The Ergodic High SNR Capacity of the Spatially-Correlated Non-Coherent MIMO Channel Within an SNR-Independent Gap

Ramy H. Gohary and Halim Yanikomeroglu

Department of Systems and Computer Engineering, Carleton University, Ottawa, ON, Canada

[1] *Abstract*—**The ergodic capacity of spatially-correlated non-coherent multiple-input multiple-output channels is not known. In this paper upper and lower bounds are derived for this capacity at asymptotically high signal-to-noise ratios (SNRs). The bounds are accurate within an approximation error that decays as 1/SNR, and the gap between these bounds depends solely on the signalling dimensions and the condition number of the transmitter correlation matrix. The upper bound on the high SNR ergodic capacity is shown to decrease monotonically with the logarithm of the condition number of the transmitter correlation matrix. Furthermore, the lower bound on this capacity is achieved by input signals in the form of the product of an isotropically distributed random Grassmannian component and a deterministic component comprising the eigenvectors and the inverse of the eigenvalues of the transmitter correlation matrix.**

I. INTRODUCTION

Multiple-input multiple-output (MIMO) communication systems can be classified into coherent systems [1] in which the receiver has access to reliable channel state information (CSI) and non-coherent systems [2] in which no CSI is available at either the transmitter or the receiver. Coherent systems are more straightforward to design [3] than their non-coherent counterparts. However, the analysis of coherent systems usually does not account for the cost of the resources that have to be expended to acquire reliable CSI. While ignoring this cost in static and slow fading scenarios is generally tolerable, ignoring it in fast fading scenarios arising in various wireless systems, including millimeter wave ones, can be rather misleading [4].

To alleviate the cost of acquiring reliable CSI in fast fading scenarios, it is desirable to use non-coherent signalling strategies, that is, strategies that do not require the receiver to have access to CSI. In [2] it was shown that, for spatially-white MIMO channels with independent identically distributed (i. i. d.) block Rayleigh fading coefficients, the input matrices that achieve the ergodic capacity can be expressed in the form of an isotropically distributed unitary component and a diagonal component with non-negative entries. A closed form expression for the asymptotically high signal-to-noise ratio (SNR) ergodic capacity of this channel was obtained in [5] when the coherence time of the channel exceeds a certain

threshold and in [6] when the coherence time is below that threshold. For the case in [5], the high SNR ergodic capacity is achieved when the diagonal component of the input signal is a scaled identity matrix and the unitary matrix component is isotropically distributed on the so-called Grassmann manifold. Various techniques for designing rate-efficient Grassmannian constellations can be found in [7] and [8]. For the case in [6], the unitary component was shown to have the same distribution as that in [5], but the optimal distribution of the diagonal component was shown to be that of the square root of the eigenvalues of a beta-distributed random matrix. In contrast with the high SNR cases, for the low SNR ones, the optimal diagonal component has at most one nonzero entry [9].

The proximity of antennas in practical wireless systems, including prospective massive MIMO ones, renders the realization of spatially-white channels generally difficult, even when the antenna spacing exceeds multiple wavelengths. This proximity induces correlation between the random entries of the channel matrix at the transmitter and receiver sides. A convenient means for characterizing this correlation is the so-called Kronecker model, whereby the spatially-correlated channel matrix is represented by a spatially-white matrix that is left and right multiplied with transmitter and receiver correlation matrices, respectively. The correlation matrices are dominated by the statistical properties of the propagation environment, and the locations, geometries and beam patterns of the transmit and receive antennas [10]. As such, these matrices vary much more slowly than the instantaneous channel parameters, and can therefore be accurately estimated and made available to the transmitter and the receiver. As will be shown herein, the impact of these matrices on the signalling methodology and the achievable rate can be significant. The case in which the correlation between the channel coefficients is described by the Kronecker model in non-coherent communications was considered in [11] for arbitrary SNRs. Therein it was shown that, in contrast with spatially-white channels [2], [5], in spatially-correlated ones the optimal input covariance can be expressed as the product of an isotropically distributed unitary component, a non-negative diagonal component and a deterministic component comprising the eigenvectors of the transmitter correlation matrix.

Although non-coherent MIMO channels that arise in practice are likely to be spatially-correlated, no closed-form ex-

pressions or bounds are currently available for their capacity, neither is there a constructive signalling strategy that approaches it. To investigate the impact of spatial correlation, we herein derive an expression for the ergodic high SNR non-coherent channel capacity for the case of in which the channel is block Rayleigh fading and the correlation is described by the Kronecker model with known transmitter and receiver correlation matrices [11]. The capacity expression that we derive is accurate within an SNR-independent gap (cf. Theorem 1 below) and an approximation error that decays with $1/\text{SNR}$ [12]. We also derive an upper bound on the gap to capacity and we show that this gap decreases monotonically with the logarithm of the transmit condition number. Finally, it is shown that the input signals that achieve the asymptotically-tight capacity is in the form of the product of an isotropically distributed random Grassmannian component and a deterministic component comprising the eigenvectors and the inverse of the eigenvalues of the transmitter correlation matrix.

