

The Optimal DoF for the Noncoherent MIMO Channel with Generic Block Fading

Khac-Hoang Ngo, Sheng Yang, Maxime Guillaud

► **To cite this version:**

Khac-Hoang Ngo, Sheng Yang, Maxime Guillaud. The Optimal DoF for the Noncoherent MIMO Channel with Generic Block Fading. 2020 IEEE Information Theory Workshop (ITW), Apr 2021, Riva del Garda (virtual), Italy. pp.1-5, 10.1109/ITW46852.2021.9457600 . hal-03420088

HAL Id: hal-03420088

<https://hal-centralesupelec.archives-ouvertes.fr/hal-03420088>

Submitted on 8 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The Optimal DoF for the Noncoherent MIMO Channel with Generic Block Fading

Khac-Hoang Ngo*, Sheng Yang†, Maxime Guillaud‡

*Department of Electrical Engineering, Chalmers University of Technology, 41296 Gothenburg, Sweden

†Laboratory of Signals and Systems, CentraleSupélec, Paris-Saclay University, 91190 Gif-sur-Yvette, France

‡Mathematical and Algorithmic Sciences Laboratory, Huawei Technologies France, 92100 Boulogne-Billancourt, France

Emails: ngok@chalmers.se, sheng.yang@centralesupelec.fr, maxime.guillaud@huawei.com

Abstract—The high-SNR capacity of the noncoherent MIMO channel has been derived for the case of independent and identically distributed (IID) Rayleigh block fading by exploiting the Gaussianity of the channel matrix. This implies the optimal degrees of freedom (DoF), i.e., the capacity pre-log factor. Nevertheless, as far as the optimal DoF is concerned, IID Rayleigh fading is apparently a sufficient but not necessary condition. In this paper, we show that the optimal DoF for the IID Rayleigh block fading channel is also the optimal DoF for a more general class of *generic* block fading channels, in which the random channel matrix has finite power and finite differential entropy. Our main contribution is a novel converse proof based on the duality approach.

Index Terms—noncoherent communications, MIMO, degrees of freedom, block fading

I. INTRODUCTION

Multiple-input multiple-output (MIMO) technology, consisting in transmitting and/or receiving with multiple antennas, has been an efficient solution to exploit the extra spatial degrees of freedom (DoF) in wireless communications. Under the ideal assumption that the channel matrix is well conditioned and known to either end of the channel, it was shown that the capacity of a point-to-point MIMO channel scales linearly with the number of antennas as $C = \min\{M, N\} \log \text{SNR} + O(1)$ in the high signal-to-noise ratio (SNR) regime, where M and N are the numbers of transmit and receive antennas, respectively [1], [2]. The DoF, defined as the pre-log of the capacity at high SNR, is $\min\{M, N\}$ in this case. In practice, however, the channel matrix varies over time and is not known *a priori*. Communication without *a priori* channel state information (CSI) is said to be *noncoherent*.

In this paper, we consider a noncoherent $M \times N$ MIMO channel. Under stationary fast Rayleigh fading, i.e., the channel changes independently after each channel use, it was shown that the channel capacity scales double-logarithmically with the SNR in the single-input single-output ($M = N = 1$) case [3]. This result was then generalized to the MIMO case in [4], where the authors showed that the capacity scales as $C = \log \log \text{SNR} + \chi(\mathbf{H}) + o(1)$ where \mathbf{H} is the channel matrix and $\chi(\mathbf{H})$ is called the fading number of the channel. This implies a zero DoF. Remarkably, the Rayleigh fading assumption was not needed in [4]. Instead, it was broadly assumed that the channel matrix has finite differential entropy and finite second moment. We refer to this fading model as *generic* fading. Under block fading, i.e., the channel matrix

is assumed to remain constant during each coherence block of T channel uses and varies independently between blocks, high-SNR approximations of the capacity have been derived for the Rayleigh fading case only [5], [6], [7], [8]. The optimal DoF was shown to be

$$d_{\text{opt}} = M^* \left(1 - \frac{M^*}{T}\right), \quad (1)$$

with $M^* := \min\{M, N, \lfloor T/2 \rfloor\}$, and can be achieved either by well-designed space-time modulations [6], [7], [8], or by simple training-based strategies [9]. The converse in these works was based on the Rayleigh fading assumption, using either a direct approximation at high SNR [5], [7] or a duality upper bound with a well chosen auxiliary output distribution [8].

In this work, we generalize the DoF result of [5], [6], [7], [8] to the *generic* fading model. Specifically, we prove that the DoF given in (1) is also the optimal DoF under generic block fading. The main technical contribution of this paper lies in the converse proof. Leveraging the duality upper bound [4], we carefully choose an auxiliary output distribution with which we derive a tight DoF upper bound.

The remainder of this paper is organized as follows. We present the channel model in Section II, and then the main result and the achievability in Section III. The converse proof is given in Section IV. Finally, we conclude the paper with a future perspective in Section V. The mathematical preliminaries for our analysis are provided in the appendix.

