

The SIMO Pre-Log Can Be Larger Than the SISO Pre-Log

Veniamin I. Morgenshtern, Giuseppe Durisi, and Helmut Bölcskei

ETH Zurich, 8092 Zurich, Switzerland

E-mail: {vmorgens, gdurisi, boelcskei}@nari.ee.ethz.ch

Abstract—We establish a lower bound on the noncoherent capacity pre-log of a temporally correlated Rayleigh block-fading single-input multiple-output (SIMO) channel. Surprisingly, when the covariance matrix of the channel satisfies a certain technical condition related to the cardinality of its smallest set of linearly dependent rows, this lower bound reveals that the capacity pre-log in the SIMO case is larger than that in the single-input single-output (SISO) case.

I. INTRODUCTION

It is well known that the *coherent-capacity* pre-log (i.e., the asymptotic ratio between capacity and the logarithm of SNR, as SNR goes to infinity) of a single-input multiple-output (SIMO) fading channel is equal to 1 and is, hence, the same as that of a single-input single-output (SISO) fading channel [1]. In the practically more relevant *noncoherent setting*, where neither transmitter nor receiver have channel-state information, but both are aware of the channel statistics, the effect of multiple antennas on the capacity¹ pre-log is understood only for a specific simple channel model, namely, the *constant block-fading* model. In this model, the channel is assumed to remain constant over a block of T symbols and to change in an independent fashion from block to block [2]. For this model, the SIMO capacity pre-log is again equal to the SISO capacity pre-log, but, differently from the coherent case, is given by $1 - 1/T$ [3], [4].

A more general way of 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 [5] and the MISO [6, Thm. 4.15] cases. The capacity bounds for the SIMO stationary-fading channel available in the literature [6, Thm. 4.13] do not allow one to determine whether the capacity pre-log in the SIMO case can be larger than that in the SISO case.

In this paper, we focus on a channel model that can be seen as lying in between the general stationary-fading model considered in [5], [6], and the simpler constant block-fading model analyzed in [2], [4]. Specifically, we assume that the fading process is independent across blocks of length T and temporally correlated within blocks, with the rank of the corresponding $T \times T$ channel covariance matrix given by² $Q < T$. For this channel model, referred to as the *correlated block-fading* model in the following, the SISO capacity pre-log is equal to $1 - Q/T$ [8].³ The SIMO and MIMO capacity pre-logs are not known in this case. A

conjecture in [8] on the MIMO capacity pre-log implies that the capacity pre-log in the SIMO case would be the same as that in the SISO case. In this paper, we disprove the conjecture in [8] by showing that in the SIMO case a capacity pre-log of $1 - 1/T$ can be obtained when the number of receive antennas is equal to Q , and the channel covariance matrix satisfies a certain technical condition detailed in Theorem 1.

Notation: Uppercase boldface letters denote matrices, and lowercase boldface letters designate vectors. The all-zero matrix of appropriate size is written as $\mathbf{0}$. The element in the i th row and j th column of a matrix \mathbf{A} is denoted as $a_{i,j}$, and the i th component of the vector \mathbf{u} is u_i . For a vector \mathbf{u} , $\text{diag}(\mathbf{u})$ denotes the diagonal matrix that has the entries of \mathbf{u} on its main diagonal. The superscripts \top and H stand for transposition and Hermitian transposition, respectively. The expectation operator is denoted as $\mathbb{E}[\cdot]$. For two matrices \mathbf{A} and \mathbf{B} , we designate the Kronecker product as $\mathbf{A} \otimes \mathbf{B}$; to simplify notation, we use the convention that the ordinary matrix product always precedes the Kronecker product, i.e., $\mathbf{AB} \otimes \mathbf{C} = (\mathbf{AB}) \otimes \mathbf{C}$. For two functions $f(x)$ and $g(x)$, the notation $f(x) = \mathcal{O}(g(x))$ means that $\lim_{x \rightarrow \infty} |f(x)|/|g(x)|$ is bounded above by a constant. We use $[n:m]$ to designate the set of natural numbers $\{n, n+1, \dots, m\}$. Let $\mathbf{g}(\mathbf{u})$ be a vector-valued function; then $\partial \mathbf{g} / \partial \mathbf{u}$ denotes the Jacobian matrix of the function $\mathbf{g}(\mathbf{u})$, i.e., the matrix that contains the partial derivative $\partial g_i / \partial u_j$ in its i th row and j th column. We write $|\mathcal{I}|$ to denote the cardinality of the set \mathcal{I} . For an $M \times N$ matrix \mathbf{A} , and two sets of indices $\mathcal{I} \subset [1:M]$ and $\mathcal{J} \subset [1:N]$, we use $[\mathbf{A}]_{\mathcal{I},\mathcal{J}}$ to denote the $|\mathcal{I}| \times |\mathcal{J}|$ submatrix of \mathbf{A} containing the elements $[a_{i,j}]_{i \in \mathcal{I}, j \in \mathcal{J}}$. Similarly, for an N -dimensional vector \mathbf{u} and a set $\mathcal{I} \subset [1:N]$, we define $\mathbf{u}_{\mathcal{I}} \triangleq [u_i]_{i \in \mathcal{I}}$. For an $M \times N$ matrix \mathbf{A} , we set $[\mathbf{A}]_{\diamond,\mathcal{J}} \triangleq [\mathbf{A}]_{[1:M],\mathcal{J}}$ and $[\mathbf{A}]_{\mathcal{I},\diamond} \triangleq [\mathbf{A}]_{\mathcal{I},[1:N]}$. Furthermore,

$$D(\mathbf{A}) \triangleq \begin{bmatrix} \text{diag}([a_{11} \dots a_{M1}]^{\top}) \\ \vdots \\ \text{diag}([a_{1N} \dots a_{MN}]^{\top}) \end{bmatrix}. \quad (1)$$

The eigenvalues of the $N \times N$ matrix \mathbf{A} are denoted by $\lambda_1(\mathbf{A}) \geq \dots \geq \lambda_N(\mathbf{A})$. The logarithm to the base 2 is written as $\log(\cdot)$. Finally, $\mathcal{CN}(\mathbf{m}, \mathbf{C})$ stands for the distribution of a jointly proper Gaussian (JPG) random vector with mean \mathbf{m} and covariance matrix \mathbf{C} .

