

# Nearest Neighbour Decoding and Pilot-Aided Channel Estimation in Stationary Gaussian Flat-Fading Channels

A. Taufiq Asyhari, Tobias Koch and Albert Guillén i Fàbregas  
University of Cambridge, Cambridge CB2 1PZ, UK  
Email: taufiq-a@ieee.org, tobi.koch@eng.cam.ac.uk, guillen@ieee.org

**Abstract**—We study the information rates of non-coherent, stationary, Gaussian, multiple-input multiple-output (MIMO) flat-fading channels that are achievable with nearest neighbour decoding and pilot-aided channel estimation. In particular, we analyse the behaviour of these achievable rates in the limit as the signal-to-noise ratio (SNR) tends to infinity. We demonstrate that nearest neighbour decoding and pilot-aided channel estimation achieves the capacity pre-log—which is defined as the limiting ratio of the capacity to the logarithm of SNR as the SNR tends to infinity—of non-coherent multiple-input single-output (MISO) flat-fading channels, and it achieves the best so far known lower bound on the capacity pre-log of non-coherent MIMO flat-fading channels.

## I. INTRODUCTION

Coherent multiple-input multiple-output (MIMO) flat-fading channels have a capacity that 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 [1], [2]. This capacity growth can be achieved using independent and identically distributed (i.i.d.) Gaussian inputs with nearest neighbour decoding. The nearest neighbour decoder is a simple decoder that selects the codeword that is closest to the channel output. In a coherent channel with additive Gaussian noise, this decoder is the maximum-likelihood decoder and is therefore optimal in the sense that it minimises the error probability (see [3] and references therein). However, the coherent channel model assumes that there is a genie that provides the fading coefficients to the decoder, which is difficult to achieve in practice. We exclude the role of the genie by studying a scheme that estimates the fading via pilot symbols. Note that with imperfect fading estimations, the nearest neighbour decoder that treats the fading estimate as if it were perfect is not necessarily optimal. Nevertheless, we show that, in some cases, nearest neighbour decoding and pilot-aided channel estimation is optimal at high SNR in the sense that it achieves the capacity pre-log. The pre-log is defined as the limiting ratio of the achievable rate to  $\log \text{SNR}$  as SNR tends to infinity. The capacity pre-log is defined in the same way but with the achievable rate replaced by the capacity.

The capacity of non-coherent fading channels, where the receiver has no knowledge of the fading coefficients, has been studied in a number of works. Building upon [4], Hassibi and Hochwald [5] studied the capacity of the block-fading channel and used pilot symbols (also known as training symbols) to obtain reasonably accurate fading estimates. Lozano and Jindal [6] provided tools for a unified treatment of pilot-based channel estimation in both block and stationary bandlimited fading channels. In these works, lower bounds on the channel capacity were obtained. Lapidoth [7] studied a single-input single-output (SISO) fading channel for more general fading processes and showed that, depending on the predictability of the fading process, the capacity growth in SNR can be, *inter alia*, logarithmically or double logarithmically. The extension of [7] to multiple-input single-output (MISO) fading channels can be found in [8]. A lower bound on the capacity of MIMO fading channels was derived by Etkin and Tse in [9].

Lapidoth and Shamai [10] and Weingarten *et al.* [11] studied non-coherent fading channels from a mismatched-decoding perspective. In particular, they studied achievable rates with Gaussian inputs and nearest neighbour decoding. In both works, it is assumed that there is a genie that provides imperfect estimates of the fading coefficients.

In our work, we add the estimation of the fading coefficients to our analysis. In particular, we study a communication system where the transmitter emits at regular intervals pilot symbols, and where the receiver performs *channel estimation* and *data detection*, separately. Based on the channel outputs corresponding to pilot transmissions, the channel estimator produces estimates for the remaining time instants using a linear minimum mean-square error (LMMSE) interpolator. Using these estimates, the data detector employs a nearest neighbour decoder to decide what the transmitted message was. We study the achievable rates of this communication scheme at high SNR. In particular, we study the pre-log for fading processes of bandlimited power spectral densities.

