# Nearest Neighbor Decoding and Pilot-Aided Channel Estimation for Fading Channels [†]

**A. Taufiq Asyhari [1],\* , Tobias Koch [2,3] and Albert Guillén i Fàbregas [4,5,6]**

[1]  School of Computing and Digital Technology, Birmingham City University, Millennium Point, Birmingham B4 7XG, UK
[2]  Signal Theory and Communications Department, Universidad Carlos III de Madrid, 28911 Leganés, Spain; koch@tsc.uc3m.es
[3]  Gregorio Marañón Health Research Institute, 28007 Madrid, Spain
[4]  Department of Information and Communication Technologies, Universitat Pompeu Fabra, 08018 Barcelona, Spain; guillen@ieee.org
[5]  Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain
[6]  Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, UK
\*  Correspondence: taufiq-a@ieee.org
[†]  The material in this paper was presented in part at the 2011 IEEE International Symposium on Information Theory (ISIT), Saint Petersburg, Russia, 31 July–5 August 2011 and the 49th Annual Allerton Conference on Communication, Control and Computing, Monticello, IL, USA, 28–30 September 2011.

**Abstract:** We study the information rates of noncoherent, stationary, Gaussian, and multiple-input multiple-output (MIMO) flat-fading channels that are achievable with nearest neighbor decoding and pilot-aided channel estimation. In particular, we investigate the behavior of these achievable rates in the limit as the signal-to-noise ratio (SNR) tends to infinity by analyzing the capacity pre-log, which is defined as the limiting ratio of the capacity to the logarithm of the SNR as the SNR tends to infinity. We demonstrate that a scheme estimating the channel using pilot symbols and detecting the message using nearest neighbor decoding (while assuming that the channel estimation is perfect) essentially achieves the capacity pre-log of noncoherent multiple-input single-output flat-fading channels, and it essentially achieves the best so far known lower bound on the capacity pre-log of noncoherent MIMO flat-fading channels. Extending the analysis to fading multiple-access channels reveals interesting relationships between the number of antennas and Doppler bandwidth in the comparative performance of joint transmission and time division multiple-access.

## 1. Introduction

The capacity of coherent multiple-input multiple-output (MIMO) channels increases with the signal-to-noise ratio (SNR) as $\min(n_t, n_r) \log\text{SNR}$, where $n_t$ and $n_r$ are the number of transmit and receive antennas, respectively, and SNR denotes the SNR per receive antenna [1,2]. The growth factor $\min(n_t, n_r)$ is sometimes referred to as the capacity pre-log [3] or spatial multiplexing gain [4–6]. This capacity growth can be achieved using a nearest neighbor decoder which selects the codeword that is closest (in Euclidean distance) to the channel output. In fact, for coherent fading channels with additive Gaussian noise, this decoder is the maximum-likelihood decoder and is therefore optimal in the sense that it minimizes the error probability (see [7] and references therein). The coherent channel model assumes that there is a genie that provides the exact fading coefficients to the decoder, an assumption that is difficult to achieve in practice. In this paper, we replace the role of the genie by

a scheme that estimates the fading coefficients via pilot symbols. This can be viewed as a particular coding strategy over a noncoherent fading channel, i.e., a channel where both communication ends do not have access to fading coefficients but may be aware of the fading statistics. Please note that with imperfect fading estimation, the nearest neighbor decoder that treats the fading estimate as if it were perfect is not necessarily optimal. Nevertheless, we show that in some relevant cases, nearest neighbor decoding with pilot-aided channel estimation achieves the capacity pre-log of noncoherent fading channels. (For noncoherent channels, the *capacity pre-log* is defined as the limiting ratio of the capacity to the logarithm of the SNR as the SNR tends to infinity.)

The capacity of noncoherent fading channels has been studied in several works. Building upon [8], Hassibi and Hochwald [9] studied the capacity of the block-fading channel and used pilot symbols (also known as training symbols) to obtain reasonably accurate fading estimates. Jindal and Lozano [10] provided tools for a unified treatment of pilot-based channel estimation in both block and stationary fading channels with bandlimited power spectral densities. In these works, lower bounds on the channel capacity were obtained. Lapidoth [3] studied a single-input single-output (SISO) fading channel for more general stationary fading processes and showed that depending on the predictability of the fading process, the capacity growth in SNR can be, *inter alia*, logarithmic or double logarithmic. The extension of [3] to multiple-input single-output (MISO) fading channels can be found in [11]. A lower bound on the capacity of stationary MIMO fading channels was derived by Etkin and Tse in [12]. With a view to next-generation (5G and beyond) communication networks, there has been an interest in capacity analyses of noncoherent massive MIMO channels with the vast majority of attempts focusing on the block-fading model [13–15].

Lapidoth and Shamai [16] and Weingarten et al. [17] studied noncoherent stationary fading channels from a mismatched-decoding perspective. In particular, they studied the achievable rates of Gaussian codebooks and nearest neighbor decoding. In both works, it is assumed that there is a genie that provides imperfect estimates of the fading coefficients.

In this work, we add the estimation of the fading coefficients to the analysis. In particular, we study a communication system where the transmitter emits pilot symbols at regular intervals, and where the receiver separately performs *channel estimation* and *data detection*. More precisely, based on the channel outputs corresponding to pilot transmissions, the channel estimator produces estimates of the fading coefficients for the remaining time instants using a linear minimum mean-squared error (LMMSE) interpolator. Using these estimates, the data detector employs a nearest neighbor decoder that detects the transmitted message. We study the achievable rates of this communication scheme at high SNR. In particular, we study the pre-log for fading processes with bandlimited power spectral densities. (The *pre-log* is defined as the limiting ratio of the achievable rate to the logarithm of the SNR as the SNR tends to infinity.)

For SISO fading channels, using simplifying arguments, Lozano [18] and Jindal and Lozano [10] showed that this scheme achieves the capacity pre-log. In particular, they express the achievable rates of this scheme in terms of the capacity of a fading channel whose SNR is reduced due to the imperfect channel estimation; (cf. [10], Equation (17)). Their expression ([10], Equation (21)) for the estimation error is based on the assumption that channel estimation is performed using infinitely many pilot symbols. However, obtaining ([10], Equation (17)) from the provided references is not straightforward, since it requires a limiting argument where both the codeword length and the number of pilot symbols tend to infinity in a controlled manner. The analysis is further complicated by the fact that for a given number of pilot symbols, the estimation error of the LMMSE interpolator is not stationary but cyclo-stationary, and it becomes stationary only as the number of pilot symbols tends to infinity. In this paper, we prove this result without any simplifying assumptions and extend it to MIMO fading channels. We show that the maximum rate pre-log with nearest neighbor decoding and pilot-aided channel estimation is given by the capacity pre-log of the coherent fading channel times the fraction of time used for the transmission of data. Hence, the loss with respect to the coherent case is solely due to the transmission of pilots used to obtain accurate fading estimates. If the inverse of

twice the bandwidth of the fading process is an integer, then for MISO channels, the above scheme achieves the capacity pre-log derived by Koch and Lapidoth [11]. For MIMO channels, the above scheme achieves the best so far known lower bound on the capacity pre-log obtained in [12]. The proof steps followed in this paper apply also to other pilot-assisted communication strategies and can be mimicked to perform rigorous analyses of their achievable rates; see, e.g., [19–22].

The rest of the paper is organized as follows. Section 2 describes the channel model and introduces our transmission scheme along with nearest neighbor decoding and pilots for channel estimation. Section 3 defines the pre-log and presents the main result. Section 4 extends the use of our scheme to a fading multiple-access channel (MAC). Sections 5 and 6 provide the proofs of our main results. Section 7 summarizes the results and concludes the paper.

## 2. System Model and Transmission Scheme

We consider a discrete-time MIMO flat-fading channel with $n_t$ transmit antennas and $n_r$ receive antennas. Thus, the channel output at time instant $k \in \mathbb{Z}$ (where $\mathbb{Z}$ denotes the set of integers) is the complex-valued $n_r$-dimensional random vector given by

$$Y_k = \sqrt{\frac{\mathsf{SNR}}{n_t}} \mathbb{H}_k x_k + Z_k. \tag{1}$$

Here $x_k \in \mathbb{C}^{n_t}$ denotes the time-$k$ channel input vector (with $\mathbb{C}$ denoting the set of complex numbers), $\mathbb{H}_k$ denotes the $(n_r \times n_t)$-dimensional random fading matrix at time $k$, and $Z_k$ denotes the $n_r$-variate random additive noise vector at time $k$.

The noise process $\{Z_k, k \in \mathbb{Z}\}$ is a sequence of independent and identically distributed (i.i.d.) complex-Gaussian random vectors with zero mean and covariance matrix $\mathsf{I}_{n_r}$, where $\mathsf{I}_{n_r}$ is the $n_r \times n_r$ identity matrix. SNR denotes the average SNR for each received antenna. The fading process $\{\mathbb{H}_k, k \in \mathbb{Z}\}$ is stationary, ergodic, and complex-Gaussian. We assume that the $n_r \cdot n_t$ processes $\{H_k(r,t), k \in \mathbb{Z}\}, r = 1, \ldots, n_r, t = 1, \ldots, n_t$ are independent and have the same law, with each process having zero mean, unit variance, and power spectral density $f_H(\lambda), -\frac{1}{2} \leq \lambda \leq \frac{1}{2}$. The assumption that the fading processes are independent is realistic for data transmission over a rich uniform scattering environment when both transmit and receive antennas have sufficient separation to ensure independent signal paths that translate into spatially-independent fading coefficients. The power spectral density $f_H(\cdot)$ is a nonnegative (measurable) function satisfying

$$\mathsf{E}\left[H_{k+m}(r,t)H_k^*(r,t)\right] = \int_{-1/2}^{1/2} e^{\mathsf{i}2\pi m\lambda} f_H(\lambda) d\lambda \tag{2}$$

where $(\cdot)^*$ denotes complex conjugation, and where $\mathsf{i} \triangleq \sqrt{-1}$. We assume that $f_H(\cdot)$ has bandwidth $\lambda_D < 1/2$, i.e., $\lambda_D$ is the smallest value such that $f_H(\lambda) = 0$ for almost every $|\lambda| > \lambda_D$. We finally assume that the fading process $\{\mathbb{H}_k, k \in \mathbb{Z}\}$ and the noise process $\{Z_k, k \in \mathbb{Z}\}$ are independent and that their joint law does not depend on $\{x_k, k \in \mathbb{Z}\}$.

The transmission involves both codewords and pilots. The former conveys the message to be transmitted, and the latter are used to facilitate the estimation of the fading coefficients at the receiver. We denote a codeword conveying a message $m$, $m \in \mathcal{M}$ at rate $R$ (where $\mathcal{M} = \{1, \ldots, \lfloor e^{nR} \rfloor\}$ is the set of possible messages, and where $\lfloor b \rfloor$ denotes the largest integer smaller than or equal to $b$) by the length-$n$ sequence of input vectors $\bar{x}_1(m), \ldots, \bar{x}_n(m)$. The codeword is selected from the codebook $\mathcal{C}$, which is drawn i.i.d. from an $n_t$-variate complex-Gaussian distribution with zero mean and identity covariance matrix, so

$$\frac{1}{n} \sum_{k=1}^{n} \mathsf{E}\left[\|\bar{X}_k(m)\|^2\right] = n_t, \quad m \in \mathcal{M} \tag{3}$$

where $\|\cdot\|$ denotes the Euclidean norm.

To estimate the fading matrix, we transmit orthogonal pilot vectors. The pilot vector $p_t \in \mathbb{C}^{n_t}$ used to estimate the fading coefficients corresponding to the $t$-th transmit antenna is given by $p_t(t) = 1$ and $p_t(t') = 0$ for $t' \neq t$. For example, the first pilot vector is $p_1 = (1, 0, \cdots, 0)^\intercal$, where $(\cdot)^\intercal$ denotes the transpose. To estimate the whole fading matrix, we thus need to send the $n_t$ pilot vectors $p_1, \ldots, p_{n_t}$.

The transmission scheme is as follows. Every $L$ time instants (for some $L \in \mathbb{N}$, where $\mathbb{N}$ is the set of positive integers), we transmit the $n_t$ pilot vectors $p_1, \ldots, p_{n_t}$. Each codeword is then split up into blocks of $L - n_t$ data vectors, which will be transmitted after the $n_t$ pilot vectors. The process of transmitting $L - n_t$ data vectors and $n_t$ pilot vectors continues until all $n$ data vectors are completed. Herein we assume that $n$ is an integer multiple of $L - n_t$. (If $n$ is not an integer multiple of $L - n_t$, then the last $L - n_t$ instants are not fully used by data vectors and include therefore time instants where we do not transmit anything. The thereby incurred loss in information rate vanishes as $n$ tends to infinity.) Prior to transmitting the first data block, and after transmitting the last data block, we introduce a guard period of $L(T - 1)$ time instants (for some $T \in \mathbb{N}$), where we transmit every $L$ time instants the $n_t$ pilot vectors $p_1, \ldots, p_{n_t}$, but we do not transmit data vectors in between. The guard period ensures that at every time instant, we can employ a channel estimator that bases its estimation on the channel outputs corresponding to the $T$ preceding and the $T$ subsequent pilot transmissions. This facilitates the analysis and, asymptotically, does not incur any loss in terms of achievable rates. The above transmission scheme is illustrated in Figure 1. The channel estimator is described in the following.
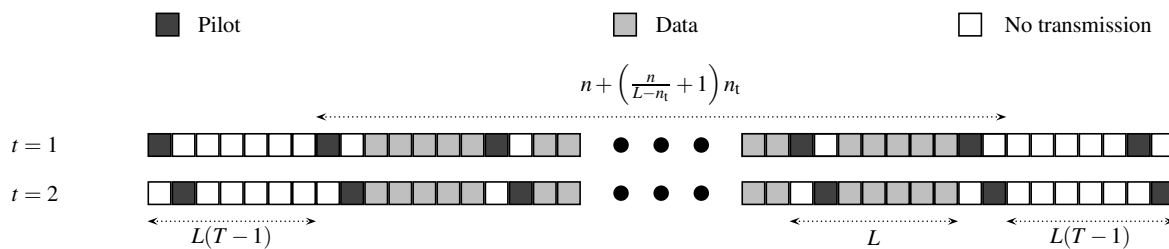


**Figure 1.** Structure of pilot and data transmission for $n_t = 2$, $L = 7$, and $T = 2$.

Please note that the total blocklength of the above transmission scheme (comprising data vectors, pilot vectors, and guard period) is given by

$$n' = n_p + n + n_g \tag{4}$$

where $n_p$ denotes the number of channel uses reserved for pilot vectors, and where $n_g$ denotes the number of channel uses during the silent guard period, i.e.,

$$n_p = \left( \frac{n}{L - n_t} + 1 + 2(T - 1) \right) n_t \tag{5}$$

$$n_g = 2(L - n_t)(T - 1). \tag{6}$$

We now turn to the decoder. Let $\mathcal{D}$ denote the set of integers reserved for the transmission of data vectors, and let $\mathcal{P}$ denote the set of integers reserved for the transmission of pilot symbols. The decoder consists of two parts: a *channel estimator* and a *data detector*. To estimate the fading coefficient at a given time instant, the channel estimator considers the channel output vectors $Y_{k'}$, $k' \in \mathcal{P}$ corresponding to the $T$ preceding and $T$ subsequent pilot transmissions and estimates $H_k(r, t)$ using a linear interpolator. The estimate $\hat{H}_k^{(T)}(r, t)$ of the fading coefficient $H_k(r, t)$ is thus given by

$$\hat{H}_k^{(T)}(r, t) = \sum_{\substack{k' = k - TL: \\ k' \in \mathcal{P}}}^{k + TL} a_{k'}(r, t) Y_{k'}(r) \tag{7}$$

where the coefficients $a_{k'}(r,t)$ are chosen in order to minimize the mean-squared error. (It has been shown in [23] that for the linear interpolator in (7), only the observations when pilots are transmitted, i.e., $Y_{k'}$, $k' \in \mathcal{P}$ are relevant for fading estimation.) In general, these coefficients depend on $k$ and $T$. However, for the sake of compactness, we do not reflect this dependence in the notation.

Please note that since the pilot vectors transmit only from one antenna, the fading coefficients corresponding to all transmit and receive antennas $(r,t)$ can be observed. Furthermore, please note that since the fading processes $\{H_k(r,t), k \in \mathbb{Z}\}$, $r = 1, \ldots, n_{\mathrm{r}}$, $t = 1, \ldots, n_{\mathrm{t}}$ are independent, estimating $H_k(r,t)$ only based on $\{Y_k(r), k \in \mathbb{Z}\}$ rather than on $\{Y_k, k \in \mathbb{Z}\}$ incurs no loss in optimality.

Since the time-lags between $\mathbb{H}_k$, $k \in \mathcal{D}$ and the observations $Y_{k'}$, $k' \in \mathcal{P}$ depend on $k$, it follows that the interpolation error

$$E_k^{(T)}(r,t) \triangleq H_k(r,t) - \hat{H}_k^{(T)}(r,t) \tag{8}$$

is not stationary but cyclo-stationary with period $L$. It can be shown that, irrespective of $r$, the variance of the interpolation error

$$\epsilon_{\ell,T}^2(r,t) \triangleq \mathsf{E}\left[\left|H_k(r,t) - \hat{H}_k^{(T)}(r,t)\right|^2\right] \tag{9}$$

tends to the following expression as $T$ tends to infinity [23]:

$$\epsilon_\ell^2(t) \triangleq \lim_{T \to \infty} \epsilon_{\ell,T}^2(r,t) \tag{10}$$

$$= 1 - \int_{-1/2}^{1/2} \frac{\mathsf{SNR}|f_{L,\ell-t+1}(\lambda)|^2}{\mathsf{SNR}f_{L,0}(\lambda) + n_{\mathrm{t}}} d\lambda \tag{11}$$

where $\ell \triangleq k \mod L$ denotes the remainder of $k/L$. Here $f_{L,\ell}(\cdot)$ is given by

$$f_{L,\ell}(\lambda) = \frac{1}{L}\sum_{\nu=0}^{L-1} \bar{f}_H\left(\frac{\lambda - \nu}{L}\right) e^{\mathrm{i}2\pi\ell\frac{\lambda-\nu}{L}}, \quad \ell = 0, \ldots, L-1 \tag{12}$$

and $\bar{f}_H(\cdot)$ is the periodic continuation of $f_H(\cdot)$, i.e., it is the periodic function of period $[-1/2, 1/2)$ that coincides with $f_H(\lambda)$ for $-1/2 \leq \lambda \leq 1/2$. If

$$L \leq \frac{1}{2\lambda_D} \tag{13}$$

then $|f_{L,\ell}(\cdot)|$ becomes

$$|f_{L,\ell}(\lambda)| = f_{L,0}(\lambda) = \frac{1}{L}f_H\left(\frac{\lambda}{L}\right). \tag{14}$$

In this case, irrespective of $\ell$ and $t$, the variance of the interpolation error is given by

$$\epsilon_\ell^2(t) = \epsilon^2 = 1 - \int_{-1/2}^{1/2} \frac{\mathsf{SNR}\left[f_H(\lambda)\right]^2}{\mathsf{SNR}f_H(\lambda) + Ln_{\mathrm{t}}} d\lambda, \quad \ell = 0, \ldots, L-1, \ t = 1, \ldots, n_{\mathrm{t}} \tag{15}$$

which vanishes as the SNR tends to infinity. Recall that $\lambda_D$ denotes the bandwidth of $f_H(\cdot)$. Thus, (13) implies that no aliasing occurs as we undersample the fading process $L$ times. Please note that in contrast to (11), the variance in (15) is independent of the transmit antenna index $t$. See Section 5.1 for a more detailed discussion.

