On Non-Coherent Fading Channels with Feedback

Amos Lapidoth Stefan M. Moser
Signal and Information Processing Laboratory
Swiss Federal Institute of Technology (ETH) Zurich
CH-8092 Zurich, Switzerland
e-mail: {lapidoth, moser}@isi.ee.ethz.ch

Abstract — **We investigate the influence of feedback on the capacity of multiple-input multiple-output (MIMO) non-coherent fading channels with memory. We derive an upper bound on the feedback capacity of regular fading channels, which is shown to coincide with a previously known upper bound on the capacity without feedback. Hence, it is concluded that whereas feedback does in general increase capacity, this increase is relatively small in the sense that the same upper bound holds for both scenarios. From this bound we derive an upper bound on the MIMO fading number in the presence of feedback and show that in the single-input single-output (SISO) case this bound is tight, *i.e.*, the SISO fading number is not changed by feedback.**

Next, we derive an upper bound on the feedback capacity of spatially independent MIMO Gaussian fading channels. From this bound we derive a new upper bound on the corresponding fading number and show that in the multiple-input single-output (MISO) case this bound is tight.

Finally, in the case of a non-regular SISO Gaussian fading process we show that feedback does not increase the pre-log, *i.e.*, the ratio of the capacity to the logarithm of the signal-to-noise ratio $\log \mathrm{SNR}$ is not changed in the limit when the SNR tends to infinity.

I. Introduction

In [1], [2] the capacity of fading channels with memory was investigated. It was shown there that for any finite-energy *regular* fading process $\{\mathbb{H}_k\}$ (*i.e.*, one having a finite differential entropy rate) the channel capacity C grows asymptotically only double-logarithmically in the signal-to-noise ratio (SNR). The *fading number* χ was introduced as the second term in the high-SNR expansion of the capacity

$$\chi(\{\mathbb{H}_k\}) \triangleq \varlimsup_{\mathrm{SNR}\uparrow\infty} \{\mathsf{C}(\mathrm{SNR}) - \log\log \mathrm{SNR}\}. \tag{1}$$

The fading number was computed for the single-input single-output (SISO) case in [1] and then extended to the single-input multiple-output (SIMO) case in [2]. For a memoryless channel, the multiple-input single-output (MISO) fading number was derived, too [1].

In [1] the following upper bound on the channel capacity of a regular fading channel with memory was derived:

$$\mathsf{C}(\mathrm{SNR}) \leq \mathsf{C}_{\mathrm{IID}}(\mathrm{SNR}) + I\big(\mathbb{H}_0; \mathbb{H}_{-\infty}^{-1}\big), \quad \mathrm{SNR} \geq 0, \tag{2}$$

where $\mathsf{C}_{\mathrm{IID}}(\mathrm{SNR})$ denotes the channel capacity without memory.

For the more restricted case of Gaussian fading with peak-power constraints, the analysis was extended in [3], [4], and [5] to *non-regular* fading processes (*i.e.*, processes of differential entropy rate negative infinity). It was shown that while for regular fading processes the channel capacity grows like $\log\log \mathrm{SNR}$, this is not necessarily the case for non-regular fading. Depending on the spectrum $\mathsf{F}(\cdot)$ of the non-regular Gaussian fading process, the asymptotic behavior of channel capacity can be varied, *e.g.*, double-logarithmic, logarithmic, or a fractional power thereof. The *pre-log* Π^{PP} was defined as

$$\Pi^{\mathrm{PP}} \triangleq \lim_{\mathrm{SNR}\uparrow\infty} \frac{\mathsf{C}^{\mathrm{PP}}(\mathrm{SNR})}{\log \mathrm{SNR}}, \tag{3}$$

where C^{PP} denotes capacity with a peak-power constraint. In the SISO case its value was computed, see (33).

Furthermore, in [6], [7] the following upper bound on the capacity of a multiple-input multiple-output (MIMO) Gaussian fading[1] channel with memory was derived:

$$\mathsf{C}^{\mathrm{PP}}(\mathrm{SNR}) \leq \mathsf{C}_{\mathrm{IID}}^{\mathrm{PP}}(\mathrm{SNR}) + n_{\mathrm{R}} \log \frac{1 + \frac{1}{\mathrm{SNR}}}{\epsilon^2\left(\frac{1}{\mathrm{SNR}}\right) + \frac{1}{\mathrm{SNR}}}, \tag{4}$$

where $\mathsf{C}_{\mathrm{IID}}^{\mathrm{PP}}(\mathrm{SNR})$ denotes the channel capacity without memory and where $\epsilon^2(\delta^2)$ denotes the error in predicting one component of the fading from a variance-δ^2 noisy observation of its infinite past.

In this paper we extend these results to the situation where there is a *noiseless feedback link* from the receiver to the transmitter. We show that, in spite of memory in the fading process, feedback does not significantly increase channel capacity. In fact, we will show that the upper bounds (2) and (4) also hold in the presence of feedback. Consequently, all asymptotic and non-asymptotic[2] upper bounds on channel capacity of regular fading channels that were derived in [1], [2] are valid in the presence of feedback, too.

We will also derive an upper bound on the fading number of general regular MIMO fading processes in presence of feedback and we will show that in the SISO case this upper bound is tight.

In the case of Gaussian fading processes we will improve this upper bound on the fading number and will show that this improved bound is in the MISO case tight. Furthermore, we will show that for non-regular SISO Gaussian fading feedback does not increase the pre-log.

Some of the results presented here can also be found in [8], but not all.

II. Channel Model

We consider a communication system as depicted in Figure 1. A message M is transmitted over a MIMO fading chan-

[1] The Gaussian fading may be regular or non-regular, however, it is assumed to be spatially independent with identical spectra.

[2] *I.e.*, bounds that are valid for all SNR.