Notation: Random and deterministic matrices will be denoted by boldface and regular face upper case letters, respectively. The $m \times m$ identity matrix will be denoted by I_m, and the $O(\cdot)$ notation will be used to imply that $g_1(x) = O\big(g_2(x)\big)$ if $g_1(x)/g_2(x) \to a$ as $x \to \infty$, a is a constant.

II. SYSTEM MODEL

We consider a frequency-flat block Rayleigh fading channel in which the signals emitted from the transmit antennas are correlated and those impinging on the receiver antennas are also correlated. This situation arises in practice when the antenna elements are not sufficiently spaced for the channel gains to be statistically independent. Denoting the number of transmit and receive antennas by M and N, respectively, the channel matrix can be expressed as $\boldsymbol{H} = A^{1/2}\boldsymbol{H}_w B^{1/2}$, where $A \in \mathbb{C}^{M \times M}$ and $B \in \mathbb{C}^{N \times N}$ are the Hermitian positive semidefinite transmitter and receiver correlation matrices, respectively, and $\boldsymbol{H}_w \in \mathbb{C}^{M \times N}$ is a random matrix with zero-mean unit-variance i. i. d. circularly-symmetric complex Gaussian entries. To avoid degenerate scenarios, we will assume that both A and B are full rank.

In the block fading model, the channel assumes a statistically independent realization over each block of T channel uses and remains essentially constant within each block. This model is realistic in perfectly interleaved frequency-hopping systems [5] and is useful in characterizing high mobility scenarios in current Long Term Evolution (LTE) cellular systems. In such cases, the received signal matrix can be expressed as

$$\boldsymbol{Y} = \boldsymbol{X}A^{1/2}\boldsymbol{H}_w B^{1/2} + \boldsymbol{V}, \tag{1}$$

where $\boldsymbol{X} \in \mathbb{C}^{T \times M}$ represents the transmitted signal matrix, and $\boldsymbol{V} \in \mathbb{C}^{T \times N}$ represents the additive noise matrix; the entries of $\boldsymbol{V}$ are i. i. d. zero mean unit variance circularly-symmetric complex Gaussian random variables. The matrices A and B are normalized so that $\text{Tr}(A) = \text{Tr}(B) = 1$, and, assuming that the power budget of the transmitter is P, we have the following power constraint:

$$\text{E}\{\text{Tr}(\boldsymbol{X}\boldsymbol{X}^\dagger)\} \leq TP. \tag{2}$$

The correlation matrices A and B can be estimated from the physical properties of the propagation environment and the antenna pattern [13], and can hence be assumed to be known. The communication scenario that we consider is non-coherent, thereby the matrix $\boldsymbol{H}_w$ in (1) is not known to either the transmitter or the receiver.

III. BACKGROUND

Let $\text{vec}(\cdot)$ be the operator that stacks the columns of the matrix argument on top of each other [14]. Conditioned on $\boldsymbol{X}$, the matrix $\boldsymbol{Y}$ in (1) is Gaussian distributed with $\text{vec}(\boldsymbol{Y}) = (B^{1/2} \otimes \boldsymbol{X}A^{1/2})\,\text{vec}(\boldsymbol{H}_w) + \text{vec}(\boldsymbol{V})$, and

$$\text{E}\{\text{vec}(\boldsymbol{Y})\,\text{vec}^\dagger(\boldsymbol{Y})|\boldsymbol{X}\} = B \otimes \boldsymbol{X}A\boldsymbol{X}^\dagger + I_{NT}, \tag{3}$$

which implies that

$$p(\boldsymbol{Y}|\boldsymbol{X}) =$$
$$\frac{\exp\Big(-\text{vec}^\dagger(\boldsymbol{Y})\big(B \otimes \boldsymbol{X}A\boldsymbol{X}^\dagger + I_{NT}\big)^{-1}\text{vec}(\boldsymbol{Y})\Big)}{\pi^{TN}\det\big(B \otimes \boldsymbol{X}A\boldsymbol{X}^\dagger + I_{NT}\big)}. \tag{4}$$

Using this expression with the mixed product property in [14], it can be readily verified that for any deterministic $T \times T$ unitary matrix Φ, $p(\Phi\boldsymbol{Y}|\Phi\boldsymbol{X}) = p(\boldsymbol{Y}|\boldsymbol{X})$, which satisfies Property 4 in [2]. This with Lemma 1 in [2] were used in [11] to show that the matrix $\boldsymbol{X}$ that achieves the ergodic non-coherent capacity of the channel in (1) can be expressed as

$$\boldsymbol{X} = \boldsymbol{Q}_X D U_A^\dagger, \tag{5}$$

where $\boldsymbol{Q}_X$ is an isotropically distributed[2] unitary matrix, $\boldsymbol{D}$ is a random diagonal matrix with non-negative entries and U_A is the matrix containing the eigenvectors of A, i.e., $A = U_A \Lambda_A U_A^\dagger$, where Λ_A is the diagonal matrix containing the eigenvalues of A. Building on [11], we derive an asymptotically tight expression for the ergodic non-coherent capacity of the channel in (1) as P goes to infinity.

