The Fading Number of Multiple-Input Multiple-Output Fading Channels with Memory

Stefan M. Moser
Department of Communication Engineering
National Chiao Tung University (NCTU)
Hsinchu, Taiwan
Email: stefan.moser@ieee.org

***Abstract*—The fading number of a general (not necessarily Gaussian) regular multiple-input multiple-output (MIMO) fading channel with arbitrary temporal and spatial memory is derived. The channel is assumed to be non-coherent, *i.e.*, neither receiver nor transmitter have knowledge about the channel state, but they only know the probability law of the fading process. The fading number is the second term in the asymptotic expansion of channel capacity when the signal-to-noise ratio (SNR) tends to infinity.**

It is shown that the fading number can be achieved by an input that is the product of two independent processes: a stationary and circularly symmetric direction- (or unit-) vector process whose distribution needs to be chosen such that it maximizes the fading number, and a non-negative magnitude process that is independent and identically distributed (IID) and that escapes to infinity.

Additionally, in the more general context of an arbitrary stationary channel model satisfying some weak conditions on the channel law, it is shown that the optimal input distribution is stationary apart from some edge effects.

I. Introduction

In recent years there has been an ever increasing interest in the fundamental theoretical understanding of wireless mobile communication systems, and in particular in the *channel capacity* which gives an ultimate limit on the information rate that can be transmitted reliably over these channels if we do not constrain delay and computing complexity.

Unfortunately, it turns out that the capacity, especially in the high signal-to-noise ratio (SNR) regime, is highly sensitive to some of the basic assumptions made in the modeling of the channel. For example, there is a tremendous difference in the high-SNR capacity depending on the assumptions made about the channel state information that is directly or indirectly available to the receiver. If the channel state is *perfectly* known to the receiver (*coherent detection*), the capacity grows logarithmically in the SNR similar to the situation without fading [1]. If the channel state is not available directly, but needs to be estimated by the receiver based on the received sequence of channel output symbols (*non-coherent detection*), the capacity depends highly on the assumptions made about the fading process: for *regular* fading[1] the capacity grows only double-logarithmically in the SNR [2], [3], *i.e.*, at high SNR these channels become extremely power-inefficient in the sense that for every additional bit capacity the SNR needs to be squared or, respectively, on a dB-scale the SNR needs to be doubled! For *non-regular* Gaussian fading the high-SNR behavior of capacity depends on the specific power spectral density and can be anything between the logarithmic and the double-logarithmic growth [4].

In an attempt to specify the threshold between the efficient low- to medium-SNR regime and the highly inefficient high-SNR regime of regular fading channels, [3], [5] define the *fading number* χ as the second term in the high-SNR asymptotic expansion of capacity, *i.e.*, the capacity at high SNR can be written as

$$C(\text{SNR}) = \log\log \text{SNR} + \chi + o(1) \tag{1}$$

where $o(1)$ denotes some terms that tend to zero as $\text{SNR} \to \infty$. We define *high-SNR* to be the region where the $o(1)$-terms in (1) are negligible. Note that due to the extremely slow growth of $\log\log \text{SNR}$, the fading number is usually the dominant term in the lower range of the high-SNR regime. Hence, it is of great practical interest to have a system with large fading number.

So far, the fading number has been successfully derived in some special cases only: the case of single-input multiple-output (SIMO) fading channels with memory has been solved in [6], [5], the fading number of memoryless multiple-input single-output (MISO) fading channels has been derived in [3], [5], and very recently the memoryless multiple-input multiple-output (MIMO) case was solved in [7].

In this paper we present the fading number for the remaining cases of MISO and MIMO fading channels with memory. The rest of this paper is structured as follows: after some remarks about notation we will define the channel model in the following section. In Section III we give some auxiliary results that are interesting also in a more general context. Section IV contains the main result and an outline of the proof. Before we conclude in Section VI we specialize the main result to some interesting cases in Section V.

