circ \psi)^{-1}(\{0\}))$, there exist $\mathrm{m}_\mathrm{p}, \mathrm{n}_\mathrm{p} \in \mathbb{N}_0^K$ and a real analytic function $g_\mathrm{p}(\cdot)$ that is bounded and nonvanishing on $\mathcal{C}(0, \epsilon_\mathrm{p})$ such that

$$|(f \circ \psi \circ \varphi_\mathrm{p})(\mathrm{v})| = \mathrm{v}^{\mathrm{m}_\mathrm{p}}, \quad \text{for all} \ \ \mathrm{v} \in \mathcal{C}(0, \epsilon_\mathrm{p})$$

and the determinant of the Jacobian of the mapping $(\psi \circ \varphi_\mathrm{p})(\cdot)$ satisfies

$$\det\left(\frac{\partial(\psi \circ \varphi_\mathrm{p})}{\partial \mathrm{v}}\right) = g_\mathrm{p}(\mathrm{v})\mathrm{v}^{\mathrm{n}_\mathrm{p}}, \quad \text{for all} \ \ \mathrm{v} \in \mathcal{C}(0, \epsilon_\mathrm{p}).$$

Thanks to Theorem 10, in the *neighborhood of zero*, every real analytic function that satisfies $f(\cdot) \not\equiv 0$ and $f(0) = 0$ can be written as a product of a monomial and a *nonvanishing* real analytic function. In order to bound the integral in (45), we will need to represent $f(\cdot)$ in this form in the neighborhood of *every* point in the domain of integration. This representation can be obtained by analyzing two cases separately. For points x such that $f(\mathrm{x}) \neq 0$, by real-analyticity and, hence, continuity, it follows that $f(\cdot)$ is already nonvanishing in the neighborhood of x and is hence trivially representable as a product of a monomial and a *nonvanishing* real analytic function. For points x such that $f(\mathrm{x}) = 0$, the desired representation can be obtained by appropriately shifting the origin in Theorem 10. The following straightforward corollary to Theorem 10 conveniently formalizes these statements in a unified fashion.

Corollary 11. *Let $f(\cdot) \not\equiv 0$ be a real analytic function from a neighborhood of $\mathrm{u} \in \mathbb{R}^K$, denoted $\Omega \subseteq \mathbb{R}^K$, to $\mathbb{R}$. Then, there exists a triple $(\mathcal{W}, \mathcal{M}, \psi(\cdot))$, such that*
(a) $\mathcal{W} \subset \Omega$ is an open set in $\mathbb{R}^K$ with $\mathrm{u} \in \mathcal{W}$,
(b) $\mathcal{M}$ is a K-dimensional real analytic manifold [11, Def. 2.10] with coordinate charts $\{\mathcal{M}_\mathrm{p}, \varphi_\mathrm{p} : \mathcal{C}(0, \epsilon_\mathrm{p}) \to \mathcal{M}_\mathrm{p}\}$ for each point $\mathrm{p} \in \mathcal{M}$, where $\mathcal{M}_\mathrm{p}$ is an open set with $\mathrm{p} \in \mathcal{M}_\mathrm{p}$ and $\varphi_\mathrm{p}(\cdot)$ is an isomorphism between $\mathcal{C}(0, \epsilon_\mathrm{p})$ and $\mathcal{M}_\mathrm{p}$ with $\varphi_\mathrm{p}(0) = \mathrm{p}$.
(c) $\psi : \mathcal{M} \to \mathcal{W}$ is a real analytic map, that satisfies the following conditions:
(i) The map $\psi(\cdot)$ is proper, i.e., the inverse image of any compact set under $\psi(\cdot)$ is compact.
(ii) The map $(\psi \circ \varphi_\mathrm{p})(\cdot)$ is an isomorphism between $\mathcal{C}(0, \epsilon_\mathrm{p}) \setminus (f \circ \psi \circ \varphi_\mathrm{p})^{-1}(\{0\})$ and $\psi(\mathcal{M}_\mathrm{p}) \setminus f^{-1}(\{0\})$.
(iii) For every point $\mathrm{p} \in \mathcal{M}$, there exist $\mathrm{m}_\mathrm{p}, \mathrm{n}_\mathrm{p} \in \mathbb{N}_0^K$ and real analytic functions $h_\mathrm{p}(\cdot)$ and $g_\mathrm{p}(\cdot)$ that are bounded and nonvanishing on $\mathcal{C}(0, \epsilon_\mathrm{p})$ such that