IV. THE ERGODIC HIGH SNR NON-COHERENT CAPACITY WITHIN AN SNR-INDEPENDENT GAP

Let $C(P)$ denote the ergodic non-coherent capacity of the channel in (1), for a given power budget P in (2). Hence,

$$C(P) = \max_{p(\boldsymbol{X}),\ \text{E}\{\text{Tr}(\boldsymbol{X}\boldsymbol{X}^\dagger)\} \leq TP} \frac{1}{T}\big(h(\boldsymbol{Y}) - h(\boldsymbol{Y}|\boldsymbol{X})\big). \tag{6}$$

Our goal is to evaluate $C(P)$ as P goes to infinity.

A. Evaluating $h(\boldsymbol{Y}|\boldsymbol{X})$

To evaluate the second term in (6), we use (3) and the fact that, conditioned on $\boldsymbol{X}$, $\boldsymbol{Y}$ is Gaussian distributed to write

$$h(\boldsymbol{Y}|\boldsymbol{X}) = NT\log\pi e + \text{E}\{\log\det\big(B \otimes \boldsymbol{X}A\boldsymbol{X}^\dagger + I_{NT}\big)\}$$

$$= NT\log\pi e + \sum_{i=1}^{M}\sum_{j=1}^{N}\text{E}\{\log(\lambda_{B_j}\lambda_{A_i}d_i^2 + 1)\}, \tag{7}$$

---

[2]An $m \times n$ matrix $\boldsymbol{Q}$ with $m \geq n$ is said to be isotropically distributed if $p(\boldsymbol{Q}) = p(\Phi\boldsymbol{Q})$ for any $m \times m$ deterministic unitary matrix Φ.

where in writing (7) we invoked (5) and the eigen properties of the Kronecker product [15] and used λ_{A_i} and λ_{B_j} to denote the i-th and j-th eigenvalues of A and B, respectively, and d_i to denote the i-th diagonal entry of $\boldsymbol{D}$. As the transmit power goes to infinity, it can be seen that (7) can be expressed as

$$
\begin{aligned}
h(\boldsymbol{Y}|\boldsymbol{X}) &= NT\log\pi e + \sum_{i=1}^{M}\sum_{j=1}^{N}\mathrm{E}\{\log\lambda_{B_j}\lambda_{A_i}d_i^2\} + O(1/P)\\
&= NT\log\pi e + N\log\det A + M\log\det B\\
&\quad + N\,\mathrm{E}\{\log\det\boldsymbol{D}^2\} + O(1/P).
\end{aligned}
\tag{8}
$$

This approximation is valid only when $\boldsymbol{D}$ is full rank and its entries scale with P. This condition will be shown to hold for the matrix $\boldsymbol{D}$ to achieve the upper and lower bounds on the high SNR ergodic capacity derived in Section IV-C below. For simplicity, we will henceforth focus on the case of equal number of transmit and receive antennas, i.e., $M = N$, and the coherence time $T \geq 2M$.

B. Asymptotic Bound on $h(\boldsymbol{Y})$

To provide a high SNR approximation of the first term in (6), we will invoke the results in [5, Section III-B] for the case of $M = N$ and $T \geq 2M$.

As P goes to infinity, the differential entropy, $h(\boldsymbol{Y})$, is dominated by the differential entropy of $\boldsymbol{X}A^{1/2}\boldsymbol{H}_w B^{1/2}$. In particular, using the results in [12] we have

$$
h(\boldsymbol{Y}) = h(\boldsymbol{X}A^{1/2}\boldsymbol{H}_w B^{1/2}) + O(1/P).
\tag{9}
$$

Herein we assume that $\boldsymbol{D}$, and subsequently $\boldsymbol{X}$, are full rank. We will later show that the distribution of $\boldsymbol{D}$ that achieves the asymptotic capacity satisfies this condition.

Expressing the matrix A in terms of its eigen decomposition, and noting that $\boldsymbol{H}_w$ is isotropically distributed, we can write $\boldsymbol{H}_w \stackrel{d}{=} U_A^\dagger\boldsymbol{H}_w$, where '$\stackrel{d}{=}$' denotes equality in distribution. Using this fact and invoking (5) yields

$$
\begin{aligned}
h(\boldsymbol{X}A^{1/2}\boldsymbol{H}_w B^{1/2}) &= h(\boldsymbol{Q}_X\boldsymbol{D}\Lambda_A^{1/2}\boldsymbol{H}_w B^{1/2})\\
&= h\big((B^{1/2}\otimes I_T)\,\mathrm{vec}(\boldsymbol{Q}_X\boldsymbol{D}\Lambda_A^{1/2}\boldsymbol{H}_w)\big)\\
&= h(\boldsymbol{Q}_X\boldsymbol{D}\Lambda_A^{1/2}\boldsymbol{H}_w) + T\log\det B,
\end{aligned}
\tag{10}
$$

where (10) follows from [16, corollary 9.6.4].