We will often split a complex vector $\mathbf{v} \in \mathbb{C}^m$ up into its magnitude $\|\mathbf{v}\|$ and its *direction* $\hat{\mathbf{v}} \triangleq \frac{\mathbf{v}}{\|\mathbf{v}\|}$ where we reserve this notation exclusively for unit vectors, *i.e.*, throughout the paper every vector carrying a hat, $\hat{\mathbf{v}}$ or $\hat{\mathbf{V}}$, denotes a (deterministic or random, respectively) vector of unit length $\|\hat{\mathbf{v}}\| = \|\hat{\mathbf{V}}\| = 1$. To be able to work with such *direction*

[1]For a mathematical definition of regularity see Section II.

vectors we shall need a differential entropy-like quantity for random vectors that take value on the unit sphere in $\mathbb{C}^m$: let λ denote the area measure on the unit sphere in $\mathbb{C}^m$. If a random vector $\hat{\mathbf{V}}$ takes value in the unit sphere and has the density $p^{\lambda}_{\hat{\mathbf{V}}}(\hat{\mathbf{v}})$ with respect to λ, then we shall let

$$h_{\lambda}(\hat{\mathbf{V}}) \triangleq -\mathsf{E}\left[\log p^{\lambda}_{\hat{\mathbf{V}}}(\hat{\mathbf{V}})\right] \tag{2}$$

if the expectation is defined.

All rates specified in this paper are in nats per channel use, *i.e.*, $\log(\cdot)$ denotes the natural logarithmic function.

II. The Channel Model

We consider a channel with n_{T} transmit antennas and n_{R} receive antennas whose time-k output $\mathbf{Y}_k \in \mathbb{C}^{n_{\mathrm{R}}}$ is given by

$$\mathbf{Y}_k = \mathbb{H}_k \mathbf{x}_k + \mathbf{Z}_k \tag{3}$$

where $\mathbf{x}_k \in \mathbb{C}^{n_{\mathrm{T}}}$ denotes the time-k channel input vector; the random matrix $\mathbb{H}_k \in \mathbb{C}^{n_{\mathrm{R}} \times n_{\mathrm{T}}}$ denotes the time-k fading matrix; and the random vector $\mathbf{Z}_k \in \mathbb{C}^{n_{\mathrm{R}}}$ denotes additive noise. We assume that the random vectors $\{\mathbf{Z}_k\}$ are spatially and temporally white, zero-mean, circularly symmetric, complex Gaussian random variables, *i.e.*, $\mathbf{Z}_k$ are independent and identically distributed (IID) $\sim \mathcal{N}_{\mathbb{C}}\big(\mathbf{0}, \sigma^2 \mathsf{I}\big)$ for some $\sigma^2 > 0$. Here I denotes the identity matrix.

As for the multi-variate fading process $\{\mathbb{H}_k\}$, we shall only assume that it is stationary, ergodic, of finite second moment $\mathsf{E}\left[\|\mathbb{H}_k\|_{\mathrm{F}}^2\right] < \infty$ (where $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm), and of finite differential entropy rate $h(\{\mathbb{H}_k\}) > -\infty$ (the *regularity* assumption). Hence, we do not necessarily assume a fading process that is Gaussian distributed. Furthermore, we assume that the fading process $\{\mathbb{H}_k\}$ and the additive noise process $\{\mathbf{Z}_k\}$ are independent and of a joint law that does not depend on the channel input $\{\mathbf{x}_k\}$.

As for the input, we consider two different constraints: a peak-power constraint and an average-power constraint. We use $\mathcal{E}$ to denote the maximal allowed instantaneous power in the former case, and to denote the allowed average power in the latter case. For both cases we set $\textsc{snr} \triangleq \frac{\mathcal{E}}{\sigma^2}$.

The capacity $\mathsf{C}(\textsc{snr})$ of the channel (3) is given by

$$\mathsf{C}(\textsc{snr}) = \lim_{n\to\infty} \frac{1}{n} \sup I\left(\mathbf{X}_1^n; \mathbf{Y}_1^n\right) \tag{4}$$

where the supremum is over the set of all probability distributions on $\mathbf{X}_1^n$ satisfying the constraints, *i.e.*,

$$\|\mathbf{X}_k\|^2 \leq \mathcal{E}, \quad \text{almost surely}, \quad k = 1, 2, \ldots, n \tag{5}$$

for a peak-power constraint, or

$$\frac{1}{n}\sum_{k=1}^{n} \mathsf{E}\left[\|\mathbf{X}_k\|^2\right] \leq \mathcal{E} \tag{6}$$

for an average-power constraint.

From [3, Theorem 4.2], [5, Theorem 6.10] we have

$$\varlimsup_{\textsc{snr}\uparrow\infty} \Big\{\mathsf{C}(\textsc{snr}) - \log\log\textsc{snr}\Big\} < \infty. \tag{7}$$

The fading number χ is now defined as in [3, Definition 4.6], [5, Definition 6.13] by

$$\chi(\{\mathbb{H}_k\}) \triangleq \varlimsup_{\textsc{snr}\uparrow\infty} \Big\{\mathsf{C}(\textsc{snr}) - \log\log\textsc{snr}\Big\}. \tag{8}$$

Prima facie the fading number depends on whether a peak-power constraint (5) or an average-power constraint (6) is imposed on the input. However, it will turn out that the MIMO fading number with memory is identical for both cases.