$$|(f \circ \psi \circ \varphi_\mathrm{p})(\mathrm{v})| = h_\mathrm{p}(\mathrm{v})\mathrm{v}^{\mathrm{m}_\mathrm{p}}, \quad \text{for all} \ \ \mathrm{v} \in \mathcal{C}(0, \epsilon_\mathrm{p}) \tag{64}$$

and the determinant of the Jacobian of the mapping $(\psi \circ \varphi_\mathrm{p})(\cdot)$ satisfies

$$\det\left(\frac{\partial(\psi \circ \varphi_\mathrm{p})}{\partial \mathrm{v}}\right) = g_\mathrm{p}(\mathrm{v})\mathrm{v}^{\mathrm{n}_\mathrm{p}}, \quad \text{for all} \ \ \mathrm{v} \in \mathcal{C}(0, \epsilon_\mathrm{p}).$$

Proof: First consider u such that $f(\mathrm{u}) \neq 0$. As already mentioned, in this case the statement of the corollary is a pure formality since $f(\cdot)$ itself is a nonvanishing real analytic function in the neighborhood of u. Formally, since $f(\cdot)$ is real analytic and, hence, continuous, there exists an open cube $\mathcal{C}(\mathrm{u}, \epsilon)$ on which $f(\cdot)$ is uniformly bounded and satisfies $f(\mathrm{v}) \neq 0$ for all $\mathrm{v} \in \mathcal{C}(\mathrm{u}, \epsilon)$. In this case, the corollary, therefore, follows immediately by choosing $\mathcal{M} \triangleq \mathcal{C}(\mathrm{u}, \epsilon)$, $\mathcal{W} \triangleq \mathcal{C}(\mathrm{u}, \epsilon)$, setting $\psi(\cdot)$ to be the identity map, defining $\mathcal{M}_\mathrm{p} \triangleq \mathcal{M}$ for all $\mathrm{p} \in \mathcal{M}$, and setting $\varphi_\mathrm{p}(\mathrm{v}) \triangleq \mathrm{v} + \mathrm{p}$ for all $\mathrm{v} \in \mathcal{C}(0, \epsilon)$.

Next, consider the more complicated case $f(\mathrm{u}) = 0$. The main idea is to apply Theorem 10 to the function $\tilde{f}(\mathrm{t}) \triangleq f(\mathrm{t} + \mathrm{u})$, $\mathrm{t} \in \Omega - \mathrm{u}$. Theorem 10 implies that there exists a triple $(\tilde{\mathcal{W}}, \tilde{\mathcal{M}}, \tilde{\psi})$ that satisfies (a)–(c) and (i)–(iii) in Theorem 10 for $\tilde{f}(\cdot)$. Now let

$$\mathcal{W} \triangleq \tilde{\mathcal{W}} + \mathrm{u}$$
$$\mathcal{M} \triangleq \tilde{\mathcal{M}}$$
$$\psi(\cdot) \triangleq \tilde{\psi}(\cdot) + \mathrm{u}.$$

Then (a)–(c) and (i) in the statement of Corollary 11 follow immediately from (a)–(c) and (i) in Theorem 10.

Condition (ii) in the statement of Corollary 11 follows from (ii) in Theorem 10 and the fact that $\varphi_\mathrm{p}(\cdot)$ is an isomorphism between $\mathcal{C}(0, \epsilon_\mathrm{p})$ and $\mathcal{M}_\mathrm{p}$.