For SISO fading channels, using some simplifying arguments, Lozano [12] and Jindal and Lozano [6] showed that this scheme achieves the capacity pre-log. In this paper, we prove this result without any simplifying assumptions and extend it to MIMO fading channels. If the inverse of twice the bandwidth of the fading process is an integer, then for MISO channels, the above scheme is optimal in the sense that

The work of A. T. Asyhari has been partly supported by the Yousef Jameel Scholarship at University of Cambridge. The work of T. Koch has received funding from the European's Seventh Framework Programme (FP7/2007–2013) under grant agreement No. 252663.

it achieves the capacity pre-log derived by Koch and Lapidoth [8]. For MIMO channels, the above scheme achieves the best so far known lower bound on the capacity pre-log obtained in [9].

The paper is organised as follows. Section II describes the channel model and introduces the encoding and decoding scheme. Section III defines the pre-log and presents the main result. And Section IV outlines the proof of this result.

## II. SYSTEM MODEL

We consider a discrete-time  $n_r \times n_t$  MIMO flat-fading channel, whose 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

$$\mathbf{Y}_k = \sqrt{\frac{\text{SNR}}{n_t}} \mathbb{H}_k \mathbf{x}_k + \mathbf{Z}_k. \quad (1)$$

Here  $\mathbf{x}_k \in \mathbb{C}^{n_t}$  denotes the time- $k$  channel input vector (with  $\mathbb{C}$  denoting the set of complex numbers);  $\mathbb{H}_k \in \mathbb{C}^{n_r \times n_t}$  denotes the fading matrix at time  $k$ ; and  $\mathbf{Z}_k \in \mathbb{C}^{n_r}$  denotes the additive noise vector at time  $k$ .

The noise process  $\{\mathbf{Z}_k, k \in \mathbb{Z}\}$  is a sequence of independent and identically distributed (i.i.d.) complex Gaussian random vectors of zero mean and covariance matrix  $\mathbf{I}_{n_r}$ , where  $\mathbf{I}_{n_r}$  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 Gaussian. We assume that the  $n_r \cdot n_t$  processes  $\{H_k(r, t), k \in \mathbb{Z}, r = 1, \dots, n_r, t = 1, \dots, n_t\}$  are independent and have the same law, with each process having zero-mean, unit-variance and power spectral density  $f_H(\lambda)$ ,  $-\frac{1}{2} \leq \lambda \leq \frac{1}{2}$ . Thus,  $f_H(\cdot)$  is a non-negative function satisfying

$$\mathbb{E} \left[ H_{k+m}(r, t) H_k^\dagger(r, t) \right] = \int_{-1/2}^{1/2} e^{i2\pi m\lambda} f_H(\lambda) d\lambda \quad (2)$$

where  $(\cdot)^\dagger$  denotes complex conjugation. We further assume that the power spectral density  $f_H(\cdot)$  has bandwidth  $\lambda_D < 1/2$ , i.e.,  $f_H(\lambda) = 0$  for  $|\lambda| > \lambda_D$  and  $f_H(\lambda) > 0$  otherwise.

We finally assume that the fading process  $\{\mathbb{H}_k, k \in \mathbb{Z}\}$  and the noise process  $\{\mathbf{Z}_k, k \in \mathbb{Z}\}$  are independent and that their joint law does not depend on  $\{\mathbf{x}_k, k \in \mathbb{Z}\}$ .

The transmission involves both codewords and pilots. The former convey the message to be transmitted, and the latter are used to facilitate the estimation of the fading coefficients at the receiver. The codeword is selected from the codebook  $\mathcal{C}$ , which is drawn i.i.d. from a zero-mean unit-variance complex Gaussian distribution. The codeword is assumed to satisfy the average-power constraint

$$\frac{1}{N} \sum_{n=1}^N \mathbb{E} [\|\bar{\mathbf{X}}_n(m)\|^2] \leq n_t, \quad m \in \mathcal{M} \quad (3)$$

where  $\mathcal{M} = \{1, \dots, e^{NR}\}$  is the set of possible messages, and  $N$  and  $R$  denote the codeword length and the coding rate.