II. SYSTEM MODEL

We consider a SIMO channel with Q receive antennas. The fading in each *component channel* follows the correlated block-fading model described in the previous section, namely, it is independent across blocks of length T , and correlated within

¹In the remainder of the paper, we consider the noncoherent setting only. Consequently, we will refer to capacity in the noncoherent setting simply as capacity. Furthermore, we shall assume Rayleigh fading throughout.

²When $Q = T$, capacity is known to grow double-logarithmically in SNR [7], and, hence, the capacity pre-log is zero.

³The constant block-fading model is obviously a special case ($Q = 1$) of the correlated block-fading model.

blocks, with the rank of the corresponding channel covariance matrix given by $Q < T$. Note that we assume the rank of the channel covariance matrix to be equal to the number of receive antennas. Our analysis relies heavily on this assumption. Across component channels, the fading is independent and identically distributed. The input-output (I/O) relation (within any block) for the m th component channel can be written as

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

where the vector $\mathbf{x} = [x_1 \cdots x_T]^T \in \mathbb{C}^T$ contains the T -dimensional signal transmitted within the block, and the vectors $\mathbf{y}_m, \mathbf{w}_m \in \mathbb{C}^T$ contain the corresponding received signal and additive noise, respectively, at the m th antenna. Finally, $\mathbf{h}_m \in \mathbb{C}^T$ contains the channel coefficients between the transmit antenna and the m th receive antenna. We assume that $\mathbf{w}_m \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_T)$ and $\mathbf{h}_m \sim \mathcal{CN}(\mathbf{0}, \mathbf{P}\mathbf{P}^H)$ are mutually independent (and independent across m) and that $\mathbf{P} \in \mathbb{C}^{T \times Q}$ (which is the same for all blocks) has rank $Q < T$. It will turn out convenient to write the channel-coefficient vector in whitened form as $\mathbf{h}_m = \mathbf{P}\mathbf{s}_m$, where $\mathbf{s}_m \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_Q)$. Finally, we assume that \mathbf{s}_m and \mathbf{w}_m change in an independent fashion from block to block.

If we define $\mathbf{y}^T \triangleq [\mathbf{y}_1^T \cdots \mathbf{y}_Q^T]$, $\mathbf{s}^T \triangleq [\mathbf{s}_1^T \cdots \mathbf{s}_Q^T]$, $\mathbf{w}^T \triangleq [\mathbf{w}_1^T \cdots \mathbf{w}_Q^T]$, and $\mathbf{X} \triangleq \text{diag}(\mathbf{x})$, we can write the channel I/O relation in the following—more compact—form

$$\mathbf{y} = \sqrt{\rho} (\mathbf{I}_Q \otimes \mathbf{X}\mathbf{P}) \mathbf{s} + \mathbf{w}. \quad (2)$$

The capacity of the channel (2) is defined as

$$C(\rho) \triangleq (1/T) \sup_{f_{\mathbf{x}}(\cdot)} I(\mathbf{x}; \mathbf{y}) \quad (3)$$

where the supremum is taken over all input distributions $f_{\mathbf{x}}(\cdot)$ that satisfy the average-power constraint $\mathbb{E}[\|\mathbf{x}\|^2] \leq T$.

III. INTUITIVE ANALYSIS

In this section, we describe a “back-of-the-envelope” method for guessing the capacity pre-log. A formal justification of this procedure is provided in Section IV.

The capacity pre-log characterizes the asymptotic behavior of the fading-channel capacity at high SNR, i.e., in the regime where the additive noise can “effectively” be ignored. In order to guess the capacity pre-log, we therefore consider the problem of identifying the transmit symbols x_i , $i \in [1:T]$, from the *noise-free* (and rescaled) observation

$$\hat{\mathbf{y}} \triangleq (\mathbf{I}_Q \otimes \mathbf{X}\mathbf{P}) \mathbf{s}. \quad (4)$$

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

We start by noting that the unknown variables in (4) 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 $z_i \triangleq 1/x_i$, $i \in [1:T]$, (we make the technical assumption $0 < |x_i| < \infty$, $i \in [1:T]$, in the remainder of this section) transforms (4) into a system of equations that is linear in \mathbf{s} and z_i , $i \in [1:T]$. Since the transformation $z_i \triangleq 1/x_i$ is invertible for $0 < x_i < \infty$,

uniqueness of the solution of the linear system of equations in \mathbf{s} and z_i , $i \in [1:T]$, is equivalent to uniqueness of the solution of the quadratic system of equations in \mathbf{s} and x_i , $i \in [1:T]$. For simplicity of exposition and concreteness, we consider the special case $T = 3$ and $Q = 2$. A direct computation reveals that (4) is equivalent to

$$\begin{bmatrix} p_{11} & p_{12} & 0 & 0 & \hat{y}_1 & 0 & 0 \\ p_{21} & p_{22} & 0 & 0 & 0 & \hat{y}_2 & 0 \\ p_{31} & p_{32} & 0 & 0 & 0 & 0 & \hat{y}_3 \\ 0 & 0 & p_{11} & p_{12} & \hat{y}_4 & 0 & 0 \\ 0 & 0 & p_{21} & p_{22} & 0 & \hat{y}_5 & 0 \\ 0 & 0 & p_{31} & p_{32} & 0 & 0 & \hat{y}_6 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \\ -z_1 \\ -z_2 \\ -z_3 \end{bmatrix} = \mathbf{0}. \quad (5)$$