To verify (iii) in the statement of Corollary 11, consider the following two cases separately. First, let $\mathrm{p} \in \mathcal{M}$ such that $(f \circ \psi)(\mathrm{p}) = 0$. Then (iii) in the statement of Corollary 11 follows from (iii) in Theorem 10 and the fact that

$$(f \circ \psi \circ \varphi_\mathrm{p})(\mathrm{v}) = (\tilde{f} \circ \tilde{\psi} \circ \varphi_\mathrm{p})(\mathrm{v}), \quad \text{for all } \mathrm{v} \in \mathcal{C}(0, \epsilon_\mathrm{p})$$
$$\det\left(\frac{\partial(\psi \circ \varphi_\mathrm{p})}{\partial \mathrm{v}}\right) = \det\left(\frac{\partial(\tilde{\psi} \circ \varphi_\mathrm{p})}{\partial \mathrm{v}}\right), \quad \text{for all } \mathrm{v} \in \mathcal{C}(0, \epsilon_\mathrm{p}).$$

Second, let $\mathrm{p} \in \mathcal{M}$ with $(f \circ \psi)(\mathrm{p}) \neq 0$. As $(\tilde{f} \circ \tilde{\psi})(\mathrm{p}) = (f \circ \psi)(\mathrm{p})$, this implies that $(\tilde{f} \circ \tilde{\psi})(\mathrm{p}) \neq 0$. Since $\tilde{f}(\cdot)$ is a continuous function (as a translation of $f(\cdot)$ that is real analytic and hence continuous), there exists an $\epsilon_\mathrm{p} > 0$ such that $\tilde{f}(\cdot)$ is bounded and nonvanishing on the open cube $\mathcal{C}(\tilde{\psi}(\mathrm{p}), \epsilon_\mathrm{p})$. Now (ii) in Theorem 10 implies that $\tilde{\psi}(\cdot)$ is an isomorphism, i.e.,

$$\tilde{\psi} : \tilde{\psi}^{-1}(\mathcal{C}(\tilde{\psi}(\mathrm{p}), \epsilon_\mathrm{p})) \to \mathcal{C}(\tilde{\psi}(\mathrm{p}), \epsilon_\mathrm{p}).$$

Define $\varphi_\mathrm{p}(\mathrm{v}) \triangleq \tilde{\psi}^{-1}(\mathrm{v} + \tilde{\psi}(\mathrm{p}))$ for $\mathrm{v} \in \mathcal{C}(0, \epsilon_\mathrm{p})$. Then $\varphi_\mathrm{p}(0) = \mathrm{p}$ and

$$f(\psi \circ \varphi_\mathrm{p})(\mathrm{v}) = (\tilde{f} \circ \tilde{\psi} \circ \varphi_\mathrm{p})(\mathrm{v})$$
$$= \tilde{f}(\mathrm{v} + \tilde{\psi}(\mathrm{p})), \quad \text{for all } \mathrm{v} \in \mathcal{C}(0, \epsilon_\mathrm{p}).$$

Therefore, we can simply set $h_\mathrm{p}(\mathrm{v}) \triangleq \tilde{f}(\mathrm{v} + \tilde{\psi}(\mathrm{p}))$ and the representation (64) is obtained. Furthermore, since $\psi(\varphi_\mathrm{p}(\mathrm{v})) = \tilde{\psi}(\varphi_\mathrm{p}(\mathrm{v})) + \mathrm{u} = \tilde{\psi}(\tilde{\psi}^{-1}(\mathrm{v} + \tilde{\psi}(\mathrm{p}))) + \mathrm{u} = \mathrm{v} + \tilde{\psi}(\mathrm{p}) + \mathrm{u}$, we have

$$\det\left(\frac{\partial(\psi \circ \varphi_\mathrm{p})}{\partial \mathrm{v}}\right) = 1, \quad \text{for all } \mathrm{v} \in \mathcal{C}(0, \epsilon_\mathrm{p}). \qquad \blacksquare$$

We now have all the ingredients required to prove Theorem 7.