The channel estimator feeds the sequence of fading estimates $\{\hat{\mathbb{H}}_k^{(T)}, k \in \mathcal{D}\}$ (which is composed of the matrix entries $\{\hat{H}_k^{(T)}(r,t), k \in \mathcal{D}\}$) to the data detector. We shall denote its realization by $\{\hat{\mathsf{H}}_k^{(T)}, k \in \mathcal{D}\}$. Based on the channel outputs $\{y_k, k \in \mathcal{D}\}$ and fading estimates $\{\hat{\mathsf{H}}_k^{(T)}, k \in \mathcal{D}\}$, the data

detector uses a nearest neighbor decoder to guess which message was transmitted. Thus, the decoder decides on the message $\hat{m}$ that satisfies

$$\hat{m} = \arg \min_{m \in \mathcal{M}} D(m) \tag{16}$$

where

$$D(m) \triangleq \sum_{k \in \mathcal{D}^{(n')}} \left\| \boldsymbol{y}_k - \sqrt{\frac{\mathsf{SNR}}{n_\mathrm{t}}}\, \hat{\mathsf{H}}_k^{(T)} \boldsymbol{x}_k(m) \right\|^2. \tag{17}$$

On the RHS of (17), assuming that the first pilot symbol is transmitted at time $k = 0$, we defined

$$\mathcal{D}^{(n')} \triangleq \{0, \ldots, n' - 1\} \cap \mathcal{D} \tag{18}$$

as the set of time indices where data vectors corresponding to a codeword of length $n'$ are transmitted. (For comparison, $\mathcal{D}$ represents the set of *all* integers that are reserved for the transmission of data vectors).

## 3. The Pre-Log

We say that a rate

$$R(\mathsf{SNR}) \triangleq \frac{\log |\mathcal{M}|}{n} \tag{19}$$

is achievable if there exists a code of length $n$ with $|\mathcal{M}|$ codewords such that the error probability tends to zero as $n$ tends to infinity. In this work, we study the set of rates that are achievable with nearest neighbor decoding and pilot-aided channel estimation. We focus on the achievable rates at high SNR. In particular, we are interested in the maximum achievable pre-log, defined as

$$\Pi_{R^*} \triangleq \limsup_{\mathsf{SNR} \to \infty} \frac{R^*(\mathsf{SNR})}{\log \mathsf{SNR}} \tag{20}$$

where $R^*(\mathsf{SNR})$ is the maximum rate achievable with nearest neighbor decoding and pilot-aided channel estimation, maximized over all possible encoders.

The capacity pre-log—which is given by (20) but with $R^*(\mathsf{SNR})$ replaced by the capacity $C(\mathsf{SNR})$—of SISO fading channels was computed by Lapidoth [3] as

$$\Pi_C = \mu\Big(\{\lambda : f_H(\lambda) = 0\}\Big) \tag{21}$$

where $\mu(\cdot)$ denotes the Lebesgue measure on the interval $[-1/2, 1/2]$. (The capacity is defined as the supremum of all achievable rates maximized over all possible encoders and decoders.) Koch and Lapidoth [11] extended this result to MISO fading channels and showed that if the fading processes $\{H_k(t), k \in \mathbb{Z}\}$, $t = 1, \ldots, n_\mathrm{t}$ are independent and have the same law, then the capacity pre-log of MISO fading channels is equal to the capacity pre-log of the SISO fading channel with fading process $\{H_k(1), k \in \mathbb{Z}\}$. Using (21), the capacity pre-log of MISO fading channels with bandlimited power spectral densities of bandwidth $\lambda_D$ can be evaluated as

$$\Pi_C = 1 - 2\lambda_D. \tag{22}$$

Since $R^*(\mathsf{SNR}) \leq C(\mathsf{SNR})$, it follows that $\Pi_{R^*} \leq \Pi_C$.

To the best of our knowledge, the capacity pre-log of MIMO fading channels is unknown. For independent fading processes $\{H_k(r,t), k \in \mathbb{Z}\}, t = 1, \ldots, n_t, r = 1, \ldots, n_r$ that have the same law, the best so far known lower bound on the MIMO pre-log is due to Etkin and Tse [12], and is given by

$$\Pi_C \geq \min(n_t, n_r)\left(1 - \min(n_t, n_r)\mu\left(\{\lambda : f_H(\lambda) > 0\}\right)\right). \tag{23}$$

For power spectral densities that are bandlimited to $\lambda_D$, this becomes

$$\Pi_C \geq \min(n_t, n_r)\left(1 - \min(n_t, n_r)\, 2\lambda_D\right). \tag{24}$$

Observe that (24) specializes to (22) for $n_r = 1$.

It should be noted that the capacity pre-log for MISO and SISO fading channels was derived under a peak-power constraint on the channel inputs, whereas the lower bound on the capacity pre-log for MIMO fading channels was derived under an average-power constraint. Clearly, the capacity pre-log corresponding to a peak-power constraint can never be larger than the capacity pre-log corresponding to an average-power constraint. It is believed that the two pre-logs are in fact identical (see the conclusions in [3]).

In this paper, we show that a communication scheme that employs nearest neighbor decoding and pilot-aided channel estimation achieves the following pre-log.

**Theorem 1.** *Consider the Gaussian MIMO flat-fading channel with $n_t$ transmit antennas and $n_r$ receive antennas (1). Then, the transmission and decoding scheme described in Section 2 achieves*

$$\Pi_{R^*} \geq \min(n_t, n_r)\left(1 - \frac{\min(n_t, n_r)}{L^*}\right) \tag{25}$$

*where $L^* = \left\lfloor \frac{1}{2\lambda_D} \right\rfloor$.*

**Proof.** See Section 5.  □

**Remark 1.** *We derive Theorem 1 for i.i.d. Gaussian codebooks, which satisfy the average-power constraint (3). Nevertheless, it can be shown that Theorem 1 continues to hold when the channel inputs satisfy a peak-power constraint. More specifically, we show in Section 5.3 that for an input distribution with power constraint $\mathsf{E}[\|\bar{X}\|^2] \leq n_t$ to achieve the pre-log (25), it is sufficient that its probability density function $p_{\bar{X}}(\cdot)$ satisfies*

$$p_{\bar{X}}(\bar{x}) \leq \frac{K}{\pi^{n_t}} e^{-\|\bar{x}\|^2}, \quad \bar{x} \in \mathbb{C}^{n_t} \tag{26}$$

*for some $K$ satisfying*

$$\lim_{\mathsf{SNR} \to \infty} \frac{\log K}{\log \mathsf{SNR}} = 0. \tag{27}$$

*The condition (26) is satisfied, for example, by i.i.d., truncated, Gaussian inputs, i.e., by inputs for which the $n_t$ elements in $\bar{X}$ are i.i.d. and*

$$p_{\bar{X}}(\bar{x}) = \begin{cases} \frac{1}{\hat{K}\pi^{n_t}} e^{-\|\bar{x}\|^2}, & \text{if } |\bar{x}(t)| \leq 1, 1 \leq t \leq n_t \\ 0, & \text{otherwise} \end{cases} \tag{28}$$

*with*

$$\hat{K} = \left(\int_{|\bar{x}| \leq 1} \frac{1}{\pi} e^{-|\bar{x}|^2} d\bar{x}\right)^{n_t}. \tag{29}$$

If $1/(2\lambda_D)$ is an integer, then (25) becomes

$$\Pi_{R^*} \geq \min(n_t, n_r)\left(1 - \min(n_t, n_r)\, 2\lambda_D\right). \tag{30}$$

Thus, in this case nearest neighbor decoding together with pilot-aided channel estimation achieves the capacity pre-log of MISO fading channels (22) as well as the lower bound on the capacity pre-log of MIMO fading channels (24).

Suppose that both transmitter and receiver use the same number of antennas, namely, $n_t' \triangleq n_r' \triangleq \min(n_t, n_r)$. Then, as the codeword length tends to infinity, we have from (4)–(6) that the fraction of time consumed for the transmission of pilots is given by

$$\lim_{n\to\infty} \frac{n_p}{n'} = \lim_{n\to\infty} \frac{\left(\frac{n}{L-n_t'} + 1 + 2(T-1)\right) n_t'}{\left(\frac{n}{L-n_t'} + 1 + 2(T-1)\right) n_t' + n + 2(L-n_t')(T-1)} = \frac{n_t'}{L}. \tag{31}$$

Consequently, by rewriting the pre-log (25) as

$$\Pi_{R^*} \geq n_t'\left(1 - \frac{n_t'}{L}\right), \quad L \leq \frac{1}{2\lambda_D} \tag{32}$$

we observe that the loss compared to the capacity pre-log $n_t' = \min(n_t, n_r)$ of the coherent fading channel is given by the fraction of time used for the transmission of pilots. This implies that the nearest neighbor decoder in combination with the channel estimator described in Section 2 is optimal at high SNR in the sense that it achieves the capacity pre-log of the coherent fading channel. Moreover, the achievable pre-log in Theorem 1 is the best pre-log that can be achieved by any scheme employing $n_t'$ pilot vectors.

To achieve the pre-log in Theorem 1, we assume that the training period $L$ satisfies $L \leq \frac{1}{2\lambda_D}$, in which case the variance of the interpolation error (15), namely

$$\epsilon^2 = 1 - \int_{-1/2}^{1/2} \frac{\mathsf{SNR}\,[f_H(\lambda)]^2}{\mathsf{SNR} f_H(\lambda) + Ln_t} d\lambda \approx \frac{2\lambda_D Ln_t}{\mathsf{SNR}} \tag{33}$$

vanishes as the reciprocal of the SNR. The achievable pre-log is then maximized by maximizing $L \leq \frac{1}{2\lambda_D}$. Please note that as a criterion of "perfect side information" for nearest neighbor decoding in fading channels, Lapidoth and Shamai [16] suggested that the variance of the fading estimation error should be negligible compared to the reciprocal of the SNR. The condition $L \leq \frac{1}{2\lambda_D}$ can thus be viewed as a sufficient condition for obtaining "nearly perfect side information" in the sense that the variance of the interpolation error is of the same order as the reciprocal of the SNR.

Of course, one could increase the training period $L$ beyond $\frac{1}{2\lambda_D}$. By increasing $L$, we reduce the rate loss due to the transmission of pilots as indicated in (32) at the cost of obtaining a larger fading estimation error, which in turn may reduce the reliability of the nearest neighbor decoder. To understand this trade-off better, we next briefly discuss the achievable pre-log when $L > \frac{1}{2\lambda_D}$. Indeed, for $L > \frac{1}{2\lambda_D}$, the variance of the interpolation error follows from (11) as

$$\epsilon_\ell^2(t) = 1 - \int_{-1/2}^{1/2} \frac{\mathsf{SNR}\,\left|f_{L,\ell-t+1}(\lambda)\right|^2}{\mathsf{SNR} f_{L,0}(\lambda) + n_t} d\lambda \tag{34}$$

$$= \int_{-1/2}^{1/2} \frac{n_t f_{L,0}(\lambda)}{\mathsf{SNR} f_{L,0}(\lambda) + n_t} d\lambda + \int_{-1/2}^{1/2} \frac{\mathsf{SNR}\left([f_{L,0}(\lambda)]^2 - \left|f_{L,\ell-t+1}(\lambda)\right|^2\right)}{\mathsf{SNR} f_{L,0}(\lambda) + n_t} d\lambda. \tag{35}$$

The former integral

$$\int_{-1/2}^{1/2} \frac{n_t f_{L,0}(\lambda)}{\mathsf{SNR} f_{L,0}(\lambda) + n_t} d\lambda \approx \frac{n_t}{\mathsf{SNR}} \tag{36}$$

vanishes as the reciprocal of the SNR. However, we prove in Appendix B that, as the SNR tends to infinity, the latter integral

$$\int_{-1/2}^{1/2} \frac{\mathsf{SNR}\left([f_{L,0}(\lambda)]^2 - |f_{L,\ell-t+1}(\lambda)|^2\right)}{\mathsf{SNR}f_{L,0}(\lambda) + n_{\mathrm{t}}} d\lambda \tag{37}$$

is bounded away from zero. This implies that the interpolation error (35) does not vanish as the SNR tends to infinity, and the pre-log achievable with the scheme described in Section 2 is zero. It thus follows that the condition $L \leq \frac{1}{2\lambda_D}$ is necessary in order to achieve a positive pre-log.

Comparing (24) and (25) with the capacity pre-log $\min(n_{\mathrm{t}}, n_{\mathrm{r}})$ of the coherent fading channel, we observe that, for a fading process of bandwidth $\lambda_D$, the penalty for not knowing the fading coefficients is roughly $(\min(n_{\mathrm{t}}, n_{\mathrm{r}}))^2 \cdot 2\lambda_D$. Consequently, the lower bound (25) does not grow linearly with $\min(n_{\mathrm{t}}, n_{\mathrm{r}})$, but it is a quadratic function of $\min(n_{\mathrm{t}}, n_{\mathrm{r}})$ that achieves its maximum at

$$\min(n_{\mathrm{t}}, n_{\mathrm{r}}) = \frac{L^*}{2}. \tag{38}$$

This gives rise to the lower bound

$$\Pi_{R^*} \geq \frac{L^*}{4} \tag{39}$$

which cannot be larger than $1/(8\lambda_D)$. The same holds for the lower bound (23).

## 4. Fading Multiple-Access Channels

In this section, we extend the use of nearest neighbor decoding with pilot-aided channel estimation to the fading MAC depicted in Figure 2. We are interested in the pre-log region that can be achieved with this scheme.
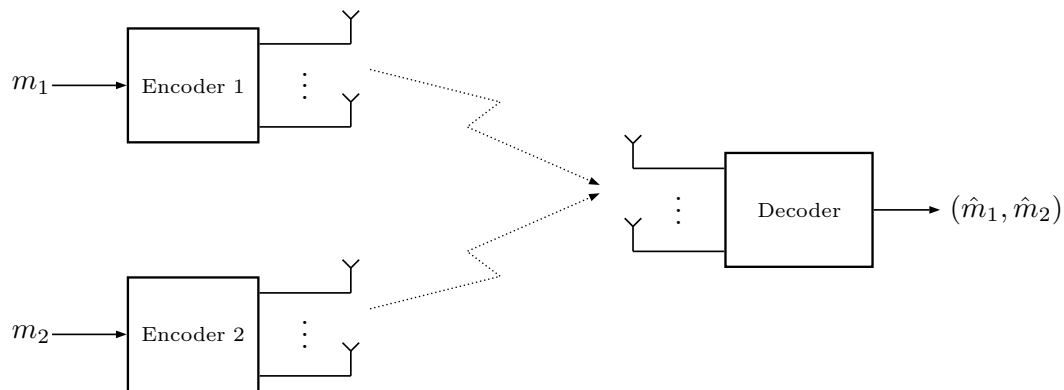


**Figure 2.** The two-user MIMO fading MAC system model.

We consider a two-user MIMO fading MAC, where two terminals wish to communicate with a third one, and where the channels between the terminals are MIMO fading channels. Extension to more than two users is straightforward. The first user has $n_{\mathrm{t},1}$ antennas, the second user has $n_{\mathrm{t},2}$ antennas, and the receiver has $n_{\mathrm{r}}$ antennas. The channel output at time instant $k \in \mathbb{Z}$ is a complex-valued $n_{\mathrm{r}}$-dimensional random vector given by

$$\boldsymbol{Y}_k = \sqrt{\mathsf{SNR}}\,\mathbb{H}_{1,k}\boldsymbol{x}_{1,k} + \sqrt{\mathsf{SNR}}\,\mathbb{H}_{2,k}\boldsymbol{x}_{2,k} + \boldsymbol{Z}_k. \tag{40}$$

Here $\boldsymbol{x}_{s,k} \in \mathbb{C}^{n_{\mathrm{t},s}}$ denotes the time-$k$ channel input vector corresponding to user $s$, $s = 1,2$; $\mathbb{H}_{s,k}$ denotes the $(n_{\mathrm{r}} \times n_{\mathrm{t},s})$-dimensional fading matrix at time $k$ corresponding to user $s$, $s = 1,2$; SNR denotes the average SNR for each transmit antenna; and $\boldsymbol{Z}_k$ denotes the $n_{\mathrm{r}}$-variate additive

noise vector at time $k$. The fading processes $\{\mathbb{H}_{s,k}, k \in \mathbb{Z}\}$, $s = 1,2$ are independent of each other and of the noise process $\{Z_k, k \in \mathbb{Z}\}$ and have the same distribution as the fading process considered in the point-to-point channel (Section 2). The noise process $\{Z_k, k \in \mathbb{Z}\}$ is a sequence of i.i.d. complex-Gaussian vectors with zero mean and covariance matrix $I_{n_r}$.

Both users transmit codewords and pilot symbols over the channel (40). To transmit the messages $m_s \in \{1, \ldots, \lfloor e^{nR_s} \rfloor\}$, $s = 1,2$ (where $m_1$ and $m_2$ are drawn independently), each user's encoder selects a codeword of length $n$ from a codebook $\mathcal{C}_s$, where $\mathcal{C}_s$, $s = 1,2$ are drawn i.i.d. from an $n_{t,s}$-variate, zero-mean, complex-Gaussian distribution of covariance matrix $I_{n_{t,s}}$. Similar to the single-user case, orthogonal pilot vectors are used. The pilot vector $\boldsymbol{p}_{s,t} \in \mathbb{C}^{n_{t,s}}$, $s = 1,2$, $t = 1, \ldots, n_{t,s}$ used to estimate the fading coefficients from transmit antenna $t$ of user $s$ is given by $p_{s,t}(t) = 1$ and $p_{s,t}(t') = 0$ for $t' \neq t$. For example, the first pilot vector of user $s$ is given by $(1, 0, \ldots, 0)^\intercal$. To estimate the fading matrices $\mathbb{H}_{1,k}$ and $\mathbb{H}_{2,k}$, each training period requires the transmission of $(n_{t,1} + n_{t,2})$ pilot vectors $\boldsymbol{p}_{1,1}, \ldots, \boldsymbol{p}_{1,n_{t,1}}, \boldsymbol{p}_{2,1}, \ldots, \boldsymbol{p}_{2,n_{t,2}}$.