To evaluate $h(\boldsymbol{Q}_X\boldsymbol{D}\Lambda_A^{1/2}\boldsymbol{H}_w)$, we follow the methodology in [5] to express this entropy in the coordinate system corresponding to the QR decomposition. In particular, because $\boldsymbol{Q}_X \in \mathbb{C}^{T\times M}$ is isotropically distributed and $T > M$, the matrix $\boldsymbol{Q}_X\boldsymbol{\Psi}$ is also isotropically distributed, where $\boldsymbol{\Psi} \in \mathbb{C}^{M\times M}$ is unitary with arbitrary distribution, but for later purposes, $\boldsymbol{\Psi}$ is chosen to be isotropically distributed. Now, the (unique) QR decomposition of $\boldsymbol{D}\Lambda_A^{1/2}\boldsymbol{H}_w$ can be expressed as $\boldsymbol{\Phi}\boldsymbol{R}$, where $\boldsymbol{\Phi} \in \mathbb{C}^{M\times M}$ is unitary, but not necessarily isotropically distributed, and $\boldsymbol{R}$ is a random upper triangular matrix. Since $\boldsymbol{\Psi}$ is isotropically distributed, the matrix $\boldsymbol{\Psi}\boldsymbol{\Phi}$ is also isotropically distributed [5]. This implies that $\boldsymbol{Q}_X\boldsymbol{D}\Lambda_A^{1/2}\boldsymbol{H}_w \stackrel{d}{=} \boldsymbol{Q}_X\boldsymbol{\Psi}\boldsymbol{\Phi}\boldsymbol{R} \stackrel{d}{=} \boldsymbol{Q}_X\boldsymbol{R}$. However, because

$\boldsymbol{Q}_X$ is isotropically distributed, it is statistically independent of $\boldsymbol{R}$ and we have

$$
h(\boldsymbol{Q}_X\boldsymbol{D}\Lambda_A^{1/2}\boldsymbol{H}_w) = h(\boldsymbol{Q}_X) + h(\boldsymbol{R}) + \mathrm{E}\{\log J_1\},
\tag{11}
$$

where $h(\boldsymbol{Q}_X)$ is computed with respect to the Haar measure, $h(\boldsymbol{R})$ is computed with respect to the space of upper triangular matrices, and J_1 is the Jacobian of the transformation from Cartesian to QR coordinates and is given by [5], [17] $J_1 = \prod_{i=1}^{M}\boldsymbol{R}_{ii}^{2(T-i)+1}$. Similarly, the differential entropy of $\boldsymbol{\Psi}\boldsymbol{D}\Lambda_A^{1/2}\boldsymbol{H}_w = \boldsymbol{\Psi}\boldsymbol{\Phi}\boldsymbol{R} \stackrel{d}{=} \boldsymbol{\Psi}\boldsymbol{R}$ can be expressed as

$$
h(\boldsymbol{\Psi}\boldsymbol{D}\Lambda_A^{1/2}\boldsymbol{H}_w) = h(\boldsymbol{\Psi}) + h(\boldsymbol{R}) + \mathrm{E}\{\log J_2\},
\tag{12}
$$

where $J_2 = \prod_{i=1}^{M}\boldsymbol{R}_{ii}^{2(M-i)+1}$.

Substituting for $h(\boldsymbol{R})$ from (12) into (11) yields

$$
\begin{aligned}
h(\boldsymbol{Q}_X\boldsymbol{D}\Lambda_A^{1/2}\boldsymbol{H}_w) &= h(\boldsymbol{\Psi}\boldsymbol{D}\Lambda_A^{1/2}\boldsymbol{H}_w) + h(\boldsymbol{Q}_X)\\
&\quad - h(\boldsymbol{\Psi}) + \mathrm{E}\{\log J_1/J_2\}\\
&= h(\boldsymbol{\Psi}\boldsymbol{D}\Lambda_A^{1/2}\boldsymbol{H}_w) + \log|\mathbb{G}_M(\mathbb{C}^T)|\\
&\quad + (T-M)\,\mathrm{E}\{\log\det\boldsymbol{H}_w^\dagger\boldsymbol{D}^2\Lambda_A\boldsymbol{H}_w\},
\end{aligned}
\tag{13}
$$

where $\mathbb{G}_M(\mathbb{C}^T)$ is the Grassmann manifold of M-dimensional subspaces in an ambient T-dimensional Euclidean space. In (13) we used the fact that $\boldsymbol{Q}_X$ and $\boldsymbol{\Psi}$ are isotropically distributed to express $h(\boldsymbol{Q}_X) - h(\boldsymbol{\Psi})$ as the logarithm of the volume of $\mathbb{G}_M(\mathbb{C}^T)$ [5], i.e., $|\mathbb{G}_M(\mathbb{C}^T)| = \frac{\prod_{i=T-M+1}^{T}2\pi^i(i-1)!}{\prod_{i=1}^{M}2\pi^i(i-1)!}$.