To estimate the fading matrix, we transmit orthogonal pilot vectors. The pilot vector  $\mathbf{p}_t$  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  $\mathbf{p}_1 = (1, 0, \dots, 0)^\top$ , where  $(\cdot)^\top$  denotes the transpose. To estimate the whole fading matrix, we thus need to send the  $n_t$  pilot vectors  $\mathbf{p}_1, \dots, \mathbf{p}_{n_t}$ .

The transmission scheme is as follows. Every  $L$  time instants (for some  $L \in \mathbb{Z}$ ), we transmit the  $n_t$  pilot vectors  $\mathbf{p}_1, \dots, \mathbf{p}_{n_t}$ . 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  $L - n_t$ .<sup>1</sup> 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  $T \in \mathbb{Z}$ ), where we transmit every  $L$  time instants the  $n_t$  pilot vectors  $\mathbf{p}_1, \dots, \mathbf{p}_{n_t}$ , 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  $T$  past and the  $T$  future pilot transmissions. This facilitates the analysis and does not incur a loss in terms of achievable rate. The above transmission scheme is illustrated in Figure 1. The channel estimator is described below.

Note that the total block-length of the above transmission scheme (comprising data vectors, pilot vectors and guard period) is given by

$$N' = N_p + N + N_{\text{un}} \quad (4)$$

where  $N_p$  denotes the number of channel uses for pilot vectors, and where  $N_{\text{un}}$  denotes the number of channel uses during the silent guard period, i.e.,

$$N_p = \left( \frac{N}{L - n_t} + 1 + 2(T - 1) \right) n_t \quad (5)$$

$$N_{\text{un}} = 2(L - n_t)(T - 1). \quad (6)$$

We now turn to the decoder. Let  $\mathcal{D}$  denote the set of time indices where data vectors of a codeword are transmitted, and let  $\mathcal{P}$  denote the set of time indices where pilots are transmitted. The decoder consists of two parts: a *channel estimator* and a *data detector*. The channel estimator considers the channel output vectors  $\mathbf{Y}_k, k \in \mathcal{P}$  corresponding to the past and future  $T$  pilot transmissions and estimates  $H_k(r, t)$  using a linear interpolator, i.e., the estimate  $\hat{H}_k^{(T)}(r, t)$  of the fading coefficient  $H_k(r, t)$  is given by

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

where the coefficients  $a_{k'}(r, t)$  are chosen in order to minimize the mean-squared error.

<sup>1</sup>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 contain therefore time instants where we do not transmit anything. The thereby incurred loss in information rate vanishes as  $N$  tends to infinity.

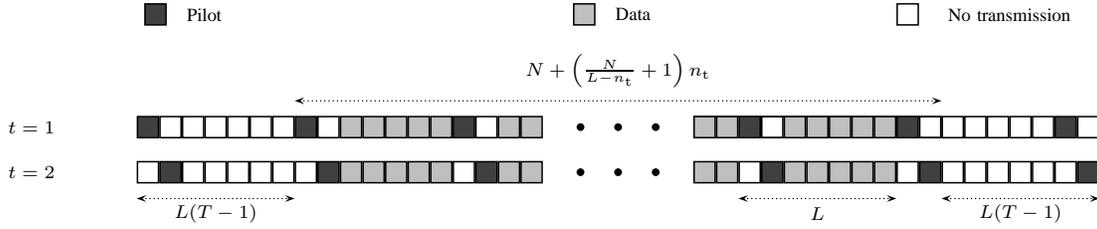


Fig. 1. Structure of pilot and data transmission for  $n_t = 2$ ,  $L = 7$  and  $T = 2$ .

