

Gaussian Fading Is the Worst Fading

Tobias Koch, *Member, IEEE*, and Amos Lapidoth, *Fellow, IEEE*

Abstract—The capacity of peak-power limited, single-antenna, noncoherent, flat-fading channels with memory is considered. The emphasis is on the capacity pre-log, i.e., on the limiting ratio of channel capacity to the logarithm of the signal-to-noise ratio (SNR), as the SNR tends to infinity. It is shown that, among all stationary and ergodic fading processes of a given spectral distribution function and whose law has no mass point at zero, the Gaussian process gives rise to the smallest pre-log. The assumption that the law of the fading process has no mass point at zero is essential in the sense that there exist stationary and ergodic fading processes whose law has a mass point at zero and that give rise to a smaller pre-log than the Gaussian process of equal spectral distribution function. An extension of these results to multiple-input single-output (MISO) fading channels with memory is also presented.

Index Terms—Channel capacity, channels with memory, fading channels, flat fading, high signal-to-noise ratio (SNR), multiplexing gain, noncoherent, time-selective.

I. INTRODUCTION

WE study the capacity of peak-power limited, single-antenna, discrete-time, flat-fading channels with memory. A noncoherent channel model is considered where the transmitter and receiver are both aware of the law of the fading process, but not of its realization. Our focus is on the capacity at high signal-to-noise ratio (SNR). Specifically, we study the capacity pre-log, which is defined as the limiting ratio of channel capacity to the logarithm of the SNR, as the SNR tends to infinity.

The capacity pre-log of *Gaussian* fading channels was derived in [1] (see also [2]). It was shown that the pre-log is given by the Lebesgue measure of the set of harmonics where the derivative of the spectral distribution function that characterizes the memory of the fading process is zero. To the best of our knowledge, the capacity pre-log of *non-Gaussian* fading channels is unknown.

In this paper, we demonstrate that the Gaussian assumption in the analysis of fading channels at high SNR is conservative in the sense that for a large class of fading processes the Gaussian process is the worst. More precisely, we show that among all stationary and ergodic fading processes of a given spectral distribution function and whose law has no mass point at zero, the Gaussian process gives rise to the smallest pre-log.

Manuscript received February 19, 2009; revised October 30, 2009. Current version published March 10, 2010. The material in this paper was presented in part at the IEEE International Symposium on Information Theory (ISIT), Seattle, WA, July 2006.

The authors are with the Department of Information Technology and Electrical Engineering, ETH Zurich, 8092 Zurich, Switzerland.

Communicated by G. Taricco, Associate Editor for Communications.

Digital Object Identifier 10.1109/TIT.2009.2039036

This paper is organized as follows. Section II describes the channel model. Section III defines channel capacity and the capacity pre-log. Section IV presents our main results. Section V provides the proofs of these results. Section VI discusses the extension of our results to multiple-input single-output (MISO) fading channels with memory. Section VII concludes the paper with a summary and a discussion of our results.

II. CHANNEL MODEL

Let \mathbb{C} and \mathbb{Z} denote the set of complex numbers and the set of integers, respectively. We consider a single-antenna flat-fading channel with memory where the time- k channel output $Y_k \in \mathbb{C}$ corresponding to the time- k channel input $x_k \in \mathbb{C}$ is given by

$$Y_k = H_k x_k + Z_k, \quad k \in \mathbb{Z}. \quad (1)$$

Here the random processes $\{Z_k, k \in \mathbb{Z}\}$ and $\{H_k, k \in \mathbb{Z}\}$ take value in \mathbb{C} and model the additive and multiplicative noises, respectively. It is assumed that these processes are statistically independent and of a joint law that does not depend on the input sequence $\{x_k, k \in \mathbb{Z}\}$.

The additive noise $\{Z_k, k \in \mathbb{Z}\}$ is a sequence of independent and identically distributed (i.i.d.) zero-mean, variance- σ^2 , circularly symmetric, complex Gaussian random variables. The multiplicative noise (“fading”) $\{H_k, k \in \mathbb{Z}\}$ is a mean- d , unit-variance, stationary, and ergodic stochastic process of spectral distribution function $F(\lambda)$, $-1/2 \leq \lambda \leq 1/2$, i.e., $F(\cdot)$ is a bounded and nondecreasing function on $[-1/2, 1/2]$ satisfying [3, p. 474, Th. 3.2]

$$\mathbb{E}[(H_{k+m} - d)(H_k - d)^*] = \int_{-1/2}^{1/2} e^{i2\pi m\lambda} dF(\lambda), \quad (k \in \mathbb{Z}, m \in \mathbb{Z})$$

where $i = \sqrt{-1}$, and where A^* denotes the complex conjugate of A . Since $F(\cdot)$ is monotonic, it is almost everywhere differentiable, and we denote its derivative by $F'(\cdot)$. (At the discontinuity points of $F(\cdot)$ the derivative $F'(\cdot)$ is undefined.) For example, if the fading process $\{H_k, k \in \mathbb{Z}\}$ is i.i.d., then