The solution of this linear system of equations is not unique, as we have 6 equations in 7 unknowns. The $x_i = 1/z_i$, $i \in [1:3]$, can, therefore, not be determined uniquely from $\hat{\mathbf{y}}$. However, if we transmit one pilot symbol and two data symbols, the system of equations becomes solvable. Take for example $x_1 = 1$ and let the receiver know the value of this (pilot) symbol. Then (5) reduces to the following inhomogeneous system of 6 equations in 6 unknowns

$$\underbrace{\begin{bmatrix} p_{11} & p_{12} & 0 & 0 & 0 & 0 \\ p_{21} & p_{22} & 0 & 0 & \hat{y}_2 & 0 \\ p_{31} & p_{32} & 0 & 0 & 0 & \hat{y}_3 \\ 0 & 0 & p_{11} & p_{12} & 0 & 0 \\ 0 & 0 & p_{21} & p_{22} & \hat{y}_5 & 0 \\ 0 & 0 & p_{31} & p_{32} & 0 & \hat{y}_6 \end{bmatrix}}_{\mathbf{B}} \begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \\ -z_2 \\ -z_3 \end{bmatrix} = \begin{bmatrix} \hat{y}_1 \\ 0 \\ 0 \\ \hat{y}_4 \\ 0 \\ 0 \end{bmatrix}. \quad (6)$$

This system of equations has a unique solution if $\det \mathbf{B} \neq 0$. We prove in Appendix A that under the technical condition on \mathbf{P} specified in Theorem 1 below, we, indeed, have that $\det \mathbf{B} \neq 0$ for almost all⁴ $\hat{y}_2, \hat{y}_3, \hat{y}_5, \hat{y}_6$. It, therefore, follows that for almost all $\hat{\mathbf{y}}$, the system of equations (6) has a unique solution. Consequently, we can recover z_2 and z_3 , and, hence, $x_2 = 1/z_2$ and $x_3 = 1/z_3$. Summarizing our findings, we expect that the capacity pre-log of the channel (2), for the special case $T = 3$ and $Q = 2$, is equal to $2/3$. This is larger than the capacity pre-log of the corresponding SISO channel (i.e., the capacity pre-log of one of the component channels), which is equal to $1 - Q/T = 1/3$ [8].

In general, we expect that under some technical conditions on \mathbf{P} the capacity pre-log of the SIMO channel as defined in Section II is equal to $(T - 1)/T = 1 - 1/T$. This is exactly what we intend to show rigorously in the next section.

IV. A LOWER BOUND ON THE CAPACITY PRE-LOG

The main result of this paper is the following theorem.

Theorem 1: Assume that there exists a subset of indices $\mathcal{I} \subset [1:T]$ of cardinality $|\mathcal{I}| = Q + 1$ such that the $[(Q + 1) \times Q]$ -dimensional submatrix $\tilde{\mathbf{P}} \triangleq [\mathbf{P}]_{\mathcal{I}, \diamond}$ of the matrix \mathbf{P} in (2) satisfies the following *Property (A)*: Any set of Q rows of $\tilde{\mathbf{P}}$ is linearly independent. Then, the capacity of the SIMO channel (2) can be lower-bounded as

$$C(\rho) \geq (1 - 1/T) \log(\rho) + \mathcal{O}(1), \quad \rho \rightarrow \infty. \quad (7)$$

⁴Except for a set of measure zero.

Remark 1: For the special case $T = Q + 1$, (7) yields a lower bound on the capacity pre-log that is tight. A matching upper bound can be obtained through steps similar to those in the proof of [8, Prop. 4]. Establishing tight upper bounds on the capacity pre-log for general values of T , however, seems to be an open problem. Different tools than those used in [8] are probably needed.

Remark 2: When $Q = 1$, the channel in (2) reduces to a SISO constant block-fading channel, and the lower bound (7) yields the correct capacity pre-log of [4], [3].

Remark 3: Property (A) is not very restrictive and is satisfied by many practically relevant matrices \mathbf{P} . For example, removing any set of $T - Q$ columns from a $T \times T$ discrete Fourier transform (DFT) matrix, results in a matrix that satisfies Property (A) when T is prime [9]. DFT covariance matrices occur naturally in basis-expansion models for time-varying channels [8].

Proof: We choose an input distribution for which the entries x_i , $i \in [1:T]$, of \mathbf{x} , are independent and identically distributed (i.i.d.), have zero mean and unit variance, and satisfy $\mathbb{E}[\log |x_i|] > -\infty$ and $h(x_i) > -\infty$. For example, we can take $x_i \sim \mathcal{CN}(0, 1)$. We then lower-bound $I(\mathbf{x}; \mathbf{y})$ in (3), evaluated for this input distribution. More precisely, we use $I(\mathbf{x}; \mathbf{y}) = h(\mathbf{y}) - h(\mathbf{y} | \mathbf{x})$ and bound the two differential entropy terms separately. Note that the class of input distributions for which (7) holds is large. This does not come as a surprise, as we are interested in the capacity pre-log only.