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. Further note that, since the fading processes  $\{H_k(r, t), k \in \mathbb{Z}\}$ ,  $r = 1, \dots, n_r$ ,  $t = 1, \dots, n_t$  are independent, estimating  $H_k(r, t)$  only based on  $\{Y_k(r), k \in \mathbb{Z}\}$  rather than on  $\{\mathbf{Y}_k, k \in \mathbb{Z}\}$  incurs no loss in optimality.

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

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

is not stationary but cyclo-stationary with period  $L$ . Nevertheless, it can be shown that, irrespective of  $(r, t)$ , the variance of the interpolation error

$$\sigma_{e,T}^2(\ell, r, t) = \mathbb{E} \left[ \left| H_k(r, t) - \hat{H}_k^{(T)}(r, t) \right|^2 \right] \quad (9)$$

tends to the following expressions as  $T$  tends to infinity [13]

$$\sigma_e^2(\ell) \triangleq \lim_{T \rightarrow \infty} \sigma_{e,T}^2(\ell, r, t) \quad (10)$$

$$= 1 - \int_{-1/2}^{1/2} \frac{\text{SNR} |f_{H_L, \ell}(\lambda)|^2}{\text{SNR} f_{H_L, 0}(\lambda) + 1} d\lambda \quad (11)$$

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

$$f_{H_L, \ell}(\lambda) = \frac{1}{L} \sum_{j=0}^{L-1} \bar{f}_H \left( \frac{\lambda - j}{L} \right) e^{i2\pi \ell \frac{\lambda - j}{L}} \quad (12)$$

and  $\bar{f}_H(\cdot)$  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

$$L \leq \frac{1}{2\lambda_D} \quad (13)$$

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

$$|f_{H_L, \ell}(\lambda)| = f_{H_L, 0}(\lambda) = \frac{1}{L} f_H \left( \frac{\lambda}{L} \right), \quad -\frac{1}{2} \leq \lambda \leq \frac{1}{2}. \quad (14)$$

In this case the interpolation error is given by

$$\sigma_e^2(\ell) = 1 - \int_{-1/2}^{1/2} \frac{\text{SNR} (f_H(\lambda))^2}{\text{SNR} f_H(\lambda) + L} d\lambda, \quad \ell = 0, \dots, L-1 \quad (15)$$

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.

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 realisation by  $\{\hat{\mathbb{H}}_k^{(T)}, k \in \mathcal{D}\}$ . Based on the channel outputs  $\{\mathbf{y}_k, k \in \mathcal{D}\}$  and fading estimates  $\{\hat{\mathbb{H}}_k^{(T)}, k \in \mathcal{D}\}$ , the data detector uses a nearest neighbour decoder to guess which message was transmitted. Thus, the decoder decides on the message  $\hat{m}$  that satisfies

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

where

$$D(m) \triangleq \sum_{k \in \mathcal{D}} \left\| \mathbf{y}_k - \sqrt{\text{SNR}/n_t} \hat{\mathbb{H}}_k^{(T)} \mathbf{x}_k(m) \right\|^2 \quad (17)$$

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

### III. THE PRE-LOG

We say that a rate is achievable if the error probability tends to zero as the codeword length tends to infinity. In this work, we study the maximum rate  $R^*(\text{SNR})$  that is achievable with nearest neighbour 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

$$\Pi_{R^*} \triangleq \limsup_{\text{SNR} \rightarrow \infty} \frac{R^*(\text{SNR})}{\log \text{SNR}}. \quad (18)$$