Assuming that transmission from both users is synchronized, the transmission scheme extends the point-to-point setup in Section 2 to the two-user MAC setup as illustrated in Figure 3. Every $L$ time instants (for some $L \geq n_{t,1} + n_{t,2}$, $L \in \mathbb{N}$), user 1 first transmits the $n_{t,1}$ pilot vectors $\boldsymbol{p}_{1,1}, \ldots, \boldsymbol{p}_{1,n_{t,1}}$. Once the transmission of the $n_{t,1}$ pilot vectors ends, user 2 transmits its $n_{t,2}$ pilot vectors $\boldsymbol{p}_{2,1}, \ldots, \boldsymbol{p}_{2,n_{t,2}}$. The codewords for both users are then split up into blocks of $(L - n_{t,1} - n_{t,2})$ data vectors, which are transmitted simultaneously after the $(n_{t,1} + n_{t,2})$ pilot vectors. The process of transmitting $(L - n_{t,1} - n_{t,2})$ data vectors and $(n_{t,1} + n_{t,2})$ pilot vectors continues until all $n$ data symbols are completed. Herein we assume that $n$ is an integer multiple of $(L - n_{t,1} - n_{t,2})$. (As in the point-to-point setup, in the limit as $n$ tends to infinity, this assumption is not critical in terms of achievable rates.) Prior to transmitting the first data block, and after transmitting the last data block, a guard period of $L(T-1)$ time instants (for some $T \in \mathbb{N}$) is introduced for the purpose of channel estimation, where we transmit every $L$ time instants the $(n_{t,1} + n_{t,2})$ pilot vectors but we do not transmit data vectors in between. Please note that codewords from both users are *jointly* transmitted at the same time instants whereas pilots from both users do not interfere and are *separately* transmitted at different time instants. The total blocklength of this transmission scheme (comprising data vectors, pilot vectors, and guard period) is given by

$$n' = n_p + n + n_g \tag{41}$$

where $n_p$ and $n_g$ are

$$n_p = \left( \frac{n}{L - n_{t,1} - n_{t,2}} + 1 + 2(T-1) \right)(n_{t,1} + n_{t,2}) \tag{42}$$
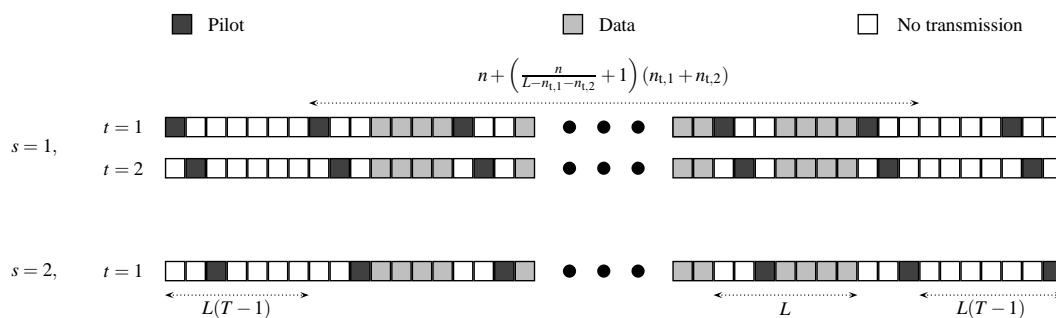
$$n_g = 2(L - n_{t,1} - n_{t,2})(T-1). \tag{43}$$



**Figure 3.** Structure of joint-transmission scheme for $n_{t,1} = 2$, $n_{t,2} = 1$, $L = 7$, and $T = 2$.

Similar to the single-user case, the receiver guesses which messages have been transmitted using a two-part decoder that consists of a channel estimator and a data detector. The channel estimator

first obtains matrix-valued fading estimates $\{\hat{\mathbb{H}}_{s,k}^{(T)}, k \in \mathcal{D}\}$, $s = 1, 2$ from the received pilots $Y_{k'}$, $k' \in \mathcal{P}$ using the same linear interpolator as (7). From the received codeword $\{y_k, k \in \mathcal{D}\}$ and the channel-estimate matrices $\{\hat{\mathsf{H}}_{s,k}^{(T)}, k \in \mathcal{D}\}$, $s = 1, 2$ (which are the realizations of $\{\hat{\mathbb{H}}_{s,k}^{(T)}, k \in \mathcal{D}\}$, $s = 1, 2$), the decoder chooses the pair of messages $(\hat{m}_1, \hat{m}_2)$ that minimizes the distance metric

$$(\hat{m}_1, \hat{m}_2) = \arg \min_{(m_1, m_2)} D(m_1, m_2) \tag{44}$$

where

$$D(m_1, m_2) \triangleq \sum_{k \in \mathcal{D}^{(n')}} \left\| y_k - \sqrt{\mathsf{SNR}} \, \hat{\mathsf{H}}_{1,k}^{(T)} x_{1,k}(m_1) - \sqrt{\mathsf{SNR}} \, \hat{\mathsf{H}}_{2,k}^{(T)} x_{2,k}(m_2) \right\|^2 \tag{45}$$

and where $\mathcal{D}^{(n')}$ is defined in the same way as (18). In the following, we shall refer to the above communication scheme as the *joint-transmission scheme*.

We shall compare the joint-transmission scheme with a time-division multiple-access (TDMA) scheme, where each user transmits its message using the transmission scheme illustrated in Figure 4. Specifically, during the first $\beta n'$ channel uses (for some $0 \leq \beta \leq 1$, and where $n'$ is given in (41)), user 1 transmits its codeword according to the transmission scheme given in Section 2 (see also Figure 4), while user 2 is silent. Then, during the next $(1 - \beta)n'$ channel uses, user 2 transmits its codeword according to the same transmission scheme, while user 1 is silent. In both cases, the receiver guesses the corresponding message $m_s$, $s = 1, 2$ using a nearest neighbor decoder and pilot-aided channel estimation.
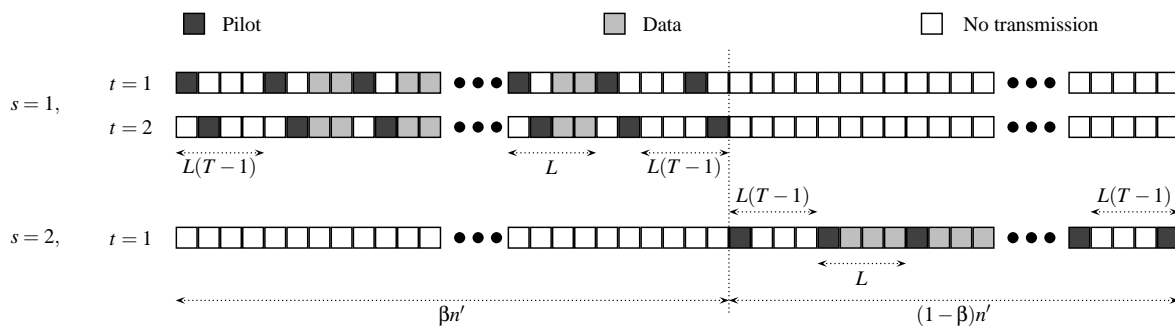


**Figure 4.** Structure of TDMA scheme for $n_{\mathrm{t},1} = 2$, $n_{\mathrm{t},2} = 1$, $L = 4$, and $T = 2$.

*4.1. The MAC Pre-Log*

Let $R_1^*(\mathsf{SNR})$, $R_2^*(\mathsf{SNR})$, and $R_{1+2}^*(\mathsf{SNR})$ be the maximum achievable rate of user 1, the maximum achievable rate of user 2, and the maximum achievable sum-rate, respectively. The achievable-rate region is given by the set [24]

$$\mathcal{R}(\mathsf{SNR}) = \left\{ (R_1, R_2) \colon R_1 \leq R_1^*(\mathsf{SNR}), \ R_2 \leq R_2^*(\mathsf{SNR}), \ R_1 + R_2 \leq R_{1+2}^*(\mathsf{SNR}) \right\}. \tag{46}$$

We are interested in the pre-logs of all rate pairs $(R_1(\mathsf{SNR}), R_2(\mathsf{SNR}))$ in $\mathcal{R}(\mathsf{SNR})$, defined as the limiting ratios of $R_1(\mathsf{SNR})$ and $R_2(\mathsf{SNR})$ to the logarithm of the SNR as the SNR tends to infinity. More precisely, the pre-log region is defined as the set of all pre-log pairs $(\Pi_{R_1}, \Pi_{R_2})$ for which there exists a sequence of rate pairs $(R_1(\mathsf{SNR}), R_2(\mathsf{SNR}))$ that, for every SNR, lies in $\mathcal{R}(\mathsf{SNR})$ and satisfies

$$\limsup_{\mathsf{SNR} \to \infty} \frac{R_1(\mathsf{SNR})}{\log \mathsf{SNR}} = \Pi_{R_1} \tag{47}$$

$$\limsup_{\mathsf{SNR} \to \infty} \frac{R_2(\mathsf{SNR})}{\log \mathsf{SNR}} = \Pi_{R_2}. \tag{48}$$

Let the maximum achievable pre-logs be defined as

$$\Pi_{R_1^*} \triangleq \limsup_{\text{SNR}\to\infty} \frac{R_1^*(\text{SNR})}{\log \text{SNR}} \tag{49}$$

$$\Pi_{R_2^*} \triangleq \limsup_{\text{SNR}\to\infty} \frac{R_2^*(\text{SNR})}{\log \text{SNR}} \tag{50}$$

$$\Pi_{R_{1+2}^*} \triangleq \limsup_{\text{SNR}\to\infty} \frac{R_{1+2}^*(\text{SNR})}{\log \text{SNR}} \tag{51}$$

and define the capacity pre-logs $\Pi_{C_1}$, $\Pi_{C_2}$, and $\Pi_{C_{1+2}}$ in the same way but with $R_1^*(\text{SNR})$, $R_2^*(\text{SNR})$, and $R_{1+2}^*(\text{SNR})$ replaced by the respective capacities $C_1(\text{SNR})$, $C_2(\text{SNR})$, and $C_{1+2}(\text{SNR})$. If the ratios of the rates to $\log \text{SNR}$ in (47)–(51) converge as $\text{SNR} \to \infty$, i.e., if the limits superior are, in fact, limits, then the pre-log region is given by the set

$$\Pi_{\mathcal{R}} = \left\{ (\Pi_{R_1}, \Pi_{R_2}) \colon \Pi_{R_1} \leq \Pi_{R_1^*}, \; \Pi_{R_2} \leq \Pi_{R_2^*}, \; \Pi_{R_1} + \Pi_{R_2} \leq \Pi_{R_{1+2}^*} \right\}. \tag{52}$$

Indeed, $\mathcal{R}(\text{SNR})$ includes all rate pairs $(R_1(\text{SNR}), R_2(\text{SNR}))$ satisfying

$$\frac{R_1(\text{SNR})}{\log \text{SNR}} \leq \frac{R_1^*(\text{SNR})}{\log \text{SNR}} \tag{53}$$

$$\frac{R_2(\text{SNR})}{\log \text{SNR}} \leq \frac{R_2^*(\text{SNR})}{\log \text{SNR}} \tag{54}$$

$$\frac{R_1(\text{SNR})}{\log \text{SNR}} + \frac{R_2(\text{SNR})}{\log \text{SNR}} \leq \frac{R_{1+2}^*(\text{SNR})}{\log \text{SNR}}. \tag{55}$$

This implies that, for every pre-log pair $(\Pi_{R_1}, \Pi_{R_2})$ in $\Pi_{\mathcal{R}}$, we can find a sequence of rate pairs $(R_1(\text{SNR}), R_2(\text{SNR}))$ in $\mathcal{R}(\text{SNR})$ that achieve (47)–(48). Conversely, if the pre-log pair $(\Pi_{R_1}, \Pi_{R_2})$ does not lie in $\Pi_{\mathcal{R}}$, then there exists a sufficiently large $\text{SNR}_0$ such that, for all $\text{SNR} \geq \text{SNR}_0$, at least one of the three conditions (53)–(55) is violated. Consequently, we cannot find a sequence of rate pairs $(R_1(\text{SNR}), R_2(\text{SNR}))$ in $\mathcal{R}(\text{SNR})$ that satisfies (47)–(48).

We next present our result on the pre-log region of the two-user MIMO fading MAC achievable with the joint-transmission scheme.

**Theorem 2.** *Consider the MIMO fading MAC (40). Then, the joint-transmission scheme achieves the pre-log region*

$$\left\{ (\Pi_{R_1}, \Pi_{R_2}) \colon \; \Pi_{R_1} \leq \min\left(n_{\mathrm{r}}, n_{\mathrm{t},1}\right)\left(1 - \frac{n_{\mathrm{t},1} + n_{\mathrm{t},2}}{L^*}\right), \right.$$

$$\Pi_{R_2} \leq \min\left(n_{\mathrm{r}}, n_{\mathrm{t},2}\right)\left(1 - \frac{n_{\mathrm{t},1} + n_{\mathrm{t},2}}{L^*}\right),$$

$$\left. \Pi_{R_1} + \Pi_{R_2} \leq \min\left(n_{\mathrm{r}}, n_{\mathrm{t},1} + n_{\mathrm{t},2}\right)\left(1 - \frac{n_{\mathrm{t},1} + n_{\mathrm{t},2}}{L^*}\right) \right\} \tag{56}$$

*where* $L^* = \left\lfloor \frac{1}{2\lambda_D} \right\rfloor$.

**Proof.** See Section 6. □

The pre-log region given in Theorem 2 is the largest region achievable with any transmission scheme that uses $(n_{\mathrm{t},1} + n_{\mathrm{t},2})/L^*$ of the time instants for transmitting pilot symbols. Indeed, even if

the channel estimator would be able to estimate the fading coefficients perfectly, and even if we could decode the data symbols using a maximum-likelihood decoder, the capacity pre-log region (without pilot transmission) would be given by the set [1,2,24]

$$\left\{ (\Pi_{R_1}, \Pi_{R_2}): \Pi_{R_1} \leq \min(n_r, n_{t,1}), \Pi_{R_2} \leq \min(n_r, n_{t,2}), \Pi_{R_1} + \Pi_{R_2} \leq \min(n_r, n_{t,1} + n_{t,2}) \right\} \quad (57)$$

which, after multiplying by $1 - (n_{t,1} + n_{t,2})/L^*$ to account for the transmission of pilot symbols, becomes (56). Thus, in order to improve upon (56), one would need to design a transmission scheme that employs less than $(n_{t,1} + n_{t,2})/L^*$ pilot symbols per channel use.

**Remark 2** (TDMA Pre-Log). *Consider the MIMO fading MAC* (40). *Then, the TDMA scheme employing nearest neighbor decoding and pilot-aided channel estimation achieves the pre-log region*

$$\left\{ (\Pi_{R_1}, \Pi_{R_2}): \ \Pi_{R_1} \leq \beta \min (n_r, n_{t,1}) \left(1 - \frac{n_{t,1}}{L^*}\right), \right.$$

$$\left. \Pi_{R_2} \leq (1 - \beta) \min (n_r, n_{t,2}) \left(1 - \frac{n_{t,2}}{L^*}\right), 0 \leq \beta \leq 1 \right\} \quad (58)$$

*where* $L^* = \left\lfloor \frac{1}{2\lambda_D} \right\rfloor$. *This follows directly from the pre-log of the point-to-point MIMO fading channel (Theorem 1) with the number of transmit antennas given by $n_{t,1}$ and $n_{t,2}$, respectively.*

Please note that the sum of the pre-logs $\Pi_{R_1} + \Pi_{R_2}$ is upper-bounded by the capacity pre-log of the point-to-point MIMO fading channel with $(n_{t,1} + n_{t,2})$ transmit antennas and $n_r$ receive antennas, since the latter channel corresponds to the case where the transmitting terminals can cooperate. While the capacity pre-log of general point-to-point MIMO fading channels remains an open problem, the capacity pre-log of point-to-point MISO fading channels is known, cf. (22). It thus follows that, for $n_r = n_{t,1} = n_{t,2} = 1$, we have

$$\Pi_{R_1} + \Pi_{R_2} \leq \Pi_{C_{1+2}} = 1 - 2\lambda_D \quad (59)$$

which together with the single-user constraints

$$\Pi_{R_1} \leq \Pi_{C_1} = 1 - 2\lambda_D \quad (60)$$
$$\Pi_{R_2} \leq \Pi_{C_2} = 1 - 2\lambda_D \quad (61)$$

implies that TDMA achieves the capacity pre-log region of the SISO fading MAC. The next section provides a more detailed comparison between the joint-transmission scheme and TDMA.

*4.2. Joint Transmission Versus TDMA*

In this section, we discuss how the joint-transmission scheme performs compared to TDMA. To this end, we compare the sum-rate pre-log $\Pi_{R^*_{1+2}}$ of the joint-transmission scheme (Theorem 2) with the sum-rate pre-log of the TDMA scheme employing nearest neighbor decoding and pilot-aided channel estimation (Remark 2) as well as with the sum-rate pre-log of the coherent TDMA scheme, where the receiver has knowledge of the realizations of the fading processes $\{\mathbb{H}_{s,k}, k \in \mathbb{Z}\}, s = 1, 2$. In the latter case, the sum-rate pre-log is given by

$$\Pi_{R^*_{1+2}} = \beta \min(n_r, n_{t,1}) + (1 - \beta) \min(n_r, n_{t,2}). \quad (62)$$

The following corollary presents a sufficient condition on $L^*$ under which the sum-rate pre-log of the joint-transmission scheme is strictly larger than that of the coherent TDMA scheme (62), as well as a sufficient condition on $L^*$ under which the sum-rate pre-log of the joint-transmission scheme is strictly smaller than the sum-rate pre-log of the TDMA scheme given in Remark 2. Since (62) is an upper bound on the sum-rate pre-log of any TDMA scheme over the MIMO fading MAC (40), and since the sum-rate pre-log given in Remark 2 is a lower bound on the sum-rate pre-log of the best TDMA scheme, it follows that the sufficient conditions presented in Corollary 1 hold also for the best TDMA scheme.

**Corollary 1.** *Consider the MIMO fading MAC* (40). *The joint-transmission scheme achieves a larger sum-rate pre-log than any TDMA scheme if*

$$L^* > \frac{\min(n_{\mathrm{r}}, n_{\mathrm{t},1} + n_{\mathrm{t},2})(n_{\mathrm{t},1} + n_{\mathrm{t},2})}{\min(n_{\mathrm{r}}, n_{\mathrm{t},1} + n_{\mathrm{t},2}) - \min(n_{\mathrm{r}}, \max(n_{\mathrm{t},1}, n_{\mathrm{t},2}))} \tag{63}$$

*where we define $a/0 \triangleq \infty$ for every $a > 0$. Conversely, the best TDMA scheme achieves a larger sum-rate pre-log than the joint-transmission scheme if*

$$L^* < \frac{\min(n_{\mathrm{r}}, n_{\mathrm{t},1} + n_{\mathrm{t},2})(n_{\mathrm{t},1} + n_{\mathrm{t},2})}{\min(n_{\mathrm{r}}, n_{\mathrm{t},1} + n_{\mathrm{t},2}) - \min(n_{\mathrm{r}}, n_{\mathrm{t},1}, n_{\mathrm{t},2})} - \frac{\min(n_{\mathrm{t},1} n_{\mathrm{r}}, n_{\mathrm{t},1}{}^2, n_{\mathrm{t},2} n_{\mathrm{r}}, n_{\mathrm{t},2}{}^2)}{\min(n_{\mathrm{r}}, n_{\mathrm{t},1} + n_{\mathrm{t},2}) - \min(n_{\mathrm{r}}, n_{\mathrm{t},1}, n_{\mathrm{t},2})}. \tag{64}$$

Recall that $L^*$ is inversely proportional to the bandwidth of the power spectral density $f_H(\cdot)$, which in turn is inversely proportional to the coherence time of the fading channel. Corollary 1 thus demonstrates that the joint-transmission scheme tends to be superior to TDMA when the coherence time of the channel is large. In contrast, TDMA is superior to the joint-transmission scheme when the coherence time of the channel is small. Intuitively, this can be explained by observing that, compared to TDMA, the joint-transmission scheme uses the antennas at the transmitters and at the receiver more efficiently, but requires more pilot symbols to estimate the fading coefficients. Thus, when the coherence time is large, the number of pilot symbols required to estimate the fading is small, so the gain in achievable rate by using the antennas more efficiently dominates the loss incurred by requiring more pilot symbols. On the other hand, when the coherence time is small, the number of pilot symbols required to estimate the fading coefficients is large and the loss in achievable rate incurred by requiring more pilot symbols dominates the gain by using the antennas more efficiently.

We next evaluate (63) and (64) for some particular values of $n_{\mathrm{r}}$, $n_{\mathrm{t},1}$, and $n_{\mathrm{t},2}$.