As \mathbf{y} conditional on \mathbf{x} is JPG, the conditional differential entropy $h(\mathbf{y} | \mathbf{x})$ can be upper-bounded in a straightforward fashion as follows:

$$\begin{aligned} h(\mathbf{y} | \mathbf{x}) &= QT \log(\pi e) \\ &+ \mathbb{E}_{\mathbf{x}} [\log \det(\mathbf{I}_{QT} + \rho(\mathbf{I}_Q \otimes \mathbf{X}\mathbf{P})) \mathbb{E}_{\mathbf{s}} [\mathbf{s}\mathbf{s}^H] (\mathbf{I}_Q \otimes \mathbf{P}^H \mathbf{X}^H)] \\ &\leq Q^2 \log(\rho) + \mathcal{O}(1) \end{aligned} \quad (8)$$

where the inequality holds because \mathbf{P} has rank Q .

Finding a tight lower bound on $h(\mathbf{y})$ is the main difficulty of the proof. In fact, the differential entropy of \mathbf{y} is often intractable even for simple input distributions. The main technical contribution of this paper is presented in Section IV-A below, where we show that if Property (A) is satisfied and if the input distribution satisfies the conditions specified at the beginning of this proof, we have

$$h(\mathbf{y}) \geq (T - 1 + Q^2) \log(\rho) + c \quad (9)$$

where c , here and in the remainder of the paper, stands for a constant⁵ that is independent of ρ . Combining (8) and (9) then yields the desired result. Note that in order to establish (8) it is sufficient to use that \mathbf{P} has rank Q , whereas the more restrictive Property (A) is crucial to establish (9). ■

A. A Lower Bound on $h(\mathbf{y})$

The main idea of our approach is to relate $h(\mathbf{y})$ to $h(\mathbf{s}, \mathbf{x}) = h(\mathbf{s}) + h(\mathbf{x})$, which is generally much simpler to compute than $h(\mathbf{y})$. It is possible to relate the entropies of two random vectors in a simple way if the vectors are of the same dimension and are connected by a *deterministic one-to-one* (in the sense of [10, p.7]) function. This is not the case for \mathbf{y} and $[\mathbf{s}^T \mathbf{x}^T]^T$. It

is, however, possible to show that if x_1 is a fixed parameter, then there is a deterministic one-to-one function between $[\mathbf{s}^T \mathbf{x}_{[2:T]}^T]^T$ and a specific subset $\mathcal{J} \subset [1:TQ]$ of components of the noiseless version $\hat{\mathbf{y}}$ of the output vector. This allows us to relate $h(\hat{\mathbf{y}}_{\mathcal{J}} | x_1)$ to $h(\mathbf{s}, \mathbf{x}_{[2:T]} | x_1) = h(\mathbf{s}, \mathbf{x}_{[2:T]})$, which will turn out to be sufficient for our purposes as $h(\mathbf{y})$ can be linked to $h(\hat{\mathbf{y}}_{\mathcal{J}} | x_1)$ according to (12) below. We now describe the details of the proof program outlined above.

Lemma 2: Assume that the matrix \mathbf{P} satisfies the conditions of Theorem 1 and take the submatrix $\tilde{\mathbf{P}}$ defined in Theorem 1 to consist of the first $Q + 1$ rows of \mathbf{P} for simplicity.⁶ Let

$$\begin{aligned} \mathcal{J} \triangleq & [1:T] \cup [T+1:T+Q+1] \\ & \cup [2T+1:2T+Q+1] \cup \dots \\ & \cup [(Q-1)T+1:(Q-1)T+Q+1] \end{aligned} \quad (10)$$

where $|\mathcal{J}| = T - 1 + Q^2$, and consider the vector-valued function $\hat{\mathbf{y}}_{\mathcal{J}} : \mathbb{C}^{T-1+Q^2} \rightarrow \mathbb{C}^{T-1+Q^2}$

$$\hat{\mathbf{y}}_{\mathcal{J}}(\mathbf{s}, \mathbf{x}_{[2:T]}) = ((\mathbf{I}_Q \otimes \mathbf{X}\mathbf{P}) \mathbf{s})_{\mathcal{J}} \quad (11)$$

parametrized by $x_1 \neq 0$. To simplify the notation we will not indicate this parametrization explicitly. The function $\hat{\mathbf{y}}_{\mathcal{J}}(\cdot)$ is one-to-one almost everywhere (a.e.) on \mathbb{C}^{T-1+Q^2} .

Proof: See Appendix A. ■

The following comments on Lemma 2 are in order. For $T = 3$ and $Q = 2$ as in the simple example in Section III, $\mathcal{J} = [1:6]$, so that $\hat{\mathbf{y}}_{\mathcal{J}} = \hat{\mathbf{y}}$. Therefore, the one-to-one correspondence established in this lemma simply means that (4) has a unique solution for fixed $x_1 \neq 0$. For the proof of the lemma, it is crucial that $x_1 \neq 0$ is fixed. In fact, one can check that if none of the components of \mathbf{x} is fixed, the resulting equivalent of the function $\hat{\mathbf{y}}_{\mathcal{J}}(\cdot)$ cannot be one-to-one, no matter how the set \mathcal{J} is chosen. Fixing $x_1 \neq 0$ in order to make the function $\hat{\mathbf{y}}_{\mathcal{J}}(\cdot)$ be one-to-one corresponds to transmitting a pilot, as done in the simple example in Section III by setting $x_1 = 1$. The cardinality of the set \mathcal{J} , which determines the lower bound on the capacity pre-log, as we shall see below, is dictated by the requirement that $\hat{\mathbf{y}}_{\mathcal{J}}$ and $[\mathbf{s}^T \mathbf{x}_{[2:T]}^T]^T$ are of the same dimension, which implies that \mathcal{J} must contain $T - 1 + Q^2$ elements. The specific choice of \mathcal{J} in (10) simplifies the proof of the lemma (see Appendix B).