III. Preliminary Results

The proof of the main result relies on some observations that hold in more general context and are therefore interesting by themselves. We state here two of these observations without proof.

A. Capacity-Achieving Input Distributions and Stationarity

One of the main assumption about our channel model is that the fading process and the additive noise are *stationary*. From an intuitive point of view it is clear that a stationary channel model should have a capacity-achieving input distribution that is also stationary. Unfortunately, we are not aware of a rigorous proof of this claim. However, we give here a slightly less strong statement that basically says that capacity can be approached up to an $\epsilon > 0$ by a distribution that looks stationary apart from edge effects:

Theorem 1: Assume some general channel model with input $\mathbf{x}_k \in \mathbb{C}^{n_{\mathrm{T}}}$ and output $\mathbf{Y}_k \in \mathbb{C}^{n_{\mathrm{R}}}$. Let the channel model be stationary, *i.e.*, for every choice of $n \in \mathbb{N}$ and distribution $Q \in \mathcal{P}(\mathbb{C}^{n_{\mathrm{T}}\times n})$ on $\mathbf{X}_1^n$ the mutual information $I\big(\mathbf{X}_1^n; \mathbf{Y}_1^n\big)$ does not change when shifting the input block over time. Assume an average-power constraint (6) and let the channel model be such that a zero input yields a zero mutual information: $I\big(\mathbf{0}; \mathbf{Y}_k\big) = 0$.

Now fix some non-negative integer κ and some power $\mathcal{E}$ with corresponding $\textsc{snr} \triangleq \mathcal{E}/\sigma^2$. Then for every fixed $\epsilon > 0$ there corresponds some positive integer $\eta = \eta(\mathcal{E}, \epsilon)$ and some joint distributions $Q^{\kappa+1}_{\mathcal{E},\epsilon} \in \mathcal{P}(\mathbb{C}^{n_{\mathrm{T}}\times(\kappa+1)})$ such that for a blocklength n sufficiently large there exists some input $\mathbf{X}_1^n$ satisfying the following:

1) The input $\mathbf{X}_1^n$ nearly achieves capacity in the sense that

$$\frac{1}{n} I\big(\mathbf{X}_1^n; \mathbf{Y}_1^n\big) \geq \mathsf{C}(\mathcal{E}) - \epsilon. \tag{9}$$

2) For every integer μ with $0 \leq \mu \leq \kappa$, every length-$(\mu+1)$ block of adjacent vectors $(\mathbf{X}_\ell, \ldots, \mathbf{X}_{\ell+\mu})$ taken from

$$\mathbf{X}_\eta, \mathbf{X}_{\eta+1}, \ldots, \mathbf{X}_{n-2\eta+2} \tag{10}$$

has the same joint distribution $Q^{\mu+1}_{\mathcal{E},\epsilon}$, where this distribution $Q^{\mu+1}_{\mathcal{E},\epsilon}$ is given as corresponding marginal distribution of $Q^{\kappa+1}_{\mathcal{E},\epsilon}$.

3) In particular, all vectors in (10) have the same marginal $Q^1_{\mathcal{E},\epsilon}$.

4) The marginal distribution $Q^1_{\mathcal{E},\epsilon}$ gives rise to a second moment $\mathcal{E}$:

$$\mathsf{E}\left[\|\mathbf{X}_\ell\|^2\right] = \mathcal{E}, \qquad \ell = \eta, \ldots, n - 2\eta + 2. \tag{11}$$

5) The first $\eta-1$ vectors and the last $2(\eta-1)$ vectors satisfy the power constraint possibly strictly:

$$\mathsf{E}\left[\|\mathbf{X}_\ell\|^2\right] \leq \mathcal{E},\ \ell \in \{1,\ldots,\eta-1\}\cup\{n-2\eta+3,\ldots,n\}. \tag{12}$$

Proof: The proof is based on a shift-and-mix argument similar to a proof given in [6] using the fact that a deterministic zero at the input yields zero information. ∎