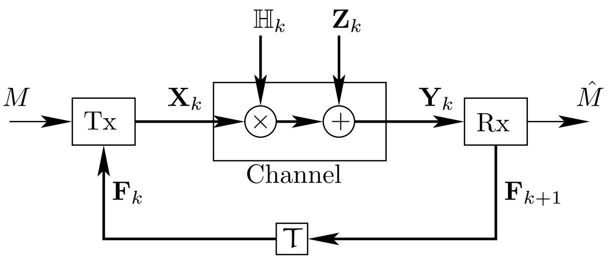

Fig. 1: A regular fading channel with noiseless feedback from the receiver to the transmitter.

nel with memory. The channel output $\mathbf{Y}_k \in \mathbb{C}^{n_\mathrm{R}}$ at time k is given by

$$\mathbf{Y}_k = \mathbb{H}_k \mathbf{x}_k + \mathbf{Z}_k, \tag{5}$$

where $\mathbf{x}_k \in \mathbb{C}^{n_\mathrm{T}}$ denotes the time-k channel input vector; the complex $n_\mathrm{R} \times n_\mathrm{T}$ random matrix $\mathbb{H}_k$ denotes the time-k fading matrix; and the random vector $\mathbf{Z}_k \in \mathbb{C}^{n_\mathrm{R}}$ denotes the time-k additive noise vector.

The additive noise process $\{\mathbf{Z}_k\}$ is assumed to be a spatially and temporally independent and identically distributed (IID), zero-mean, circularly-symmetric, complex Gaussian process, *i.e.*,

$$\{\mathbf{Z}_k\} \sim \text{IID } \mathcal{N}_\mathbb{C}\big(\mathbf{0}, \sigma^2 \mathsf{I}\big) \tag{6}$$

for some $\sigma^2 > 0$.

Neither transmitter nor receiver knows the realization of the fading process $\{\mathbb{H}_k\}$; they only know its law. We assume that $\{\mathbb{H}_k\}$ is stationary, ergodic, and of finite energy

$$\mathsf{E}\left[\|\mathbb{H}_k\|_\mathrm{F}^2\right] < \infty. \tag{7}$$

Sometimes we will further require that $\{\mathbb{H}_k\}$ be *regular*, *i.e.*, that it be of finite differential entropy rate

$$h(\{\mathbb{H}_k\}) > -\infty. \tag{8}$$

We will clearly specify when this additional assumption is made.

Furthermore, we assume a feedback link from the receiver to the transmitter. We allow the feedback to be noiseless, however, to preserve the causality of the system, we require it to be delayed by one time-step, so that the feedback random vector $\mathbf{F}_k$ that is available to the transmitter at time k consists of all past channel output vectors, *i.e.*,

$$\mathbf{F}_k = \mathbf{Y}_1^{k-1}. \tag{9}$$

Hence, the channel input $\mathbf{x}_k$ is a deterministic function of the message M and the feedback $\mathbf{Y}_1^{k-1}$.

We consider two types of power constraints: an average-power constraint and a peak-power constraint. Under the former we require that

$$\frac{1}{n}\sum_{k=1}^{n} \mathsf{E}\Big[\big\|\mathbf{X}_k\big(M, \mathbf{Y}_1^{k-1}\big)\big\|^2\Big] \le \mathcal{E}, \tag{10}$$

where n denotes the blocklength and where all messages are assumed equiprobable.

Under the peak-power constraint we replace (10) with the almost sure constraint that for every message m

$$\big\|\mathbf{X}_k\big(m, \mathbf{Y}_1^{k-1}\big)\big\|^2 \le \mathcal{E}, \quad \text{a.s., } 1 \le k \le n. \tag{11}$$

The signal-to-noise ratio SNR is defined for both constraints by

$$\mathrm{SNR} \triangleq \frac{\mathcal{E}}{\sigma^2}. \tag{12}$$

The subject of our investigation is the channel capacity (and in the asymptotic case the fading number or the pre-log, respectively) in presence of feedback.

To clarify notation we will use a subscript "FB" whenever feedback is available. Furthermore, we will use the superscript "Avg" whenever a result holds under an average-power constraint and the superscript "PP" for a result holding under a peak-power constraint. Quantities without these superscripts are valid in both situations.

III. Regular Fading and Feedback

The following main result is valid for all values of the SNR:

Theorem 1. *Let a* regular *(not necessarily Gaussian) fading channel be defined as described in Section II above with a fading process that is stationary, ergodic, of finite energy, and of finite differential entropy rate.*

Then the channel capacity in the presence of noiseless feedback from the receiver to the transmitter C_FB is upper bounded by

$$\mathsf{C}_\mathrm{FB}(\mathrm{SNR}) \le \mathsf{C}_\mathrm{IID}(\mathrm{SNR}) + I\big(\mathbb{H}_0; \mathbb{H}_{-\infty}^{-1}\big), \tag{13}$$

where $\mathsf{C}_\mathrm{IID}(\mathrm{SNR})$ denotes the channel capacity without feedback or memory.

Proof. For a proof see [8]. ☐

Remark 2. *This result even holds in the (hypothetical) case when one allows the input vector $\mathbf{x}_k$ to also depend on the past fading realizations $\mathbb{H}_1^{k-1}$. Furthermore, it holds irrespective whether an average-power constraint* (10) *or a peak-power constraint* (11) *is imposed.*

Remark 3. *From Theorem 1 it immediately follows that any upper bound to $\mathsf{C}_\mathrm{IID}(\mathrm{SNR})$ automatically is also an upper bound to*

$$\mathsf{C}_\mathrm{FB}(\mathrm{SNR}) - I\big(\mathbb{H}_0; \mathbb{H}_{-\infty}^{-1}\big). \tag{14}$$

Consequently, all upper bounds on $\mathsf{C}_\mathrm{IID}(\mathrm{SNR})$ derived in [1], [2] can be used to upper bound $\mathsf{C}_\mathrm{FB}(\mathrm{SNR})$. In particular, even in presence of noiseless feedback the capacity of regular fading channels grows asymptotically only double-logarithmically in the SNR.

Also, since feedback does not hurt (the transmitter may simply ignore it and achieve the same results as without feedback), any lower bound on $\mathsf{C}(\mathrm{SNR})$ is also a lower bound on $\mathsf{C}_\mathrm{FB}(\mathrm{SNR})$.