### 4.2.1. Receiver Employs Less Antennas than Transmitters

Suppose that $n_{\mathrm{r}} \leq \min(n_{\mathrm{t},1}, n_{\mathrm{t},2})$. Then, the right-hand sides (RHSs) of (63) and (64) become $\infty$, so every finite $L^*$ satisfies (64). Thus, if the number of receive antennas is smaller than the number of transmit antennas, then, irrespective of $L^*$, TDMA is superior to the joint-transmission scheme.

### 4.2.2. Receiver Employs More Antennas than Transmitters

Suppose that $n_{\mathrm{r}} \geq n_{\mathrm{t},1} + n_{\mathrm{t},2}$, and suppose that $n_{\mathrm{t},1} = n_{\mathrm{t},2} = n_{\mathrm{t}}$. Then, (63) and (64) become

$$L^* > 4n_{\mathrm{t}} \tag{65}$$

and

$$L^* < 3n_{\mathrm{t}}. \tag{66}$$

Thus, if $L^*$ is greater than $4n_{\mathrm{t}}$, then the joint-transmission scheme is superior to TDMA. In contrast, if $L^*$ is smaller than $3n_{\mathrm{t}}$, then TDMA is superior. This is illustrated in Figure 5 for the case where $n_{\mathrm{r}} = 2$ and $n_{\mathrm{t},1} = n_{\mathrm{t},2} = 1$. Please note that if $L^*$ is between $3n_{\mathrm{t}}$ and $4n_{\mathrm{t}}$, then the joint-transmission scheme is superior to the TDMA scheme presented in Remark 2, but it may be inferior to the best TDMA scheme.
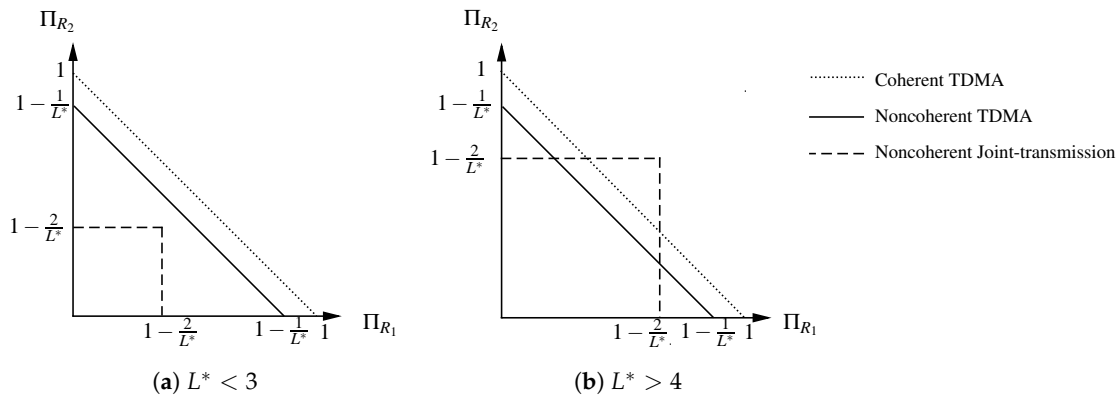
**Figure 5.** Pre-log regions for a fading MAC with $n_{\mathrm{r}} = 2$ and $n_{\mathrm{t},1} = n_{\mathrm{t},2} = 1$ for different values of $L^*$. Depicted are the pre-log region for the joint-transmission scheme as given in Theorem 2 (dashed line), the pre-log region of the TDMA scheme as given in Remark 2 (solid line), and the pre-log region of the coherent TDMA scheme (62) (dotted line).

### 4.2.3. A Case in Between

Suppose that $n_{\mathrm{r}} \leq n_{\mathrm{t},1} + n_{\mathrm{t},2}$ and $n_{\mathrm{t},2} < n_{\mathrm{r}} \leq n_{\mathrm{t},1}$. Then, (63) and (64) become

$$L^* > \infty \tag{67}$$

and

$$L^* < n_{\mathrm{t},2} + \frac{n_{\mathrm{r}} n_{\mathrm{t},1}}{n_{\mathrm{r}} - n_{\mathrm{t},2}}. \tag{68}$$

Thus, in this case the joint-transmission scheme is always inferior to the coherent TDMA scheme (62), but it can be superior to the TDMA scheme in Remark 2.

### 4.3. Typical Values of $L^*$

We briefly discuss the range of values of $L^*$ that may occur in practical scenarios. To this end, we first recall that $L^* = \lfloor 1/(2\lambda_D) \rfloor$, and that $\lambda_D$ is the bandwidth of the fading power spectral density $f_H(\cdot)$, which can be associated with the Doppler spread of the channel as [12]

$$\lambda_D = \frac{f_m}{W_c}. \tag{69}$$

Here $f_m$ is the maximum Doppler shift given by

$$f_m = \frac{v}{c} f_c \tag{70}$$

where $v$ is the speed of the mobile device, $c = 3 \cdot 10^8$ m/s is the speed of light, and $f_c$ is the carrier frequency; and $W_c$ is the coherence bandwidth of the channel approximated as [12,25]

$$W_c \approx \frac{1}{5\sigma_\tau} \tag{71}$$

where $\sigma_\tau$ is the delay spread. Following the order-of-magnitude computations of Etkin and Tse [12], we determine typical values of $\lambda_D$ for indoor, urban, and hilly area environments and for carrier frequencies ranging from 800 MHz to 5 GHz and tabulate the results in Table 1.

**Table 1.** Typical values of $L^*$ for various environments with $f_c$ ranging from 800 MHz to 5 GHz. The values of the delay spread are taken from [12,25] for indoor and urban environments and from [26] for hilly area environments.

| Environment | Delay Spread $\sigma_\tau$ | Mobile Speed $v$ | $\lambda_D \approx 5\sigma_\tau \frac{v}{c} f_c$ | $L^*$ |
|---|---|---|---|---|
| Indoor | 10–100 ns | 5 km/h | $2 \times 10^{-7}$–$10^{-5}$ | $5 \times 10^4$–$2.5 \times 10^6$ |
| Urban | 1–2 μs | 5 km/h | $2 \times 10^{-5}$–$2 \times 10^{-4}$ | $2.5 \times 10^3$–$2.5 \times 10^4$ |
| Urban | 1–2 μs | 75 km/h | $2 \times 10^{-4}$–0.004 | 125–$2.5 \times 10^3$ |
| Hilly area | 3–10 μs | 200 km/h | 0.002–0.05 | 10–250 |

For indoor environments and mobile speeds of 5 km/h, we have that $L^*$ is typically larger than $5 \times 10^4$. For urban environments, $L^*$ is typically larger than $2.5 \times 10^3$ for mobile speeds of 5 km/h and larger than 125 for mobile speeds of 75 km/h. For hilly area environments and mobile speeds of 200 km/h, $L^*$ ranges typically from 10 to 250. Thus, for most practical scenarios, $L^*$ is typically large. It therefore follows that, if $n_r \geq n_{t,1} + n_{t,2}$, the condition (63) is satisfied unless $n_{t,1} + n_{t,2}$ is very large. For example, if the receiver employs more antennas than the transmitters, and if $n_{t,1} = n_{t,2} = n_t$, then $L^* > 4n_t$ is satisfied even for urban environments and mobile speeds of 75 km/h, as long as $n_t < 30$. Only for hilly area environments and mobile speeds of 200 km/h, this condition may not be satisfied for a practical number of transmit antennas. Thus, if the number of antennas at the receiver is sufficiently large, then the joint-transmission scheme is superior to TDMA in most practical scenarios. On the other hand, if $n_r \leq \min(n_{t,1}, n_{t,2})$, then TDMA is always superior to the joint-transmission scheme, irrespective of how large $L^*$ is. This suggests that one should use more antennas at the receiver than at the transmitters.

## 5. Proof of Theorem 1

Theorem 1 is proved as follows. We first characterize the estimation error from the linear interpolator (7). We then compute the rates achievable with the communication scheme described in Section 2. Finally, we analyze the pre-log corresponding to these rates.

### 5.1. Linear Interpolator

We first note that the estimate of $H_k(r,t)$ is given by (7), namely,

$$\hat{H}_k^{(T)}(r,t) = \sum_{\substack{k'=k-TL: \\ k' \in \mathcal{P}}}^{k+TL} a_{k'}(r,t) Y_{k'}(r), \quad k \in \mathcal{D}. \tag{72}$$

We denote the interpolation error by $E_k^{(T)}(r,t) = H_k(r,t) - \hat{H}_k^{(T)}(r,t)$.

For future reference, and for any $k \in \mathbb{Z}$, we express $k = jL + \ell$, so $\ell = k \mod L$. Assuming that the first pilot symbol is transmitted at $k = 0$, it follows that $\ell = 0, \ldots, n_t - 1$ for $k \in \mathcal{P}$ and $\ell = n_t, \ldots, L-1$ for $k \in \mathcal{D}$. The statistical properties of the channel estimator for a given window size $T$ are summarized in the following lemma.

**Lemma 1.** *For a given $T$, the linear interpolator* (72) *has the following properties.*

1.  *For each $t = 1, \ldots, n_t$, $r = 1, \ldots, n_r$, and $\ell = n_t, \ldots, L-1$, the estimate $\hat{H}_{jL+\ell}^{(T)}(r,t)$ and the corresponding estimation error $E_{jL+\ell}^{(T)}(r,t)$ are independent zero-mean complex-Gaussian random variables.*

2.  (a) *For a given transmit antenna $t$ and $\ell \in \{n_t, \ldots, L-1\}$, the $n_r$ processes*

$$\{(\hat{H}_{jL+\ell}^{(T)}(1,t), E_{jL+\ell}^{(T)}(1,t)), j \in \mathbb{Z}\}, \ldots, \{(\hat{H}_{jL+\ell}^{(T)}(n_r,t), E_{jL+\ell}^{(T)}(n_r,t)), j \in \mathbb{Z}\} \tag{73}$$

*are independent and have the same law.*

(b)   *For a given receive antenna r and $\ell \in \{n_t, \ldots, L-1\}$, the $n_t$ processes*

$$\{(\hat{H}_{jL+\ell}^{(T)}(r,1), E_{jL+\ell}^{(T)}(r,1)), j \in \mathbb{Z}\}, \ldots, \{(\hat{H}_{jL+\ell}^{(T)}(r,n_t), E_{jL+\ell}^{(T)}(r,n_t)), j \in \mathbb{Z}\} \tag{74}$$

   *are independent but have different laws.*

3.   *For each $\ell = n_t, \ldots, L-1$, the process $\{(\hat{\mathbb{H}}_{jL+\ell}^{(T)}, \mathbb{H}_{jL+\ell}, \mathbf{Z}_{jL+\ell}, \mathbf{X}_{jL+\ell}), j \in \mathbb{Z}\}$ is jointly stationary and ergodic.*

4.   *For $\ell = n_t, \ldots, L-1$, it holds that*

$$\mathsf{E}\left[\mathbf{Z}_\ell^\dagger \hat{\mathbb{H}}_\ell^{(T)} \mathbf{X}_\ell\right] = 0 \tag{75}$$

*where $(\cdot)^\dagger$ denotes the conjugate transpose.*

**Proof.** See Appendix A.   □

*5.2. Achievable Rates and Pre-Logs*

In the following proof, we only consider the case where $n_t = n_r$. The more general case of $n_t \neq n_r$ follows then by using only $n_r$ transmit antennas or by ignoring $n_r - n_t$ antennas at the receiver. This yields a lower bound on the maximum achievable rate and does not incur a loss with respect to the pre-log. Indeed, it can be shown that the nearest neighbor decoder described in Section 2 achieves the pre-log $\min(n_r, n_t)$. Thus, increasing $n_t$ beyond $n_r$ or $n_r$ beyond $n_t$ does not improve the pre-log achievable by such a decoder. In fact, increasing $n_t$ beyond $n_r$ requires the transmission of more pilot symbols and does therefore even reduce the pre-log achievable with the communication scheme described in Section 2.

To prove Theorem 1, we analyze the generalized mutual information (GMI) [27] for the channel and communication scheme in Section 2. The GMI, denoted by $I_T^{\text{gmi}}(\text{SNR})$, specifies the highest information rate for which the average probability of error, averaged over the ensemble of i.i.d. Gaussian codebooks, tends to zero as the codeword length $n$ tends to infinity (see [7,16,17] and references therein). The GMI for stationary Gaussian fading channels employing nearest neighbor decoding has been evaluated in [16,17] for the case where a genie provides the receiver with an estimate of the fading process. However, the estimate considered in [16,17] is assumed to be jointly stationary and ergodic with $\{(\mathbb{H}_k, \mathbf{X}_k, \mathbf{Z}_k), k \in \mathbb{Z}\}$, which is not satisfied by $\{\hat{\mathbb{H}}_k^{(T)}, k \in \mathcal{D}\}$. We therefore need to adapt the work in [16,17] to our channel model. For completeness, we present all the main steps here, even though they are very similar to the ones in [16,17].

We prove Theorem 1 as follows:

1.   We compute a lower bound on $I_T^{\text{gmi}}(\text{SNR})$ for a fixed window size $T$ (Section 5.2.1).
2.   We analyze the behavior of this lower bound as $T$ tends to infinity (Section 5.2.2).
3.   We evaluate the limiting ratio of this lower bound to $\log \text{SNR}$ as SNR tends to infinity (Section 5.2.3).

5.2.1. $I_T^{\text{gmi}}(\text{SNR})$ for a Fixed $T$

We analyze the GMI for a fixed $T$ using a random coding upper bound on the average error probability. Please note that due to the symmetry of the codebook construction, it suffices to consider the error behavior conditioned on the event that message 1 was transmitted. Let $\mathcal{E}(m')$ denote the event that $D(m') \leq D(1)$, where $D(\cdot)$ was defined in (17). The ensemble-average error probability—where the average is over the ensemble of i.i.d. Gaussian codes—corresponding to message $m = 1$ is thus given by

$$\bar{P}_e(1) = \Pr\left\{\bigcup_{m' \neq 1} \mathcal{E}(m')\right\}. \tag{76}$$

To evaluate the GMI from the RHS of (76), we next define some useful quantities. Recall the channel and transmission model in Section 2. Without loss of generality, assume that the first pilot vector is transmitted at time $k = 0$. Let

$$F(\mathsf{SNR}) \triangleq n_{\mathrm{r}} + \frac{\mathsf{SNR}}{(L - n_{\mathrm{t}})n_{\mathrm{t}}} \sum_{\ell = n_{\mathrm{t}}}^{(L-1)} \mathsf{E}\left[ \left\| \mathbb{E}_\ell^{(T)} \right\|_F^2 \right] \tag{77}$$

where $\mathbb{E}_\ell^{(T)}$ is a random matrix whose row-*r* column-*t* entry is given by $E_\ell^{(T)}(r, t)$, and where $\| \cdot \|_F$ denotes the Frobenius norm. For some arbitrary $\delta > 0$, we further define the typical set

$$\mathcal{T}_\delta \triangleq \left\{ \left( \boldsymbol{x}_k, \boldsymbol{y}_k, \hat{\mathsf{H}}_k^{(T)} \right), k = 0, \ldots, n' - 1 : \left| \frac{1}{n} \sum_{k \in \mathcal{D}^{(n')}} \left\| \boldsymbol{y}_k - \sqrt{\frac{\mathsf{SNR}}{n_{\mathrm{t}}}} \, \hat{\mathsf{H}}_k^{(T)} \boldsymbol{x}_k \right\|^2 - F(\mathsf{SNR}) \right| < \delta \right\} \tag{78}$$

with $\mathcal{D}^{(n')} = \{0, \ldots, n' - 1\} \cap \mathcal{D}$ and $n' = n_{\mathrm{p}} + n + n_{\mathrm{g}}$, as given in (18) and (4), respectively. Then, we have the following convergence as $n$ tends to infinity.

**Lemma 2.** *For the channel model and communication scheme described in Section 2, we have that*

$$\lim_{n \to \infty} \Pr \left\{ \left( \boldsymbol{X}^{n'}, \boldsymbol{Y}^{n'}, \hat{\mathbb{H}}^{(T),n'} \right) \in \mathcal{T}_\delta \right\} = 1, \quad \forall \delta > 0 \tag{79}$$

*where we have used the notation $U^{n'}$ to denote the sequence $U_0, \ldots, U_{n'-1}$.*

**Proof.** We have

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k \in \mathcal{D}^{(n')}} \left\| \boldsymbol{y}_k - \sqrt{\frac{\mathsf{SNR}}{n_{\mathrm{t}}}} \, \hat{\mathsf{H}}_k^{(T)} \boldsymbol{x}_k \right\|^2$$

$$= \lim_{n \to \infty} \frac{1}{n} \sum_{k \in \mathcal{D}^{(n')}} \left\| \sqrt{\frac{\mathsf{SNR}}{n_{\mathrm{t}}}} \left( \mathsf{H}_k - \hat{\mathsf{H}}_k^{(T)} \right) \boldsymbol{x}_k + \boldsymbol{z}_k \right\|^2 \tag{80}$$

$$= \frac{1}{L - n_{\mathrm{t}}} \sum_{\ell = n_{\mathrm{t}}}^{L-1} \lim_{n \to \infty} \frac{L - n_{\mathrm{t}}}{n} \sum_{j=0}^{\frac{n}{L-n_{\mathrm{t}}}-1} \left\| \sqrt{\frac{\mathsf{SNR}}{n_{\mathrm{t}}}} \left( \mathsf{H}_{jL+\ell} - \hat{\mathsf{H}}_{jL+\ell}^{(T)} \right) \boldsymbol{x}_{jL+\ell} + \boldsymbol{z}_{jL+\ell} \right\|^2 \tag{81}$$

$$= \frac{1}{L - n_{\mathrm{t}}} \sum_{\ell = n_{\mathrm{t}}}^{L-1} \mathsf{E} \left[ \left\| \sqrt{\frac{\mathsf{SNR}}{n_{\mathrm{t}}}} \left( \mathbb{H}_\ell - \hat{\mathbb{H}}_\ell^{(T)} \right) \bar{\boldsymbol{X}}_\ell + \boldsymbol{Z}_\ell \right\|^2 \right], \quad \text{almost surely} \tag{82}$$

$$= \frac{1}{L - n_{\mathrm{t}}} \sum_{\ell = n_{\mathrm{t}}}^{L-1} \left( n_{\mathrm{r}} + \frac{\mathsf{SNR}}{n_{\mathrm{t}}} \mathsf{E} \left[ \left\| \mathbb{E}_\ell^{(T)} \bar{\boldsymbol{X}}_\ell \right\|^2 \right] \right) \tag{83}$$

$$= F(\mathsf{SNR}). \tag{84}$$

Herein (82) follows from (Part 3) of Lemma 1 and the ergodic theorem ([28], Chapter 7); (83) follows from (Part 4) of Lemma 1; and (84) follows since $\bar{\boldsymbol{X}}_\ell$ has zero mean and covariance matrix $\mathsf{I}_{n_{\mathrm{t}}}$, and is independent from $\mathbb{E}_\ell^{(T)}$ (since $\{\mathbb{E}_k^{(T)}, k \in \mathcal{D}\}$ is a function of $\{(\mathbb{H}_k, \boldsymbol{Z}_k), k \in \mathbb{Z}\}$). It thus follows that, as $n \to \infty$,

$$\frac{1}{n} \sum_{k \in \mathcal{D}^{(n')}} \left\| \boldsymbol{y}_k - \sqrt{\frac{\mathsf{SNR}}{n_{\mathrm{t}}}} \, \hat{\mathsf{H}}_k^{(T)} \boldsymbol{x}_k \right\|^2 \tag{85}$$

converges to $F(\mathsf{SNR})$ almost surely, which in turn implies that it also converges in probability, which is (79). □

Considering the typical set (78), and following the derivation in [16,17], the error probability $\bar{P}_e(1)$ in (76) can be upper-bounded as

$$\bar{P}_e(1) \leq e^{nR} \cdot \Pr\left\{ \frac{1}{n} \cdot D(m') < F(\mathsf{SNR}) + \delta \,\middle|\, \left(\mathbf{X}^{n'}(1), \mathbf{Y}^{n'}, \hat{\mathbb{H}}^{(T),n'}\right) \in \mathcal{T}_\delta \right\}$$
$$+ \Pr\left\{ \left(\mathbf{X}^{n'}(1), \mathbf{Y}^{n'}, \hat{\mathbb{H}}^{(T),n'}\right) \in \mathcal{T}_\delta^c \right\}, \quad m' \neq 1 \tag{86}$$

where $\mathcal{T}_\delta^c$ denotes the complement of $\mathcal{T}_\delta$. It follows from Lemma 2 that the second term on the RHS of (86) can be made arbitrarily small by letting $n$ tend to infinity.