Remark 2: Neglecting the edge-effects for the moment, Theorem 1 basically says that, for every $\mu \leq \kappa$, every block of $\mu+1$ *adjacent* vectors has the same distribution independent of the time shift. From this immediately follows that the distribution of every subset of (not necessarily adjacent) vectors of a $\mu+1$ block does not change when the vectors are shifted in time (simply marginalize those vectors out that are not member of the subset). Hence, Theorem 1 *almost* proves that the capacity-achieving input distribution is stationary: the only problems are the edge effects and the fixed (but freely selectable) value of κ.[2]

B. Capacity-Achieving Input Distributions and Circular Symmetry

The second preliminary remark concerns circular symmetry. We say that a a vector random process $\{\mathbf{W}_k\}$ is *circularly symmetric* if

$$\{\mathbf{W}_k\} \stackrel{\mathscr{L}}{=} \{\mathbf{W}_k e^{\mathrm{i}\Theta_k}\}, \tag{13}$$

where $\stackrel{\mathscr{L}}{=}$ stands for "equal in law" and where the process $\{\Theta_k\}$ is IID $\sim \mathcal{U}([0,2\pi])$ and independent of $\{\mathbf{W}_k\}$. Note that this is not to be confused with *isotropically distributed*, which means that a vector has equal probability to point in every direction.

Remark 3: Note an important subtlety of this definition: being circularly symmetric does not only imply that for every time k the corresponding random vector $\mathbf{W}_k$ is circularly symmetric, but also that from past vectors $\mathbf{W}_{-\infty}^{k-1}$ one cannot gain any knowledge about the present phase, *i.e.*, the phase is IID.

Lemma 4: Assume a channel as given in (3). Then the capacity-achieving input process can be assumed to be circularly symmetric, *i.e.*, the input $\{\mathbf{X}_k\}$ can be replaced by $\{\mathbf{X}_k e^{\mathrm{i}\Theta_k}\}$, where $\{\Theta_k\}$ is IID $\sim \mathcal{U}([0,2\pi])$ and independent of every other random quantity.

Remark 5: The proof of Lemma 4 relies only on the fact that the additive noise is assumed to be circularly symmetric. Hence, for the lemma to hold the noise need not be Gaussian distributed and may even have memory as long as it is circularly symmetric.

IV. The Fading Number of MIMO Fading Channels with Memory

Theorem 6: Consider a MIMO fading channel with memory (3) where the stationary and ergodic fading process $\{\mathbb{H}_k\}$ takes value in $\mathbb{C}^{n_\mathrm{R}\times n_\mathrm{T}}$ and satisfies $h(\{\mathbb{H}_k\}) > -\infty$ and $\mathsf{E}\left[\|\mathbb{H}_k\|_\mathrm{F}^2\right] < \infty$. Then, irrespective of whether a peak-power constraint (5) or an average-power constraint (6) is imposed on the input, the fading number $\chi(\{\mathbb{H}_k\})$ is given by

$$\begin{aligned}\chi(\{\mathbb{H}_k\}) &= \sup_{\substack{Q_{\{\hat{\mathbf{X}}_k\}}\\ \text{stationary}\\ \text{circ. sym.}}} \Bigg\{ h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|}\,\middle|\,\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=-\infty}^{-1}\right) \\ &\qquad + n_\mathrm{R}\mathsf{E}\Big[\log\|\mathbb{H}_0\hat{\mathbf{X}}_0\|^2\Big] - \log 2 \\ &\qquad - h\big(\mathbb{H}_0\hat{\mathbf{X}}_0 \,\big|\, \{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=-\infty}^{-1}, \hat{\mathbf{X}}_{-\infty}^0\big)\Bigg\}\end{aligned} \tag{14}$$

where the maximization is over all stochastic unit-vector processes $\{\hat{\mathbf{X}}_k\}$ that are stationary and circularly symmetric.

Moreover, the fading number is achievable by a stationary input that can be expressed as a product of two *independent* processes:

$$\mathbf{X}_k = R_k \cdot \hat{\mathbf{X}}_k, \tag{15}$$

where $\{\hat{\mathbf{X}}_k\} \in \mathbb{C}^{n_\mathrm{T}}$ is a stationary and circularly symmetric unit-vector process with the probability distribution that maximizes (14), and where $\{R_k\} \in \mathbb{R}_0^+$ is a scalar non-negative IID random process such that

$$\log R_k^2 \sim \mathcal{U}\left([\log\log\mathcal{E}, \log\mathcal{E}]\right). \tag{16}$$

Note that this input satisfies the peak-power constraint (5) (and therefore also the average-power constraint (6)).