Theorem 1 immediately specializes to the asymptotic case $\mathrm{SNR} \uparrow \infty$:

Corollary 4. *Let a regular fading channel be defined as in Theorem 1. Then the fading number in the presence of noiseless feedback $\chi_\mathrm{FB}(\{\mathbb{H}_k\})$ is upper bounded as follows:*

$$\chi_\mathrm{FB}(\{\mathbb{H}_k\}) \le \chi_\mathrm{IID}(\mathbb{H}_0) + I\big(\mathbb{H}_0; \mathbb{H}_{-\infty}^{-1}\big), \tag{15}$$

where $\chi_\mathrm{IID}(\mathbb{H}_0)$ denotes the fading number without memory (or feedback).

Since the fading number of a SISO fading channel with memory (but without feedback) can be written as

$$\chi(\{H_k\}) = \chi_{\text{IID}}(H_0) + I(H_0; H_{-\infty}^{-1}), \tag{16}$$

(see [1]) the right-hand side of (15) can be achieved without feedback. Hence in the SISO case (15) is tight:

Corollary 5. *The fading number of a SISO regular fading channel with memory is not changed by a noiseless feedback link from the receiver to the transmitter:*

$$\chi_{\text{FB}}(\{H_k\}) = \chi(\{H_k\}) \tag{17}$$
$$= \log \pi + \mathsf{E}\left[\log |H_0|^2\right] - h(\{H_k\}). \tag{18}$$

The intuition behind this result is that feedback only helps in improving the input-power allocation and that this improvement is negligible in the log log-regime.

Remark 6. *The result of Corollary 5 can be generalized to the situation of* partial side-information. *In this scenario a side-information process* $\{\mathbf{S}_k\}$*, where* $\{H_k, \mathbf{S}_k\}$ *is jointly stationary, ergodic, and of finite mutual information rate*

$$I\big(\{H_k\}; \{\mathbf{S}_k\}\big) \triangleq \lim_{n\uparrow\infty} \frac{1}{n} I\big(H_1^n; \mathbf{S}_1^n\big) < \infty, \tag{19}$$

is known to the receiver. It can then be shown that the disclosure of the side-information to the transmitter and a noiseless feedback link from the receiver to the transmitter does not increase the asymptotic SISO channel capacity, i.e.,

$$\lim_{\text{SNR}\uparrow\infty} \{\mathsf{C}_{\text{FB,SI@Tx}}(\text{SNR}) - \mathsf{C}(\text{SNR})\} = 0, \tag{20}$$

where the above holds irrespective of whether an average-power constraint (10) *or a peak-power constraint* (11) *is imposed.*

That is, the fading number in the presence of noiseless feedback and transmitter side-information is given by

$$\chi_{\text{FB,SI@Tx}}\big(\{H_k\}\big|\{\mathbf{S}_k\}\big)$$
$$= \chi\big(\{H_k\}\big|\{\mathbf{S}_k\}\big) \tag{21}$$
$$= \log \pi + \mathsf{E}\left[\log |H_0|^2\right] - h\big(\{H_k\}\big|\{\mathbf{S}_k\}\big), \tag{22}$$

where $\chi\big(\{H_k\}\big|\{\mathbf{S}_k\}\big)$ *denotes the fading number of a fading channel with memory and partial side-information at the receiver only [1].*

IV. Gaussian Fading and Feedback

In this section we restrict ourselves to the family of Gaussian fading processes. This allows us to significantly tighten the bounds. As a first result we will improve the upper bound (13):

Theorem 7. *Let the mean-*D *Gaussian MIMO fading* $\{\mathbb{H}_k\}$ *be such that the process* $\{\mathbb{H}_k - \mathsf{D}\}$ *is spatially IID with each component being a zero-mean, unit-variance, circularly-symmetric, stationary and ergodic, complex Gaussian process with spectral distribution function* $\mathsf{F}(\cdot)$*. Then the capacity under a peak-power constraint and in presence of feedback* $\mathsf{C}_{\text{FB}}^{\text{PP}}(\text{SNR})$ *is upper bounded by*

$$\mathsf{C}_{\text{FB}}^{\text{PP}}(\text{SNR}) \le \mathsf{C}_{\text{IID}}^{\text{PP}}(\text{SNR}) + n_{\text{R}} \log \frac{1 + \frac{1}{\text{SNR}}}{\epsilon^2\left(\frac{1}{\text{SNR}}\right) + \frac{1}{\text{SNR}}}. \tag{23}$$

Here, $\mathsf{C}_{\text{IID}}^{\text{PP}}(\text{SNR})$ *denotes the capacity in the memoryless fading case (without feedback), and* $\epsilon^2(\delta^2)$ *denotes the error in predicting a component of the fading matrix from a variance-*δ^2 *noisy observation of its infinite past,* i.e.,

$$\epsilon^2(\delta^2) = \exp\left(\int_{-1/2}^{1/2} \log\left(\mathsf{F}'(\lambda) + \delta^2\right) \mathrm{d}\lambda\right) - \delta^2. \tag{24}$$

Proof. A proof is given in Appendix A. □

Applied to the case of regular fading Theorem 7 leads immediately to the following upper bound on the MIMO Gaussian fading number with memory and feedback under a peak-power constraint:

Corollary 8. *Assume a mean-*D *Gaussian MIMO fading* $\{\mathbb{H}_k\}$ *such that the process* $\{\mathbb{H}_k - \mathsf{D}\}$ *is spatially IID with each component as described in Theorem 7. Further assume that the fading process is regular. Then the fading number under a peak-power constraint and in presence of feedback* $\chi_{\text{FB}}^{\text{PP}}(\{\mathbb{H}_k\})$ *is upper bounded as follows:*

$$\chi_{\text{FB}}^{\text{PP}}(\{\mathbb{H}_k\}) \le \chi_{\text{IID}}^{\text{PP}}(\mathbb{H}_0) + n_{\text{R}} \log \frac{1}{\epsilon^2}, \tag{25}$$

where $\chi_{\text{IID}}^{\text{PP}}(\mathbb{H}_0)$ *denotes the fading number in the memoryless fading case, and* ϵ^2 *denotes the error in predicting a component of the fading matrix from its infinite past:*

$$\epsilon^2 = \exp\left(\int_{-1/2}^{1/2} \log \mathsf{F}'(\lambda)\, \mathrm{d}\lambda\right). \tag{26}$$