Notation: For random quantities, we use non-italic letters with sans-serif fonts, e.g., a scalar \mathbf{x} , a vector \mathbf{v} , and a matrix \mathbf{M} . Deterministic quantities are denoted with italic letters, e.g., a scalar x , a vector \mathbf{v} , and a matrix \mathbf{M} . The Euclidean norm is denoted by $\|\mathbf{v}\|$ and the Frobenius norm by $\|\mathbf{M}\|_{\text{F}}$. The trace, transpose and conjugate transpose of \mathbf{M} are denoted $\text{tr}\{\mathbf{M}\}$, \mathbf{M}^{T} and \mathbf{M}^{H} , respectively. $\{\lambda_i(\mathbf{M})\}$ denote the eigenvalues of \mathbf{M} in decreasing order. We use $\text{diag}(x_1, \dots, x_N)$ to denote the diagonal matrix with diagonal entries x_1, \dots, x_N , and $H(\cdot)$, $h(\cdot)$, and $D(\|\cdot\|)$ to denote the entropy, differential entropy, and Kullback-Leibler (KL) divergence, respectively. Logarithms are in base 2; $\mathbb{1}\{\cdot\}$ is the indicator function; $\log^+(x) := \max\{\log(x), 0\}$; $(x)^+ := \max\{x, 0\}$; $\Gamma(x) = \int_0^{\infty} z^{x-1} e^{-z} dz$ is the Gamma function; $\Gamma_m(a) := \pi^{m(m-1)/2} \prod_{k=1}^m \Gamma(a - k + 1)$ is the complex multivariate Gamma function; $\mathcal{I}_{\mu, \nu} := \int_0^{\infty} \frac{x^{\mu}}{(1+x^2)^{\nu}} dx$ (see Lemma 3 in the appendix).

II. CHANNEL MODEL

We consider a MIMO channel consisting of a transmitter equipped with M antennas and a receiver with N antennas. The channel between the transmitter and the receiver is flat and block fading with coherence time of T channel uses. That is, the channel matrix $\mathbf{H} \in \mathbb{C}^{N \times M}$ containing the fading coefficients from the M transmit antennas to the N receive antennas remains unchanged during each block of length T and changes independently between blocks. The realizations of \mathbf{H} are *unknown* to both the transmitter and the receiver. During a coherence block b , the received signal is

$$\mathbf{Y}[b] = \mathbf{H}[b]\mathbf{X}[b] + \mathbf{Z}[b], \quad b = 1, 2, \dots,$$

where $\mathbf{Z}[b] \in \mathbb{C}^{N \times T}$ is the additive white Gaussian noise (AWGN) with independent and identically distributed (IID) $\mathcal{N}_{\mathbb{C}}(0, 1)$ entries and $\mathbf{X}[b]$ is the transmitted signal satisfying the power constraint

$$\frac{1}{n_b} \sum_{b=1}^{n_b} \|\mathbf{X}[b]\|_{\text{F}}^2 \leq PT,$$

where n_b is the number of blocks spanned by a codeword. The parameter P is referred to as the SNR of the channel. Hereafter, we omit the block index b whenever confusion is not likely.

Since the channel is block memoryless, the channel capacity is given by $C(P) = \frac{1}{T} \max_{\mathbf{p}_{\mathbf{X}}: \mathbb{E}[\|\mathbf{X}\|_{\text{F}}^2] \leq PT} I(\mathbf{X}; \mathbf{Y})$ bits per channel use. Then we say that d_{opt} is the optimal DoF with $d_{\text{opt}} := \lim_{P \rightarrow \infty} \frac{C(P)}{\log P}$. We assume that the channel matrix \mathbf{H} is drawn from a generic distribution satisfying the following conditions:

$$h(\mathbf{H}) > -\infty, \quad \mathbb{E}[\|\mathbf{H}\|_{\text{F}}^2] < \infty.$$

That is, the channel matrix has finite differential entropy and finite second moment. This class of fading model includes as a special case the IID Rayleigh fading model in which \mathbf{H} contains IID $\mathcal{N}_{\mathbb{C}}(0, 1)$ entries considered in [5], [6], [7], [8].

For notational convenience, we define some parameters related to the channel's coherence time T , the number of transmit antennas M , and the number of receive antennas N as $\underline{S} := \min\{N, T\}$, $\bar{S} := \max\{N, T\}$, $L := \min\{M, N, T\}$, and $M' := \min\{M, N\}$ for future reference.

III. MAIN RESULT: THE OPTIMAL DOF

The optimal DoF of the noncoherent MIMO generic block fading channel described above is stated in Theorem 1.

Theorem 1. *For the noncoherent $M \times N$ MIMO channel in generic, flat, and block fading with coherence interval T , if $T = 1$, the optimal DoF is zero; otherwise, the optimal DoF is given by*

$$d_{\text{opt}} = M^* \left(1 - \frac{M^*}{T}\right) \quad (2)$$

with $M^* := \min\{M, N, \lfloor T/2 \rfloor\}$.

The zero optimal DoF result for $T = 1$ (fast fading) has been shown in [4] and is included in Theorem 1 for completeness.

In this case, the channel capacity scales double-logarithmically with the SNR.

Corollary 1. *In the single input and/or single output case ($\min\{M, N\} = 1$) or the $T = 2$ case, the optimal DoF is $d_{\text{opt}} = 1 - \frac{1}{T}$.*

Remark 1. *Theorem 1 generalizes the optimal DoF of the noncoherent IID Rayleigh block fading channel given in [7], [8]. This results show that the optimal DoF (2) holds even for non-Rayleigh fading channels as long as the channel matrix has finite differential entropy and finite power.*

For $T > 1$, the optimal DoF is achieved by using only M^* antennas and a simple pilot-based scheme: let the transmitter send pilot symbols in M^* channel uses of a coherence block, and send data symbols in the remaining $T - M^*$ channel uses; the receiver estimates the channel based on the received pilot symbols and detects coherently the data symbols based on the channel estimate. A performance analysis of this pilot-based scheme following the same lines of [7, Section V], [9] shows that the DoF (2) is indeed achievable. We present next the converse proof.