The capacity pre-log—which is given by (18) but with  $R^*(\text{SNR})$  replaced by the capacity  $C(\text{SNR})$ —of SISO fading channels was computed by Lapidoth [7] as

$$\Pi_C = \mu(\{\lambda: f_H(\lambda) = 0\}) \quad (19)$$

where  $\mu(\cdot)$  denotes the Lebesgue measure on the interval  $[-1/2, 1/2]$ . Koch and Lapidoth [8] extended this result to MISO fading channels and showed that if the fading processes  $\{H_k(t), k \in \mathbb{Z}\}$ ,  $t = 1, \dots, n_t$  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 (19), the capacity pre-log of MISO fading channels with power spectral density of bandwidth  $\lambda_D$  can be evaluated as

$$\Pi_C = 1 - 2\lambda_D. \quad (20)$$

Since  $R^*(\text{SNR}) \leq C(\text{SNR})$ , 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  $\{H_k(r, t), k \in \mathbb{Z}\}$ ,  $t = 1, \dots, n_t$ ,  $r = 1, \dots, n_r$  that have the same law, the best so far known lower bound on the MIMO

pre-log is due to Etkin and Tse [9]

$$\Pi_C \geq \min(n_t, n_r) \left( 1 - \min(n_t, n_r) \mu(\{\lambda: f_H(\lambda) > 0\}) \right). \quad (21)$$

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

$$\Pi_C \geq \min(n_t, n_r) (1 - \min(n_t, n_r) 2\lambda_D). \quad (22)$$

Observe that (22) specialises to (20) 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 conclusion in [7]).

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

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

$$\Pi_{R^*} \geq \min(n_t, n_r) \left( 1 - \frac{\min(n_t, n_r)}{L^*} \right) \quad (23)$$

where  $L^*$  is the largest integer satisfying  $L^* \leq \frac{1}{2\lambda_D}$ .

*Proof:* Due to page limitations, only an outline of the proof is given in Section IV. ■

*Remark 1:* We derive Theorem 1 for i.i.d. Gaussian inputs satisfying the average-power constraint (3). Nevertheless, using truncated Gaussian inputs, it can be shown that Theorem 1 also holds when the channel inputs have to satisfy a peak-power constraint, i.e., with probability one  $|\bar{\mathbf{X}}_k| \leq 1$ .

If  $1/(2\lambda_D)$  is an integer, then (23) becomes

$$\Pi_{R^*} \geq \min(n_t, n_r) (1 - \min(n_t, n_r) 2\lambda_D). \quad (24)$$

Thus, in this case nearest neighbour decoding together with pilot-aided channel estimation achieves the capacity pre-log of MISO fading channels (20), as well as the lower bound on the capacity pre-log of MIMO fading channels (22).

Comparing (23) and (22) with the capacity pre-log  $\min(n_t, n_r)$  for coherent fading channels [1], [2], we observe that, for a fading process of bandwidth  $\lambda_D$ , the penalty for not knowing the fading coefficients is roughly  $(\min(n_t, n_r))^2 2\lambda_D$ . Consequently, the lower bound (23) does not grow linearly with  $\min(n_t, n_r)$ , but it is a quadratic function of  $\min(n_t, n_r)$  that achieves its maximum at

$$\min(n_t, n_r) = \frac{L^*}{2}. \quad (25)$$

This gives rise to the lower bound

$$\Pi_{R^*} \geq \frac{L^*}{4} \quad (26)$$

which cannot be larger than  $1/(8\lambda_D)$ . The same holds for the lower bound (21).

#### IV. PROOF OUTLINE

We first note that it suffices to consider the case where  $n_t = n_r$ . If  $n_t > n_r$ , then we employ only  $n_r$  transmit antennas, and if  $n_r > n_t$ , then we ignore  $n_r - n_t$  antennas at the receiver. This yields in both cases a lower bound on the achievable rate.

To prove Theorem 1, we analyse the generalized mutual information (GMI) for the above channel and communication scheme. The GMI, denoted by  $I^{\text{gmi}}(\text{SNR})$ , 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 [3], [10], [11] and references therein).