Proof. This follows by subtracting log log SNR from both sides of (23) and then letting the SNR tend to infinity. □

Remark 9. *Note that* (25) *is for MIMO typically tighter than* (15)*, indeed, for such Gaussian channels,*

$$I\big(\mathbb{H}_0; \mathbb{H}_{-\infty}^{-1}\big) = n_{\text{T}} n_{\text{R}} \log \frac{1}{\epsilon^2}. \tag{27}$$

A similar result can be proved in the case of an average-power constraint:

Theorem 10. *Let the mean-*D *Gaussian MIMO fading* $\{\mathbb{H}_k\}$ *be such that the process* $\{\mathbb{H}_k - \mathsf{D}\}$ *is spatially IID with each component being a zero-mean, unit-variance, circularly-symmetric, stationary and ergodic, regular complex Gaussian process with spectral distribution function* $\mathsf{F}(\cdot)$*. Then the fading number under an average-power constraint and in presence of feedback* $\chi_{\text{FB}}^{\text{Avg}}(\{\mathbb{H}_k\})$ *is upper bounded by*

$$\chi_{\text{FB}}^{\text{Avg}}(\{\mathbb{H}_k\}) \le \chi_{\text{IID}}^{\text{Avg}}(\mathbb{H}_0) + n_{\text{R}} \log \frac{1}{\epsilon^2}, \tag{28}$$

where $\chi_{\text{IID}}^{\text{Avg}}(\mathbb{H}_0)$ *denotes the fading number in the memoryless fading case, and* ϵ^2 *denotes the error in predicting a component of the fading matrix from its infinite past* (26)*.*

Proof. A proof is given in Appendix B. □

Since the fading number of a mean-$\mathbf{d}$ Gaussian MISO fading process $\{\mathbf{H}_k\}$ where $\{\mathbf{H}_k - \mathbf{d}\}$ is spatially IID can be written as

$$\chi(\{\mathbf{H}_k^{\mathsf{T}}\}) = \chi_{\text{IID}}(\{\mathbf{H}_k^{\mathsf{T}}\}) + \log \frac{1}{\epsilon^2} \tag{29}$$

(see [5, Corollary 5.7], [1]), it follows from Theorem 7 and Theorem 10 that feedback does not increase the MISO Gaussian fading number:

Corollary 11. *Let the mean-$\mathbf{d}$ Gaussian MISO fading be such that the process $\{\mathbf{H}_k - \mathbf{d}\}$ is spatially IID with each component being a zero-mean, unit-variance, circularly-symmetric, regular complex Gaussian process with spectral distribution function $\mathsf{F}(\cdot)$. Then the fading number in presence of feedback $\chi_{\text{FB}}(\{\mathbf{H}_k^{\mathsf{T}}\})$ is given by*

$$\chi_{\text{FB}}(\{\mathbf{H}_k^{\mathsf{T}}\}) = \chi(\{\mathbf{H}_k^{\mathsf{T}}\}) \tag{30}$$

$$= -1 + \log d_*^2 - \text{Ei}\big(-d_*^2\big) + \log \frac{1}{\epsilon^2}, \tag{31}$$

with

$$d_* = \max_{\|\hat{\mathbf{x}}\|=1} \big|\mathbf{d}^{\mathsf{T}}\hat{\mathbf{x}}\big| \tag{32}$$

and where the prediction error ϵ^2 is defined in (26).

Next, we consider the pre-log Π^{PP} as defined in (3). We will restrict ourselves here to the SISO case.

In [3], [4] it was shown that the pre-log is determined by the ratio of the total length of the frequency bands where the spectral density is null to the total frequencies:

$$\Pi^{\text{PP}} = \mu\big(\{\lambda : \mathsf{F}'(\lambda) = 0\}\big), \tag{33}$$

where $\mu(\cdot)$ denotes the Lebesgue measure on the interval $(-1/2, 1/2]$.

We now extend these results to the case where there is noiseless feedback from the receiver to the transmitter.

Theorem 12. *Suppose that the SISO fading $\{H_k\}$ and the specular component d are such that $\{H_k - d\}$ is a zero-mean, unit-variance, finite-energy, circularly-symmetric, stationary and ergodic, complex Gaussian process. Then, under a peak-power constraint, noiseless feedback from the receiver to the transmitter does not increase the asymptotic channel capacity in the sense that,*

$$\lim_{\text{SNR}\uparrow\infty} \frac{\mathsf{C}_{\text{FB}}^{\text{PP}}(\text{SNR})}{\log \text{SNR}} = \lim_{\text{SNR}\uparrow\infty} \frac{\mathsf{C}^{\text{PP}}(\text{SNR})}{\log \text{SNR}}. \tag{34}$$

That is, the pre-log in the presence of noiseless feedback $\Pi_{\text{FB}}^{\text{PP}}$ is given by (33).

Proof. For a proof see [8]. □

A. A Proof of Theorem 7

Using standard arguments like, *e.g.*, shown in [8], the information rate $\mathsf{R}_{\text{FB}}^{\text{PP}}(\text{SNR})$ of the system shown in Figure 1 can be upper bounded by

$$\mathsf{R}_{\text{FB}}^{\text{PP}}(\text{SNR}) \leq \lim_{n\uparrow\infty} \frac{1}{n} I\big(M; \mathbf{Y}_1^n\big). \tag{35}$$