Proof: The proof is rather long and technical. We will give here only an outline. The proof consists of two parts: in a first part we derive an upper bound on the fading number assuming an average-power constraint (6) on the channel input. In a second part we then derive a lower bound on the fading number by assuming one particular input distribution that satisfies the peak-power constraint (5). We then show that both bounds coincide. Since a peak-power constraint is more restrictive than the corresponding average-power constraint the theorem follows.

a) Outline of Upper Bound: Similarly to the proof of the SIMO fading number [6], [5] we use the chain rule to write

$$\frac{1}{n}I\big(\mathbf{X}_1^n;\mathbf{Y}_1^n\big) = \frac{1}{n}\sum_{k=1}^n I\big(\mathbf{X}_1^n;\mathbf{Y}_k\,\big|\,\mathbf{Y}_1^{k-1}\big) \tag{17}$$

and then split each term on the RHS of the above into terms that are memoryless and terms that take care of the memory:

$$\begin{aligned}&I\big(\mathbf{X}_1^n;\mathbf{Y}_k\,\big|\,\mathbf{Y}_1^{k-1}\big) \\ &\quad\leq I\big(\mathbf{X}_k;\mathbf{Y}_k\big) - I\left(\frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|};\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{k-1}\right) \\ &\qquad + I\big(\mathbb{H}_k\hat{\mathbf{X}}_k;\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{k-1}\,\big|\,\hat{\mathbf{X}}_1^n\big).\end{aligned} \tag{18}$$

Note that in the situation of multiple-antennas at both transmitter and receiver it is not possible to gain full knowledge about all fading coefficients even if both $\mathbf{X}$ and $\mathbf{Y}$ are known!

[2] As a matter of fact one can choose κ arbitrarily large, however, note that the size of the edges where the lemma does not hold depends on κ!

Note further that, in order to be able to discard the noise, we rely on the observation that the capacity-achieving input distribution escapes to infinity [6], [5].

In a next step we now could use the bounding techniques known from [3], [5] to get a bound on the memoryless MIMO fading channel. Unfortunately, it turns out that this will lead to a non-tight bound. Instead we split the term $I(\mathbf{X}_k;\mathbf{Y}_k)$ up into magnitude term and a term that takes care of the direction:

$$I(\mathbf{X}_k;\mathbf{Y}_k) \le I(\mathbf{X}_k;\|\mathbf{Y}_k\|) + I\left(\mathbf{X}_k;\frac{\mathbf{Y}_k}{\|\mathbf{Y}_k\|}\,\middle|\,\|\mathbf{Y}_k\|\right) \tag{19}$$

and show that

$$I\left(\mathbf{X}_k;\frac{\mathbf{Y}_k}{\|\mathbf{Y}_k\|}\,\middle|\,\|\mathbf{Y}_k\|\right) \le I\left(\|\mathbb{H}_k\hat{\mathbf{X}}_k\|,\hat{\mathbf{X}}_k;\frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}\right). \tag{20}$$

The first term in (19) now (almost) looks like a MISO fading channel where we can fix the fact that the output is non-negative by multiplying $\|\mathbf{Y}_k\|$ by a independent circularly symmetric phase (note that this does not change the mutual information).

In order to have a bound that is independent of the unknown capacity-achieving input distribution, in a final step we maximize the bound over all input distributions. Here we can rely on Theorem 1 which says that we can restrict ourselves to stationary input distributions.

Hence, the upper bound basically looks like

$$\begin{aligned}\chi(\{\mathbb{H}_k\}) \le{} & \text{“}\chi_{\text{MISO,IID}}\text{”}\big(\|\mathbb{H}\mathbf{X}+\mathbf{Z}\|e^{\mathrm{i}\Theta}\big)\\ &+ I\left(\|\mathbb{H}_k\hat{\mathbf{X}}_k\|,\hat{\mathbf{X}}_k;\frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|}\right)\\ &- I\left(\frac{\mathbb{H}_k\hat{\mathbf{X}}_k}{\|\mathbb{H}_k\hat{\mathbf{X}}_k\|};\left\{\frac{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell}{\|\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\|}\right\}_{\ell=1}^{k-1}\right)\\ &+ I\big(\mathbb{H}_k\hat{\mathbf{X}}_k;\{\mathbb{H}_\ell\hat{\mathbf{X}}_\ell\}_{\ell=1}^{k-1}\,\big|\,\hat{\mathbf{X}}_1^n\big).\end{aligned} \tag{21}$$

Here the first term corresponds to an expression that is similar to the memoryless fading number, the second term takes care of the contribution of the direction of the input, and the last two terms take care of the contribution of the memory.