Lemma 2 can be used to relate the conditional differential entropy $h(\hat{\mathbf{y}}_{\mathcal{J}} | x_1)$ to $h(\mathbf{s}, \mathbf{x}_{[2:T]})$. Before doing so, we establish a simple lower bound on $h(\mathbf{y})$ that is explicit in $h(\hat{\mathbf{y}}_{\mathcal{J}} | x_1)$. Let \mathcal{N} be the complement of \mathcal{J} in $[1:QT]$. Then

$$\begin{aligned} h(\mathbf{y}) &= h(\mathbf{y}_{\mathcal{J}}, \mathbf{y}_{\mathcal{N}}) = h(\mathbf{y}_{\mathcal{J}}) + h(\mathbf{y}_{\mathcal{N}} | \mathbf{y}_{\mathcal{J}}) \\ &\geq h(\sqrt{\rho} \hat{\mathbf{y}}_{\mathcal{J}} + \mathbf{w}_{\mathcal{J}}) + h(\mathbf{y}_{\mathcal{N}} | \mathbf{s}, \mathbf{x}, \mathbf{y}_{\mathcal{J}}) \\ &\geq h(\sqrt{\rho} \hat{\mathbf{y}}_{\mathcal{J}} + \mathbf{w}_{\mathcal{J}} | \mathbf{w}_{\mathcal{J}}) + h(\mathbf{w}_{\mathcal{N}}) \\ &\geq |\mathcal{J}| \log(\rho) + h(\hat{\mathbf{y}}_{\mathcal{J}} | x_1) + c. \end{aligned} \quad (12)$$

Through this chain of inequalities, we got rid of the noise \mathbf{w} . This corresponds to considering the noise-free I/O relation (4) in the intuitive explanation given in Section III. Inserting $|\mathcal{J}| = T - 1 + Q^2$ into (12), we obtain the desired result (9) provided

⁶This assumption will be made in the remainder of the paper, without explicitly mentioning it again.

⁵The value of this constant can change at each appearance.

that $h(\hat{\mathbf{y}}_{\mathcal{J}} | x_1)$ is finite, which will be proved by means of the following lemma.

Lemma 3 (Transformation of differential entropy): Assume that $\mathbf{g} : \mathbb{C}^N \rightarrow \mathbb{C}^N$ is a continuous vector-valued function that is one-to-one a.e. on \mathbb{C}^N . Let $\mathbf{u} \in \mathbb{C}^N$ be a random vector and $\mathbf{v} = \mathbf{g}(\mathbf{u})$. Then

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

Proof: The proof follows from the change-of-variable theorem for integrals [10, Thm. 7.26]. \blacksquare

Let $f_{x_1}(x)$ denote the density of x_1 . Then

$$\begin{aligned} h(\hat{\mathbf{y}}_{\mathcal{J}} | x_1) &= \int f_{x_1}(x) h(\hat{\mathbf{y}}_{\mathcal{J}} | x_1 = x) dx \\ &= h(\mathbf{s}, \mathbf{x}_{[2:T]} | x_1) + \mathbb{E}_{\mathbf{s}, \mathbf{x}} \left[\log \left| \det \frac{\partial \hat{\mathbf{y}}_{\mathcal{J}}}{\partial (\mathbf{s}, \mathbf{x}_{[2:T]})} \right| \right] \end{aligned} \quad (13)$$

where in the second equality we applied Lemma 3 to $h(\hat{\mathbf{y}}_{\mathcal{J}} | x_1 = x)$, using that the function $\hat{\mathbf{y}}_{\mathcal{J}}(\cdot)$ in (11) is continuous and is one-to-one a.e. as shown in Lemma 2. The first term on the RHS of (13) satisfies

$$h(\mathbf{s}, \mathbf{x}_{[2:T]} | x_1) = h(\mathbf{s}) + h(\mathbf{x}_{[2:T]}) > -\infty$$

where the inequality follows because the $x_i, i \in [2:T]$, are i.i.d. and have finite differential entropy. It therefore remains to show that the second term on the RHS of (13) is finite as well. As the RHS of (11) is linear in \mathbf{s} , we have that

$$\partial \hat{\mathbf{y}}_{\mathcal{J}} / \partial \mathbf{s} = [\mathbf{I}_Q \otimes \mathbf{X}\mathbf{P}]_{\mathcal{J}, \diamond}.$$

Furthermore, using [11, Eq. (5), Sec. 7.2], the RHS of (11) can be rewritten as $\hat{\mathbf{y}}_{\mathcal{J}}(\mathbf{s}, \mathbf{x}_{[2:T]}) = ((\mathbf{S} \otimes \mathbf{I}_T)D(\mathbf{P})\mathbf{x})_{\mathcal{J}}$, where the operator $D(\cdot)$ was defined in (1) and

$$\mathbf{S}^T = [\mathbf{s}_1 \dots \mathbf{s}_Q]. \quad (14)$$

Hence, we have that

$$\partial \hat{\mathbf{y}}_{\mathcal{J}} / \partial \mathbf{x}_{[2:T]} = [(\mathbf{S} \otimes \mathbf{I}_T)D(\mathbf{P})]_{\mathcal{J}, [2:T]}.$$