The GMI characterizes the rate of exponential decay of the expression

$$\Pr\left\{ \frac{1}{n} \cdot D(m') < F(\mathsf{SNR}) + \delta \,\middle|\, \left(\mathbf{X}^{n'}(1), \mathbf{Y}^{n'}, \hat{\mathbb{H}}^{(T),n'}\right) \in \mathcal{T}_\delta \right\}, \quad m' \neq 1 \tag{87}$$

as $n \to \infty$ [16,17]. The computation of the GMI requires the conditional log moment-generating function of the metric $D(m')$ associated with the wrong message output $m' \neq 1$, conditioned on the channel outputs and on the fading estimates, which is defined as

$$\kappa_n\left(\theta, \mathbf{y}^{n'}, \hat{\mathsf{H}}^{(T),n'}\right) \triangleq \log \mathsf{E}\left[ \exp\left( \frac{\theta}{n} \sum_{k \in \mathcal{D}^{(n')}} D_k(m') \right) \,\middle|\, \left\{ (\mathbf{y}_k, \hat{\mathsf{H}}_k^{(T)}), \, k \in \mathcal{D}^{(n')} \right\} \right] \tag{88}$$

where

$$D_k(m') \triangleq \left\| \mathbf{y}_k - \sqrt{\frac{\mathsf{SNR}}{n_\mathrm{t}}} \hat{\mathsf{H}}_k^{(T)} \mathbf{x}_k(m') \right\|^2. \tag{89}$$

Proceeding along the lines of [16,17], we can express the conditional log moment-generating function in (88) as the sum of conditional log moment-generating functions for the individual vector metrics $D_k(m')$, $k \in \mathcal{D}^{(n')}$, i.e.,

$$\kappa_n\left(\theta, \mathbf{y}^{n'}, \hat{\mathsf{H}}^{(T),n'}\right)$$
$$= \sum_{k \in \mathcal{D}^{(n')}} \log \mathsf{E}\left[ \exp\left( \frac{\theta}{n} D_k(m') \right) \,\middle|\, \mathbf{y}_k, \hat{\mathsf{H}}_k^{(T)} \right] \tag{90}$$
$$= \sum_{k \in \mathcal{D}^{(n')}} \left( \frac{\theta}{n} \mathbf{y}_k^\dagger \left( \mathsf{I}_{n_\mathrm{r}} - \frac{\theta}{n}\frac{\mathsf{SNR}}{n_\mathrm{t}} \hat{\mathsf{H}}_k^{(T)} \hat{\mathsf{H}}_k^{\dagger(T)} \right)^{-1} \mathbf{y}_k - \log \det\left( \mathsf{I}_{n_\mathrm{r}} - \frac{\theta}{n}\frac{\mathsf{SNR}}{n_\mathrm{t}} \hat{\mathsf{H}}_k^{(T)} \hat{\mathsf{H}}_k^{\dagger(T)} \right) \right). \tag{91}$$

We then have that, for all $\theta < 0$,

$$\lim_{n\to\infty} \frac{1}{n} \cdot \kappa_n\left(n\theta, \mathbf{y}^{n'}, \hat{\mathsf{H}}^{(T),n'}\right)$$
$$= \lim_{n\to\infty} \frac{1}{n} \sum_{k \in \mathcal{D}^{(n')}} \theta \mathbf{y}_k^\dagger \left( \mathsf{I}_{n_\mathrm{r}} - \theta\frac{\mathsf{SNR}}{n_\mathrm{t}} \hat{\mathsf{H}}_k^{(T)} \hat{\mathsf{H}}_k^{\dagger(T)} \right)^{-1} \mathbf{y}_k$$
$$- \lim_{n\to\infty} \frac{1}{n} \sum_{k \in \mathcal{D}^{(n')}} \log \det\left( \mathsf{I}_{n_\mathrm{r}} - \theta\frac{\mathsf{SNR}}{n_\mathrm{t}} \hat{\mathsf{H}}_k^{(T)} \hat{\mathsf{H}}_k^{\dagger(T)} \right) \tag{92}$$
$$= \frac{1}{L - n_\mathrm{t}} \sum_{\ell = n_\mathrm{t}}^{L-1} \lim_{n\to\infty} \frac{L - n_\mathrm{t}}{n} \sum_{j=0}^{\frac{n}{L-n_\mathrm{t}}-1} \theta \mathbf{y}_{jL+\ell}^\dagger \left( \mathsf{I}_{n_\mathrm{r}} - \theta\frac{\mathsf{SNR}}{n_\mathrm{t}} \hat{\mathsf{H}}_{jL+\ell}^{(T)} \hat{\mathsf{H}}_{jL+\ell}^{\dagger(T)} \right)^{-1} \mathbf{y}_{jL+\ell}$$
$$- \frac{1}{L - n_\mathrm{t}} \sum_{\ell = n_\mathrm{t}}^{L-1} \lim_{n\to\infty} \frac{L - n_\mathrm{t}}{n} \sum_{j=0}^{\frac{n}{L-n_\mathrm{t}}-1} \log \det\left( \mathsf{I}_{n_\mathrm{r}} - \theta\frac{\mathsf{SNR}}{n_\mathrm{t}} \hat{\mathsf{H}}_{jL+\ell}^{(T)} \hat{\mathsf{H}}_{jL+\ell}^{\dagger(T)} \right) \tag{93}$$

$$
= \frac{1}{L - n_{\mathrm t}} \sum_{\ell=n_{\mathrm t}}^{L-1} \mathsf{E}\left[ \theta \mathbf{Y}_\ell^\dagger \cdot \left( \mathsf{I}_{n_{\mathrm r}} - \theta \frac{\mathsf{SNR}}{n_{\mathrm t}} \hat{\mathbb{H}}_\ell^{(T)} \hat{\mathbb{H}}_\ell^{\dagger(T)} \right)^{-1} \cdot \mathbf{Y}_\ell \right]
$$

$$
- \frac{1}{L - n_{\mathrm t}} \sum_{\ell=n_{\mathrm t}}^{L-1} \mathsf{E}\left[ \log \det \left( \mathsf{I}_{n_{\mathrm r}} - \theta \frac{\mathsf{SNR}}{n_{\mathrm t}} \hat{\mathbb{H}}_\ell^{(T)} \hat{\mathbb{H}}_\ell^{\dagger(T)} \right) \right], \quad \text{almost surely} \tag{94}
$$

$$
\triangleq \kappa(\theta, \mathsf{SNR}) \tag{95}
$$

where the last step should be regarded as the definition of $\kappa(\theta, \mathsf{SNR})$. The convergence in (94) is due to the ergodicity of $\{(\mathbf{Y}_{jL+\ell}, \hat{\mathbb{H}}_{jL+\ell}^{(T)}), j \in \mathbb{Z}\}$, $\ell = n_{\mathrm t}, \ldots, L-1$ (see (Part 3) of Lemma 1) and the ergodic theorem.

Following the same steps as in [16,17], we can then show that for all $\delta' > 0$, the ensemble-average error probability can be bounded as

$$
\bar{P}_e(1) \le \exp(nR)\exp\left( -n\left( I_T^{\mathrm{gmi}}(\mathsf{SNR}) - \delta' \right) \right) + \varepsilon(\delta', n) \tag{96}
$$

for some $\varepsilon(\delta', n)$ satisfying

$$
\lim_{n \to \infty} \varepsilon(\delta', n) = 0, \quad \delta' > 0. \tag{97}
$$

On the RHS of (96), $I_T^{\mathrm{gmi}}(\mathsf{SNR})$ denotes the GMI as a function of the SNR for a fixed $T$, which is given by

$$
I_T^{\mathrm{gmi}}(\mathsf{SNR}) = \frac{L - n_{\mathrm t}}{L} \left( \sup_{\theta < 0} (\theta F(\mathsf{SNR}) - \kappa(\theta, \mathsf{SNR})) \right). \tag{98}
$$

Herein the pre-factor $(L - n_{\mathrm t})/L$ equals the fraction of time instants used for data transmission. The bound (96) implies that for rates below $I_T^{\mathrm{gmi}}(\mathsf{SNR})$, the communication scheme described in Section 2 has vanishing error probability as $n$ tends to infinity.

Combining (77) and (94) with (98) yields

$$
I_T^{\mathrm{gmi}}(\mathsf{SNR}) = \sup_{\theta < 0} \frac{1}{L} \sum_{\ell=n_{\mathrm t}}^{L-1} \left\{ \theta \left( n_{\mathrm r} + \frac{\mathsf{SNR}}{n_{\mathrm t}} \mathsf{E}\left[ \left\| \mathbb{E}_\ell^{(T)} \right\|_F^2 \right] \right) + \mathsf{E}\left[ \log \det \left( \mathsf{I}_{n_{\mathrm r}} - \theta \frac{\mathsf{SNR}}{n_{\mathrm t}} \hat{\mathbb{H}}_\ell^{(T)} \hat{\mathbb{H}}_\ell^{\dagger(T)} \right) \right] \right.
$$

$$
\left. - \mathsf{E}\left[ \theta \mathbf{Y}_\ell^\dagger \left( \mathsf{I}_{n_{\mathrm r}} - \theta \frac{\mathsf{SNR}}{n_{\mathrm t}} \hat{\mathbb{H}}_\ell^{(T)} \hat{\mathbb{H}}_\ell^{\dagger(T)} \right)^{-1} \mathbf{Y}_\ell \right] \right\}. \tag{99}
$$

Following the steps used in ([29], Appendix D), it can be shown that, for $\theta < 0$,

$$
\mathsf{E}\left[ \theta \mathbf{Y}_\ell^\dagger \left( \mathsf{I}_{n_{\mathrm r}} - \theta \frac{\mathsf{SNR}}{n_{\mathrm t}} \hat{\mathbb{H}}_\ell^{(T)} \hat{\mathbb{H}}_\ell^{\dagger(T)} \right)^{-1} \mathbf{Y}_\ell \right] \le 0. \tag{100}
$$

As observed in ([29], Appendix D), a good lower bound on $I_T^{\mathrm{gmi}}(\mathsf{SNR})$ for high SNR follows by choosing

$$
\theta = \frac{-1}{n_{\mathrm r} + \mathsf{SNR}\, n_{\mathrm r} \epsilon_{*,T}^2} \tag{101}
$$

where

$$
\epsilon_{*,T}^2 = \max_{\substack{r=1,\ldots,n_{\mathrm r},\\ t=1,\ldots,n_{\mathrm t},\\ \ell=n_{\mathrm t},\ldots,L-1}} \epsilon_{\ell,T}^2(r,t). \tag{102}
$$

Hence, substituting the choice of $\theta$ in (101), and applying (100) to the RHS of (99), we obtain the following lower bound on $I_T^{\text{gmi}}(\text{SNR})$:

$$I_T^{\text{gmi}}(\text{SNR}) \geq \frac{1}{L} \sum_{\ell=n_t}^{L-1} \left\{ \mathsf{E}\left[ \log\det\left( \mathsf{I}_{n_r} + \frac{\text{SNR}}{n_t n_r + n_t n_r \text{SNR} \epsilon_{*,T}^2} \hat{\mathbb{H}}_\ell^{(T)} \hat{\mathbb{H}}_\ell^{\dagger(T)} \right) \right] - 1 \right\}. \tag{103}$$

### 5.2.2. $I_T^{\text{gmi}}(\text{SNR})$ as $T \to \infty$

We next analyze the RHS of (103) in the limit as $T$ tends to infinity. To this end, we note that, for $L \leq \frac{1}{2\lambda_D}$, the variance of the interpolation error tends to (15), namely

$$\epsilon_\ell^2(t) = 1 - \int_{-1/2}^{1/2} \frac{\text{SNR}\,[f_H(\lambda)]^2}{\text{SNR} f_H(\lambda) + L n_t} d\lambda \tag{104}$$

as $T$ tends to infinity, irrespective of $\ell$ and $t$. We shall therefore omit the subscript and argument and write $\epsilon^2$ instead of $\epsilon_\ell^2(t)$. Please note that for a fixed $T$, the entries of

$$\frac{1}{\sqrt{n_t n_r + n_t n_r \text{SNR} \epsilon_{*,T}^2}} \hat{\mathbb{H}}_\ell^{(T)} \tag{105}$$

are independent but not i.i.d., which follows from Part 2) of Lemma 1. However, as $T$ tends to infinity, their distribution becomes identical due to (104) and hence they converge in distribution to

$$\frac{1}{\sqrt{n_t n_r + n_t n_r \text{SNR} \epsilon_{*,T}^2}} \hat{\mathbb{H}}_\ell^{(T)} \xrightarrow{d} \frac{1}{\sqrt{n_t n_r + n_t n_r \text{SNR} \epsilon^2}} \bar{\mathbb{H}} \tag{106}$$

where the entries of $\bar{\mathbb{H}}$ are i.i.d. complex-Gaussian random variables with zero mean and variance $(1 - \epsilon^2)$.

Next note that

$$\log\det\left( \mathsf{I}_{n_r} + \frac{\text{SNR}}{n_t n_r + n_t n_r \text{SNR} \epsilon_{*,T}^2} \hat{\mathbb{H}}_\ell^{(T)} \hat{\mathbb{H}}_\ell^{\dagger(T)} \right) \tag{107}$$

is a nonnegative, continuous function with respect to the entries of the matrix

$$\frac{1}{n_t n_r + n_t n_r \text{SNR} \epsilon_{*,T}^2} \hat{\mathbb{H}}_\ell^{(T)} \hat{\mathbb{H}}_\ell^{\dagger(T)}. \tag{108}$$

It therefore follows from Portmanteau's lemma [30] that, as $T \to \infty$, the RHS of (103) can be lower-bounded by

$$\lim_{T \to \infty} \frac{1}{L} \sum_{\ell=n_t}^{L-1} \left\{ \mathsf{E}\left[ \log\det\left( \mathsf{I}_{n_r} + \frac{\text{SNR}}{n_t n_r + n_t n_r \text{SNR} \epsilon_{*,T}^2} \hat{\mathbb{H}}_\ell^{(T)} \hat{\mathbb{H}}_\ell^{\dagger(T)} \right) \right] - 1 \right\}$$

$$\geq \frac{L - n_t}{L} \left\{ \mathsf{E}\left[ \log\det\left( \mathsf{I}_{n_r} + \frac{\text{SNR}}{n_t n_r + n_t n_r \text{SNR} \epsilon^2} \bar{\mathbb{H}}\bar{\mathbb{H}}^\dagger \right) \right] - 1 \right\} \tag{109}$$

$$\geq \frac{L - n_t}{L} \left( \mathsf{E}\left[ \log\det\left( \frac{\text{SNR}}{n_t n_r + n_t n_r \text{SNR} \epsilon^2} \bar{\mathbb{H}}\bar{\mathbb{H}}^\dagger \right) \right] - 1 \right) \tag{110}$$

where the last inequality follows from the lower bound $\log\det(\mathsf{I} + \mathsf{A}) \geq \log\det \mathsf{A}$.

Combining (110) with (103), and using that, by assumption, $n_t = n_r$, we obtain that

$$I^{\text{gmi}}(\text{SNR}) \triangleq \lim_{T \to \infty} I_T^{\text{gmi}}(\text{SNR}) \tag{111}$$

$$\geq \frac{L - n_t}{L} \left( n_t \log \text{SNR} - n_t \log \left( n_t^2 + n_t^2 \text{SNR} \epsilon^2 \right) + \mathsf{E} \left[ \log \det \bar{\mathbb{H}} \bar{\mathbb{H}}^\dagger \right] - 1 \right). \tag{112}$$

### 5.2.3. The Pre-Log

It remains to compute a lower bound on the pre-log. To this end, we compute the limiting ratio of the RHS of (112) to $\log \text{SNR}$ as SNR tends to infinity. We first consider

$$\text{SNR}\, \epsilon^2 = \text{SNR} \left( 1 - \int_{-1/2}^{1/2} \frac{\text{SNR}\, [f_H(\lambda)]^2}{\text{SNR} f_H(\lambda) + L n_t} d\lambda \right) \tag{113}$$

$$= \int_{-1/2}^{1/2} \frac{\text{SNR} f_H(\lambda) L n_t}{\text{SNR} f_H(\lambda) + L n_t} d\lambda. \tag{114}$$

Since the integrand is bounded by

$$0 \leq \frac{\text{SNR} f_H(\lambda) L n_t}{\text{SNR} f_H(\lambda) + L} \leq L n_t \tag{115}$$

it follows that $0 \leq \text{SNR}\, \epsilon^2 \leq L n_t$, which implies that

$$\lim_{\text{SNR} \to \infty} \frac{\log \left( n_t^2 + n_t^2 \text{SNR}\, \epsilon^2 \right)}{\log \text{SNR}} = 0. \tag{116}$$

We next consider the term $\mathsf{E} \left[ \log \det \bar{\mathbb{H}} \bar{\mathbb{H}}^\dagger \right] - 1$. Please note that by ([31], Lemma A.2) and the assumption $n_t = n_r$, we have

$$\mathsf{E} \left[ \log \det \bar{\mathbb{H}} \bar{\mathbb{H}}^\dagger \right] - 1 = n_t \log(1 - \epsilon^2) + \sum_{b=0}^{n_t - 1} \psi(n_t - b) - 1 \tag{117}$$

where $\psi(\cdot)$ is Euler's digamma function [32]. Furthermore, since

$$0 \leq \frac{\text{SNR}\, [f_H(\lambda)]^2}{\text{SNR} f_H(\lambda) + L n_t} \leq f_H(\lambda) \tag{118}$$

we have by the Dominated Convergence Theorem [28] that

$$\lim_{\text{SNR} \to \infty} \epsilon^2 = \lim_{\text{SNR} \to \infty} \left( 1 - \int_{-1/2}^{1/2} \frac{\text{SNR}\, [f_H(\lambda)]^2}{\text{SNR} f_H(\lambda) + L n_t} d\lambda \right) = 0 \tag{119}$$

so $\log(1 - \epsilon^2)$ vanishes as the SNR tends to infinity. Combining (119) with (117) yields

$$\lim_{\text{SNR} \to \infty} \frac{\mathsf{E} \left[ \log \det \bar{\mathbb{H}} \bar{\mathbb{H}}^\dagger \right] - 1}{\log \text{SNR}} = 0. \tag{120}$$

It follows from (112), (116), and (120) that

$$\Pi_{R^*} \geq n_t \left( 1 - \frac{n_t}{L} \right) \tag{121}$$

$$= \min(n_t, n_r) \left( 1 - \frac{\min(n_t, n_r)}{L} \right), \quad L \leq \frac{1}{2\lambda_D} \tag{122}$$

where we have used that $n_t = n_r = \min(n_t, n_r)$. Please note that the condition $L \leq \frac{1}{2\lambda_D}$ is necessary since otherwise (104) would not hold. This proves Theorem 1.