Note that we have skipped over a lot of problems like, *e.g.*, the edge effects of Theorem 1, the order of the limits of $n\to\infty$ and $\mathcal{E}\to\infty$, the fact that the escaping-to-infinity property only comes into play in the limit when $\mathcal{E}$ tends to infinity, or the care that is needed when dealing with the “almost MISO fading number.”

b) Outline of Lower Bound: To derive a lower bound we choose a specific input distribution which naturally yields a lower bound to channel capacity. Let $\{\mathbf{X}_k\}$ be of the form

$$\mathbf{X}_k = R_k\cdot\hat{\mathbf{X}}_k. \tag{22}$$

Here $\{\hat{\mathbf{X}}_k\}$ is a sequence of random unit-vectors forming a stochastic process that is stationary, circularly symmetric, and of a distribution that achieves the maximum in (14). The random variables $R_k\in\mathbb{R}_0^+$ are IID and satisfy (16). Finally we assume that $\{R_k\}\perp\!\!\!\perp\{\hat{\mathbf{X}}_k\}$.

We then again start using the chain rule to write

$$\frac{1}{n}I(\mathbf{X}_1^n;\mathbf{Y}_1^n) = \frac{1}{n}\sum_{k=1}^n I\big(\mathbf{X}_k;\mathbf{Y}_1^n\,\big|\,\mathbf{X}_1^{k-1}\big), \tag{23}$$

and then treat each term separately:

$$\begin{aligned}I\big(\mathbf{X}_k;\mathbf{Y}_1^n\,\big|\,\mathbf{X}_1^{k-1}\big) \ge{} & I\big(\mathbf{X}_k;\mathbf{Y}_k\,\big|\,\hat{\mathbf{Y}}_{k+1}^n,\mathbf{Y}_1^{k-1},\hat{\mathbf{X}}_1^{k-1}\big)\\ &+ I\big(\hat{\mathbf{X}}_k;\hat{\mathbf{Y}}_{k+1}^n\,\big|\,\mathbf{Y}_1^{k-1},\hat{\mathbf{X}}_1^{k-1}\big).\end{aligned} \tag{24}$$

Note that the first term basically is the memoryless situation based on the side-information of past and future terms. To simplify the notation let's call this side-information $\mathbf{S}_k$:

$$\mathbf{S}_k \triangleq \big(\hat{\mathbf{Y}}_{k+1}^n,\mathbf{Y}_1^{k-1},\hat{\mathbf{X}}_1^{k-1}\big). \tag{25}$$

Contrary to the derivation of the upper bound that has been based on the memoryless MISO case, we will base the derivation of the lower bound on memoryless SIMO, *i.e.*, we split the first term in (24) into two parts:

$$I\big(\mathbf{X}_k;\mathbf{Y}_k\,\big|\,\mathbf{S}_k\big) = I\big(\hat{\mathbf{X}}_k;\mathbf{Y}_k\,\big|\,\mathbf{S}_k\big) + I\big(R_k;\mathbf{Y}_k\,\big|\,\hat{\mathbf{X}}_k,\mathbf{S}_k\big). \tag{26}$$

Now we have the problem that the second term does not correspond exactly to the SIMO situation since the input of the channel is real instead of complex. This is fixed by various arithmetic changes which at the end yield the following bound:

$$\begin{aligned}&I\big(\mathbf{X}_k;\mathbf{Y}_k\,\big|\,\mathbf{S}_k\big)\\ &\quad\approx I\big(R_k e^{\mathrm{i}\Theta_k};\mathbf{Y}_k e^{\mathrm{i}\Theta_k}\,\big|\,\hat{\mathbf{X}}_k,\mathbf{S}_k\big)\\ &\qquad + h_\lambda\big(\hat{\mathbf{Y}}_k\,\big|\,\mathbf{S}_k\big) - h_\lambda\big(\hat{\mathbf{Y}}_k e^{\mathrm{i}\Theta_k}\,\big|\,\hat{\mathbf{X}}_k,\mathbf{S}_k\big).\end{aligned} \tag{27}$$