Proof: For each $\mathrm{u} \in \Delta$, Corollary 11 implies that there exists a triple $(\mathcal{W}_\mathrm{u}, \mathcal{M}_\mathrm{u}, \psi_\mathrm{u})$ such that $\mathcal{W}_\mathrm{u} \subseteq \Omega$ is an open set

containing u, $\mathcal{M}_\mathrm{u}$ is a real analytic manifold, and $\psi_\mathrm{u} : \mathcal{M}_\mathrm{u} \to \mathcal{W}_\mathrm{u}$ is a proper map. Furthermore, for each $\mathrm{p} \in \mathcal{M}_\mathrm{u}$ there exists a coordinate chart $\{\mathcal{M}_{\mathrm{u},\mathrm{p}}, \varphi_{\mathrm{u},\mathrm{p}} : \mathcal{C}(0, \epsilon_{\mathrm{u},\mathrm{p}}) \to \mathcal{M}_{\mathrm{u},\mathrm{p}}\}$, where $\mathcal{M}_{\mathrm{u},\mathrm{p}}$ is an open set with $\mathrm{p} \in \mathcal{M}_{\mathrm{u},\mathrm{p}}$ and $\varphi_{\mathrm{u},\mathrm{p}}(\cdot)$ is an isomorphism between $\mathcal{C}(0, \epsilon_{\mathrm{u},\mathrm{p}})$ and $\mathcal{M}_{\mathrm{u},\mathrm{p}}$ with $\varphi_{\mathrm{u},\mathrm{p}}(0) = \mathrm{p}$, such that $(\psi_\mathrm{u} \circ \varphi_{\mathrm{u},\mathrm{p}})(\cdot)$ is a real analytic map [11, p.49] on $\mathcal{C}(0, \epsilon_{\mathrm{u},\mathrm{p}})$ and

$$|(f \circ \psi_\mathrm{u} \circ \varphi_{\mathrm{u},\mathrm{p}})(\mathrm{v})| = h_{\mathrm{u},\mathrm{p}}(\mathrm{v}) \mathrm{v}^{\mathrm{m}_{\mathrm{u},\mathrm{p}}}$$
$$\det\left(\frac{\partial(\psi_\mathrm{u} \circ \varphi_{\mathrm{u},\mathrm{p}})}{\partial \mathrm{v}}\right) = g_{\mathrm{u},\mathrm{p}}(\mathrm{v}) \mathrm{v}^{\mathrm{n}_{\mathrm{u},\mathrm{p}}}$$

for all $\mathrm{v} \in \mathcal{C}(0, \epsilon_{\mathrm{u},\mathrm{p}})$, where $g_{\mathrm{u},\mathrm{p}}(\cdot)$ and $h_{\mathrm{u},\mathrm{p}}(\cdot)$ are real analytic functions that are nonvanishing on $\mathcal{C}(0, \epsilon_{\mathrm{u},\mathrm{p}})$. Now, for each $\mathrm{u} \in \Delta$ we choose an open neighborhood of u, denoted as $\mathcal{W}'_\mathrm{u}$, and a compact neighborhood of u, denoted Δ_u, such that $\mathrm{u} \in \mathcal{W}'_\mathrm{u} \subset \Delta_\mathrm{u} \subset \mathcal{W}_\mathrm{u}$. Since Δ is a compact set [27, 2.31] there exists a finite set of vectors $\{\mathrm{u}_1, \ldots, \mathrm{u}_N\}$ with $\mathrm{u}_i \in \Delta$ such that

$$\Delta \subset \bigcup_{i \in [1:N]} \mathcal{W}'_i \subset \bigcup_{i \in [1:N]} \Delta_i,$$