*5.3. A Note on the Input Distribution*

The pre-log in Theorem 1 is derived using codebooks whose entries are drawn i.i.d. from an $n_t$-variate Gaussian distribution with zero mean and identity covariance matrix. However, Gaussian inputs are not necessary to achieve the pre-log (25). In fact, as we shall argue next, the pre-log (25) can be achieved by any i.i.d. inputs with a probability density function satisfying $\mathsf{E}[\|\bar{X}\|^2] \leq n_t$ and (26) and (27), namely,

$$p_{\bar{X}}(\bar{x}) \leq \frac{K}{\pi^{n_t}} e^{-\|\bar{x}\|^2}, \quad \bar{x} \in \mathbb{C}^{n_t} \tag{123}$$

$$\lim_{\mathsf{SNR}\to\infty} \frac{\log K}{\log \mathsf{SNR}} = 0. \tag{124}$$

Indeed, since the inputs have a density, they also satisfy $\mathsf{E}[\|\bar{X}\|^2] > 0$. To show that the conditions (26) and (27) suffice to achieve (25), we follow the steps in Section 5.2 but with $F(\mathsf{SNR})$ replaced by

$$F(\mathsf{SNR}) = n_r + \frac{\mathsf{SNR}}{(L - n_t)n_t} \sum_{\ell=n_t}^{L-1} \mathsf{E}\left[\left\|\mathbb{E}_\ell^{(T)} \bar{X}_\ell\right\|_F^2\right]. \tag{125}$$

We then upper-bound $F(\mathsf{SNR})$ and $\kappa(\theta, \mathsf{SNR})$ as follows. Using that for any two matrices A and B we have $\|AB\|_F^2 \leq \|A\|_F^2 \cdot \|B\|_F^2$ [33, Section 5.6], and using that $\mathbb{E}_\ell^{(T)}$ and $\bar{X}_\ell$ are independent, we can upper-bound $F(\mathsf{SNR})$ by

$$F(\mathsf{SNR}) \leq n_r + \frac{\mathsf{SNR}}{(L - n_t)n_t} \sum_{\ell=n_t}^{L-1} \mathsf{E}\left[\left\|\mathbb{E}_\ell^{(T)}\right\|_F^2\right] \cdot \mathsf{E}\left[\|\bar{X}_\ell\|^2\right]. \tag{126}$$

As for $\kappa(\theta, \mathsf{SNR})$, we have

$$\mathsf{E}\left[\exp\left(\frac{\theta}{n} D_k(m')\right) \middle| y_k, \hat{\mathsf{H}}_k^{(T)}\right]$$

$$= \int_{\bar{x}_k} p_{\bar{X}}(\bar{x}_k) \exp\left(\frac{\theta}{n}\left\|y_k - \sqrt{\frac{\mathsf{SNR}}{n_t}} \hat{\mathsf{H}}_k^{(T)} \bar{x}_k\right\|^2\right) d\bar{x}_k \tag{127}$$

$$\leq \int_{\bar{x}_k} \frac{K}{\pi^{n_t}} \exp\left(-\|\bar{x}_k\|^2 + \frac{\theta}{n}\left\|y_k - \sqrt{\frac{\mathsf{SNR}}{n_t}} \hat{\mathsf{H}}_k^{(T)} \bar{x}_k\right\|^2\right) d\bar{x}_k \tag{128}$$

$$= \frac{K}{\det\left(\mathsf{I}_{n_r} - \frac{\theta}{n}\frac{\mathsf{SNR}}{n_t}\hat{\mathsf{H}}_k^{(T)}\hat{\mathsf{H}}_k^{\dagger(T)}\right)} \exp\left(\frac{\theta}{n} y_k^\dagger \left(\mathsf{I}_{n_r} - \frac{\theta}{n}\frac{\mathsf{SNR}}{n_t}\hat{\mathsf{H}}_k^{(T)}\hat{\mathsf{H}}_k^{\dagger(T)}\right)^{-1} y_k\right). \tag{129}$$

Here (128) follows from (123), and (129) follows by evaluating the integral as in ([17], Appendix A). By following the steps in Section 5.2, and by choosing

$$\theta = \frac{-1}{n_r + \mathsf{SNR}\, n_r \epsilon_{*,T}^2 \mathsf{E}\left[\|\bar{X}\|^2\right]} \tag{130}$$

where $\epsilon_{*,T}^2$ is given in (102), we obtain from (126) and (129) that

$$I_T^{\text{gmi}}(\mathsf{SNR}) \geq \frac{1}{L} \sum_{\ell=n_t}^{L-1} \left\{ \mathsf{E}\left[ \log \det \left( \mathsf{I}_{n_r} + \frac{\mathsf{SNR}}{n_t n_r + n_t n_r \mathsf{SNR}\epsilon_{*,T}^2 \mathsf{E}\left[\|\bar{X}\|^2\right]} \hat{\mathbb{H}}_\ell^{(T)} \hat{\mathbb{H}}_\ell^{\dagger(T)} \right) \right] \right\}$$
$$- \frac{L-n_t}{L}\left(1 + \log K\right). \tag{131}$$

Taking the limit as $T$ tends to infinity, and repeating the steps in Section 5.2, it follows that

$$I^{\text{gmi}}(\mathsf{SNR}) = \lim_{T \to \infty} I_T^{\text{gmi}}(\mathsf{SNR}) \tag{132}$$

$$\geq \frac{L-n_t}{L}\left( \mathsf{E}\left[ \log \det \left( \frac{\mathsf{SNR}}{n_t n_r + n_t n_r \mathsf{SNR}\,\epsilon^2 \mathsf{E}\left[\|\bar{X}\|^2\right]} \bar{\mathbb{H}}\bar{\mathbb{H}}^\dagger \right) \right] - 1 - \log K \right) \tag{133}$$

$$= \frac{L-n_t}{L}\left( n_t \log \mathsf{SNR} - n_t \log \left( n_t{}^2 + n_t{}^2 \mathsf{SNR}\,\epsilon^2 \mathsf{E}\left[\|\bar{X}\|^2\right] \right) \right.$$
$$\left. + \mathsf{E}\left[ \log \det \bar{\mathbb{H}}\bar{\mathbb{H}}^\dagger \right] - 1 - \log K \right) \tag{134}$$

where we have again used the assumption $n_t = n_r$.

We conclude by evaluating the limiting ratio of the RHS of (134) to $\log \mathsf{SNR}$ as $\mathsf{SNR}$ tends to infinity. Using (115) and that $\mathsf{E}[\|\bar{X}\|^2] \leq n_t$, we obtain that

$$\lim_{\mathsf{SNR}\to\infty} \frac{\log\left( n_t{}^2 + n_t{}^2 \mathsf{SNR}\,\epsilon^2 \mathsf{E}\left[\|\bar{X}\|^2\right] \right)}{\log \mathsf{SNR}} = 0. \tag{135}$$

This in turn yields together with (120) that

$$\lim_{\mathsf{SNR}\to\infty} \frac{I^{\text{gmi}}(\mathsf{SNR})}{\log \mathsf{SNR}} \geq n_t \left( 1 - \frac{n_t}{L} \right) \tag{136}$$

provided that

$$\lim_{\mathsf{SNR}\to\infty} \frac{\log K}{\log \mathsf{SNR}} = 0. \tag{137}$$

It thus follows that any i.i.d. input distribution satisfying $\mathsf{E}[\|\bar{X}\|^2] \leq n_t$ and (26) and (27) achieves the pre-log (25).

## 6. Proof of Theorem 2

In contrast to the proof of Theorem 1, for the fading MAC, it is not sufficient to restrict ourselves to the case of $n_{t,1} = n_{t,2} = n_r$. For example, increasing $n_r$ beyond $n_{t,1}$ and $n_{t,2}$ does not increase the single-rate pre-logs $\Pi_{R_1^*}$ and $\Pi_{R_2^*}$, but it does increase the pre-log of the achievable sum-rate $\Pi_{R_{1+2}^*}$. For the proof of Theorem 2, we therefore consider a general setup of $n_{t,1}$, $n_{t,2}$, and $n_r$.

We derive the achievable pre-logs for the MAC case by following similar steps as in the point-to-point case. We first consider the average error probability, averaged over the ensemble of i.i.d. Gaussian codebooks. Let $\bar{P}_e$ and $\bar{P}_e(m_1, m_2)$ be the ensemble-average error probability and the ensemble-average error probability when messages $m_1$ and $m_2$ are transmitted, respectively. Due to the symmetry of the codebook construction, $\bar{P}_e$ is equal to $\bar{P}_e(1,1)$ and it therefore suffices to consider $\bar{P}_e(1,1)$ to derive the achievable rates. Let $\mathcal{E}(m_1', m_2')$ denote the event that $D(m_1', m_2') \leq D(1,1)$,

where $D(\cdot, \cdot)$ was defined in (45). Using the union bound, the error probability $\bar{P}_e(1,1)$ can be upper-bounded as

$$\bar{P}_e(1,1) = \Pr\left\{ \bigcup_{(m_1', m_2') \neq (1,1)} \mathcal{E}(m_1', m_2') \right\} \tag{138}$$

$$\leq \Pr\left\{ \bigcup_{m_1' \neq 1} \mathcal{E}(m_1', 1) \right\} + \Pr\left\{ \bigcup_{m_2' \neq 1} \mathcal{E}(1, m_2') \right\} + \Pr\left\{ \bigcup_{m_1' \neq 1} \bigcup_{m_2' \neq 1} \mathcal{E}(m_1', m_2') \right\}. \tag{139}$$

We next analyze the three probabilities on the RHS of (139). Let the matrix $\mathbb{E}_{s,k}^{(T)}$, $s = 1, 2$ with entries $E_{s,k}^{(T)}(r,t)$ be the estimation-error matrix in estimating $\mathbb{H}_{s,k}$, i.e.,

$$\mathbb{E}_{s,k}^{(T)} = \mathbb{H}_{s,k} - \hat{\mathbb{H}}_{s,k}^{(T)}. \tag{140}$$

To facilitate the analysis, we first generalize $F(\mathsf{SNR})$ and $\mathcal{T}_\delta$, defined in the point-to-point case in (77) and (78), to the MAC case:

$$F(\mathsf{SNR}) \triangleq n_{\mathrm{r}} + \frac{\mathsf{SNR}}{L - n_{\mathrm{t},1} - n_{\mathrm{t},2}} \sum_{\ell = n_{\mathrm{t},1} + n_{\mathrm{t},2}}^{L-1} \mathsf{E}\left[ \left\| \mathbb{E}_{1,\ell}^{(T)} \right\|_F^2 + \left\| \mathbb{E}_{2,\ell}^{(T)} \right\|_F^2 \right], \tag{141}$$

$$\mathcal{T}_\delta \triangleq \left\{ \left( x_{s,k}, y_k, \hat{\mathsf{H}}_{s,k}^{(T)} \right), k = 0, \ldots, n' - 1, s = 1, 2 : \right.$$

$$\left. \left| \frac{1}{n} \sum_{k \in \mathcal{D}^{(n')}} \left\| y_k - \sqrt{\mathsf{SNR}}\, \hat{\mathsf{H}}_{1,k}^{(T)} x_{1,k} - \sqrt{\mathsf{SNR}}\, \hat{\mathsf{H}}_{2,k}^{(T)} x_{2,k} \right\|^2 - F(\mathsf{SNR}) \right| < \delta \right\} \tag{142}$$

for some $\delta > 0$, with $n'$ given in (41) and $\mathcal{D}^{(n')} = \{0, \ldots, n' - 1\} \cap \mathcal{D}$. Using $F(\mathsf{SNR})$ and the typical set $\mathcal{T}_\delta$, we continue by evaluating the GMI for each of the three probabilities on the RHS of (139), which correspond to the error events $(m_1' \neq 1, m_2' = 1)$, $(m_1' = 1, m_2' \neq 1)$, and $(m_1' \neq 1, m_2' \neq 1)$.

*6.1. Error Event $(m_1' \neq 1, m_2' = 1)$*

Following the steps in Section 5.2 to derive (86), we can upper-bound the ensemble-average error probability for the error event $\mathcal{E}(m_1', 1)$, $m_1' \neq 1$ as

$$\Pr\left\{ \bigcup_{m_1' \neq 1} \mathcal{E}(m_1', 1) \right\}$$

$$\leq e^{nR_1} \cdot \Pr\left\{ \frac{1}{n} \cdot D(m_1', 1) < F(\mathsf{SNR}) + \delta \,\middle|\, \left\{ \left( \mathbf{X}_s^{n'}(1), \mathbf{Y}^{n'}, \hat{\mathbb{H}}_s^{(T),n'} \right), s = 1, 2 \right\} \in \mathcal{T}_\delta \right\}$$

$$+ \Pr\left\{ \left\{ \left( \mathbf{X}_s^{n'}(1), \mathbf{Y}^{n'}, \hat{\mathbb{H}}_s^{(T),n'} \right), s = 1, 2 \right\} \in \mathcal{T}_\delta^c \right\}, \quad m_1' \neq 1. \tag{143}$$

Note that the second probability on the RHS of (143) vanishes as $n$ tends to infinity, which can be shown along the lines of the proof of Lemma 2.

The GMI of user 1 gives the rate of exponential decay of the term

$$\Pr\left\{ \frac{1}{n} \cdot D(m_1', 1) < F(\mathsf{SNR}) + \delta \,\middle|\, \left\{ \left( \mathbf{X}_s^{n'}(1), \mathbf{Y}^{n'}, \hat{\mathbb{H}}_s^{(T),n'} \right), s = 1, 2 \right\} \in \mathcal{T}_\delta \right\} \tag{144}$$

as $n \to \infty$. Its evaluation requires the expression of the log moment-generating function of the metric $D(m'_1, 1)$, conditioned on the channel outputs, on $m'_2 = 1$, and on the fading estimates, which is defined as

$$\kappa_{1,n}\left(\theta, \mathbf{y}^{n'}, \mathbf{x}_2^{n'}(1), \hat{\mathsf{H}}_1^{(T),n'}, \hat{\mathsf{H}}_2^{(T),n'}\right)$$

$$\triangleq \log \mathsf{E}\left[\exp\left(\frac{\theta}{n}\sum_{k\in\mathcal{D}^{(n')}} D_k(m'_1, 1)\right) \middle| \left\{\left(\mathbf{y}_k, \mathbf{x}_{2,k}(1), \hat{\mathsf{H}}_{1,k}^{(T)}, \hat{\mathsf{H}}_{2,k}^{(T)}\right), k \in \mathcal{D}^{(n')}\right\}\right] \tag{145}$$

where

$$D_k(m'_1, m'_2) \triangleq \left\|\mathbf{y}_k - \sqrt{\mathsf{SNR}}\,\hat{\mathsf{H}}_{1,k}^{(T)}\mathbf{x}_{1,k}(m'_1) - \sqrt{\mathsf{SNR}}\,\hat{\mathsf{H}}_{2,k}^{(T)}\mathbf{x}_{2,k}(m'_2)\right\|^2. \tag{146}$$

Following the steps used in Section 5.2 to obtain (90) and (91), it can be shown that

$$\kappa_{1,n}\left(\theta, \mathbf{y}^{n'}, \mathbf{x}_2^{n'}(1), \hat{\mathsf{H}}_1^{(T),n'}, \hat{\mathsf{H}}_2^{(T),n'}\right)$$

$$= \sum_{k\in\mathcal{D}^{(n')}}\left\{\frac{\theta}{n}\left(\mathbf{y}_k - \sqrt{\mathsf{SNR}}\,\hat{\mathsf{H}}_{2,k}^{(T)}\mathbf{x}_{2,k}(1)\right)^\dagger\left(\mathsf{I}_{n_\mathrm{r}} - \frac{\theta}{n}\mathsf{SNR}\,\hat{\mathsf{H}}_{1,k}^{(T)}\hat{\mathsf{H}}_{1,k}^{\dagger(T)}\right)^{-1}\left(\mathbf{y}_k - \sqrt{\mathsf{SNR}}\,\hat{\mathsf{H}}_{2,k}^{(T)}\mathbf{x}_{2,k}(1)\right)\right.$$

$$\left. - \log\det\left(\mathsf{I}_{n_\mathrm{r}} - \frac{\theta}{n}\mathsf{SNR}\,\hat{\mathsf{H}}_{1,k}^{(T)}\hat{\mathsf{H}}_{1,k}^{\dagger(T)}\right)\right\}. \tag{147}$$

Then, following the steps used in Section 5.2 to derive (92)–(94), we obtain that, for all $\theta < 0$,

$$\lim_{n\to\infty}\frac{1}{n}\cdot\kappa_{1,n}\left(n\theta, \mathbf{y}^{n'}, \mathbf{x}_2^{n'}(1), \hat{\mathsf{H}}_1^{(T),n'}, \hat{\mathsf{H}}_2^{(T),n'}\right)$$

$$= \frac{1}{L - n_{\mathrm{t},1} - n_{\mathrm{t},2}}\sum_{\ell=n_{\mathrm{t},1}+n_{\mathrm{t},2}}^{L-1}\left(g_{1,\ell}(\theta, \mathsf{SNR}) - \mathsf{E}\left[\log\det\left(\mathsf{I}_{n_\mathrm{r}} - \theta\,\mathsf{SNR}\,\hat{\mathbb{H}}_{1,\ell}^{(T)}\hat{\mathbb{H}}_{1,\ell}^{\dagger(T)}\right)\right]\right) \tag{148}$$

$$\triangleq \kappa_1(\theta, \mathsf{SNR}), \quad \text{almost surely} \tag{149}$$

where the last step should be regarded as the definition of $\kappa_1(\theta, \mathsf{SNR})$. In (148), we define

$$g_{1,\ell}(\theta, \mathsf{SNR}) \triangleq \mathsf{E}\left[\theta\left(\mathbf{Y}_\ell - \sqrt{\mathsf{SNR}}\,\hat{\mathbb{H}}_{2,\ell}^{(T)}\mathbf{X}_{2,\ell}\right)^\dagger\left(\mathsf{I}_{n_\mathrm{r}} - \theta\,\mathsf{SNR}\,\hat{\mathbb{H}}_{1,\ell}^{(T)}\hat{\mathbb{H}}_{1,\ell}^{\dagger(T)}\right)^{-1}\left(\mathbf{Y}_\ell - \sqrt{\mathsf{SNR}}\,\hat{\mathbb{H}}_{2,\ell}^{(T)}\mathbf{X}_{2,\ell}\right)\right]. \tag{150}$$

Following the derivation in [16,17], we can then upper-bound

$$\Pr\left\{\bigcup_{m'_1\neq 1}\mathcal{E}(m'_1, 1)\right\} \leq \exp\left(nR_1\right)\exp\left(-n\left(I_{1,T}^{\mathrm{gmi}}(\mathsf{SNR}) - \delta'\right)\right) + \varepsilon_1(\delta', n) \tag{151}$$

for any $\delta' > 0$, and for some $\varepsilon_1(\delta', n)$ satisfying

$$\lim_{n\to\infty}\varepsilon_1(\delta', n) = 0, \quad \delta' > 0. \tag{152}$$

On the RHS of (151), $I_{1,T}^{\mathrm{gmi}}(\mathsf{SNR})$ denotes the GMI for user 1 as a function of the SNR for a fixed $T$, i.e.,

$$I_{1,T}^{\mathrm{gmi}}(\mathsf{SNR}) = \frac{L - n_{\mathrm{t},1} - n_{\mathrm{t},2}}{L}\left(\sup_{\theta<0}\left(\theta F(\mathsf{SNR}) - \kappa_1(\theta, \mathsf{SNR})\right)\right). \tag{153}$$

The pre-factor $(L - n_{\mathrm{t},1} - n_{\mathrm{t},2})/L$ equals the fraction of time used for data transmission. The bound (151) implies that, for all rates below $I_{1,T}^{\mathrm{gmi}}(\mathsf{SNR})$, the error probability in decoding user 1's message for the scheme described in Section 4 vanishes as $n$ tends to infinity.