Note that our choice of R_k guarantees that $R_k e^{\mathrm{i}\Theta_k}$ achieves the fading number of memoryless SIMO fading with side-information. Hence, we get

$$\begin{aligned}&\chi(\{\mathbb{H}_k\})\\ &\ge \chi_{\text{IID}}\big(\mathbb{H}_0\hat{\mathbf{X}}_0\,\big|\,\hat{\mathbf{X}}_0,\mathbf{S}_0\big) + h_\lambda\big(\hat{\mathbf{Y}}_0\,\big|\,\mathbf{S}_0\big)\\ &\quad - h_\lambda\big(\hat{\mathbf{Y}}_0 e^{\mathrm{i}\Theta_0}\,\big|\,\hat{\mathbf{X}}_0,\mathbf{S}_0\big) + I\big(\hat{\mathbf{X}}_0;\hat{\mathbf{Y}}_1^\infty\,\big|\,\mathbf{Y}_{-\infty}^{-1},\hat{\mathbf{X}}_{-\infty}^{-1}\big)\\ &= h_\lambda\left(\frac{\mathbb{H}_0\hat{\mathbf{X}}_0}{\|\mathbb{H}_0\hat{\mathbf{X}}_0\|}e^{\mathrm{i}\Theta_0}\,\middle|\,\hat{\mathbf{X}}_0,\mathbf{S}_0\right) + n_{\mathrm{R}}\mathsf{E}\Big[\log\|\mathbb{H}_0\hat{\mathbf{X}}_0\|^2\Big]\\ &\quad - \log 2 - h\big(\mathbb{H}_0\hat{\mathbf{X}}_0\,\big|\,\hat{\mathbf{X}}_0,\mathbf{S}_0\big) + h_\lambda\big(\hat{\mathbf{Y}}_0\,\big|\,\mathbf{S}_0\big)\\ &\quad - h_\lambda\big(\hat{\mathbf{Y}}_0 e^{\mathrm{i}\Theta_0}\,\big|\,\hat{\mathbf{X}}_0,\mathbf{S}_0\big) + I\big(\hat{\mathbf{X}}_0;\hat{\mathbf{Y}}_1^\infty\,\big|\,\mathbf{Y}_{-\infty}^{-1},\hat{\mathbf{X}}_{-\infty}^{-1}\big).\end{aligned} \tag{28}$$

In this outline we have again simplified things considerably, *e.g.*, we have interchanged the order of the limits of $n\to\infty$ and $\mathcal{E}\to\infty$, and we have neglected the influence of the noise in various places.

The result now follows by showing that the lower bound is equivalent to the upper bound. This follows from some arithmetic changes and from stationarity. ■

V. Some Special Cases

A. MISO Fading With Memory

Next we are going to study the fading number of MISO fading with memory which has been unknown so far. If we specialize Theorem 6 to the situation of only one antenna at the receiver we get the following:

Corollary 7: Consider a MISO fading channel with memory where the stationary and ergodic fading process $\{\mathbf{H}_k\}$ takes value in $\mathbb{C}^{n_\text{T}}$ and satisfies $h(\{\mathbf{H}_k\}) > -\infty$ and $\mathsf{E}\big[\|\mathbf{H}_k\|^2\big] < \infty$. Then, irrespective of whether a peak-power constraint (5) or an average-power constraint (6) is imposed on the input, the fading number $\chi\big(\{\mathbf{H}_k^\mathsf{T}\}\big)$ is given by

$$\chi\big(\{\mathbf{H}_k^\mathsf{T}\}\big) = \sup_{\substack{Q_{\{\hat{\mathbf{X}}_k\}} \\ \text{stationary}}} \left\{ \log\pi + \mathsf{E}\Big[\log|\mathbf{H}_0^\mathsf{T}\hat{\mathbf{X}}_0|^2\Big] - h\big(\mathbf{H}_0^\mathsf{T}\hat{\mathbf{X}}_0 \,\big|\, \{\mathbf{H}_\ell^\mathsf{T}\hat{\mathbf{X}}_\ell\}_{\ell=-\infty}^{-1}, \hat{\mathbf{X}}_{-\infty}^{0}\big) \right\} \tag{29}$$

where the maximization is over all stochastic unit-vector processes $\{\hat{\mathbf{X}}_k\}$ that are stationary.

Remark 8: Note that in contrast to the situation without memory where the optimal input is a beam-forming input using a deterministic direction that maximizes the fading number, here beam-forming is in general not optimal anymore.