Let  $\mathbb{E}_k^{(T)}$  denote the estimation error in estimating  $\mathbb{H}_k$ , i.e.,  $\mathbb{E}_k^{(T)}$  is composed of the matrix entries  $E_k^{(T)}(r, t)$  (8). Then, for the above channel model, the GMI can be evaluated as

$$I^{\text{gmi}}(\text{SNR}) = \sup_{\theta \leq 0} \left( \theta B(\text{SNR}) - \kappa(\theta, \text{SNR}) \right) \quad (27)$$

where

$$B(\text{SNR}) = \frac{1}{L} \sum_{\ell=1}^{L-n_t} \mathbb{E} \left[ n_r + \sqrt{\text{SNR}/n_t} \left\| \mathbb{E}_\ell^{(T)} \right\|_F^2 \right] \quad (28)$$

(with  $\|\cdot\|_F$  denoting the Frobenius norm); and where  $\kappa(\theta, \text{SNR})$  is the conditional log moment-generating function of the metric  $D(m')$  associated with an incorrect message—conditioned on the channel outputs and on the fading estimates—which is given by

$$\begin{aligned} \kappa(\theta, \text{SNR}) &= \frac{1}{L} \sum_{\ell=1}^{L-n_t} \mathbb{E} \left[ \theta \mathbf{Y}_\ell^\dagger \left( \mathbf{I}_{n_r} - \theta \frac{\text{SNR}}{n_t} \hat{\mathbb{H}}_\ell^{(T)} \hat{\mathbb{H}}_\ell^{\dagger(T)} \right)^{-1} \mathbf{Y}_\ell \right] \\ &\quad - \frac{1}{L} \sum_{\ell=1}^{L-n_t} \mathbb{E} \left[ \log \det \left( \mathbf{I}_{n_r} - \theta \frac{\text{SNR}}{n_t} \hat{\mathbb{H}}_\ell^{(T)} \hat{\mathbb{H}}_\ell^{\dagger(T)} \right) \right]. \end{aligned} \quad (29)$$

Following [14] it can be shown that for  $\theta \leq 0$

$$\mathbb{E} \left[ \theta \mathbf{Y}_\ell^\dagger \left( \mathbf{I}_{n_r} - \theta \frac{\text{SNR}}{n_t} \hat{\mathbb{H}}_\ell^{(T)} \hat{\mathbb{H}}_\ell^{\dagger(T)} \right)^{-1} \mathbf{Y}_\ell \right] \leq 0. \quad (30)$$

As observed in [14], the choice  $\theta = -\frac{1}{n_r + (\text{SNR}/n_t)n_t n_r \sigma_{e^*, T}^2}$  yields a good lower bound at high SNR. Here

$$\sigma_{e^*, T}^2 = \max_{r, t, \ell} \mathbb{E} \left[ \left| E_\ell^{(T)}(r, t) \right|^2 \right]. \quad (31)$$

Substituting this choice to the right-hand side (RHS) of (27), and applying (30) to upper-bound  $\kappa(\theta, \text{SNR})$ , we obtain

$$\begin{aligned} I^{\text{gmi}}(\text{SNR}) &\geq \frac{1}{L} \sum_{\ell=1}^{L-n_t} \mathbb{E} \left[ \log \det \left( \mathbf{I}_{n_r} + \frac{\text{SNR} \hat{\mathbb{H}}_\ell^{(T)} \hat{\mathbb{H}}_\ell^{\dagger(T)}}{n_t n_r + n_t n_r \text{SNR} \sigma_{e^*, T}^2} \right) \right] \\ &\quad - \frac{L - n_t}{L}. \end{aligned} \quad (32)$$

We continue by analysing the RHS of (32) in the limit as the size of 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 interpolation error tends to (15), namely

$$\sigma_{e^*}^2 = \lim_{T \rightarrow \infty} \sigma_{e^*,T}^2 = 1 - \int_{-1/2}^{1/2} \frac{\text{SNR}(f_H(\lambda))^2}{\text{SNR}f_H(\lambda) + L} d\lambda. \quad (33)$$