We hence continue with bounding $\frac{1}{n} I\big(M; \mathbf{Y}_1^n\big)$ as follows:

$$\frac{1}{n} I\big(M; \mathbf{Y}_1^n\big)$$

$$= \frac{1}{n}\sum_{k=1}^{n} I\big(M; \mathbf{Y}_k \big| \mathbf{Y}_1^{k-1}\big) \tag{36}$$

$$= \frac{1}{n}\sum_{k=1}^{n} I\big(M, \mathbf{X}_1^k; \mathbf{Y}_k \big| \mathbf{Y}_1^{k-1}\big) \tag{37}$$

$$= \frac{1}{n}\sum_{k=1}^{n} \Big(I\big(\mathbf{X}_1^k; \mathbf{Y}_k \big| \mathbf{Y}_1^{k-1}\big) + I\big(M; \mathbf{Y}_k \big| \mathbf{Y}_1^{k-1}, \mathbf{X}_1^k\big)\Big) \tag{38}$$

$$= \frac{1}{n}\sum_{k=1}^{n} I\big(\mathbf{X}_1^k; \mathbf{Y}_k \big| \mathbf{Y}_1^{k-1}\big) \tag{39}$$

$$= \frac{1}{n}\sum_{k=1}^{n} \Big(h\big(\mathbf{Y}_k \big| \mathbf{Y}_1^{k-1}\big) - h\big(\mathbf{Y}_k \big| \mathbf{Y}_1^{k-1}, \mathbf{X}_1^k\big)\Big) \tag{40}$$

$$\leq \frac{1}{n}\sum_{k=1}^{n} \sup_{p_{\mathbf{X}_1^k}} \Big\{ h\big(\mathbf{Y}_k\big) - h\big(\mathbf{Y}_k \big| \mathbf{Y}_1^{k-1}, \mathbf{X}_1^k\big)\Big\} \tag{41}$$

$$= \frac{1}{n}\sum_{k=1}^{n} \sup_{p_{\mathbf{X}_k}} \sup_{p_{\mathbf{X}_1^{k-1}|\mathbf{X}_k}} \Big\{ h\big(\mathbf{Y}_k\big) - h\big(\mathbf{Y}_k \big| \mathbf{X}_k, \mathbf{Y}_1^{k-1}, \mathbf{X}_1^{k-1}\big)\Big\} \tag{42}$$

$$= \frac{1}{n}\sum_{k=1}^{n} \sup_{p_{\mathbf{X}_k}} \Big\{ h\big(\mathbf{Y}_k\big) - \inf_{p_{\mathbf{X}_1^{k-1}|\mathbf{X}_k}} h\big(\mathbf{Y}_k \big| \mathbf{X}_k, \{\mathbb{H}_\ell \mathbf{X}_\ell + \mathbf{Z}_\ell\}_{\ell=1}^{k-1}, \mathbf{X}_1^{k-1}\big)\Big\}. \tag{43}$$

Here, the first equality follows from the chain rule; in the second we use the fact that $\mathbf{X}_1^k$ is a function of M and the feedback $\mathbf{Y}_1^{k-1}$; in the third the chain rule is applied again; the fourth equality follows because

$$I\big(M; \mathbf{Y}_k \big| \mathbf{Y}_1^{k-1}, \mathbf{X}_1^k\big) = 0 \tag{44}$$

which can be proved, *e.g.*, graphically using a technique based on *causal interpretations* [9], [10]. The inequality follows from introducing a supremum over the distributions of $\mathbf{X}_1^k$ and from the fact that conditioning does not increase entropy.

To further lower bound the second term in the sum we introduce the matrix-valued random process $\{\mathbb{W}_k\} \in \mathbb{C}^{n_{\text{R}} \times n_{\text{T}}}$ which is temporally and spatially IID with $\mathcal{N}_{\mathbb{C}}(0,1)$ entries:

$$\inf_{p_{\mathbf{X}_1^{k-1}|\mathbf{X}_k}} h\big(\mathbf{Y}_k \big| \mathbf{X}_k, \{\mathbb{H}_\ell \mathbf{X}_\ell + \mathbf{Z}_\ell\}_{\ell=1}^{k-1}, \mathbf{X}_1^{k-1}\big)$$

$$= \inf_{p_{\mathbf{X}_1^{k-1}|\mathbf{X}_k}} h\left(\mathbf{Y}_k \middle| \mathbf{X}_k, \left\{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell + \frac{1}{\|\mathbf{X}_\ell\|}\mathbf{Z}_\ell\right\}_{\ell=1}^{k-1}, \mathbf{X}_1^{k-1}\right) \tag{45}$$

$$= \inf_{p_{\hat{\mathbf{X}}_1^{k-1}|\mathbf{X}_k}} h\left(\mathbf{Y}_k \middle| \mathbf{X}_k, \left\{\mathbb{H}_\ell \hat{\mathbf{X}}_\ell + \frac{1}{\sqrt{\mathcal{E}}}\mathbf{Z}_\ell\right\}_{\ell=1}^{k-1}, \hat{\mathbf{X}}_1^{k-1}\right) \tag{46}$$

$$= \inf_{p_{\hat{\mathbf{X}}_1^{k-1}|\mathbf{X}_k}} h\left(\mathbf{Y}_k \middle| \mathbf{X}_k, \left\{\left(\mathbb{H}_\ell + \frac{\sigma}{\sqrt{\mathcal{E}}}\mathbb{W}_\ell\right)\hat{\mathbf{X}}_\ell\right\}_{\ell=1}^{k-1}, \hat{\mathbf{X}}_1^{k-1}\right) \tag{47}$$

$$\geq \inf_{p_{\hat{\mathbf{X}}_1^{k-1}|\mathbf{X}_k}} h\left(\mathbf{Y}_k \middle| \mathbf{X}_k, \left\{\left(\mathbb{H}_\ell + \frac{\sigma}{\sqrt{\mathcal{E}}}\mathbb{W}_\ell\right)\hat{\mathbf{X}}_\ell\right\}_{\ell=1}^{k-1}, \left\{\mathbb{H}_\ell + \frac{\sigma}{\sqrt{\mathcal{E}}}\mathbb{W}_\ell\right\}_{\ell=1}^{k-1}, \hat{\mathbf{X}}_1^{k-1}\right) \tag{48}$$

$$= h\left(\mathbf{Y}_k \middle| \mathbf{X}_k, \left\{\mathbb{H}_\ell + \frac{\sigma}{\sqrt{\mathcal{E}}}\mathbb{W}_\ell\right\}_{\ell=1}^{k-1}, \hat{\mathbf{X}}_1^{k-1}\right) \tag{49}$$

$$= h\left(\mathbf{Y}_k \middle| \mathbf{X}_k, \left\{\mathbb{H}_\ell + \frac{\sigma}{\sqrt{\mathcal{E}}}\mathbb{W}_\ell\right\}_{\ell=1}^{k-1}\right). \tag{50}$$