IV. THE CONVERSE PROOF

In this converse proof, we shall make use of the mathematical preliminaries (Lemmas 1, 2, and 3) in the appendix. The channel input-output mutual information is expressed as

$$I(\mathbf{X}; \mathbf{Y}) = h(\mathbf{Y}) - h(\mathbf{Y} | \mathbf{X}). \quad (3)$$

By using Lemma 1 with $\mathbf{W} = [\mathbf{H} \ \mathbf{Z}]$ and $\mathbf{A} = [\mathbf{X} \ \mathbf{I}_T]^\top$ for each realization $\check{\mathbf{X}}$ of \mathbf{X} , the entropy $h(\mathbf{Y} | \mathbf{X})$ is given by

$$h(\mathbf{Y} | \mathbf{X}) = N \mathbb{E} [\log \det(\mathbf{I}_T + \mathbf{X}^* \mathbf{X})] + \mathbb{E}[h(\check{\mathbf{H}})], \quad (4)$$

where $\check{\mathbf{H}}$ contains the first T columns of $[\mathbf{H} \ \mathbf{Z}] \mathbf{U}_{[\mathbf{X} \ \mathbf{I}_T]^\top}$ with $\mathbf{U}_{[\mathbf{X} \ \mathbf{I}_T]^\top}$ being an $(M+T) \times (M+T)$ unitary matrix containing the left singular values of $[\mathbf{X} \ \mathbf{I}_T]^\top$. In particular, under IID Rayleigh fading, $\check{\mathbf{H}}$ is an $N \times T$ Gaussian matrix with IID $\mathcal{N}_{\mathbb{C}}(0, 1)$ entries, thus $h(\check{\mathbf{H}}) = NT \log(\pi e)$.

To bound $h(\mathbf{Y})$, we use the duality approach [4] as follows

$$\begin{aligned} h(\mathbf{Y}) &= \mathbb{E}[-\log p_{\mathbf{Y}}(\mathbf{Y})] \\ &= \mathbb{E}[-\log q_{\mathbf{Y}}(\mathbf{Y})] - D(p_{\mathbf{Y}} \| q_{\mathbf{Y}}) \\ &\leq \mathbb{E}[-\log q_{\mathbf{Y}}(\mathbf{Y})], \end{aligned} \quad (5)$$

due to the nonnegativity of the KL divergence $D(p_{\mathbf{Y}} \| q_{\mathbf{Y}})$. Here, the distribution $p_{\mathbf{Y}}$ is imposed by the input, channel, and noise distributions, while $q_{\mathbf{Y}}$ is any distribution in $\mathbb{C}^{N \times T}$. Note that a proper choice of $q_{\mathbf{Y}}$ is the key to a tight upper bound. Let us consider the singular value decomposition (SVD) of \mathbf{Y} :

$$\mathbf{Y} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^{\text{H}},$$

where $\mathbf{U} \in \mathbb{C}^{N \times \underline{S}}$ and $\mathbf{V} \in \mathbb{C}^{T \times \underline{S}}$ are (truncated) unitary matrices, and $\mathbf{\Sigma} = \text{diag}(\sigma_1, \dots, \sigma_{\underline{S}})$ contains the singular values of \mathbf{Y} in decreasing order. To make the SVD unique, we further assume that the diagonal elements of \mathbf{V} are real and nonnegative [8]. Then \mathbf{U} belongs to the Stiefel manifold

$\mathcal{S}(\mathbb{C}^N, \underline{S})$, while \mathbf{V} belongs to a submanifold $\tilde{\mathcal{S}}(\mathbb{C}^T, \underline{S})$ of $\mathcal{S}(\mathbb{C}^T, \underline{S})$. The Jacobian of this SVD transformation is given by [7, App. A]

$$\begin{aligned}
J_{\tilde{\mathcal{S}}, \underline{S}}(\sigma_1, \dots, \sigma_{\underline{S}}) &= \prod_{i=1}^{\underline{S}} \sigma_i^{2(\underline{S}-i)+1} \prod_{i < j}^{\underline{S}} (\sigma_i^2 - \sigma_j^2)^2 \\
&= \prod_{i=1}^{\underline{S}} \sigma_i^{2(\underline{S}-i)+1} \prod_{i=1}^{\underline{S}} \prod_{j=i+1}^{\underline{S}} (\sigma_i^2 - \sigma_j^2)^2 \\
&\leq \prod_{i=1}^{\underline{S}} \sigma_i^{2(\underline{S}-i)+1} \prod_{i=1}^{\underline{S}} \sigma_i^{4(\underline{S}-i)} \\
&= \prod_{i=1}^{\underline{S}} \sigma_i^{2T+2N-4i+1}, \quad (6) \\
&=: \hat{J}_{\tilde{\mathcal{S}}, \underline{S}}(\sigma_1, \dots, \sigma_{\underline{S}})
\end{aligned}$$

where the inequality is due to the decreasing order of $\sigma_1, \dots, \sigma_{\underline{S}}$. We choose $q_{\mathbf{Y}}$ such that \mathbf{U} , \mathbf{V} , and $\mathbf{\Sigma}$ are mutually independent with the following distributions.