B. Spatially IID Gaussian MIMO Fading Channels with Temporal Memory

Assume a fading process $\mathbb{H}_k = \mathsf{D} + \tilde{\mathbb{H}}_k$ where all components of the matrix process $\{\tilde{\mathbb{H}}_k\}$ are independent and identically distributed zero-mean, unit-variance Gaussian random processes with spectral distribution function $\mathsf{F}(\cdot)$, *i.e.*, the fading components are spatially IID, but have temporal memory. Note that for some constant unitary $n_\text{R}\times n_\text{R}$ matrix U and some constant unitary $n_\text{T}\times n_\text{T}$ matrix V the law of $\mathsf{U}\tilde{\mathbb{H}}_k\mathsf{V}$ is identical to the law of $\tilde{\mathbb{H}}_k$. Therefore, without loss of generality, we may restrict ourselves to matrices D that are "diagonal" with the singular values of D, $|d_1| \geq |d_2| \geq \ldots \geq |d_{\min\{n_\text{R},n_\text{T}\}}| > 0$, on its diagonal.

Proposition 9: The fading number of a spatially IID Gaussian MIMO fading channel with temporal memory is upper-bounded as follows:

$$\chi\big(\{\mathbb{H}_k\}\big) \leq \min\{n_\text{R},n_\text{T}\}\log\frac{\delta^2}{\min\{n_\text{R},n_\text{T}\}} - n_\text{R}\log\epsilon_{\text{pred}}^2 + n_\text{R}\log n_\text{R} - n_\text{R} - \log\Gamma(n_\text{R}) \tag{30}$$

where

$$\delta^2 \triangleq \big(|d_1|^2\cdot\dots\cdot|d_{\min\{n_\text{R},n_\text{T}\}}|^2\big)^{1/\min\{n_\text{R},n_\text{T}\}} \cdot\left(1+\frac{1}{|d_1|^2}+\cdots+\frac{1}{|d_{\min\{n_\text{R},n_\text{T}\}}|^2}\right) \tag{31}$$

and where ϵ_{pred}^2 denotes the prediction error when predicting the value of one component of $\mathbb{H}_0$ after having observed the infinite past.

VI. Discussion & Conclusion

We have derived the fading number in the most general situation of MIMO fading with memory where the fading process is not limited to a Gaussian distribution, but may be any stationary, ergodic, and regular distribution of finite energy. In particular we allow both temporal and spatial memory. The MIMO fading number is achievable by an input process that can be written as product of two independent processes: an IID non-negative "magnitude" process and a stationary and circularly symmetric "direction" process. The former has the same logarithmically uniform distribution (16) that has been used in previous publications about the fading number [3], [5], [6]. The "direction" process depends on the particular law of the fading process, *i.e.*, it needs to be chosen such as to maximize the fading number. The expression of the fading number is therefore not given in a completely closed form but still contains a maximization. However, one has to be aware that we have not specified the fading process in closed form either, *i.e.*, we do not believe it possible to further simply (14) without making more detailed assumptions about $\{\mathbb{H}_k\}$. We are still working on a fully closed-form expression for the important special situation of Gaussian fading processes.

The proof of the main result is strongly based on a new theorem showing that the capacity-achieving input distribution of a stationary channel model can (almost) be assumed to be stationary. Even though this result is very intuitive, we are not aware of any proof in the literature. We believe this result to be of importance also in many other situations.

We also have specialized the result to MISO fading with memory and shown that in contrast to the memoryless situation this fading number is in general *not* achieved by beam-forming.

Acknowledgments

This work was supported by the Industrial Technology Research Institute (ITRI), Zhudong, Taiwan, under JRC NCTU-ITRI and by the National Science Council under NSC 95-2221-E-009-046.

References

[1] İ. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov.–Dec. 1999.

[2] G. Taricco and M. Elia, "Capacity of fading channel with no side information," *Electron. Lett.*, vol. 33, no. 16, pp. 1368–1370, Jul. 31, 1997.

[3] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat fading channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2426–2467, Oct. 2003.

[4] A. Lapidoth, "On the asymptotic capacity of stationary Gaussian fading channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 2, pp. 437–446, Feb. 2005.

[5] S. M. Moser, "Duality-based bounds on channel capacity," Ph.D. dissertation, Swiss Fed. Inst. of Techn., Zurich, Oct. 2004, Diss. ETH No. 15769. [Online]. Available: http://moser.cm.nctu.edu.tw/

[6] A. Lapidoth and S. M. Moser, "The fading number of single-input multiple-output fading channels with memory," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 437–453, Feb. 2006.

[7] S. M. Moser, "The fading number of memoryless multiple-input multiple-output fading channels," Apr. 2007, to app. in *IEEE Trans. Inform. Theory*. [Online]. Available: http://moser.cm.nctu.edu.tw/publications.html