Here in the first equality we have divided $\mathbf{Y}_\ell$ by the known magnitude of $\mathbf{X}_\ell$ and we have introduced $\hat{\mathbf{X}}_\ell \triangleq \frac{\mathbf{X}_\ell}{\|\mathbf{X}_\ell\|}$; for the following step note that the differential entropy is minimized by choosing the magnitude of $\mathbf{X}_\ell$ largest in order to reduce the influence of the noise $\mathbf{Z}_\ell$; in the third equality we use the fact that $\sigma \mathbb{W}_\ell \hat{\mathbf{X}}_\ell \sim \mathbf{Z}_\ell$ independent of the distribution of

$\hat{\mathbf{X}}_\ell$; in the following inequality we rely on conditioning that reduces entropy; in the second last equality we drop the superfluous terms $\big(\mathbb{H}_\ell + \frac{\sigma}{\sqrt{\mathcal{E}}}\mathbb{W}_\ell\big)\hat{\mathbf{X}}_\ell$ since they can be computed from $\mathbb{H}_\ell + \frac{\sigma}{\sqrt{\mathcal{E}}}\mathbb{W}_\ell$ and $\hat{\mathbf{X}}_\ell$; and in the last equality we note that conditional on $\mathbf{X}_k$ the output $\mathbf{Y}_k$ is independent of the past inputs.

Using (50) in (43) we get

$$
\begin{aligned}
&\frac{1}{n} I\big(M;\mathbf{Y}_1^n\big) \\
&\quad\le \frac{1}{n}\sum_{k=1}^{n} \sup_{p_{\mathbf{X}_k}} I\left(\mathbf{Y}_k;\mathbf{X}_k,\left\{\mathbb{H}_\ell + \frac{\sigma}{\sqrt{\mathcal{E}}}\mathbb{W}_\ell\right\}_{\ell=1}^{k-1}\right) && (51)\\
&\quad\le \frac{1}{n}\sum_{k=1}^{n} \sup_{p_{\mathbf{X}_k}} I\left(\mathbf{Y}_k;\mathbf{X}_k,\left\{\mathbb{H}_\ell + \frac{\sigma}{\sqrt{\mathcal{E}}}\mathbb{W}_\ell\right\}_{\ell=-\infty}^{k-1}\right) && (52)\\
&\quad= \frac{1}{n}\sum_{k=1}^{n} \sup_{p_{\mathbf{X}_0}} I\left(\mathbf{Y}_0;\mathbf{X}_0,\left\{\mathbb{H}_\ell + \frac{\sigma}{\sqrt{\mathcal{E}}}\mathbb{W}_\ell\right\}_{\ell=-\infty}^{-1}\right) && (53)\\
&\quad= \sup_{p_{\mathbf{X}_0}}\left\{ I\big(\mathbf{Y}_0;\mathbf{X}_0\big) + I\left(\mathbf{Y}_0;\left\{\mathbb{H}_\ell + \frac{\sigma}{\sqrt{\mathcal{E}}}\mathbb{W}_\ell\right\}_{\ell=-\infty}^{-1}\,\middle|\,\mathbf{X}_0\right)\right\} && (54)\\
&\quad\le \sup_{p_{\mathbf{X}_0}} I\big(\mathbf{Y}_0;\mathbf{X}_0\big) + \sup_{p_{\mathbf{X}_0}} I\left(\mathbb{H}_0\mathbf{X}_0 + \mathbf{Z}_0;\left\{\mathbb{H}_\ell + \frac{\sigma}{\sqrt{\mathcal{E}}}\mathbb{W}_\ell\right\}_{\ell=-\infty}^{-1}\,\middle|\,\mathbf{X}_0\right) && (55)\\
&\quad\le \sup_{p_{\mathbf{X}_0}} I\big(\mathbf{Y}_0;\mathbf{X}_0\big) + \sup_{\mathbf{x}:\|\mathbf{x}\|^2\le\mathcal{E}} I\left(\mathbb{H}_0\mathbf{x} + \mathbf{Z}_0;\left\{\mathbb{H}_\ell + \frac{\sigma}{\sqrt{\mathcal{E}}}\mathbb{W}_\ell\right\}_{\ell=-\infty}^{-1}\right) && (56)\\
&\quad= \mathrm{C}_{\mathrm{IID}}^{\mathrm{PP}}(\mathrm{SNR}) + \sup_{\|\hat{\mathbf{x}}\|=1} I\left(\mathbb{H}_0\hat{\mathbf{x}} + \frac{1}{\sqrt{\mathcal{E}}}\mathbf{Z}_0;\left\{\mathbb{H}_\ell + \frac{\sigma}{\sqrt{\mathcal{E}}}\mathbb{W}_\ell\right\}_{\ell=-\infty}^{-1}\right). && (57)
\end{aligned}
$$

Here, in the second inequality we have added the infinite past; in the subsequent equality we used stationarity; and in the last equality $\mathrm{C}_{\mathrm{IID}}^{\mathrm{PP}}(\mathrm{SNR})$ denotes the capacity in the situation without memory and without feedback.

Since we assumed that the rows of the fading matrix are independent, we may write the second term of (57) componentwise:

$$
\begin{aligned}
&I\left(\mathbb{H}_0\hat{\mathbf{x}} + \frac{1}{\sqrt{\mathcal{E}}}\mathbf{Z}_0;\left\{\mathbb{H}_\ell + \frac{\sigma}{\sqrt{\mathcal{E}}}\mathbb{W}_\ell\right\}_{\ell=-\infty}^{-1}\right) \\
&\quad= \sum_{r=1}^{n_\mathrm{R}} I\left(\sum_{t=1}^{n_\mathrm{T}} H_0^{(r,t)}\hat{x}^{(t)} + \frac{1}{\sqrt{\mathcal{E}}}Z_0^{(r)};\left\{H_\ell^{(r,t)} + \frac{\sigma}{\sqrt{\mathcal{E}}}W_\ell^{(r,t)}\right\}_{\substack{1\le t\le n_\mathrm{T}\\ -\infty<\ell\le -1}}\right) && (58)\\
&\quad= \sum_{r=1}^{n_\mathrm{R}}\left(h\left(\sum_{t=1}^{n_\mathrm{T}} H_0^{(r,t)}\hat{x}^{(t)} + \frac{1}{\sqrt{\mathcal{E}}}Z_0^{(r)}\right) - h\left(\sum_{t=1}^{n_\mathrm{T}} H_0^{(r,t)}\hat{x}^{(t)} + \frac{1}{\sqrt{\mathcal{E}}}Z_0^{(r)}\,\middle|\,\left\{H_\ell^{(r,t)} + \frac{\sigma}{\sqrt{\mathcal{E}}}W_\ell^{(r,t)}\right\}_{\substack{1\le t\le n_\mathrm{T}\\ -\infty<\ell\le -1}}\right)\right). && (59)
\end{aligned}
$$