- Since the signal power is not captured in the singular vectors, as far as the DoF is concerned, the choice of distribution on the manifold for \mathbf{U} and \mathbf{V} can be arbitrary as long as $\mathbb{E}[-\log q_{\mathbf{U}}(\mathbf{U})]$ and $\mathbb{E}[-\log q_{\mathbf{V}}(\mathbf{V})]$ are finite. Here, for a closed-form expression, we let \mathbf{U} and \mathbf{V} be uniformly distributed in the Stiefel manifold $\mathcal{S}(\mathbb{C}^N, \underline{S})$ and submanifold $\tilde{\mathcal{S}}(\mathbb{C}^T, \underline{S})$, respectively. That is,

$$q_{\mathbf{U}}(\mathbf{U}) = \frac{1}{|\mathcal{S}(\mathbb{C}^N, \underline{S})|} \mathbb{1}\{\mathbf{U} \in \mathcal{S}(\mathbb{C}^N, \underline{S})\}, \quad (7)$$

$$q_{\mathbf{V}}(\mathbf{V}) = \frac{1}{|\tilde{\mathcal{S}}(\mathbb{C}^T, \underline{S})|} \mathbb{1}\{\mathbf{V} \in \tilde{\mathcal{S}}(\mathbb{C}^T, \underline{S})\}, \quad (8)$$

where the volumes of $\mathcal{S}(\mathbb{C}^N, \underline{S})$ and $\tilde{\mathcal{S}}(\mathbb{C}^T, \underline{S})$ are given by $|\mathcal{S}(\mathbb{C}^n, m)| = \frac{2^m \pi^{mn}}{\Gamma_m(n)}$ and $|\tilde{\mathcal{S}}(\mathbb{C}^n, m)| = \frac{|\mathcal{S}(\mathbb{C}^n, m)|}{(2\pi)^m} = \frac{\pi^{m(n-1)}}{\Gamma_m(n)}$, respectively [10, Sec. V].

- On the other hand, the choice of $q_{\mathbf{\Sigma}}$ is crucial in deriving a tight DoF upper bound. Our choice is made so that, after taking the Jacobian of the SVD transformation into account, $\mathbb{E}[-\log q_{\mathbf{Y}}(\mathbf{Y})]$ depends on $\{\sigma_i^2\}$ only through $\mathbb{E}[\log(1 + \sigma_i^2)]$, which can be straightforwardly upper bounded in terms of $\log P$. Specifically, we let the singular values of \mathbf{Y} follow the distribution with the pdf

$$\begin{aligned}
q_{\sigma_1, \dots, \sigma_{\underline{S}}}(\sigma_1, \dots, \sigma_{\underline{S}}) &= \beta \frac{\hat{J}_{\tilde{\mathcal{S}}, \underline{S}}(\sigma_1, \dots, \sigma_{\underline{S}})}{\prod_{i=1}^{\underline{S}} (1 + \sigma_i^2)^{\alpha_i}} \\
&= \beta \prod_{i=1}^{\underline{S}} \frac{\sigma_i^{2T+2N-4i+1}}{(1 + \sigma_i^2)^{\alpha_i}}, \quad (9)
\end{aligned}$$

where β is a scaling factor. Lemma 3 implies that with $2\alpha_i = (2T + 2N - 4i + 1) + 1 + \epsilon$, that is, $\alpha_i = T + N - 2i + 1 + \frac{\epsilon}{2}$, $i \in [\underline{S}]$, for any $\epsilon > 0$, then $\prod_{i=1}^{\underline{S}} \frac{\sigma_i^{2T+2N-4i+1}}{(1 + \sigma_i^2)^{\alpha_i}}$ is integrable, i.e., there exists β such that $\beta \prod_{i=1}^{\underline{S}} \frac{\sigma_i^{2T+2N-4i+1}}{(1 + \sigma_i^2)^{\alpha_i}}$ is a pdf. Specifically, β is given by $\beta = \prod_{i=1}^{\underline{S}} \mathcal{I}_{2T+2N-4i+1, \alpha_i}^{-1}$.

Having specified $q_{\mathbf{Y}}$, we now proceed to compute $\mathbb{E}[-\log q_{\mathbf{Y}}(\mathbf{Y})]$ using the change of variables as

$$\begin{aligned}
&\mathbb{E}[-\log q_{\mathbf{Y}}(\mathbf{Y})] \\
&= \mathbb{E}[-\log q_{\mathbf{U}, \mathbf{\Sigma}, \mathbf{V}}(\mathbf{U}, \mathbf{\Sigma}, \mathbf{V})] + \mathbb{E}\left[\log\left(J_{\tilde{\mathcal{S}}, \underline{S}}(\sigma_1, \dots, \sigma_{\underline{S}})\right)\right] \\
&= \mathbb{E}[-\log q_{\mathbf{U}}(\mathbf{U})] + \mathbb{E}[-\log q_{\mathbf{V}}(\mathbf{V})] \\
&\quad + \mathbb{E}[-\log q_{\sigma_1, \dots, \sigma_{\underline{S}}}(\sigma_1, \dots, \sigma_{\underline{S}})] \\
&\quad + \mathbb{E}\left[\log\left(J_{\tilde{\mathcal{S}}, \underline{S}}(\sigma_1, \dots, \sigma_{\underline{S}})\right)\right]. \quad (10)
\end{aligned}$$

Plugging (6), (7), (8), and (9) into (10), we obtain

$$\begin{aligned}
\mathbb{E}[-\log q_{\mathbf{Y}}(\mathbf{Y})] &\leq \log |\mathcal{S}(\mathbb{C}^N, \underline{S})| + \log |\tilde{\mathcal{S}}(\mathbb{C}^T, \underline{S})| - \log \beta \\
&\quad + \sum_{i=1}^{\underline{S}} \alpha_i \mathbb{E}[\log(1 + \sigma_i^2)]. \quad (11)
\end{aligned}$$