Similarly, since by the orthogonality principle  $\hat{\mathbb{H}}_\ell^{(T)}$  and  $\mathbb{E}_\ell^{(T)}$  are independent, and since all entries in  $\mathbb{H}_\ell$  have unit variance, it follows that

$$\sigma_h^2 = \lim_{T \rightarrow \infty} (1 - \sigma_{e^*,T}^2) = \int_{-1/2}^{1/2} \frac{\text{SNR}(f_H(\lambda))^2}{\text{SNR}f_H(\lambda) + L} d\lambda. \quad (34)$$

We thus have by (34) that, irrespective of  $\ell$ , the estimate  $\hat{\mathbb{H}}_\ell^{(T)}$  tends to  $\bar{\mathbb{H}}$  in distribution

$$\frac{\hat{\mathbb{H}}_\ell^{(T)} \hat{\mathbb{H}}_\ell^{\dagger(T)}}{n_t n_r + n_t n_r \text{SNR} \sigma_{e^*,T}^2} \xrightarrow{d} \frac{\bar{\mathbb{H}} \bar{\mathbb{H}}^\dagger}{n_t n_r + n_t n_r \text{SNR} \sigma_{e^*}^2} \quad (35)$$

as  $T$  tends to infinity, where the entries of  $\bar{\mathbb{H}}$  are i.i.d., circularly-symmetric, complex Gaussian random variables with zero mean and variance  $1 - \sigma_{e^*}^2$ . Consequently, since the function  $A \mapsto \log \det(I + A)$  is continuous and bounded from below, we obtain from Portmanteau's Lemma [15] that

$$\begin{aligned} \lim_{T \rightarrow \infty} \mathbb{E} \left[ \log \det \left( I_{n_r} + \frac{\text{SNR} \hat{\mathbb{H}}_\ell^{(T)} \hat{\mathbb{H}}_\ell^{\dagger(T)}}{n_t n_r + n_t n_r \text{SNR} \sigma_{e^*,T}^2} \right) \right] \\ \geq \mathbb{E} \left[ \log \det \left( I_{n_r} + \frac{\text{SNR} \bar{\mathbb{H}} \bar{\mathbb{H}}^\dagger}{n_t n_r + n_t n_r \text{SNR} \sigma_{e^*}^2} \right) \right] \end{aligned} \quad (36)$$

which yields the following lower bound on the GMI:

$$\begin{aligned} \lim_{T \rightarrow \infty} I^{\text{gmi}}(\text{SNR}) \\ \geq \frac{L - n_t}{L} \mathbb{E} \left[ \log \det \left( I_{n_r} + \frac{\text{SNR} \bar{\mathbb{H}} \bar{\mathbb{H}}^\dagger}{n_t n_r + n_t n_r \text{SNR} \sigma_{e^*}^2} \right) \right] - \frac{L - n_t}{L} \quad (37) \\ \geq \frac{L - n_t}{L} \left( \mathbb{E} \left[ \log \det \left( \frac{\text{SNR} \bar{\mathbb{H}} \bar{\mathbb{H}}^\dagger}{n_t n_r + n_t n_r \text{SNR} \sigma_{e^*}^2} \right) \right] - 1 \right) \quad (38) \\ = \frac{L - n_t}{L} \left( n_t \log \text{SNR} - n_t \log (n_t^2 + n_t^2 \text{SNR} \sigma_{e^*}^2) \right. \\ \left. + \mathbb{E} \left[ \log \det \bar{\mathbb{H}} \bar{\mathbb{H}}^\dagger \right] - 1 \right). \quad (39) \end{aligned}$$

Here the second step follows by lower-bounding  $\log \det(I + A) \geq \log \det A$ ; and the third step follows by evaluating the determinant and by using that, by our assumption,  $n_t = n_r$ .