Note that the random variable $\sum_{t=1}^{n_\mathrm{T}} H_0^{(r,t)}\hat{x}^{(t)} + \frac{1}{\sqrt{\mathcal{E}}}Z_0^{(r)}$ is Gaussian with a variance $1 + 1/\mathrm{SNR}$ (for $\|\hat{\mathbf{x}}\| = 1$) and that conditioned on $\left\{H_\ell^{(r,t)} + \frac{\sigma}{\sqrt{\mathcal{E}}}W_\ell^{(r,t)}\right\}_{1\le t\le n_\mathrm{T}, -\infty<\ell\le -1}$ it is Gaussian with a variance $\epsilon^2(1/\mathrm{SNR}) + 1/\mathrm{SNR}$ where $\epsilon^2(\delta^2)$ denotes the error in predicting a component of $\mathbb{H}_0$ from a variance-δ^2 noisy observation of its infinite past.[3] Hence, we get

$$
\begin{aligned}
\mathsf{R}_{\mathrm{FB}}^{\mathrm{PP}}(\mathrm{SNR}) &\le \mathrm{C}_{\mathrm{IID}}^{\mathrm{PP}}(\mathrm{SNR}) + \sum_{r=1}^{n_\mathrm{R}} \log\frac{1+\frac{1}{\mathrm{SNR}}}{\epsilon^2\left(\frac{1}{\mathrm{SNR}}\right) + \frac{1}{\mathrm{SNR}}} && (60)\\
&= \mathrm{C}_{\mathrm{IID}}^{\mathrm{PP}}(\mathrm{SNR}) + n_\mathrm{R}\log\frac{1+\frac{1}{\mathrm{SNR}}}{\epsilon^2\left(\frac{1}{\mathrm{SNR}}\right) + \frac{1}{\mathrm{SNR}}}. && (61)
\end{aligned}
$$

B. A Proof of Theorem 10

Analogously to the proof of Theorem 7 one can show that (35) holds. We will therefore continue with bounding $\frac{1}{n}I\big(M;\mathbf{Y}_1^n\big)$:

$$
\begin{aligned}
&\frac{1}{n} I\big(M;\mathbf{Y}_1^n\big) \\
&\quad= \frac{1}{n}\sum_{k=1}^{n} I\big(M;\mathbf{Y}_k\,\big|\,\mathbf{Y}_1^{k-1}\big) && (62)\\
&\quad= \frac{1}{n}\sum_{k=1}^{n}\Big(I\big(M,\mathbf{Y}_1^{k-1};\mathbf{Y}_k\big) - I\big(\mathbf{Y}_1^{k-1};\mathbf{Y}_k\big)\Big) && (63)\\
&\quad\le \frac{1}{n}\sum_{k=1}^{n} I\big(M,\mathbf{Y}_1^{k-1};\mathbf{Y}_k\big) && (64)\\
&\quad\le \frac{1}{n}\sum_{k=1}^{n} I\big(M,\mathbf{Y}_1^{k-1},\mathbb{H}_1^{k-1};\mathbf{Y}_k\big) && (65)\\
&\quad= \frac{1}{n}\sum_{k=1}^{n} I\big(M,\mathbf{Y}_1^{k-1},\mathbb{H}_1^{k-1},\mathbf{X}_k;\mathbf{Y}_k\big) && (66)\\
&\quad= \frac{1}{n}\sum_{k=1}^{n}\Big(I\big(\mathbb{H}_1^{k-1},\mathbf{X}_k;\mathbf{Y}_k\big) + I\big(M,\mathbf{Y}_1^{k-1};\mathbf{Y}_k\,\big|\,\mathbb{H}_1^{k-1},\mathbf{X}_k\big)\Big) && (67)\\
&\quad= \frac{1}{n}\sum_{k=1}^{n} I\big(\mathbb{H}_1^{k-1},\mathbf{X}_k;\mathbf{Y}_k\big) && (68)\\
&\quad= \frac{1}{n}\sum_{k=1}^{n}\Big(I\big(\mathbf{X}_k;\mathbf{Y}_k\big) + I\big(\mathbb{H}_1^{k-1};\mathbf{Y}_k\,\big|\,\mathbf{X}_k\big)\Big) && (69)\\
&\quad\le \frac{1}{n}\sum_{k=1}^{n}\left(\mathrm{C}_{\mathrm{IID}}^{\mathrm{Avg}}\left(\frac{\mathsf{E}\big[\|\mathbf{X}_k\|^2\big]}{\sigma^2}\right) + I\big(\mathbb{H}_1^{k-1};\mathbf{Y}_k\,\big|\,\mathbf{X}_k\big)\right) && (70)\\
&\quad\le \mathrm{C}_{\mathrm{IID}}^{\mathrm{Avg}}(\mathrm{SNR}) + \frac{1}{n}\sum_{k=1}^{n} I\big(\mathbb{H}_1^{k-1};\mathbf{Y}_k\,\big|\,\mathbf{X}_k\big). && (71)
\end{aligned}
$$

Here the first two equalities follow from the chain rule; the subsequent inequality from the non-negativity of mutual information; the following inequality from adding random matrices; the subsequent equality follows since $\mathbf{X}_k$ is a deterministic function of M and $\mathbf{Y}_1^{k-1}$ (and hypothetically also $\mathbb{H}_1^{k-1}$); then we have used the chain rule again; (68) follows since[4]

$$
I\big(M,\mathbf{Y}_1^{k-1};\mathbf{Y}_k\,\big|\,\mathbb{H}_1^{k-1},\mathbf{X}_k\big) = 0; \qquad (72)
$$

[3]Here we use the assumption that the columns of $\{\mathbb{H}_k\}$ are spatially IID.