Substituting the bounds of $h(\mathbf{Y})$ in (4) and $h(\mathbf{Y}|\mathbf{X})$ in (5), (11) into (3), we have the following bound

$$\begin{aligned}
I(\mathbf{X}; \mathbf{Y}) &\leq \underbrace{\sum_{i=1}^{\underline{S}} \alpha_i \mathbb{E}[\log(1 + \sigma_i^2)] - N \mathbb{E}[\log \det(\mathbf{I}_T + \mathbf{X}^H \mathbf{X})]}_{c_1} \\
&\quad + \log |\mathcal{S}(\mathbb{C}^N, \underline{S})| + \log |\tilde{\mathcal{S}}(\mathbb{C}^T, \underline{S})| - \log \beta - \mathbb{E}[h(\check{\mathbf{H}})] \\
&= \underbrace{\sum_{i=1}^{\underline{S}} (\alpha_i - N) \mathbb{E}[\log(1 + \sigma_i^2)]}_{c_1} \\
&\quad + \underbrace{N \left(\sum_{i=1}^{\underline{S}} \mathbb{E}[\log(1 + \sigma_i^2)] - \mathbb{E}[\log \det(\mathbf{I}_T + \mathbf{X}^H \mathbf{X})] \right)}_{c_2} \\
&\quad + \log |\mathcal{S}(\mathbb{C}^N, \underline{S})| + \log |\tilde{\mathcal{S}}(\mathbb{C}^T, \underline{S})| - \log \beta \\
&\quad - \mathbb{E}[h(\check{\mathbf{H}})]. \quad (12)
\end{aligned}$$

To proceed, we bound c_1 and c_2 . For c_2 , $\mathbb{E}[\log \det(\mathbf{I}_T + \mathbf{X}^H \mathbf{X})]$ is bounded in terms of the singular values of \mathbf{Y} as follows

$$\begin{aligned}
&\sum_{i=1}^{\underline{S}} \mathbb{E}[\log(1 + \sigma_i^2)] \\
&= \mathbb{E}[\log \det(\mathbf{I}_N + \mathbf{Y} \mathbf{Y}^H)] \\
&= \mathbb{E}[\log \det(\mathbf{I}_N + \mathbf{H} \mathbf{X} \mathbf{X}^H \mathbf{H}^H + \mathbf{H} \mathbf{X} \mathbf{Z}^H + \mathbf{Z} \mathbf{X}^H \mathbf{H}^H + \mathbf{Z} \mathbf{Z}^H)] \\
&\leq \mathbb{E}[\log \det(\mathbf{I}_N + \mathbb{E}_{\mathbf{Z}}[\mathbf{H} \mathbf{X} \mathbf{X}^H \mathbf{H}^H + \mathbf{H} \mathbf{X} \mathbf{Z}^H + \mathbf{Z} \mathbf{X}^H \mathbf{H}^H + \mathbf{Z} \mathbf{Z}^H])] \quad (13)
\end{aligned}$$

$$\begin{aligned}
&= \mathbb{E}[\log \det((\mathbf{I} + T) \mathbf{I}_N + \mathbf{H} \mathbf{X} \mathbf{X}^H \mathbf{H}^H)] \quad (14) \\
&= \mathbb{E}[\log \det(\mathbf{I}_M + (T + 1)^{-1} \mathbf{X} \mathbf{X}^H \mathbf{H}^H \mathbf{H})] + N \log(T + 1) \\
&\leq \sum_{i=1}^L \mathbb{E}\left[\log\left(1 + \lambda_i((T + 1)^{-1} \mathbf{X} \mathbf{X}^H \mathbf{H}^H \mathbf{H})\right)\right] + N \log(T + 1) \quad (15)
\end{aligned}$$

$$\begin{aligned}
&\leq \sum_{i=1}^L \mathbb{E}\left[\log\left(1 + (T + 1)^{-1} \lambda_i(\mathbf{X} \mathbf{X}^H) \lambda_1(\mathbf{H}^H \mathbf{H})\right)\right] + N \log(T + 1) \quad (16)
\end{aligned}$$

$$\leq \sum_{i=1}^L \mathbb{E}_{\mathbf{X}} \left[\log \left(1 + \lambda_i(\mathbf{X}\mathbf{X}^H) \frac{\mathbb{E}_{\mathbf{H}}[\|\mathbf{H}\|_{\mathbb{F}}^2]}{T+1} \right) \right] + N \log(T+1) \quad (17)$$

$$\leq \sum_{i=1}^L \mathbb{E} [\log(1 + \lambda_i(\mathbf{X}\mathbf{X}^H))] + \log^+ \frac{\mathbb{E}[\|\mathbf{H}\|_{\mathbb{F}}^2]}{T+1} + N \log(T+1) \quad (18)$$

$$\begin{aligned} &\leq \sum_{i=1}^{\min\{M,T\}} \mathbb{E} [\log(1 + \lambda_i(\mathbf{X}\mathbf{X}^H))] + \log^+ \frac{\mathbb{E}[\|\mathbf{H}\|_{\mathbb{F}}^2]}{T+1} \\ &\quad + N \log(T+1) \\ &= \mathbb{E} [\log \det(\mathbf{I}_T + \mathbf{X}^H \mathbf{X})] + \log^+ \frac{\mathbb{E}[\|\mathbf{H}\|_{\mathbb{F}}^2]}{T+1} + N \log(T+1), \end{aligned}$$