To obtain an upper bound on $h(\boldsymbol{\Psi}\boldsymbol{D}\Lambda_A^{1/2}\boldsymbol{H}_w)$, we note that

$$
\begin{aligned}
&\mathrm{E}\{(I_M\otimes\boldsymbol{\Psi}\boldsymbol{D}\Lambda_A^{1/2})\,\mathrm{vec}(\boldsymbol{H}_w)\,\mathrm{vec}^\dagger(\boldsymbol{H}_w)(I_M\otimes\Lambda_A^{1/2}\boldsymbol{D}\boldsymbol{\Psi}^\dagger)\}\\
&\qquad = I_M\otimes\mathrm{E}\{\boldsymbol{\Psi}\boldsymbol{D}^2\Lambda_A\boldsymbol{\Psi}^\dagger\}
\end{aligned}
\tag{14}
$$

To evaluate (14) we use the following lemma.

Lemma 1 ([18]): For any isotropically distributed unitary matrix $\boldsymbol{\Phi} \in \mathbb{C}^{n\times n}$, $\mathrm{E}\{[\boldsymbol{\Phi}]_{ij}^*[\boldsymbol{\Phi}]_{k\ell}\} = \frac{1}{n}\delta_{ik}\delta_{j\ell}$, where δ_{ik} is the Kronecker delta. $\square$

This implies that $\mathrm{E}\{\boldsymbol{\Psi}\boldsymbol{D}^2\Lambda_A\boldsymbol{\Psi}^\dagger\} = \frac{1}{M}\mathrm{Tr}\big(\mathrm{E}\{\boldsymbol{D}^2\}\Lambda_A\big)I_M$, whence an upper bound on $h(\boldsymbol{\Psi}\boldsymbol{D}\Lambda_A^{1/2}\boldsymbol{H}_w)$ can be readily obtained by applying Theorem 9.6.5 in [16], which yields

$$
h(\boldsymbol{\Psi}\boldsymbol{D}\Lambda_A^{1/2}\boldsymbol{H}_w) \leq \log\det\frac{\pi e}{M}\mathrm{Tr}\big(\mathrm{E}\{\boldsymbol{D}^2\}\Lambda_A\big)I_{M^2}
\tag{15}
$$

$$
= M^2\log\frac{\pi e}{M}\mathrm{Tr}\big(\mathrm{E}\{\boldsymbol{D}^2\}\Lambda_A\big),
\tag{16}
$$

and equality holds if and only if $\boldsymbol{\Psi}\boldsymbol{D}\Lambda_A^{1/2}\boldsymbol{H}_w$ is Gaussian distributed with covariance $\frac{1}{M}\mathrm{Tr}\big(\mathrm{E}\{\boldsymbol{D}^2\}\Lambda_A\big)I_{M^2}$, that is, if and only if the entries of $\boldsymbol{\Psi}\boldsymbol{D}\Lambda_A^{1/2}\boldsymbol{H}_w$ are i. i. d. zero mean Gaussian random variables with variance $\frac{1}{M}\mathrm{Tr}\big(\mathrm{E}\{\boldsymbol{D}^2\}\Lambda_A\big)$.

C. Upper and Lower Bounds on Capacity

1) Upper Bound: To obtain an upper bound on capacity, we invoke (2) to bound the right hand side of (16). In particular, let $\lambda_{A_1} \geq \cdots \geq \lambda_{A_M} > 0$, then we have

$$
h(\boldsymbol{\Psi}\boldsymbol{D}\Lambda_A^{1/2}\boldsymbol{H}_w) \leq M^2\log\frac{\pi eT}{M}\lambda_{A_1}P,
\tag{17}
$$

Notice that this bound is not achievable unless A and $\mathrm{E}\{\boldsymbol{D}^2\}$ are scaled identities, i.e., $A = \frac{1}{M}I_M$, $\mathrm{E}\{\boldsymbol{D}^2\} = \frac{PT}{M}I_M$.