To summarize, the Jacobian matrix in (13) is given by

$$\frac{\partial \hat{\mathbf{y}}_{\mathcal{J}}}{\partial (\mathbf{s}, \mathbf{x}_{[2:T]})} = [[\mathbf{I}_Q \otimes \mathbf{X}\mathbf{P}]_{\mathcal{J}, \diamond} \quad [(\mathbf{S} \otimes \mathbf{I}_T)D(\mathbf{P})]_{\mathcal{J}, [2:T]}]. \quad (15)$$

As shown in Appendix B, the determinant of this Jacobian matrix can be factorized as follows

$$\begin{aligned} &|\det [[\mathbf{I}_Q \otimes \mathbf{X}\mathbf{P}]_{\mathcal{J}, \diamond} \quad [(\mathbf{S} \otimes \mathbf{I}_T)D(\mathbf{P})]_{\mathcal{J}, [2:T]}]| \\ &= \prod_{j \in [Q+2:T]} \left| \sum_{q=1}^Q s_{1q} p_{jq} \right| |\det \mathbf{M}_1(\mathbf{X})| |\det \mathbf{M}_2(\mathbf{S})| \\ &\quad \times |\det \mathbf{M}_3(\mathbf{P})| |\det \mathbf{M}_4(\mathbf{S})| |\det \mathbf{M}_5(\mathbf{X})| \end{aligned} \quad (16)$$

where

$$\begin{aligned} \mathbf{M}_1(\mathbf{X}) &\triangleq \mathbf{I}_Q \otimes [\mathbf{X}]_{[1:Q+1], [1:Q+1]} \\ \mathbf{M}_2(\mathbf{S}) &\triangleq \mathbf{S} \otimes \mathbf{I}_{Q+1} \\ \mathbf{M}_3(\mathbf{P}) &\triangleq \left[\begin{array}{c} \mathbf{I}_Q \otimes \tilde{\mathbf{P}} \\ D(\tilde{\mathbf{P}}) \end{array} \right]_{\diamond, [2:Q+1]} \end{aligned}$$

$$\begin{aligned} \mathbf{M}_4(\mathbf{S}) &\triangleq \begin{bmatrix} (\mathbf{S}^{-1} \otimes \mathbf{I}_Q) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_Q \end{bmatrix} \\ \mathbf{M}_5(\mathbf{X}) &\triangleq \begin{bmatrix} \mathbf{I}_{Q^2} & \mathbf{0} \\ \mathbf{0} & [\mathbf{X}]_{[2:Q+1], [2:Q+1]}^{-1} \end{bmatrix}. \end{aligned}$$

Hence, we can rewrite the second term on the RHS of (13) as

$$\begin{aligned} &\mathbb{E}_{\mathbf{s}, \mathbf{x}} \left[\log \left| \det \frac{\partial \hat{\mathbf{y}}_{\mathcal{J}}}{\partial (\mathbf{s}, \mathbf{x}_{[2:T]})} \right| \right] \\ &= \sum_{j \in [Q+2:T]} \mathbb{E}_{\mathbf{s}} \left[\log \left| \sum_{q=1}^Q s_{1q} p_{jq} \right| \right] + \mathbb{E}_{\mathbf{x}} [\log |\det \mathbf{M}_1(\mathbf{X})|] \\ &\quad + \mathbb{E}_{\mathbf{s}} [\log |\det \mathbf{M}_2(\mathbf{S})|] + \log |\det \mathbf{M}_3(\mathbf{P})| \\ &\quad + \mathbb{E}_{\mathbf{s}} [\log |\det \mathbf{M}_4(\mathbf{S})|] + \mathbb{E}_{\mathbf{x}} [\log |\det \mathbf{M}_5(\mathbf{X})|]. \end{aligned} \quad (17)$$

The first and the third term on the RHS of (17) are finite because \mathbf{S} has i.i.d. Gaussian components. The fifth term is finite for the same reason, because $\log |\det \mathbf{S}^{-1}| = -\log |\det \mathbf{S}|$. The second and the sixth term are finite because $\mathbb{E}[\log |x_i|] > -\infty, i \in [1:T]$, by assumption. Finally, we show in Appendix C that the matrix $\mathbf{M}_3(\mathbf{P})$ has full rank if Property (A) is satisfied. This then implies that the fourth term on the RHS of (17) is also finite.

V. CONCLUSION AND FURTHER WORK

In this paper, we analyzed the noncoherent-capacity pre-log of a temporally correlated block-fading channel. We showed that, surprisingly, the capacity pre-log in the SIMO case can be larger than that in the SISO case. This result was established for the special case of the number of receive antennas being equal to the rank of the channel covariance matrix. Interesting open issues include extending the lower bound in Theorem 1 to an arbitrary number of receive antennas and finding a tight upper bound on the capacity pre-log.

APPENDIX A

Proof of Lemma 2: We need to show that the function $\hat{\mathbf{y}}_{\mathcal{J}}(\mathbf{s}, \mathbf{x}_{[2:T]})$ is one-to-one a.e. Hence, we can exclude sets of measure zero from its domain. In particular, we shall consider the restriction of the function $\hat{\mathbf{y}}_{\mathcal{J}}(\mathbf{s}, \mathbf{x}_{[2:T]})$ to the set of pairs $(\mathbf{s}, \mathbf{x}_{[2:T]})$, which satisfy (i) $0 < |x_i| < \infty$ for all $i \in [2:T]$; (ii) the matrix \mathbf{S} , defined in (14) is invertible; (iii) the sum $\sum_{q=1}^Q s_{1q} p_{jq}$ is nonzero for all $j \in [Q+2:T]$.