where (13) follows from Jensen's inequality since the log det function is concave on the set of positive definite matrices; (14) holds because $\mathbb{E}[\mathbf{Z}] = \mathbf{0}$ and $\mathbb{E}[\mathbf{Z}\mathbf{Z}^H] = T\mathbf{I}_N$; (15) holds because the rank of $\mathbf{X}\mathbf{X}^H\mathbf{H}^H\mathbf{H}$ is upper bounded by $L := \min\{M, N, T\}$; (16) follows from Lemma 2; (17) is due to $\lambda_1(\mathbf{H}^H\mathbf{H}) \leq \|\mathbf{H}\|_{\mathbb{F}}^2$ and Jensen's inequality; and (18) follows from

$$\begin{aligned} \log(1 + ax) &\leq \log(\max\{1, a\} + \max\{1, a\}x) \\ &= \log(1 + x) + \log^+ a, \forall x \geq 0, a \geq 0. \end{aligned}$$

Therefore,

$$c_2 \leq N \log^+ \frac{\mathbb{E}[\|\mathbf{H}\|_{\mathbb{F}}^2]}{T+1} + N^2 \log(T+1). \quad (19)$$

For c_1 , we use Jensen's inequality to write

$$\begin{aligned} c_1 &\leq \sum_{i=1}^{\underline{S}} (\alpha_i - N)^+ \log(1 + \mathbb{E}[\sigma_i^2]) \\ &= \sum_{i=1}^{\underline{S}} (T - 2i + 1 + \epsilon/2)^+ \log(1 + \mathbb{E}[\sigma_i^2]), \quad (20) \end{aligned}$$

where we recall that $\alpha_i = T + N - 2i + 1 + \frac{\epsilon}{2}$, $i \in [\underline{S}]$. For $i = 1, \dots, L$, we bound $\mathbb{E}[\sigma_i^2]$ as

$$\begin{aligned} \mathbb{E}[\sigma_i^2] &\leq \mathbb{E}[\text{tr}(\mathbf{Y}\mathbf{Y}^H)] \\ &= \mathbb{E}[\text{tr}(\mathbf{H}\mathbf{X}\mathbf{X}^H\mathbf{H}^H + \mathbf{H}\mathbf{X}\mathbf{Z}^H + \mathbf{Z}\mathbf{X}^H\mathbf{H}^H + \mathbf{Z}\mathbf{Z}^H)] \\ &= \mathbb{E}[\text{tr}(\mathbf{X}\mathbf{X}^H\mathbf{H}^H\mathbf{H})] + \mathbb{E}[\text{tr}(\mathbf{Z}\mathbf{Z}^H)] \\ &= \sum_{i=1}^L \mathbb{E}[\lambda_i(\mathbf{X}\mathbf{X}^H\mathbf{H}^H\mathbf{H})] + NT \\ &\leq \sum_{i=1}^L \mathbb{E}[\lambda_i(\mathbf{X}\mathbf{X}^H)\lambda_1(\mathbf{H}^H\mathbf{H})] + NT \quad (21) \end{aligned}$$

$$\begin{aligned} &\leq \sum_{i=1}^{\min\{M,T\}} \mathbb{E}[\lambda_i(\mathbf{X}\mathbf{X}^H)] \mathbb{E}[\|\mathbf{H}\|_{\mathbb{F}}^2] + NT \quad (22) \\ &= \mathbb{E}[\|\mathbf{X}\|_{\mathbb{F}}^2] \mathbb{E}[\|\mathbf{H}\|_{\mathbb{F}}^2] + NT \\ &\leq PT \mathbb{E}[\|\mathbf{H}\|_{\mathbb{F}}^2] + NT, \end{aligned}$$

where (21) follows from Lemma 2, and (22) is due to $L \leq \min\{M, T\}$ and $\lambda_1(\mathbf{H}^H\mathbf{H}) \leq \|\mathbf{H}\|_{\mathbb{F}}^2$. Thus

$$\begin{aligned} \log(1 + \mathbb{E}[\sigma_i^2]) &\leq \log(PT \mathbb{E}[\|\mathbf{H}\|_{\mathbb{F}}^2] + NT + 1) \\ &= \log \left(P \left(T \mathbb{E}[\|\mathbf{H}\|_{\mathbb{F}}^2] + \frac{NT+1}{P} \right) \right) \\ &= \log P + \log T + \log \mathbb{E}[\|\mathbf{H}\|_{\mathbb{F}}^2] + o(1) \quad (23) \end{aligned}$$

as $P \rightarrow \infty$. In the high-SNR regime, since the noise variance is bounded, the main contributor to the power of \mathbf{Y} is $\mathbf{H}\mathbf{X}$, which has rank at most L . Thus, it is intuitive that the $\underline{S} - L$ smallest singular values of \mathbf{Y} carry information about the noise only and are bounded. To see this, we follow the footsteps in [7, p. 377] as follows. Since $\sigma_L, \dots, \sigma_{\underline{S}}$ are the $\underline{S} - L$ smallest singular values of \mathbf{Y} , for any $(N - L) \times N$ truncated unitary matrix \mathbf{Q} , we have $\sum_{i=L+1}^{\underline{S}} \sigma_i^2 \leq \text{tr}(\mathbf{Q}\mathbf{Y}\mathbf{Y}^H\mathbf{Q}^H)$. We write $\mathbf{Z} = \mathbf{Z}_1 + \mathbf{Z}_2$, where \mathbf{Z}_1 is the projection of \mathbf{Z} onto the subspace spanned by the row vectors of $\mathbf{H}\mathbf{X}$, and \mathbf{Z}_2 contains the perpendicular components. Since the subspace $\text{Span}(\mathbf{H}\mathbf{X})$ is independent of \mathbf{Z} , the total power in \mathbf{Z}_2 is $\mathbb{E}[\text{tr}(\mathbf{Z}_2\mathbf{Z}_2^H)] = N(T - L)$. Since $\mathbf{H}\mathbf{X} + \mathbf{Z}_1$ has rank L , we can find a $(N - L) \times N$ truncated unitary matrix \mathbf{Q}_0 such that $\mathbf{Q}_0(\mathbf{H}\mathbf{X} + \mathbf{Z}_1) = \mathbf{0}$. Note that \mathbf{Q}_0 is independent of \mathbf{Z}_2 , thus