where we set $\mathcal{W}'_i \triangleq \mathcal{W}'_{\mathrm{u}_i}$ and $\Delta_i \triangleq \Delta_{\mathrm{u}_i}$ for $i \in [1:N]$. Take an $i \in [1:N]$ and set $\mathcal{M}_i \triangleq \mathcal{M}_{\mathrm{u}_i}$, $\mathcal{W}_i \triangleq \mathcal{W}_{\mathrm{u}_i}$, and $\psi_i \triangleq \psi_{\mathrm{u}_i}$. Since the mapping $\psi_i : \mathcal{M}_i \to \mathcal{W}_i$ is proper, the set $\psi_i^{-1}(\Delta_i) \subset \mathcal{M}_i$ is a compact set. Therefore, there exists a finite number M_i of points $\mathrm{p}_1, \ldots, \mathrm{p}_{M_i} \in \mathcal{M}_i$ such that

$$\psi_i^{-1}(\Delta_i) \subset \bigcup_{j \in [1:M_i]} \mathcal{M}_{i,j} \qquad (65)$$

with $\mathcal{M}_{i,j} \triangleq \mathcal{M}_{\mathrm{u}_i,\mathrm{p}_j}$. Since (65) holds for all $i \in [1:N]$, we can upper-bound the integral in (45) as follows:

$$\int_\Delta |\log(|f(\mathrm{u})|)| d\mathrm{u}$$
$$\leq \sum_{i \in [1:N]} \int_{\Delta_i} |\log(|f(\mathrm{u})|)| d\mathrm{u}$$
$$\leq \sum_{i \in [1:N]} \sum_{j \in [1:M_i]} \int_{\Delta_i \cap \psi_i(\mathcal{M}_{i,j})} |\log(|f(\mathrm{u})|)| d\mathrm{u}$$
$$\leq \sum_{i \in [1:N]} \sum_{j \in [1:M_i]} \int_{\psi_i(\mathcal{M}_{i,j})} |\log(|f(\mathrm{u})|)| d\mathrm{u}. \qquad (66)$$

Since $f(\cdot)$ is a real analytic function and, hence, $f^{-1}(\{0\})$ is a set of measure zero, we have

$$\int_{\psi_i(\mathcal{M}_{i,j})} |\log(|f(\mathrm{u})|)| d\mathrm{u} = \int_{\psi_i(\mathcal{M}_{i,j}) \setminus f^{-1}(\{0\})} |\log(|f(\mathrm{u})|)| d\mathrm{u}. \qquad (67)$$

Next, recall that according to (ii) in Corollary 11 $(\psi_i \circ \varphi_{\mathrm{p}_j})(\cdot)$ is an isomorphism between $\mathcal{C}_{i,j} \triangleq \mathcal{C}(0, \epsilon_{\mathrm{u}_i,\mathrm{p}_j}) \setminus (f \circ \psi_i \circ \varphi_{\mathrm{p}_j})^{-1}(\{0\})$ and $\psi_i(\mathcal{M}_{i,j}) \setminus f^{-1}(\{0\})$. Therefore, we can apply

the change of variables theorem [21, Theorem 7.26] to get

$$\int_{\psi_i(\mathcal{M}_{i,j})\setminus f^{-1}(\{0\})} \left|\log(|f(\mathrm{u})|)\right| d\mathrm{u}$$

$$= \int_{\mathcal{C}_{i,j}} \left|g_{i,j}(\mathrm{v})\mathrm{v}^{\mathrm{n}_{i,j}} \log\left(|h_{i,j}(\mathrm{v})\mathrm{v}^{\mathrm{m}_{i,j}}|\right)\right| d\mathrm{v}$$

$$\leq \underbrace{\sup_{\mathrm{v}\in\mathcal{C}_{i,j}} \left(|g_{i,j}(\mathrm{v})\mathrm{v}^{\mathrm{n}_{i,j}}|\right)}_{c_{i,j}} \int_{\mathcal{C}_{i,j}} \left|\log\left(|h_{i,j}(\mathrm{v})\mathrm{v}^{\mathrm{m}_{i,j}}|\right)\right| d\mathrm{v}$$

$$\overset{(a)}{=} c_{i,j} \int_{\mathcal{C}_{i,j}} \left|\log\left(|\mathrm{v}^{\mathrm{m}_{i,j}}|\right)\right| d\mathrm{v}$$
$$+ c_{i,j} \int_{\mathcal{C}_{i,j}} \left|\log\left(|h_{i,j}(\mathrm{v})|\right)\right| d\mathrm{v}$$