Using (17) with (13), (10) and (9), we have

$$h(\boldsymbol{Y}) \le M^2 \log \frac{\pi e T}{M} \lambda_{A_1} P + \log |\mathbb{G}_M(\mathbb{C}^T)|$$
$$+ (T-M) \operatorname{E}\{\log \det \boldsymbol{H}_w^\dagger \boldsymbol{D}^2 \Lambda_A \boldsymbol{H}_w\} + T \log \det B + O(1/P).$$

Using this bound and setting $M = N$ in (8), we have from (6)

$$C(P) \le M\left(\frac{M}{T}-1\right) \log \pi e + \frac{M^2}{T} \log \frac{TP}{M} + \frac{M}{T} \sum_{i=2}^M \log \frac{\lambda_{A_1}}{\lambda_{A_i}}$$
$$+ \left(1 - \frac{2M}{T}\right) \operatorname{E}\{\log \det \boldsymbol{D}^2\} + \frac{1}{T} \log |\mathbb{G}_M(\mathbb{C}^T)|$$
$$+ \left(1 - \frac{M}{T}\right) \operatorname{E}\{\log \det \Lambda_A \boldsymbol{H}_w \boldsymbol{H}_w^\dagger\}$$
$$+ \left(1 - \frac{M}{T}\right) \log \det B + O(1/P). \quad (18)$$

When $T \ge 2M$, $C(P)$ can be further bounded by using Jensen's inequality and the concavity of $\log \det(\cdot)$. In particular, $\operatorname{E}\{\log \det \boldsymbol{D}^2\} \le \log \det \operatorname{E}\{\boldsymbol{D}^2\}$, with equality if and only if $\boldsymbol{D}$ is deterministic. Subject to (2), $\operatorname{E}\{\boldsymbol{D}^2\} \le M \log \frac{TP}{M}$, which is achieved when $D = \sqrt{\frac{PT}{M}} I_M$. Hence,

$$C(P) \le M\left(1 - \frac{M}{T}\right) \log \frac{TP}{\pi e M}$$
$$+ \left(1 - \frac{2M}{T}\right) \log \det A + \left(1 - \frac{M}{T}\right) \log \det B$$
$$+ \frac{M^2}{T} \log \lambda_{A_1} + \frac{1}{T} \log |\mathbb{G}_M(\mathbb{C}^T)|$$
$$+ \left(1 - \frac{M}{T}\right) \operatorname{E}\{\log \det \boldsymbol{H}_w \boldsymbol{H}_w^\dagger\} + O(1/P). \quad (19)$$

Since $\operatorname{Tr}(A) = \operatorname{Tr}(B) = 1$, the arithmetic-geometric mean inequality can be invoked to show that $\det A \le M^{-M}$ and $\det B \le M^{-M}$, with equality if and only if $\lambda_{A_i} = \lambda_{B_i} = \frac{1}{M}$, $i = 1, \dots, M$. Using this observation yields

$$C(P) \le M\left(1 - \frac{M}{T}\right) \log \frac{TP}{\pi e}$$
$$- M\left(3 - \frac{4M}{T}\right) \log M + \frac{1}{T} \log |\mathbb{G}_M(\mathbb{C}^T)|$$
$$+ \left(1 - \frac{M}{T}\right) \operatorname{E}\{\log \det \boldsymbol{H}_w \boldsymbol{H}_w^\dagger\} + O(1/P). \quad (20)$$

Note that $C(P) < \infty$ for $P < \infty$ and does not depend on the system parameters.

2) Lower Bound: Recall that, for an arbitrary distribution of $\boldsymbol{D}$, the right hand side of (15) is achievable if and only if the entries of $\boldsymbol{\Psi} \boldsymbol{D} \Lambda_A^{1/2} \boldsymbol{H}_w$ are i. i. d. Gaussian distributed. This condition is satisfied if $\boldsymbol{D}$ is deterministic and given by $D = \sqrt{\frac{TP}{\operatorname{Tr}\Lambda_A^{-1}}} \Lambda_A^{-1/2}$, cf. (2). Although this choice of D is suboptimal, the gap of the rate it yields to capacity will be shown to be bounded by a SNR-independent constant. Substituting this D in (16) yields $h(\boldsymbol{\Psi} \boldsymbol{D} \Lambda_A^{1/2} \boldsymbol{H}_w) = M^2 \log \frac{\pi e TP}{\operatorname{Tr}\Lambda_A^{-1}}$, and

$$h(\boldsymbol{Y}) = M^2 \log \pi e + \log |\mathbb{G}_M(\mathbb{C}^T)|$$
$$+ (T - M) \operatorname{E}\{\log \det \boldsymbol{H}_w \boldsymbol{H}_w^\dagger\}$$
$$+ MT \log \frac{TP}{\operatorname{Tr}\Lambda_A^{-1}} + T \log \det B + O(1/P),$$

$$h(\boldsymbol{Y}|\boldsymbol{X})$$
$$= MT \log \pi e + M \log \det B + M^2 \log \frac{TP}{\operatorname{Tr}\Lambda_A^{-1}} + O(1/P).$$

Since this setting of D is not necessarily optimal, we have

$$C(P) \ge M\left(1 - \frac{M}{T}\right) \log \frac{TP}{\pi e \operatorname{Tr}\Lambda_A^{-1}} + \frac{1}{T} \log |\mathbb{G}_M(\mathbb{C}^T)|$$
$$+ \left(1 - \frac{M}{T}\right) \log \det B + \left(1 - \frac{M}{T}\right) \operatorname{E}\{\log \det \boldsymbol{H}_w \boldsymbol{H}_w^\dagger\}$$
$$+ O(1/P). \quad (21)$$