$$\begin{aligned} \mathbb{E} \left[\sum_{i=L+1}^{\underline{S}} \sigma_i^2 \right] &\leq \mathbb{E}[\text{tr}(\mathbf{Q}_0\mathbf{Y}\mathbf{Y}^H\mathbf{Q}_0^H)] \\ &= \mathbb{E}[\text{tr}(\mathbf{Q}_0\mathbf{Z}_2\mathbf{Z}_2^H\mathbf{Q}_0^H)] = (N - L)(T - L). \end{aligned}$$

This implies that

$$\mathbb{E}[\sigma_i^2] \leq (N - L)(T - L), \quad i = L + 1, \dots, \underline{S}. \quad (24)$$

Plugging (23) and (24) into (20), we get

$$\begin{aligned} c_1 &\leq \sum_{i=1}^L (T - 2i + 1 + \epsilon/2)^+ (\log P + \log(T \mathbb{E}[\|\mathbf{H}\|_{\mathbb{F}}^2])) \\ &\quad + \sum_{i=L+1}^{\underline{S}} (T - 2i + 1 + \epsilon/2)^+ \log(1 + (N - L)(T - L)) \\ &\quad + o(1). \quad (25) \end{aligned}$$

Substituting (19) and (25) into (12), after some manipulations, we obtain that for any $\epsilon > 0$,

$$I(\mathbf{X}; \mathbf{Y}) \leq \sum_{i=1}^L (T - 2i + 1 + \epsilon/2)^+ \log P + c_0 + o(1) \quad (26)$$

as $P \rightarrow \infty$, where

$$\begin{aligned} c_0 &= \sum_{i=1}^L (T - 2i + 1 + \epsilon/2)^+ \log(T \mathbb{E}[\|\mathbf{H}\|_{\mathbb{F}}^2]) \\ &\quad + \sum_{i=L+1}^{\underline{S}} (T - 2i + 1 + \epsilon/2)^+ \log(1 + (N - L)(T - L)) \\ &\quad + \log |\mathcal{S}(\mathbf{C}^N, \underline{S})| + \log |\tilde{\mathcal{S}}(\mathbf{C}^T, \underline{S})| - \log \beta - \mathbb{E}[h(\check{\mathbf{H}})] \\ &\quad + N \log^+ \frac{\mathbb{E}[\|\mathbf{H}\|_{\mathbb{F}}^2]}{T+1} + N^2 \log(T+1). \end{aligned}$$

We see that the high-SNR capacity pre-log is $\sum_{i=1}^L (T - 2i + 1 + \epsilon/2)^+$. Letting ϵ arbitrarily close to zero (but remaining positive), this pre-log converges to $\sum_{i=1}^L (T - 2i + 1)^+ = \sum_{i=1}^{\min\{L, \lfloor T/2 \rfloor\}} (T - 2i + 1) = \sum_{i=1}^{M^*} (T - 2i + 1) = M^*(T - M^*)$, where the first equality holds because $T - 2i + 1 < 0$ whenever $i > \lfloor T/2 \rfloor$. Thus the optimal DoF is upper-bounded by $M^*(1 - \frac{M^*}{T})$. Furthermore, as $\epsilon \rightarrow 0$,

$$\begin{aligned} c_0 &\rightarrow \underline{S} + \underline{S}(N + T - 1) \log \pi - \log(\Gamma_{\underline{S}}(N)\Gamma_{\underline{S}}(T)) \\ &+ \sum_{i=1}^{\underline{S}} \log \mathcal{I}_{2T+2N-4i+1, \alpha_i} + N^2 \log(T + 1) - \mathbb{E}[h(\tilde{\mathbf{H}})] \\ &+ N \log^+ \frac{\mathbb{E}[\|\mathbf{H}\|_{\mathbb{F}}^2]}{T + 1} + M^*(T - M^*) \log(T \mathbb{E}[\|\mathbf{H}\|_{\mathbb{F}}^2]) \\ &+ \mathbb{1}\{M' \leq \lfloor T/2 \rfloor\} (\lfloor T/2 \rfloor (T - \lfloor T/2 \rfloor) - M'(T - M')) \\ &\quad \times \log(1 + (N - M')(T - M')). \end{aligned}$$

V. CONCLUSION AND PERSPECTIVE

In this paper, we have derived the optimal DoF for the noncoherent MIMO generic block-fading channel. Our results generalize the known optimal DoF for the Rayleigh fading case to a wider class of fading in which the channel matrix has finite differential entropy and finite second moment.

In the future, it would be interesting, as in the IID Rayleigh block fading case [7], [8], to characterize the constant term after the logarithmic term in the capacity formula.¹ Note that even for IID Rayleigh fading, no high-SNR approximation (up to a vanishing term) of the channel capacity has been found for the case $1 < T < 2 \min\{M, N\}$. To this end, the *escape-to-infinity* property [4], [11] would be useful. It allows one to assume without loss of generality that the high-SNR capacity-achieving input distribution has no mass in a disk around the origin, whose radius can be made arbitrarily large.