$$\overset{(b)}{\leq} c_{i,j} \int_{-\epsilon_{i,j}}^{\epsilon_{i,j}} \cdots \int_{-\epsilon_{i,j}}^{\epsilon_{i,j}} \left|\sum_{k=1}^{K}[\mathrm{m}_{i,j}]_k \log\left(|v_k|\right)\right| dv_1 \ldots dv_K$$
$$+ \underbrace{c_{i,j} \sup_{\mathrm{v}\in\mathcal{C}_{i,j}} \left(\left|\log\left(|h_{i,j}(\mathrm{v})|\right)\right|\right) (2\epsilon_{i,j})^K}_{\hat{c}_{i,j}}$$

$$\overset{(c)}{\leq} c_{i,j} \sum_{k=1}^{K}[\mathrm{m}_{i,j}]_k \int_{-\epsilon_{i,j}}^{\epsilon_{i,j}} \cdots \int_{-\epsilon_{i,j}}^{\epsilon_{i,j}} \left|\log\left(|v_k|\right)\right| dv_1 \ldots dv_K + \hat{c}_{i,j}$$

$$= \underbrace{c_{i,j} \sum_{k=1}^{K}[\mathrm{m}_{i,j}]_k (2\epsilon_{i,j})^{(K-1)} \int_{-\epsilon_{i,j}}^{\epsilon_{i,j}} \left|\log\left(|v|\right)\right| dv}_{\tilde{c}_{i,j}} + \hat{c}_{i,j}$$

$$\overset{(d)}{<} \infty. \tag{68}$$

Here, $c_{i,j}, \tilde{c}_{i,j}, \hat{c}_{i,j} > 0$, $i \in [1\!:\!N]$, $j \in [1\!:\!M_i]$, are finite constants; in (a) we used the fact that $g_{i,j}(\cdot)$ is bounded and nonvanishing on $\mathcal{C}_{i,j}$; in (b) $[\mathrm{m}_{i,j}]_k$ denotes the kth component of the vector $\mathrm{m}_{i,j}$; in (c) we used the triangle inequality to bound the first term, the second term is finite because $h_{i,j}(\cdot)$ is bounded and nonvanishing on $\mathcal{C}_{i,j}$; and in (d) we used $\int_{-\epsilon_{i,j}}^{\epsilon_{i,j}} \log\left(|v|\right) dv < \infty$. Combining (66), (67), and (68), we complete the proof. ∎

<div align="center">REFERENCES</div>

[1] 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)*, Austin, TX, U.S.A., June 2010, pp. 320–324.

[2] 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)*, St. Petersburg, Russia, Aug. 2011, pp. 2020–2024.

[3] W. Yang, G. Durisi, V. I. Morgenshtern, and E. Riegler, "Capacity pre-log of SIMO correlated block-fading channels," in *Proc. International Symposium on Wireless Communication Systems (ISWCS)*, Aachen, Germany, Nov. 2011, pp. 869–873.

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

[5] 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.

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

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

[8] A. Lapidoth, "On the asymptotic capacity of stationary Gaussian fading channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 437–446, Feb. 2005.

[9] T. Koch, "On heating up and fading in communication channels," Ph.D. dissertation, ETH Zurich, Diss. ETH No. 18339, May 2009.

[10] 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.

[11] S. Watanabe, *Algebraic Geometry and Statistical Learning Theory*, ser. Cambridge Monographs on Applied and Computational Mathematics. Cambridge Univ. Press, 2009, vol. 25.

[12] H. Hironaka, "Resolution of singularities of an algebraic variety over a field of characteristic zero: I," *Annals of Math.*, vol. 79, no. 1, pp. 109–203, Jan. 1964.

[13] ——, "Resolution of singularities of an algebraic variety over a field of characteristic zero: II," *Annals of Math.*, vol. 79, no. 2, pp. 205–326, Mar. 1964.