[4]This statement can be proved again graphically using a technique based on *causal interpretations*, see [9], [10].

in the following equality we have used the chain rule once more; in the second last inequality we upper bounded the mutual information term by the IID capacity for a given power at time k; and in the last inequality we have used the concavity of capacity.

We continue by upper bounding the second term:

$$
\begin{aligned}
&I\big(\mathbb{H}_1^{k-1}; \mathbf{Y}_k \,\big|\, \mathbf{X}_k\big) \\
&\quad \le I\big(\mathbb{H}_1^{k-1}; \mathbb{H}_k\mathbf{X}_k \,\big|\, \mathbf{X}_k\big) && (73)\\
&\quad \le I\big(\mathbb{H}_{-\infty}^{-1}; \mathbb{H}_0\mathbf{X}_0 \,\big|\, \mathbf{X}_0\big) && (74)\\
&\quad = I\big(\mathbb{H}_{-\infty}^{-1}; \mathbb{H}_0\hat{\mathbf{X}}_0 \,\big|\, \mathbf{X}_0\big) && (75)\\
&\quad \le \sup_{\|\hat{\mathbf{x}}\|=1} I\big(\mathbb{H}_{-\infty}^{-1}; \mathbb{H}_0\hat{\mathbf{x}}\big) && (76)\\
&\quad = \sup_{\|\hat{\mathbf{x}}\|=1} \sum_{r=1}^{n_\mathrm{R}} I\left(\{H_\ell^{(r,t)}\}_{\substack{1\le t\le n_\mathrm{T}\\ -\infty<\ell\le -1}}; \sum_{t=1}^{n_\mathrm{T}} H_0^{(r,t)}\hat{x}^{(t)}\right) && (77)\\
&\quad = \sup_{\|\hat{\mathbf{x}}\|=1} \sum_{r=1}^{n_\mathrm{R}} \left(h\left(\sum_{t=1}^{n_\mathrm{T}} H_0^{(r,t)}\hat{x}^{(t)}\right) \right.\\
&\qquad\qquad \left. - h\left(\sum_{t=1}^{n_\mathrm{T}} H_0^{(r,t)}\hat{x}^{(t)} \,\middle|\, \{H_\ell^{(r,t)}\}_{\substack{1\le t\le n_\mathrm{T}\\ -\infty<\ell\le -1}}\right)\right). && (78)
\end{aligned}
$$

Here, the first inequality follows from the data processing inequality; in the subsequent inequality we have added the infinite past and then used stationarity; and the following equality follows from scaling where $\hat{\mathbf{X}}_0 \triangleq \frac{\mathbf{X}_0}{\|\mathbf{X}_0\|}$. In (77) we use the fact that the rows of the fading matrix are independent in order to write the expression componentwise.

We continue analogously to (59): since the random variable $\sum_{t=1}^{n_\mathrm{T}} H_0^{(r,t)}\hat{x}^{(t)}$ is Gaussian with variance 1 for $\|\hat{\mathbf{x}}\| = 1$ and since conditioned on $\{H_\ell^{(r,t)}\}_{1\le t\le n_\mathrm{T}, -\infty<\ell\le -1}$ it is Gaussian with variance ϵ^2 (because the columns of $\{\mathbb{H}_k\}$ are spatially IID), it follows that

$$
\begin{aligned}
\mathcal{R}_{\mathrm{FB}}^{\mathrm{Avg}}(\mathrm{SNR}) &\le \mathrm{C}_{\mathrm{IID}}^{\mathrm{Avg}}(\mathrm{SNR}) + \sum_{r=1}^{n_\mathrm{R}} \log\frac{1}{\epsilon^2} && (79)\\
&= \mathrm{C}_{\mathrm{IID}}^{\mathrm{Avg}}(\mathrm{SNR}) + n_\mathrm{R}\log\frac{1}{\epsilon^2}. && (80)
\end{aligned}
$$

Acknowledgments

The authors wish to thank Tobias Koch for fruitful discussions.

References

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

[2] ——, "The fading number of single-input multiple-output fading channels with memory," November 2004, submitted to *IEEE Transactions on Information Theory*.

[3] A. Lapidoth, "On the high SNR capacity of stationary Gaussian fading channels," in *Proceedings Forty-First Allerton Conference on Communication, Control and Computing*, Allerton House, Monticello, Illinois, October 1–3, 2003.

[4] ——, "On the asymptotic capacity of stationary Gaussian fading channels," *IEEE Transactions on Information Theory*, vol. 51, no. 2, pp. 437–446, February 2005.

[5] T. Koch, "On the asymptotic capacity of multiple-input single-output fading channels with memory," Master's thesis, Signal and Information Processing Laboratory, ETH Zurich, Switzerland, April 2004, supervised by Prof. Dr. Amos Lapidoth.

[6] T. Koch and A. Lapidoth, "Degrees of freedom in non-coherent stationary MIMO fading channels," in *Proceedings Winter School on Coding and Information Theory*, Bratislava, Slovakia, February 20–25, 2005.

[7] ——, "The fading number and degrees of freedom in non-coherent MIMO fading channels: a peace pipe," 2005, submitted to *IEEE International Symposium on Information Theory (ISIT)*.

[8] S. M. Moser, "Duality-based bounds on channel capacity," Ph.D. dissertation, Swiss Federal Institute of Technology, Zurich, October 2004, Diss. ETH No. 15769. [Online]. Available: http://www.isi.ee.ethz.ch/∼moser

[9] J. L. Massey, "Causal interpretations of random variables," *Problemy Peredachi Informatsii (Problems of Information Transmission)*, vol. 32, no. 1, pp. 131–136, January–March 1996.

[10] ——, "Correction to 'Causal interpretations of random variables'," *Problemy Peredachi Informatsii (Problems of Information Transmission)*, May 16, 1997.