D. Bounds on Gap to Capacity

Comparing (19) with (21), it can be seen that setting $\boldsymbol{D} = \sqrt{\frac{TP}{\operatorname{Tr}\Lambda_A^{-1}}} \Lambda_A^{-1/2}$ yields a gap to capacity, Δ, where

$$\Delta = (T-2M) \log \det A + M^2 \log \lambda_{A_1} + M(T-M) \log \frac{\operatorname{Tr} A^{-1}}{M}.$$

We note that, when $A = \frac{1}{M} I_M$, $\Delta = 0$. Otherwise, Δ is strictly greater than zero, and an upper bound on it can be derived as follows. Let κ_A be the transmit condition number, i.e., $\kappa_A = \lambda_{A_1}/\lambda_{A_M}$. Hence,

$$\Delta = \left(1 - 2\frac{M}{T}\right) \log \prod_{i=1}^M \frac{\lambda_{A_i}}{\lambda_{A_1}} + M\left(1 - \frac{M}{T}\right) \log \sum_{i=1}^M \frac{\lambda_{A_1}}{\lambda_{A_i}}$$
$$- M\left(1 - \frac{M}{T}\right) \log M. \quad (22)$$

To obtain a bound on Δ, we note that, because $\lambda_{A_1} \ge \cdots \ge \lambda_{A_M}$, the first term in (22) is nonpositive, and the argument of the logarithm in the second term is bounded by $M\kappa_A$. Using these observations, we have

$$\Delta \le M\left(1 - \frac{M}{T}\right) \log \kappa_A. \quad (23)$$

Hence, we have shown that the upper and lower bounds obtained on capacity are within a gap proportional to the logarithm of the transmit condition number, and that that gap does not depend on the SNR or the receiver correlation matrix.

E. Main Theorem

We now summarize the results of our analysis.

Theorem 1: For the communication scenario described by (1) and (2) when $T \ge 2M$ and $M = N$, the ergodic high SNR capacity, $C(P)$, satisfies

$$\log \frac{TP}{\pi e \operatorname{Tr}\Lambda_A^{-1}} \le \frac{C(P) - c}{M(1 - M/T)} \le \log \frac{TP\kappa_A}{\pi e \operatorname{Tr}\Lambda_A^{-1}},$$

where $c = \frac{1}{T} \log |\mathbb{G}_M(\mathbb{C}^T)| + \left(1 - \frac{M}{T}\right) \log \det B + \left(1 - \frac{M}{T}\right) \operatorname{E}\{\log \det \boldsymbol{H}_w \boldsymbol{H}_w^\dagger\} + O(1/P)$. Furthermore, the lower bound on $C(P)$ is achieved by input signal matrices of the form in (5), with $D = \sqrt{\frac{TP}{\operatorname{Tr}\Lambda_A^{-1}}} \Lambda_A^{-1/2}$. $\square$

This result asserts that the rate achieved by the signalling strategy in which the input matrices assume the form in (5), with $D = \sqrt{\frac{TP}{\operatorname{Tr}\Lambda_A^{-1}}} \Lambda_A^{-1/2}$ is within $M\left(1 - \frac{M}{T}\right) \log \kappa_A$

from capacity. This implies that: 1) The signalling strategy of Theorem 1 is optimal when the channel coefficients are possibly correlated at the receiver but independent at the transmitter; 2) For channels with transmit condition number slightly greater than 1, the rate loss due to the signalling strategy in Theorem 1 is relatively small; and 3) The rate loss due to the signalling strategy in Theorem 1 is unbounded as κ_A goes to infinity, even though $C(P)$ itself is finite as shown (20). In the latter case it is more beneficial that only the non-negligible eigenmodes of the channel are excited.

F. Effect of Transmit Condition Number on Achievable rate

The rate on the right hand side of (21) is achievable using the technique in Theorem 1. That theorem shows that, up to an $O(1/P)$ approximation error, the rate in (21) is within $M\left(1 - \frac{M}{T}\right)\log\kappa_A$ of the channel capacity. To investigate the effect of increasing κ_A, we will derive upper and lower bounds on that (achievable) rate and we will show that both bounds decrease with κ_A, which indicates that high transmit correlation is generally deleterious to the rate achieved by the signalling strategy of Theorem 1.

Transmit correlation affects the right hand side of (21) through $\log \operatorname{Tr} \Lambda_A^{-1}$. Hence, upper and lower bounds on the achievable rate can be derived by obtaining lower and upper bounds on $\log \operatorname{Tr} \Lambda_A^{-1}$. We will prove the following result.