[14] S. G. Krantz and H. R. Parks, *A Primer of Real Analytic Functions*, 2nd ed. Birkhäuser, 2002.

[15] K. Fritzsche and H. Grauert, *From Holomorphic Functions to Complex Manifolds*. Springer, 2002.

[16] T. Tao, "An uncertainty principle for cyclic groups of prime order," *Math. Res. Lett.*, vol. 12, no. 1, pp. 121–127, 2005.

[17] G. Grimmett and D. Stirzaker, *Probability and Random Processes*, 3rd ed. Oxford Univ. Press, 2001.

[18] 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.

[19] 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.

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

[21] W. Rudin, *Real and Complex Analysis*, 3rd ed. New York, NY, U.S.A.: McGraw-Hill, 1987.

[22] H. Lütkepohl, *Handbook of Matrices*. Chichester, U.K.: Wiley, 1996.

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

[24] 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., Contr., and Comput.*, Monticello, IL, Oct. 2012.

[25] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes*, 4th ed., ser. Electrical and Computer Engineering. McGraw-Hill, 2002.

[26] R. I. Muirhead, *Aspects of Multivariate Statistical Theory*. Wiley, 2005.

[27] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed. McGraw-Hill, 1976.

Veniamin I. Morgenshtern was born in Leningrad, Russia, in 1982. He studied mathematics and software engineering at Saint-Petersburg State University, Russia, where he received the Dipl. Math. degree with honors in 2004. He then joined the Communication Technology Laboratory at ETH Zurich, Switzerland, as a research assistant. In 2007 he was a visiting researcher at the University of Illinois at Urbana-Champaign, U.S.A. He graduated from ETH Zurich in 2010 with a Dr. Sc. degree. From 2010 to 2012, Dr. Morgenshtern was a postdoctoral researcher at ETH Zurich. He is currently a postdoctoral researcher with the Department of Statistics, Stanford University, Stanford, CA. His research interests are in information theory, communication theory, mathematical signal processing, and high-dimensional statistics. Dr. Morgenshtern received an ETH Medal for his dissertation. He currently holds a Swiss National Science Foundation scholarship for advanced researchers.

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.

He was a visiting researcher at the Max Planck Institute for Mathematics in the Sciences in Leipzig, Germany (Sep. 2004 – Feb. 2005), the Communication Theory Group at ETH Zurich, Switzerland (Sep. 2010 – Feb. 2011 and June 2012 – Nov. 2012), and the Department of Electrical and Computer Engineering at The Ohio State University in Columbus, Ohio (Mar. 2012). From 2005 – 2006, he was a post-doc at the Institute for Analysis and Scientific Computing, Vienna University of Technology. From 2007 – 2010, he was a senior researcher at the Telecommunications Research Center Vienna (FTW). Since 2010, he has been a post-doc at the Institute of Telecommunications at Vienna University of Technology.

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

Bernd Sturmfels received doctoral degrees in 1987 from the University of Washington, Seattle, and the Technical University Darmstadt, Germany. After postdoctoral years in Minneapolis and Linz, Austria, he taught at Cornell University, before joining UC Berkeley in 1995, where he is Professor of Mathematics, Statistics, and Computer Science. His honors include a Sloan Fellowship, and a David and Lucile Packard Fellowship, a Clay Senior Scholarship, a Humboldt Senior Research Prize, the SIAM von Neumann Lecturership, and a Sarlo Distinguished Mentoring Award. Recently, he served as Vice President of the American Mathematical Society. A leading experimentalist among mathematicians, Sturmfels has authored ten books and over 200 research articles, in the areas of combinatorics, algebraic geometry, symbolic computation, and their applications. He has mentored 35 doctoral students and numerous postdocs. His current research focuses on algebraic statistics and tropical geometry.

Wei Yang (S'09) received the B.E. degree in communication engineering and M.E. degree in communication and information systems from the Beijing University of Posts and Telecommunications, Beijing, China, in 2008 and 2011, respectively. He is currently pursuing a Ph.D. degree in electrical engineering at Chalmers University of Technology, Gothenburg, Sweden. From July to August 2012, he was a visiting student at the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA.