To show that this restriction of the function $\hat{\mathbf{y}}_{\mathcal{J}}(\cdot)$ [which, with slight abuse of notation we still call $\hat{\mathbf{y}}_{\mathcal{J}}(\cdot)$] is one-to-one, we take an element $\tilde{\mathbf{y}}$ from its range and prove that the equation

$$\hat{\mathbf{y}}_{\mathcal{J}}(\mathbf{s}', \mathbf{x}'_{[2:T]}) = \tilde{\mathbf{y}} \quad (18)$$

has a unique solution in the set of pairs $(\mathbf{s}', \mathbf{x}'_{[2:T]})$ satisfying the constraints (i)–(iii). The element $\tilde{\mathbf{y}}$ can be represented as $\tilde{\mathbf{y}} = ((\mathbf{I}_Q \otimes \mathbf{X}\mathbf{P})\mathbf{s})_{\mathcal{J}}$, with $\mathbf{X} = \text{diag}([x_1 \mathbf{x}'_{[2:T]}]^T)$, where $(\mathbf{s}, \mathbf{x}_{[2:T]})$ satisfies the constraints (i)–(iii). Hence, (18) can be rewritten in the following way

$$((\mathbf{I}_Q \otimes \mathbf{X}'\mathbf{P})\mathbf{s}')_{\mathcal{J}} = ((\mathbf{I}_Q \otimes \mathbf{X}\mathbf{P})\mathbf{s})_{\mathcal{J}} \quad (19)$$

with $\mathbf{X}' = \text{diag}([x_1 (\mathbf{x}'_{[2:T]})^T]^T)$. To prove that (19) has a unique solution, we follow the approach described in Section III and convert (19) into a linear system of equations through

a change of variables. In particular, thanks to constraint (i), we can multiply both sides of (19) by $[\mathbf{I}_Q \otimes \mathbf{X}']_{\mathcal{J},\mathcal{J}}^{-1}$ and by $[\mathbf{I}_Q \otimes \mathbf{X}]_{\mathcal{J},\mathcal{J}}^{-1}$ and perform the substitution $z'_i = 1/x'_i$, $i \in [2:T]$. Then, we define $\mathbf{z}' = [z'_2 \dots z'_T]^\top$ and manipulate the equation such that all the unknowns are on one side of the equation and all the terms depending on the constant x'_1 are on the other side. These steps yield the following inhomogeneous linear system of equations

$$\mathbf{G} \begin{bmatrix} \mathbf{s}' \\ -\mathbf{z}' \end{bmatrix} = \frac{1}{x'_1} \mathbf{u} \quad (20)$$

where

$$\mathbf{G} = \left[[\mathbf{I}_Q \otimes \mathbf{X}^{-1} \mathbf{P}]_{\mathcal{J},\diamond} \quad [(\mathbf{S} \otimes \mathbf{I}_T) D(\mathbf{P})]_{\mathcal{J},[2:T]} \right] \quad (21)$$

$$\mathbf{u} = \left[\mathbf{p}_1^\top \mathbf{s}_1 \quad \underbrace{0 \dots 0}_{T-1 \text{ times}} \quad \mathbf{p}_1^\top \mathbf{s}_2 \quad \underbrace{0 \dots 0}_Q \dots \mathbf{p}_1^\top \mathbf{s}_Q \quad \underbrace{0 \dots 0}_Q \right]^\top$$

and \mathbf{p}_1^\top denotes the first row of \mathbf{P} . The solution of (20) is unique if and only if $\det \mathbf{G} \neq 0$. To establish this, it is useful to note that the matrix \mathbf{G} has a structure similar to the matrix on the RHS of (15) [the only difference is that \mathbf{X} in (15) is replaced by \mathbf{X}^{-1} in \mathbf{G}]. Therefore, we can factorize $\det \mathbf{G}$ in exactly the same way as the RHS of (15). Finally, we invoke the constraints (i), (ii), and (iii) together with Property (A) to conclude that each term in the resulting factorization is nonzero. This completes the proof.

We point out that for $T = 3$ and $Q = 2$, the matrix \mathbf{B} defined in (6) is related to \mathbf{G} in (21) according to $\mathbf{B} = (\mathbf{I}_2 \otimes \mathbf{X}) \mathbf{G}$, and therefore $\det \mathbf{B} \neq 0$ a.e., as claimed in Section III.

APPENDIX B

Proof of (16): As a consequence of the choice of \mathcal{J} in (10), each of the last $T - Q - 1$ columns of the matrix on the RHS of (15) has exactly one nonzero element. This allows us to use the Laplace formula to expand the determinant along these columns iteratively to get

$$\begin{aligned} & \left| \det \left[[\mathbf{I}_Q \otimes \mathbf{X} \mathbf{P}]_{\mathcal{J},\diamond} \quad [(\mathbf{S} \otimes \mathbf{I}_T) D(\mathbf{P})]_{\mathcal{J},[2:T]} \right] \right| \\ &= \prod_{j \in [Q+2:T]} \left| \sum_{q=1}^Q s_{1q} p_{jq} \right| |\det(\mathbf{E})| \quad (22) \end{aligned}$$

where

$$\mathbf{E} = [\mathbf{I}_Q \otimes \tilde{\mathbf{X}} \tilde{\mathbf{P}} \quad (\mathbf{S} \otimes \mathbf{I}_{Q+1}) \tilde{\mathbf{D}}]$$

$$\tilde{\mathbf{D}} = [D(\tilde{\mathbf{P}})]_{\diamond,[2:Q+1]}, \quad \tilde{\mathbf{X}} = [\mathbf{X}]_{[1:Q+1],[1:Q+1]}$$