Lemma 2: Any positive definite matrix A, satisfies

$$\log(M - 1 + \kappa_A) \leq \log \operatorname{Tr} \Lambda_A^{-1}$$
$$\leq \log\left(2(M-1) + \frac{1}{\kappa_A} + (M-1)^2 \kappa_A\right). \quad (24)$$

Proof: To prove the first inequality, we write

$$\log \operatorname{Tr} \Lambda_A^{-1} = \log\left(\frac{1}{\lambda_{A_M}}\left(\frac{\lambda_{A_M}}{\lambda_{A_1}} + \cdots + \frac{\lambda_{A_M}}{\lambda_{A_{M-1}}} + 1\right)\right)$$
$$= \log(M - 1 + \kappa_A).$$

To prove the second inequality, we write

$$\log \operatorname{Tr} \Lambda_A^{-1} = \log\left(\frac{1}{\lambda_{A_1}}\left(1 + \frac{\lambda_{A_1}}{\lambda_{A_2}} + \cdots + \frac{\lambda_{A_1}}{\lambda_{A_M}}\right)\right)$$
$$\leq \log\left(2(M-1) + \frac{1}{\kappa_A} + (M-1)^2 \kappa_A\right).$$

Equality hold when $M = 2$. Otherwise, inequality is strict if the eigenvalues of A are distinct. ∎

It can be verified that both inequalities are monotonically increasing in κ_A. Hence, using the result of this lemma in the right hand side of (21), it can be seen that both the lower and upper bounds on the rates achievable by the strategy outlined in Theorem 1 are monotonically decreasing with the condition number of the transmitter correlation matrix, which indicates that the more ill-conditioned this matrix is, the less the rate that can be achieved with this strategy. This result has an intuitive explanation: The signalling strategy in Theorem 1 allocates more power to the weakest eigenmode of the channel. Hence, the power allocated to that eigenmode in ill-conditioned situations infringes on the power available for transmission over the stronger eigenmodes.

V. CONCLUSION

We obtained closed-form expressions for a lower and an upper bound on the high SNR ergodic capacity of spatially-correlated MIMO channels and we showed that the difference between these bounds is monotonically decreasing in the condition number of the transmitter correlation matrix; cf. Theorem 1. In particular, we obtained an expression for the high SNR ergodic capacity of the considered channel that is tight within an approximation error that decays as 1/SNR and a constant that depends solely on the signalling dimension and the condition number of the transmitter correlation matrix.

REFERENCES

[1] I. E. Telatar, "Capacity of multiantenna Gaussian channels," *Eur. Trans. Telecom.*, vol. 10, pp. 585–595, Nov. 1999.
[2] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation multiple-antenna communications in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 46, pp. 543–564, Mar. 2000.
[3] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inf. Theory*, vol. 45, pp. 1456–1467, July 1999.
[4] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inf. Theory*, vol. 49, pp. 951–963, Apr. 2003.
[5] 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, pp. 359–383, Feb. 2002.
[6] W. Yang, G. Durisi, and E. Riegler, "On the capacity of large-MIMO block-fading channels," *IEEE J. Select. Areas Commun.*, vol. 31, pp. 117–132, Feb. 2013.
[7] R. H. Gohary and T. N. Davidson, "Non-coherent MIMO communication: Grassmannian constellations and efficient detection," *IEEE Trans. Inf. Theory*, vol. 55, pp. 1176–1205, Mar. 2009.
[8] I. Kammoun, A. M. Cipriano, and J.-C. Belfiore, "Non-coherent codes over the Grassmannian," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 3657–3667, Oct. 2007.
[9] S. G. Srinivasan and M. K. Varanasi, "Optimal constellations for the low-SNR noncoherent MIMO block Rayleigh-fading channel," *IEEE Trans. Inf. Theory*, vol. 55, pp. 776–796, Feb. 2009.
[10] K. Yu, M. Bengtsson, B. Ottersten, D. McNamara, P. Karlsson, and M. Beach, "Modeling of wide-band MIMO radio channels based on NLoS indoor measurements," *IEEE Trans. Veh. Technol.*, vol. 53, pp. 655–665, May 2004.
[11] S. A. Jafar and A. Goldsmith, "Multiple-antenna capacity in correlated Rayleigh fading with channel covariance information," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 990–997, May 2005.
[12] R. H. Gohary and H. Yanikomeroglu, "On the accuracy of the high SNR approximation of the differential entropy of signals in additive Gaussian noise," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, (Florence), May 2014.
[13] G. L. Stüber, *Principles of Mobile Communications*. Boston: Kluwer Academic, 1996.
[14] A. Graham, *Kronecker Products and Matrix Calculus: with Applications*. New York: Elis Horwood Ltd., 1981.
[15] G. W. Stewart, "Perturbation bounds for the QR factorization of a matrix," *SIAM J. Num. Anal.*, vol. 14, pp. 509–518, June 1977.
[16] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[17] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*. New York: Wiley, 1982.
[18] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Found. Trend. Commun. Inf. Theory*, vol. 1, no. 1, pp. 1–182, 2004.