Mr. Yang is the recipient of a Student Paper Award at the 2012 IEEE International Symposium on Information Theory (ISIT), Cambridge, MA. His research interests are in the areas of information and communication theory.

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, Zurich, Switzerland. Since 2010, he has been an assistant professor at Chalmers University of Technology, Gothenburg, Sweden. He held visiting researcher positions at IMST (Germany), University of Pisa (Italy), and Vienna University of Technology (Austria). Dr. Durisi is co-author of a paper that won a student paper award at the International Symposium on Information Theory (ISIT 2012). 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 information theory, communication theory, and compressive sensing.

Helmut Bölcskei was born in Mödling, Austria on May 29, 1970, and received the Dipl.-Ing. and Dr. techn. degrees in electrical engineering from Vienna University of Technology, Vienna, Austria, in 1994 and 1997, respectively. In 1998 he was with Vienna University of Technology. From 1999 to 2001 he was a postdoctoral researcher in the Information Systems Laboratory, Department of Electrical Engineering, and in the Department of Statistics, Stanford University, Stanford, CA. He was in the founding team of Iospan Wireless Inc., a Silicon Valley-based startup company (acquired by Intel Corporation in 2002) specialized in multiple-input multiple-output (MIMO) wireless systems for high-speed Internet access, and was a co-founder of Celestrius AG, Zurich, Switzerland. From 2001 to 2002 he was an Assistant Professor of Electrical Engineering at the University of Illinois at Urbana-Champaign. He has been with ETH Zurich since 2002, where he is Professor of Electrical Engineering. He was a visiting researcher at Philips Research Laboratories Eindhoven, The Netherlands, ENST Paris, France, and the Heinrich Hertz Institute Berlin, Germany. His research interests are in information theory, mathematical signal processing, applied and computational harmonic analysis, and statistics.

He received the 2001 IEEE Signal Processing Society Young Author Best Paper Award, the 2006 IEEE Communications Society Leonard G. Abraham Best Paper Award, the 2010 Vodafone Innovations Award, the ETH "Golden Owl" Teaching Award, is a Fellow of the IEEE, a 2011 EURASIP Fellow, a 2013-2014 Distinguished Lecturer of the IEEE Information Theory Society, and was an Erwin Schrödinger Fellow of the Austrian National Science Foundation (FWF). He was a plenary speaker at several IEEE conferences and served as an associate editor of the IEEE Transactions on Information Theory, the IEEE Transactions on Signal Processing, the IEEE Transactions on Wireless Communications, and the EURASIP Journal on Applied Signal Processing. He is currently editor-in-chief of the IEEE Transactions on Information Theory and serves on the editorial boards of "Foundations and Trends in Networking", "Foundations and Trends in Communications and Information Theory", and the IEEE Signal Processing Magazine. He was TPC co-chair of the 2008 IEEE International Symposium on Information Theory and served on the Board of Governors of the IEEE Information Theory Society.

Shaowei Lin was born in Singapore in 1981. He received the B.Sc. degree (with honors) in mathematics from Stanford University in 2005. He then worked on MIMO communications at the Agency for Science, Technology, and Research (A*STAR) Institute for Infocomm Research (I2R) in Singapore. From 2006 to 2011, he studied algebraic statistics under Bernd Sturmfels at the University of California, Berkeley, and graduated with a Ph.D. in mathematics. His thesis was titled "Algebraic Methods for Evaluating Integrals in Bayesian Statistics." After that, he completed a one-year collaboration with the Artificial Intelligence Lab headed by Andrew Ng at Stanford University to explore mathematical challenges in deep learning. Currently, he is a scientist with the Sense and Sense-abilities program at A*STAR I2R in Singapore where he uses machine learning principles to design algorithms and protocols for wireless sensor networks. His research interests are in algebraic geometry, asymptotic theory, singular learning theory, deep learning, and neural networks.