Combining (141) and (148) with (153), we obtain that

$$I_{1,T}^{\text{gmi}}(\text{SNR}) = \sup_{\theta<0} \frac{1}{L} \sum_{\ell=n_{\text{t},1}+n_{\text{t},2}}^{L-1} \left\{ \theta \left( n_{\text{r}} + \text{SNR}\, \mathsf{E}\left[ \left\| \mathbb{E}_{1,\ell}^{(T)} \right\|_F^2 + \left\| \mathbb{E}_{2,\ell}^{(T)} \right\|_F^2 \right] \right) - g_{1,\ell}(\theta, \text{SNR}) \right.$$

$$\left. + \mathsf{E}\left[ \log \det \left( \mathsf{I}_{n_{\text{r}}} - \theta\, \text{SNR}\, \hat{\mathbb{H}}_{1,\ell}^{(T)} \hat{\mathbb{H}}_{1,\ell}^{\dagger(T)} \right) \right] \right\}. \tag{154}$$

Since the supremum over $\theta < 0$ is difficult to evaluate, we next consider a lower bound on $I_{1,T}^{\text{gmi}}(\text{SNR})$. By noting that $g_{1,\ell}(\theta, \text{SNR}) \leq 0$ for $\theta \leq 0$ (which can be shown using the technique developed in ([29], Appendix D), and by choosing

$$\theta = \frac{-1}{n_{\text{r}} + n_{\text{r}}\, (n_{\text{t},1} + n_{\text{t},2})\, \text{SNR}\, \epsilon_{*,T}^2} \tag{155}$$

where

$$\epsilon_{*,T}^2 = \max_{\substack{s=1,2,\\ r=1,\ldots,n_{\text{r}},\\ t=1,\ldots,n_{\text{t},s},\\ \ell=n_{\text{t},1}+n_{\text{t},2},\ldots,L-1}} \mathsf{E}\left[ \left| E_{s,\ell}^{(T)}(r,t) \right|^2 \right] \tag{156}$$

we obtain the following lower bound on $I_{1,T}^{\text{gmi}}(\text{SNR})$:

$$I_{1,T}^{\text{gmi}}(\text{SNR}) \geq \frac{1}{L} \sum_{\ell=n_{\text{t},1}+n_{\text{t},2}}^{L-1} \mathsf{E}\left[ \log \det \left( \mathsf{I}_{n_{\text{r}}} + \frac{\text{SNR}\, \hat{\mathbb{H}}_{1,\ell}^{(T)} \hat{\mathbb{H}}_{1,\ell}^{\dagger(T)}}{n_{\text{r}} + n_{\text{r}}\, (n_{\text{t},1} + n_{\text{t},2})\, \text{SNR}\, \epsilon_{*,T}^2} \right) - 1 \right]. \tag{157}$$

(As pointed out in Section 5, this choice of $\theta$ yields a good lower bound at high SNR.) We continue by analyzing the RHS of (157) in the limit as the observation window $T$ of the channel estimator tends to infinity. To this end, we note that, for $L \leq \frac{1}{2\lambda_D}$, the variance of the interpolation error tends to (15) (with SNR in (15) replaced by $n_{\text{t}}$SNR due to the difference between the point-to-point channel model (1) and the MAC channel model (40)), so

$$\lim_{T\to\infty} \mathsf{E}\left[ \left| E_{s,\ell}^{(T)}(r,t) \right|^2 \right] = \epsilon^2 = 1 - \int_{-1/2}^{1/2} \frac{\text{SNR}\, [f_H(\lambda)]^2}{\text{SNR} f_H(\lambda) + L}\, d\lambda \tag{158}$$

irrespective of $s, \ell, r$, and $t$. It follows that the estimate $\hat{\mathbb{H}}_{1,\ell}^{(T)}$ tends to $\bar{\mathbb{H}}_1$ in distribution as $T$ tends to infinity, which implies that

$$\frac{\hat{\mathbb{H}}_{1,\ell}^{(T)} \hat{\mathbb{H}}_{1,\ell}^{\dagger(T)}}{n_{\text{r}} + n_{\text{r}}\, (n_{\text{t},1} + n_{\text{t},2})\, \text{SNR}\, \epsilon_{*,T}^2} \xrightarrow{\,d\,} \frac{\bar{\mathbb{H}}_1 \bar{\mathbb{H}}_1^{\dagger}}{n_{\text{r}} + n_{\text{r}}\, (n_{\text{t},1} + n_{\text{t},2})\, \text{SNR}\, \epsilon^2} \tag{159}$$

where the $n_{\text{r}} \times n_{\text{t},1}$ entries of $\bar{\mathbb{H}}_1$ are i.i.d., circularly-symmetric, complex-Gaussian random variables with zero mean and variance $(1 - \epsilon^2)$. Using Portmanteau's lemma (as used in (109)), we obtain that

$$I_1^{\text{gmi}}(\text{SNR}) = \lim_{T\to\infty} I_{1,T}^{\text{gmi}}(\text{SNR}) \tag{160}$$

$$\geq \frac{L - n_{\text{t},1} - n_{\text{t},2}}{L} \left( \mathsf{E}\left[ \log \det \left( \mathsf{I}_{n_{\text{r}}} + \frac{\text{SNR}\, \bar{\mathbb{H}}_1 \bar{\mathbb{H}}_1^{\dagger}}{n_{\text{r}} + n_{\text{r}}\, (n_{\text{t},1} + n_{\text{t},2})\, \text{SNR}\, \epsilon^2} \right) \right] - 1 \right) \tag{161}$$

$$\geq \frac{L - n_{\text{t},1} - n_{\text{t},2}}{L} \min(n_{\text{r}}, n_{\text{t},1}) \left[ \log \text{SNR} - \log \left( n_{\text{r}} + n_{\text{r}}(n_{\text{t},1} + n_{\text{t},2})\, \text{SNR}\, \epsilon^2 \right) \right]$$

$$+ \frac{L - n_{\text{t},1} - n_{\text{t},2}}{L} \Psi_1 \tag{162}$$

where

$$\Psi_1 \triangleq \begin{cases} \mathsf{E}\left[\log\det \bar{\mathbb{H}}_1 \bar{\mathbb{H}}_1^\dagger\right] - 1, & n_\mathrm{r} \le n_{\mathrm{t},1} \\ \mathsf{E}\left[\log\det \bar{\mathbb{H}}_1^\dagger \bar{\mathbb{H}}_1\right] - 1, & n_\mathrm{r} > n_{\mathrm{t},1}. \end{cases} \tag{163}$$

The inequality (162) follows by lower-bounding $\log\det (\mathsf{I} + \mathsf{A}) \ge \log\det\mathsf{A}$.

By evaluating the limiting ratio of the RHS of (162) to $\log\mathsf{SNR}$ as $\mathsf{SNR}$ tends to infinity following similar steps as in Section 5.2.3, we obtain the following lower bound on the maximum achievable pre-log of user 1:

$$\Pi_{R_1^*} \ge \min(n_\mathrm{r}, n_{\mathrm{t},1})\left(1 - \frac{n_{\mathrm{t},1} + n_{\mathrm{t},2}}{L}\right), \quad L \le \frac{1}{2\lambda_D}. \tag{164}$$

As in the point-to-point case, the condition $L \le 1/(2\lambda_D)$ is necessary to obtain (15). The lower bound (164) yields one boundary of the pre-log region presented in Theorem 2.

### 6.2. Error Event $(m_1' = 1, m_2' \ne 1)$

The error event $(m_1' = 1, m_2' \ne 1)$ can be analyzed by swapping user 1 and user 2 and then using the results obtained in the previous subsection for the error event $(m_1' \ne 1, m_2' = 1)$. We thus have the lower bound

$$\Pi_{R_2^*} \ge \min(n_\mathrm{r}, n_{\mathrm{t},2})\left(1 - \frac{n_{\mathrm{t},1} + n_{\mathrm{t},2}}{L}\right), \quad L \le \frac{1}{2\lambda_D} \tag{165}$$

which yields the second boundary of the pre-log region presented in Theorem 2.

### 6.3. Error Event $(m_1' \ne 1, m_2' \ne 1)$

For the error event $(m_1' \ne 1, m_2' \ne 1)$, the analysis of the achievable sum rate follows the same analysis as in the point-to-point case (Section 5.2). More specifically, the GMI $I_{1+2,T}^{\mathrm{gmi}}(\mathsf{SNR})$ that describes the exponential decay of the term

$$\Pr\left\{\frac{1}{n} \cdot D(m_1', m_2') < F(\mathsf{SNR}) + \delta \,\middle|\, \left\{\left(\boldsymbol{X}_s^{n'}(1), \boldsymbol{Y}^{n'}, \hat{\mathbb{H}}_s^{(T),n'}\right), s = 1, 2\right\} \in \mathcal{T}_\delta\right\} \tag{166}$$

can be viewed as the GMI of an $n_\mathrm{r} \times (n_{\mathrm{t},1} + n_{\mathrm{t},2})$-dimensional point-to-point MIMO channel with fading matrix $[\mathbb{H}_{1,k}, \mathbb{H}_{2,k}]$ and fading estimate matrix $\left[\hat{\mathbb{H}}_{1,k}^{(T)}, \hat{\mathbb{H}}_{2,k}^{(T)}\right]$. The maximum achievable sum-rate pre-log can therefore be obtained by following the same steps as in Section 5.2, but with arbitrary $n_\mathrm{r}$ and $n_\mathrm{t} = n_{\mathrm{t},1} + n_{\mathrm{t},2}$. It can thus be shown that the maximum achievable sum-rate pre-log $\Pi_{R_{1+2}^*}$ is lower-bounded by

$$\Pi_{R_{1+2}^*} \ge \min(n_\mathrm{r}, n_{\mathrm{t},1} + n_{\mathrm{t},2})\left(1 - \frac{n_{\mathrm{t},1} + n_{\mathrm{t},2}}{L}\right), \quad L \le \frac{1}{2\lambda_D}. \tag{167}$$

On the RHS of (167), the term $\min(n_\mathrm{r}, n_{\mathrm{t},1} + n_{\mathrm{t},2})$ corresponds to the MIMO gain, which is given by the minimum number of receive and transmit antennas, and the term $\left(1 - \frac{n_{\mathrm{t},1} + n_{\mathrm{t},2}}{L}\right)$ corresponds to the fraction of time indices for data transmission. This yields the third boundary of the pre-log region presented in Theorem 2.

## 7. Conclusions

In this paper, we studied a communication scheme for MIMO fading channels that estimates the fading via transmission of pilot symbols at regular intervals and feeds the fading estimates to the nearest neighbor decoder. Restricting ourselves to fading processes with a bandlimited power spectral density, we studied the information rates achievable with this scheme at high SNR. Specifically, we analyzed the achievable rate pre-log, defined as the limiting ratio of the achievable rate to the logarithm of the SNR in the limit as the SNR tends to infinity.

We showed that in order to obtain fading estimates whose variance vanishes as the SNR tends to infinity, the portion of time required for pilot transmission must be greater than or equal to the number of transmit antennas times twice the bandwidth of the fading power spectral density. We demonstrated that in this case, the nearest neighbor decoder achieves the capacity pre-log of the coherent fading channel times the fraction of time used for the transmission of data. Hence, the loss with respect to the coherent case is solely due to the transmission of pilots used to obtain accurate fading estimates. Our achievability bounds are tight in the sense that any scheme using as many pilots as our proposed scheme cannot achieve a higher pre-log using a nearest neighbor decoder. Furthermore, if the inverse of twice the bandwidth of the fading process is an integer, then, for MISO channels, our scheme achieves the capacity pre-log of the noncoherent fading channel derived by Koch and Lapidoth [11]. For noncoherent MIMO channels, our scheme achieves the best so far known lower bound on the capacity pre-log obtained by Etkin and Tse [12]. Since the last result only yields a lower bound on the capacity pre-log of MIMO channels, there may exist other schemes achieving a better pre-log than our scheme.

## Appendix A. Proof of Lemma 1

We prove each part of Lemma 1 in a separate item:

1.  By the orthogonality principle [34], it follows that $\hat{H}_k^{(T)}(r,t)$ and $E_k^{(T)}(r,t)$ are uncorrelated. Noting that the pilot symbols are unity, we can write (72) as

$$\hat{H}_k^{(T)}(r,t) = \sum_{\substack{k'=k-TL: \\ k' \in \mathcal{P}}}^{k+TL} a_{k'}(r,t) \left( \sqrt{\frac{\mathsf{SNR}}{n_{\mathrm{t}}}} H_{k'}(r,t) + Z_{k'}(r) \right), \quad k \in \mathcal{D}. \tag{A1}$$

Since the processes $\{H_k(r,t), k \in \mathbb{Z}\}$ and $\{Z_k(r), k \in \mathbb{Z}\}$ are zero-mean complex-Gaussian, we have from (A1) and the orthogonality principle that $\hat{H}_k^{(T)}(r,t)$ and $E_k^{(T)}(r,t)$ are independent zero-mean complex-Gaussian random variables.

2.  Recall from Section 5.1 that the time index $k$ can be written as $k = jL + \ell$. Then, for $k \in \mathcal{D}$, we have $\ell = n_{\mathrm{t}}, \ldots, L-1$, and for $k \in \mathcal{P}$ we have $\ell = 0, \ldots, n_{\mathrm{t}} - 1$. Since the pilot vectors are transmitted sequentially from $\boldsymbol{p}_1$ to $\boldsymbol{p}_{n_{\mathrm{t}}}$, we have for $(jL + \ell) \in \mathcal{P}$ that

$$\boldsymbol{x}_{jL+\ell} = \boldsymbol{p}_{\ell+1}, \quad \ell = 0, \ldots, n_{\mathrm{t}} - 1. \tag{A2}$$

That is, the pilot vectors that help estimate the fading coefficients from transmit antenna $t$ are transmitted at time instants whose remainder after division by $L$ is equal to $t - 1$. This implies that in order to estimate $H_k(r, t)$, there is no loss in optimality by considering only the outputs $Y_{k'}(r)$ for $k' \in \mathcal{P} \cap \{k - TL, \dots, k + TL\}$ satisfying

$$k' \bmod L = t - 1. \tag{A3}$$

Indeed, the channel outputs $Y_{k'}(r)$, $k' \bmod L \neq t - 1$ correspond to $H_{k'}(r, t')$, $t' \neq t$, which are independent from $H_k(r, t)$ since we assumed that the fading processes corresponding to different transmit and receive antennas are independent. It follows that for estimation at time $k = jL + \ell$, the coefficients $a_{k'}(r, t)$ that minimize the mean-squared error depend only on $L$ and $\ell$ [23]. The fading estimate (72) can then be expressed as

$$\hat{H}_{jL+\ell}^{(T)}(r, t) = \sum_{\tau=-T}^{T-1} \alpha_{-\tau L, \ell}(r, t) Y_{(j-\tau)L+t-1}(r) \tag{A4}$$

$$= \sum_{\tau=-T}^{T-1} \alpha_{-\tau L, \ell}(r, t) \left( \sqrt{\frac{\mathsf{SNR}}{n_{\mathrm{t}}}} H_{(j-\tau)L+t-1}(r, t) + Z_{(j-\tau)L+t-1}(r) \right) \tag{A5}$$

where for a given $L$ and $\ell = n_{\mathrm{t}}, \dots, L - 1$, we defined

$$\alpha_{-\tau L, \ell}(r, t) \triangleq a_{(j-\tau)L+t-1}(r, t), \quad \tau = -T, \dots, T - 1. \tag{A6}$$

Noting again that the $n_{\mathrm{r}} \cdot n_{\mathrm{t}}$ processes $\{H_k(r, t), k \in \mathbb{Z}\}$ are independent from each other and have the same law, we obtain the following results from (A5):

(a)   For a given $t$, the time differences between the index of interest $(jL + \ell)$ and the positions of pilots $((j - \tau)L + t - 1)$ do not depend on $r$. It thus follows that for a given $t$, the optimal coefficients $\alpha_{-\tau L, \ell}(r, t)$ are identical for all $r = 1, \dots, n_{\mathrm{r}}$ [23]. This implies that for a given $t$ and $\ell$, the $n_{\mathrm{r}}$ processes

$$\{(\hat{H}_{jL+\ell}^{(T)}(1, t), E_{jL+\ell}^{(T)}(1, t)), j \in \mathbb{Z}\}, \dots, \{(\hat{H}_{jL+\ell}^{(T)}(n_{\mathrm{r}}, t), E_{jL+\ell}^{(T)}(n_{\mathrm{r}}, t)), j \in \mathbb{Z}\} \tag{A7}$$

are independent and have the same law.

(b)   For a given $r$, the time differences between the index of interest $(jL + \ell)$ and the position of pilots $((j - \tau)L + t - 1)$ depend on $t$. It thus follows from [23] that for a given $r$, the optimal coefficients $\alpha_{-\tau L, \ell}(r, t)$ are generally different for $t = 1, \dots, n_{\mathrm{t}}$. This implies that for a given $r$ and $\ell$, the $n_{\mathrm{t}}$ processes

$$\{(\hat{H}_{jL+\ell}^{(T)}(r, 1), E_{jL+\ell}^{(T)}(r, 1)), j \in \mathbb{Z}\}, \dots, \{(\hat{H}_{jL+\ell}^{(T)}(r, n_{\mathrm{t}}), E_{jL+\ell}^{(T)}(r, n_{\mathrm{t}})), j \in \mathbb{Z}\} \tag{A8}$$

are independent but have different laws.

3.   We first note that $\{\mathbb{H}_k, k \in \mathbb{Z}\}$ is an ergodic Gaussian process, which implies that it is also weakly mixing [35]. (See [36] for a definition of a weakly-mixing process.) Since $\{Z_k, k \in \mathbb{Z}\}$ is an i.i.d. Gaussian process and independent from $\{\mathbb{H}_k, k \in \mathbb{Z}\}$, it follows from [36, Proposition 1.6] that $\{(\mathbb{H}_k, Z_k), k \in \mathbb{Z}\}$ is jointly ergodic.