Next, using simple properties of the Kronecker product and exploiting the block-diagonal structure of $\tilde{\mathbf{D}}$, we factorize \mathbf{E} into a product of simple terms:

$$\begin{aligned} \mathbf{E} &= (\mathbf{I}_Q \otimes \tilde{\mathbf{X}}) (\mathbf{S} \otimes \mathbf{I}_{Q+1}) \left[(\mathbf{I}_Q \otimes \tilde{\mathbf{P}}) \quad \tilde{\mathbf{D}} \right] \\ &\times \begin{bmatrix} (\mathbf{S}^{-1} \otimes \mathbf{I}_Q) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_Q \end{bmatrix} \begin{bmatrix} \mathbf{I}_{Q^2} & \mathbf{0} \\ \mathbf{0} & [\tilde{\mathbf{X}}]_{[2:Q+1],[2:Q+1]}^{-1} \end{bmatrix}. \quad (23) \end{aligned}$$

The proof is completed by inserting (23) into (22), and using the multiplicativity of the determinant.

APPENDIX C

Lemma 4: Let \mathbf{A} be an $(N+1) \times N$ matrix. If any set of N rows of \mathbf{A} is linearly independent, then the $N(N+1) \times N(N+1)$ matrix $\hat{\mathbf{A}}$ defined as $\hat{\mathbf{A}} = [(\mathbf{I}_N \otimes \mathbf{A}) \quad [D(\mathbf{A})]_{\diamond,[2:N+1]}]$ has full rank.

Proof: The proof is by contradiction. Assume that $\hat{\mathbf{A}}$ does not have full rank. Then there exists an $N(N+1)$ -dimensional nonzero vector $\mathbf{u}^\top = [\mathbf{u}_1^\top \dots \mathbf{u}_N^\top]$, where $\mathbf{u}_n \in \mathbb{C}^{N+1}$, such that $\mathbf{u}^\top \hat{\mathbf{A}} = \mathbf{0}$. Because $\hat{\mathbf{A}} = [(\mathbf{I}_N \otimes \mathbf{A}) \quad [D(\mathbf{A})]_{\diamond,[2:N+1]}]$, we have in particular that (i) $\mathbf{u}^\top (\mathbf{I}_N \otimes \mathbf{A}) = \mathbf{0}$ and (ii) $\mathbf{u}^\top [D(\mathbf{A})]_{\diamond,[2:N+1]} = \mathbf{0}$. We next analyze these two equalities separately. Equality (i) can be restated as $\mathbf{u}_n^\top \mathbf{A} = \mathbf{0}$ for all $n \in [1:N]$, which implies that all vectors \mathbf{u}_n lie in the kernel of the $N \times (N+1)$ matrix \mathbf{A}^\top . Because \mathbf{A}^\top has rank N , its kernel must be of dimension 1. Hence, all vectors \mathbf{u}_n must be collinear, i.e., there exists a vector \mathbf{v} and a set of N constants c_n such that $\mathbf{u}_n = c_n \mathbf{v}$, for all $n \in [1:N]$. The vector \mathbf{v} and at least one of the constants c_n must be nonzero because \mathbf{u} is nonzero. Furthermore, because $\mathbf{v}^\top \mathbf{A} = \mathbf{0}$, and because any set of N rows of \mathbf{A} is linearly independent by assumption, all components of \mathbf{v} must be nonzero.

We now use this property of \mathbf{v} to analyze equality (ii), which can be restated as

$$\mathbf{u}^\top [D(\mathbf{A})]_{\diamond,[2:N+1]} = [c_1 \mathbf{v}^\top \dots c_N \mathbf{v}^\top] [D(\mathbf{A})]_{\diamond,[2:N+1]} = \mathbf{0}$$

or, after straightforward manipulations, as

$$[\text{diag}(\mathbf{v})]_{[2:N+1],[2:N+1]} [\mathbf{A}]_{[2:N+1],\diamond} [c_1 \dots c_N]^\top = \mathbf{0}.$$

Because all the components of \mathbf{v} are nonzero, this last equality implies that $[\mathbf{A}]_{[2:N+1],\diamond} [c_1 \dots c_N]^\top = \mathbf{0}$. However, this contradicts the assumption that any set of N rows of \mathbf{A} is linearly independent (recall that at least one of the constants c_n is nonzero). Hence, $\hat{\mathbf{A}}$ must have full rank. ■

REFERENCES

- [1] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov. 1999.
- [2] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 139–157, Jan. 1999.
- [3] 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.
- [4] 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.
- [5] A. Lapidoth, "On the asymptotic capacity of stationary Gaussian fading channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 437–446, Feb. 2005.
- [6] T. Koch, *On heating up and fading in communication channels*, ser. Information Theory and its Applications, A. Lapidoth, Ed. Konstanz, Germany: Hartung-Gorre Verlag, May 2009, vol. 5.
- [7] 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.
- [8] 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.
- [9] T. Tao, "An uncertainty principle for cyclic groups of prime order," *Math. Res. Lett.*, vol. 12, no. 1, pp. 121–127, 2005.
- [10] W. Rudin, *Real and Complex Analysis*, 3rd ed. New York, NY, USA: McGraw-Hill, 1987.
- [11] H. Lütkepohl, *Handbook of Matrices*. Chichester, U.K.: Wiley, 1996.