Our novel converse proof can be used for other problems, such as characterizing the optimal DoF region for the noncoherent MIMO multiple-access channel (MAC), which is not known even for the IID Rayleigh block fading case. For the two-user single-input multiple-output (SIMO) MAC in generic block fading, we have found the optimal DoF region in [12], but a generalization to the MIMO MAC was not obvious. The main challenge is to deal with inter-user interference which becomes an equivalent colored noise while decoding the signal of a user. This can be taken into account in the choice of auxiliary output distribution following the approach in the current paper.

APPENDIX MATHEMATICAL PRELIMINARIES

Lemma 1. *Let $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V} \in \mathbb{C}^{m \times t}$ have full column rank ($\mathbf{U} \in \mathbb{C}^{m \times m}$), and $\mathbf{W} \in \mathbb{C}^{n \times m}$ be a random matrix such that $h(\mathbf{W}) > -\infty$ and $\mathbb{E}[\|\mathbf{W}\|_{\mathbb{F}}^2] < \infty$. Then we have*

$$h(\mathbf{W}\mathbf{A}) = n \log \det(\mathbf{A}^H \mathbf{A}) + h(\mathbf{W}'),$$

¹In our analysis, the term c_0 in (26) would be a loose upper bound on the constant term in the channel capacity since the terms $\mathcal{I}_{2T+2N-4i+1, \alpha_i}$ —although they do not scale with the power—become very large as $\epsilon \rightarrow 0$.

where \mathbf{W}' contains the first t columns of $\mathbf{W}\mathbf{U}$. Furthermore, $h(\mathbf{W}')$ is finite, i.e., $-\infty < h(\mathbf{W}') < \infty$.

Proof. See [12, Appendix A-1]. \square

Lemma 2. *If \mathbf{A} and \mathbf{B} are $n \times n$ Hermitian positive semidefinite matrices, then*

$$\lambda_i(\mathbf{A})\lambda_n(\mathbf{B}) \leq \lambda_i(\mathbf{A}\mathbf{B}) \leq \lambda_i(\mathbf{A})\lambda_1(\mathbf{B}), \quad i \in [n],$$

where $\{\lambda_i(\mathbf{M})\}$ denote the eigenvalues of a matrix \mathbf{M} in decreasing order.

Proof. The result follows immediately by applying [13, Theorem 3] and [13, Theorem 4] with $k = 1$ therein. \square

Lemma 3. *The function $f(x) = \frac{x^\mu}{(1+x^2)^\nu}$, $x \geq 0$ is integrable for any $\mu \geq 1$ and $2\nu > \mu + 1$.*

Proof. Since $f(x)$ is a nonnegative function, we have $\int_0^\infty f(x) dx \leq \int_0^\infty g(x) dx$ if $f(x) \leq g(x), \forall x \geq 0$. Let $\nu = \frac{\mu+1}{2} + \epsilon$ with $\epsilon > 0$, we have $\frac{x^\mu}{(1+x^2)^\nu} = \frac{x^\mu}{(1+x^2)^{\frac{\mu+1}{2} + \epsilon}}$ is a decreasing function in μ for any $x \geq 0$. Thus $\frac{x^\mu}{(1+x^2)^\nu} \leq \frac{x}{(1+x^2)^{1+\epsilon}}$, $x \geq 0, \forall \mu \geq 1$. Therefore, $\int_0^\infty \frac{x^\mu}{(1+x^2)^\nu} dx \leq \int_0^\infty \frac{x}{(1+x^2)^{1+\epsilon}} dx = \frac{1}{2\epsilon} < \infty, \forall \epsilon > 0$. \square

We denote $\mathcal{I}_{\mu, \nu} := \int_0^\infty \frac{x^\mu}{(1+x^2)^\nu} dx$. Note that $\mathcal{I}_{\mu, \nu} \rightarrow \infty$ as $2\nu - \mu - 1 \rightarrow 0$.

REFERENCES

- [1] I. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecommun.*, vol. 10, pp. 585–595, Nov./Dec. 1999.
- [2] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless personal communications*, vol. 6, no. 3, pp. 311–335, 1998.
- [3] G. Taricco and M. Elia, "Capacity of fading channel with no side information," *Electronics Letters*, vol. 33, no. 16, pp. 1368–1370, Jul. 1997.
- [4] A. Lapidoth and S. 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.
- [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] W. Yang, G. Durisi, and E. Riegler, "On the capacity of large-MIMO block-fading channels," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 117–132, Feb. 2013.
- [9] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.
- [10] P. M. Marques and S. A. Abrantes, "On the derivation of the exact, closed-form capacity formulas for receiver-sided correlated MIMO channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 3, pp. 1139–1161, 2008.
- [11] G. Durisi and H. Bölcskei, "High-SNR capacity of wireless communication channels in the noncoherent setting: A primer," *AEU-International Journal of Electronics and Communications*, vol. 65, no. 8, pp. 707–712, Aug. 2011.
- [12] K.-H. Ngo, S. Yang, and M. Guillaud, "The optimal DoF region for the two-user non-coherent SIMO multiple-access channel," *arXiv preprint arXiv:1806.04102*, 2018.
- [13] B. Wang and F. Zhang, "Some inequalities for the eigenvalues of the product of positive semidefinite Hermitian matrices," *Linear algebra and its applications*, vol. 160, pp. 113–118, 1992.