We next evaluate the process $\{(\hat{\mathbb{H}}_k^{(T)}, \mathbb{H}_k, Z_k), k \in \mathcal{D}\}$. Please note that this process cannot be expressed directly as a time-invariant function of $\{(\mathbb{H}_k, Z_k), k \in \mathbb{Z}\}$. Indeed, by assuming $k = jL + \ell$, we can see from (A5) that the function to produce $\hat{\mathbb{H}}_k^{(T)}$ from $\{(\mathbb{H}_k, Z_k), k \in \mathbb{Z}\}$ depends on the time index $k$ via $\ell$. To sidestep this problem, and to facilitate the analysis, we

introduce a "dummy" matrix-valued process $\{\mathbb{A}_{k,\ell},\ k \in \mathbb{Z}\}$ of dimension $n_{\mathrm{r}} \times n_{\mathrm{t}}$, where the row-$r$ column-$t$ entry of $\mathbb{A}_{k,\ell}$ is given by

$$A_{k,\ell}(r,t) = \sum_{\tau=-T}^{T-1} \alpha_{-\tau L,\ell}(r,t) \left( \sqrt{\frac{\mathrm{SNR}}{n_{\mathrm{t}}}} H_{k-\tau L-\ell+t-1}(r,t) + Z_{k-\tau L-\ell+t-1}(r) \right). \tag{A9}$$

Here the coefficients $\alpha_{-\tau L,\ell}$, $\tau = -T,\ldots,T$ have the same value as those in (A5) for a given $L$ and $\ell$. Consequently, for every $\ell = n_{\mathrm{t}},\ldots,L-1$, the process $\{\mathbb{A}_{k,\ell},\ k \in \mathbb{Z}\}$ is a time-invariant function of $\{(\mathbb{H}_k, \mathbf{Z}_k),\ k \in \mathbb{Z}\}$ that coincides with $\hat{\mathbb{H}}_k^{(T)}$ for $k = jL + \ell$. This in turn implies that for every $\ell = n_{\mathrm{t}},\ldots,L-1$, the process $\{(\mathbb{A}_{k,\ell}, \mathbb{H}_k, \mathbf{Z}_k),\ k \in \mathbb{Z}\}$ is jointly weakly mixing. Furthermore, by the definition of weakly mixing [35–37], the process $\{(\mathbb{A}_{jL+\ell,\ell}, \mathbb{H}_{jL+\ell}, \mathbf{Z}_{jL+\ell}),\ j \in \mathbb{Z}\}$ (for every $\ell = n_{\mathrm{t}},\ldots,L-1$) is also jointly weakly mixing. Since for $k = jL + \ell$, $k \in \mathcal{D}$, the matrix $\mathbb{A}_{jL+\ell,\ell}$ is identical to $\hat{\mathbb{H}}_{jL+\ell}^{(T)}$, it follows that the process $\{(\hat{\mathbb{H}}_{jL+\ell}^{(T)}, \mathbb{H}_{jL+\ell}, \mathbf{Z}_{jL+\ell}),\ j \in \mathbb{Z}\}$ (for every $\ell = n_{\mathrm{t}},\ldots,L-1$) is jointly weakly mixing, which implies ergodicity.

We finally evaluate the joint behavior of the two processes $\{(\hat{\mathbb{H}}_{jL+\ell}^{(T)}, \mathbb{H}_{jL+\ell}, \mathbf{Z}_{jL+\ell}),\ j \in \mathbb{Z}\}$ and $\{\mathbf{X}_{jL+\ell},\ j \in \mathbb{Z}\}$ for $\ell = n_{\mathrm{t}},\ldots,L-1$. Since, for every $\ell = n_{\mathrm{t}},\ldots,L-1$, the process $\{\mathbf{X}_{jL+\ell},\ j \in \mathbb{Z}\}$ is i.i.d. and independent from $\{(\hat{\mathbb{H}}_{jL+\ell}^{(T)}, \mathbb{H}_{jL+\ell}, \mathbf{Z}_{jL+\ell}),\ j \in \mathbb{Z}\}$, we have by [38, Lemma 2] that

$$\{(\hat{\mathbb{H}}_{jL+\ell}^{(T)},\ \mathbb{H}_{jL+\ell},\ \mathbf{Z}_{jL+\ell},\ \mathbf{X}_{jL+\ell}),\ j \in \mathbb{Z}\},\quad \ell = n_{\mathrm{t}},\ldots,L-1 \tag{A10}$$

is jointly ergodic.

4.　Please note that the process $\{\hat{\mathbb{H}}_k^{(T)},\ k \in \mathcal{D}\}$ is a function of $\{(\mathbb{H}_k, \mathbf{Z}_k),k \in \mathcal{P}\}$. Since $\{\mathbf{Z}_k,\ k \in \mathcal{D}\}$ has zero mean and is independent from $\{(\mathbb{H}_k, \mathbf{Z}_k),\ k \in \mathcal{P}\}$ and $\{\mathbf{X}_k,\ k \in \mathcal{D}\}$, it follows that for every of $\ell = n_{\mathrm{t}},\ldots,L-1$ (which correspond to $k \in \mathcal{D}$),

$$\mathsf{E}\left[\mathbf{Z}_\ell^\dagger \hat{\mathbb{H}}_\ell^{(T)} \mathbf{X}_\ell\right] = 0. \tag{A11}$$

## Appendix B. Variance of the Interpolation Error for $L > \frac{1}{2\lambda_D}$

Recall that the variance of the interpolation error tends to the following value as $T$ tends to infinity:

$$\epsilon_\ell^2(t) = 1 - \int_{-1/2}^{1/2} \frac{\mathrm{SNR}\,|f_{L,\ell-t+1}(\lambda)|^2}{\mathrm{SNR}f_{L,0}(\lambda) + n_{\mathrm{t}}}\,d\lambda \tag{A12}$$

where

$$f_{L,\ell}(\lambda) = \frac{1}{L}\sum_{\nu=0}^{L-1} \bar{f}_H\left(\frac{\lambda-\nu}{L}\right) e^{\mathrm{i}2\pi\ell\frac{\lambda-\nu}{L}},\quad -\frac{1}{2} \le \lambda \le \frac{1}{2}. \tag{A13}$$

This can be lower-bounded as

$$\epsilon_\ell^2(t) = \int_{-1/2}^{1/2} \frac{n_{\mathrm{t}}f_{L,0}(\lambda)}{\mathrm{SNR}f_{L,0}(\lambda) + n_{\mathrm{t}}}\,d\lambda + \int_{-1/2}^{1/2} \frac{\mathrm{SNR}\left([f_{L,0}(\lambda)]^2 - |f_{L,\ell-t+1}(\lambda)|^2\right)}{\mathrm{SNR}f_{L,0}(\lambda) + n_{\mathrm{t}}}\,d\lambda \tag{A14}$$

$$\ge \int_{-1/2}^{1/2} \frac{\mathrm{SNR}\left([f_{L,0}(\lambda)]^2 - |f_{L,\ell-t+1}(\lambda)|^2\right)}{\mathrm{SNR}f_{L,0}(\lambda) + n_{\mathrm{t}}}\,d\lambda \tag{A15}$$

where the inequality follows because the first integral in (A14) is nonnegative. Defining $\ell' \triangleq \ell - t + 1$, we next note that

$$[f_{L,0}(\lambda)]^2 - |f_{L,\ell'}(\lambda)|^2 = \frac{1}{L^2} \sum_{\nu=0}^{L-1} \sum_{\substack{\nu'=0, \\ \nu' \neq \nu}}^{L-1} \bar{f}_H \left( \frac{\lambda - \nu}{L} \right) \bar{f}_H \left( \frac{\lambda - \nu'}{L} \right) \left[ 1 - e^{i2\pi\ell' \frac{\lambda-\nu}{L}} \cdot e^{-i2\pi\ell' \frac{\lambda-\nu'}{L}} \right] \quad \text{(A16)}$$

$$= \frac{2}{L^2} \sum_{\nu=0}^{L-1} \sum_{\nu' > \nu}^{L-1} \bar{f}_H \left( \frac{\lambda - \nu}{L} \right) \bar{f}_H \left( \frac{\lambda - \nu'}{L} \right) \left[ 1 - \cos \left( 2\pi\ell' \frac{\nu' - \nu}{L} \right) \right]. \quad \text{(A17)}$$

Since the summands are nonnegative, it follows that

$$[f_{L,0}(\lambda)]^2 - |f_{L,\ell'}(\lambda)|^2 \geq \frac{2}{L^2} \bar{f}_H \left( \frac{\lambda}{L} \right) \bar{f}_H \left( \frac{\lambda - 1}{L} \right) \left[ 1 - \cos \left( \frac{2\pi\ell'}{L} \right) \right]. \quad \text{(A18)}$$

The RHS of (A15) can thus be lower-bounded as

$$\int_{-1/2}^{1/2} \frac{\text{SNR} \left( [f_{L,0}(\lambda)]^2 - |f_{L,\ell'}(\lambda)|^2 \right)}{\text{SNR} f_{L,0}(\lambda) + n_{\text{t}}} d\lambda \geq \frac{2 \left[ 1 - \cos \left( \frac{2\pi\ell'}{L} \right) \right]}{L^2} \int_{\mathcal{L}} \frac{\text{SNR} \bar{f}_H \left( \frac{\lambda}{L} \right) \bar{f}_H \left( \frac{\lambda-1}{L} \right)}{\text{SNR} f_{L,0}(\lambda) + n_{\text{t}}} d\lambda \quad \text{(A19)}$$

where $\mathcal{L}$ denotes the interval in $[-1/2, 1/2]$ where $\bar{f}_H \left( \frac{\lambda}{L} \right)$ and $\bar{f}_H \left( \frac{\lambda-1}{L} \right)$ overlap.

We next express $L$ as

$$L = \frac{1}{2\lambda_D} + \varepsilon \quad \text{(A20)}$$

for some $\varepsilon > 0$. Then, the interval $\mathcal{L}$ is of Lebesgue measure

$$\mu(\mathcal{L}) = \min(1, 2\lambda_D \varepsilon). \quad \text{(A21)}$$

By Fatou's lemma [39], we obtain that

$$\liminf_{\text{SNR} \to \infty} \frac{2 \left[ 1 - \cos \left( \frac{2\pi\ell'}{L} \right) \right]}{L^2} \int_{\mathcal{L}} \frac{\text{SNR} \bar{f}_H \left( \frac{\lambda}{L} \right) \bar{f}_H \left( \frac{\lambda-1}{L} \right)}{\text{SNR} f_{L,0}(\lambda) + n_{\text{t}}} d\lambda$$

$$\geq \frac{2 \left[ 1 - \cos \left( \frac{2\pi\ell'}{L} \right) \right]}{L^2} \int_{\mathcal{L}} \liminf_{\text{SNR} \to \infty} \frac{\text{SNR} \bar{f}_H \left( \frac{\lambda}{L} \right) \bar{f}_H \left( \frac{\lambda-1}{L} \right)}{\text{SNR} f_{L,0}(\lambda) + n_{\text{t}}} d\lambda \quad \text{(A22)}$$

$$= \frac{2 \left[ 1 - \cos \left( \frac{2\pi\ell'}{L} \right) \right]}{L^2} \int_{\mathcal{L}} \frac{\bar{f}_H \left( \frac{\lambda}{L} \right) \bar{f}_H \left( \frac{\lambda-1}{L} \right)}{f_{L,0}(\lambda)} d\lambda. \quad \text{(A23)}$$

Since $\mathcal{L}$ is of positive Lebesgue measure, and since the integrand on the RHS of (A23) is strictly positive, it follows that [40]

$$\int_{\mathcal{L}} \frac{\bar{f}_H \left( \frac{\lambda}{L} \right) \bar{f}_H \left( \frac{\lambda-1}{L} \right)}{f_{L,0}(\lambda)} d\lambda > 0. \quad \text{(A24)}$$

Furthermore, for $\ell' = \ell - t + 1$ and $\ell = n_{\text{t}}, \ldots, L - 1$, we have

$$\cos \left( \frac{2\pi\ell'}{L} \right) < 1. \quad \text{(A25)}$$

Consequently, combining (A25) and (A24) with (A23), (A19), and (A15), we obtain the desired result

$$\liminf_{\text{SNR} \to \infty} \epsilon_\ell^2(t) > 0. \quad \text{(A26)}$$

## References

1. Foschini, G.J. Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas. *Bell Labs Tech. J.* **1996**, *1*, 41–59, doi:10.1002/bltj.2015.

2. Telatar, E. Capacity of multi-antenna Gaussian channels. *Eur. Trans. Telecommun.* **1999**, *10*, 585–595, doi:10.1002/ett.4460100604.

3. Lapidoth, A. On the asymptotic capacity of stationary Gaussian fading channels. *IEEE Trans. Inf. Theory* **2005**, *51*, 437–446, doi:10.1109/TIT.2004.840889.

4. Zheng, L.; Tse, D.N.C. Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels. *IEEE Trans. Inf. Theory* **2003**, *49*, 1073–1096, doi:10.1109/TIT.2003.810646.

5. Heath, R.W., Jr.; Lozano, A. *Foundations of MIMO Communication*; Cambridge University Press: Cambridge, UK, 2019.

6. Nikbakht, H.; Wigger, M.A.; Shamai Shitz, S. Multiplexing gains under mixed-delay constraints on Wyner's soft-handoff model. *Entropy* **2020**, *22*, 182, doi:10.3390/e22020182.

7. Lapidoth, A. Nearest neighbor decoding for additive non-Gaussian noise channels. *IEEE Trans. Inf. Theory* **1996**, *42*, 1520–1529, doi:10.1109/18.532892.

8. Marzetta, T.L. BLAST training: Estimating channel characteristics for high-capacity space-time wireless. In Proceedings of the 37th Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, USA, 22–24 September 1999.

9. Hassibi, B.; Hochwald, B.M. How much training is needed in multiple-antenna wireless links? *IEEE Trans. Inf. Theory* **2003**, *49*, 951–963, doi:10.1109/TIT.2003.809594.

10. Jindal, N.; Lozano, A. A unified treatment of optimum pilot overhead in multipath fading channels. *IEEE Trans. Commun.* **2010**, *58*, 2939–2948, doi:10.1109/TCOMM.2010.083110.090696.

11. Koch, T.; Lapidoth, A. The fading number and degrees of freedom in non-coherent MIMO fading channels: A peace pipe. In Proceedings of the IEEE International Symposium on Information Theory (ISIT 2005), Adelaide, Australia, 4–9 September 2005; doi:10.1109/ISIT.2005.1523418.

12. Etkin, R.H.; Tse, D.N.C. Degrees of freedom in some underspread MIMO fading channels. *IEEE Trans. Inf. Theory* **2006**, *52*, 1576–1608, doi:10.1109/TIT.2006.871595.

13. Nam, J.; Caire, G.; Debbah, M.; Poor, H.V. Capacity scaling of massive MIMO in strong spatial correlation regimes. *IEEE Trans. Inf. Theory* **2020**, *66*, 3040–3064, doi:10.1109/TIT.2019.2961323.

14. Gomez-Cuba, F.; Chowdhury, M.; Manolakos, A.; Erkip, E.; Goldsmith, A.J. Capacity scaling in a non-coherent wideband massive SIMO block fading channel. *IEEE Trans. Wirel. Commun.* **2019**, *18*, 5691–5704, doi:10.1109/TWC.2019.2938519.

15. Vu, M.N.; Tran, N.H.; Wijeratne, D.G.; Pham, K.; Lee, K.S.; Nguyen, D.H.N. Optimal signaling schemes and capacity of non-coherent Rician fading channels with low-resolution output quantization. *IEEE Trans. Wirel. Commun.* **2019**, *18*, 2989–3004, doi:10.1109/TWC.2019.2907952.

16. Lapidoth, A.; Shamai Shitz, S. Fading channels: How perfect need "perfect side information" be? *IEEE Trans. Inf. Theory* **2002**, *48*, 1118–1134, doi:10.1109/18.995552.

17. Weingarten, H.; Steinberg, Y.; Shamai Shitz, S. Gaussian codes and weighted nearest neighbor decoding in fading multiple-antenna channels. *IEEE Trans. Inf. Theory* **2004**, *50*, 1665–1686, doi:10.1109/TIT.2004.831744.

18. Lozano, A. Interplay of spectral efficiency, power and Doppler spectrum for reference-signal-assisted wireless communication. *IEEE Trans. Wirel. Commun.* **2008**, *7*, 5020–5029, doi:10.1109/T-WC.2008.070873.

19. Asyhari, A.T.; Koch, T.; Guillén i Fàbregas, A. Nearest neighbour decoding and pilot-aided channel estimation in stationary Gaussian flat-fading channels. In Proceedings of the IEEE International Symposium on Information Theory (ISIT 2011), St. Petersburg, Russia, 31 July–5 August 2011; doi:10.1109/ISIT.2011.6034081.

20. Asyhari, A.T.; Koch, T.; Guillén i Fàbregas, A. Nearest neighbour decoding with pilot-assisted channel estimation for fading multiple-access channels. In Proceedings of the 49th Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, USA, 28–30 September 2011; doi:10.1109/Allerton.2011.6120371.

21. Asyhari, A.T.; ten Brink, S. Orthogonal or superimposed pilots? A rate-efficient channel estimation strategy for stationary MIMO fading channels. *IEEE Trans. Wirel. Commun.* **2017**, *16*, 2776–2789, doi:10.1109/TWC.2017.2665467.

22. Verenzuela, D. Exploring alternative massive MIMO designs: Superimposed pilots and mixed-ADCs. Ph.D. Thesis, Linköping University, Linköping, Sweden, 2020.

23. Ohno, S.; Giannakis, G.B. Average-rate optimal PSAM transmissions over time-selective fading channels. *IEEE Trans. Wirel. Commun.* **2002**, *1*, 712–720, doi:10.1109/TWC.2002.804183.

24. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; Wiley: Hoboken, NJ, USA, 2006.

25. Rappaport, T.S. *Wireless Communications: Principles and Practice*, 2nd ed.; Prentice Hall PTR: Upper Saddle River, NJ, USA, 2002.

26. Saunders, S.R.; Aragón Zavala, A. *Antennas and Propagation for Wireless Communication Systems*, 2nd ed.; Wiley: Chichester, UK, 2007.

27. Merhav, N.; Kaplan, G.; Lapidoth, A.; Shamai Shitz, S. On information rates for mismatched decoders. *IEEE Trans. Inf. Theory* **1994**, *40*, 1953–1967, doi:10.1109/18.340469.

28. Durrett, R. *Probability: Theory and Examples*, 4th ed.; Cambridge University Press: Cambridge, UK, 2010.

29. Asyhari, A.T.; Guillén i Fàbregas, A. Nearest neighbor decoding in MIMO block-fading channels with imperfect CSIR. *IEEE Trans. Inf. Theory* **2012**, *58*, 1483–1517, doi:10.1109/TIT.2011.2173719.

30. Van der Vaart, A.W.; Wellner, J.A. *Weak Convergence and Empirical Processes: With Applications to Statistics*; Springer: New York, NY, USA, 1996.

31. Grant, A. Rayleigh fading multi-antenna channels. *EURASIP J. Appl. Signal Process.* **2002**, *2002*, 316–329, doi:10.1155/S1110865702000665.

32. Abramowitz, M.; Stegun, I.A. *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables*; Dover: New York, NY, USA, 1965.

33. Horn, R.A.; Johnson, C.R. *Matrix Analysis*; Cambridge University Press: Cambridge, UK, 1985.

34. Poor, H.V. *An Introduction to Signal Detection and Estimation*, 2nd ed.; Springer (A Dowden & Culver book): New York, NY, USA, 1994.

35. Sethuraman, V.; Hajek, B. Capacity per unit energy of fading channels with a peak constraint. *IEEE Trans. Inf. Theory* **2005**, *51*, 3102–3120, doi:10.1109/TIT.2005.853329.

36. Brown, J.R. *Ergodic Theory and Topological Dynamics*; Academic Press: New York, NY, USA, 1976.

37. Petersen, K. *Ergodic Theory*; Cambridge Studies in Advanced Mathematics 2; Cambridge University Press: Cambridge, UK, 1983.

38. Kim, Y.H. A coding theorem for a class of stationary channels with feedback. *IEEE Trans. Inf. Theory* **2008**, *54*, 1488–1499, doi:10.1109/TIT.2008.917685.

39. Royden, H.L. *Real Analysis*, 2nd ed.; Macmillan: New York, NY, USA, 1968.

40. Weir, A.J. *Lebesgue Integration and Measure*; Cambridge University Press: Cambridge, UK, 1973.