To compute a lower bound on the pre-log

$$\Pi_{R^*} \triangleq \lim_{\text{SNR} \rightarrow \infty} \frac{I^{\text{gmi}}(\text{SNR})}{\log \text{SNR}} \quad (40)$$

we first note that, by [16],  $\mathbb{E}[\log \det \bar{\mathbb{H}} \bar{\mathbb{H}}^\dagger]$  is finite. We further

note that

$$\text{SNR} \sigma_{e^*}^2 = \int_{-1/2}^{1/2} \frac{\text{SNR} f_H(\lambda) L}{\text{SNR} f_H(\lambda) + L} d\lambda \leq L \quad (41)$$

which implies that  $\log(n_t^2 + n_t^2 \text{SNR} \sigma_{e^*}^2)$  is finite, too. Thus, computing the ratio of the RHS of (39) to  $\log \text{SNR}$  in the limit as the SNR tends to infinity, we obtain the lower bound

$$\Pi_{R^*} \geq \left(1 - \frac{n_t}{L}\right) n_t \quad (42)$$

$$= \min(n_t, n_r) \left(1 - \frac{\min(n_t, n_r)}{L}\right), \quad L \leq \frac{1}{2\lambda_D} \quad (43)$$

where we have used that  $n_t = n_r = \min(n_t, n_r)$ . The condition  $L \leq 1/(2\lambda_D)$  is necessary since otherwise (15) would not hold. This proves Theorem 1.

## REFERENCES

- [1] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.
- [2] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecomm.*, vol. 10, no. 6, pp. 585–595, Nov./Dec. 1999.
- [3] A. Lapidoth, "Nearest neighbor decoding for additive non-gaussian noise channels," *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 1520–1529, Sep. 1996.
- [4] T. L. Marzetta, "BLAST training: Estimating channel characteristics for high-capacity space-time wireless," in *Proc. 37th Annual Allerton Conf. on Communication, Control, and Computing*, Monticello, IL, Sep. 1999, pp. 958–966.
- [5] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.
- [6] N. Jindal and A. Lozano, "A unified treatment of optimum pilot overhead in multipath fading channels," *IEEE Trans. Commun.*, vol. 58, no. 10, pp. 2939–2948, Oct. 2010.
- [7] A. Lapidoth, "On the asymptotic capacity of stationary Gaussian fading channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 437–446, Feb. 2005.
- [8] T. Koch and A. Lapidoth, "The fading number and degrees of freedom in non-coherent MIMO fading channels: A peace pipe," in *Proc. IEEE Int. Symp. Inf. Theory*, Adelaide, Australia, Sep. 2005, pp. 661–665.
- [9] R. H. Etkin and D. Tse, "Degrees of freedom in some underspread MIMO fading channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1576–1608, Apr. 2006.
- [10] A. Lapidoth and S. Shamai, "Fading channels: how perfect need "perfect side information" be?" *IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 1118–1134, May 2002.
- [11] H. Weingarten, Y. Steinberg, and S. Shamai, "Gaussian codes and weighted nearest neighbor decoding in fading multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1665–1686, Aug. 2004.
- [12] A. Lozano, "Interplay of spectral efficiency, power and Doppler spectrum for reference-signal-assisted wireless communication," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 5020–5029, Dec. 2008.
- [13] S. Ohno and G. B. Giannakis, "Average-rate optimal PSAM transmissions over time-selective fading channels," *IEEE Trans. Wireless Commun.*, vol. 1, no. 4, pp. 712–720, Oct. 2002.
- [14] A. T. Asyhari and A. Guillén i Fàbregas, "Nearest neighbour decoding in MIMO block-fading channels with imperfect CSIR," *Submitted to IEEE Trans. Inf. Theory*, Mar. 2010.
- [15] A. W. van der Vaart and J. A. Wellner, *Weak Convergence and Empirical Processes*. New York, NY: Springer-Verlag, 1996.
- [16] A. Grant, "Rayleigh fading multi-antenna channels," *EURASIP Journal on Applied Signal Processing*, vol. 3, pp. 316–329, Mar. 2002.