$$F'(\lambda) = 1, \quad -\frac{1}{2} \leq \lambda \leq \frac{1}{2}.$$

III. CHANNEL CAPACITY AND THE PRE-LOG

Channel capacity is defined as the supremum of all achievable rates. (We refer to [4, Ch. 8] for a definition of an achievable rate and for a more detailed discussion of channel capacity.) It was shown (e.g., [5, Th. 2]) that the capacity of our channel (1) under a peak-power constraint A^2 on the inputs is given by

$$C(\text{SNR}) = \lim_{n \rightarrow \infty} \frac{1}{n} \sup I(X_1^n; Y_1^n) \quad (2)$$

where SNR is defined as

$$\text{SNR} \triangleq \frac{A^2}{\sigma^2}$$

A_m^n denotes the sequence A_m, \dots, A_n ; and where the maximization is over all joint distributions on X_1, \dots, X_n satisfying with probability one

$$|X_k|^2 \leq A^2, \quad k = 1, \dots, n. \quad (3)$$

The capacity pre-log is defined as [1]

$$\Pi \triangleq \overline{\lim}_{\text{SNR} \rightarrow \infty} \frac{C(\text{SNR})}{\log \text{SNR}}. \quad (4)$$

For *Gaussian fading*, i.e., when $\{H_k - d, k \in \mathbb{Z}\}$ is a circularly symmetric, complex Gaussian process, the pre-log Π_G is given by the Lebesgue measure of the set of harmonics where the derivative of the spectral distribution function is zero, i.e.

$$\Pi_G = \mu(\{\lambda: F'(\lambda) = 0\}) \quad (5)$$

where $\mu(\cdot)$ denotes the Lebesgue measure on the interval $[-1/2, 1/2]$; see [1], [2]. (Here the subscript ‘‘G’’ stands for ‘‘Gaussian’’.) This result indicates that if the fading process is Gaussian and satisfies

$$\mu(\{\lambda: F'(\lambda) = 0\}) > 0$$

then the corresponding channel capacity grows logarithmically in the SNR. Note that otherwise the capacity can increase with the SNR in various ways. For instance, in [6] fading channels are studied that result in a capacity which increases double-logarithmically with the SNR, and in [1] spectral distribution functions are presented for which capacity grows as a fractional power of the logarithm of the SNR.

IV. MAIN RESULT

We show that, among all stationary and ergodic fading processes of a given spectral distribution function and whose law has no mass point at zero, the Gaussian process gives rise to the smallest pre-log. This is made precise in the following theorem.

Theorem 1: Consider a mean- d , unit-variance, stationary, and ergodic fading process $\{H_k, k \in \mathbb{Z}\}$ whose spectral distribution function is given by $F(\cdot)$ and whose law satisfies

$$\Pr[H_k = 0] = 0, \quad k \in \mathbb{Z}.$$

Then the corresponding capacity pre-log Π is lower bounded by

$$\Pi \geq \mu(\{\lambda: F'(\lambda) = 0\}). \quad (6)$$

Proof: See Section V-A. \square

Note 1: Theorem 1 continues to hold if $\{Z_k, k \in \mathbb{Z}\}$ is a sequence of i.i.d., variance- σ^2 , complex (not necessarily Gaussian) random variables of finite differential entropy. Thus, among all pairs of fading processes (satisfying the conditions

of Theorem 1) and i.i.d. additive noise processes of variance σ^2 , the pair where both processes are Gaussian gives rise to the smallest pre-log.

The assumption that the law of the fading process has no mass point at zero is essential in the following sense.

Note 2: There exists a mean- d , unit-variance, stationary, and ergodic fading process $\{H_k, k \in \mathbb{Z}\}$ of some spectral distribution function $F(\cdot)$ such that

$$\Pi < \mu(\{\lambda: F'(\lambda) = 0\}). \quad (7)$$

By Theorem 1, this process must satisfy

$$\Pr[H_k = 0] > 0, \quad k \in \mathbb{Z}.$$

Proof: See Section V-B. \square

Note 3: The inequality in (6) can be strict. For example, consider the *phase-noise channel* with memoryless phase noise. This channel can be viewed as a fading channel where the fading process $\{H_k, k \in \mathbb{Z}\}$ is given by

$$H_k = e^{i\Theta_k}, \quad k \in \mathbb{Z}$$

and where $\{\Theta_k, k \in \mathbb{Z}\}$ is i.i.d. with Θ_k being uniformly distributed over $[-\pi, \pi)$. This process gives rise to a pre-log $\Pi = 1/2$, whereas the Gaussian fading of equal spectral distribution function yields $\Pi_G = 0$.

Proof: For a derivation of the capacity pre-log of the phase-noise channel see Section V-C. \square

V. PROOFS

This section provides the proofs of our main results. For a proof of Theorem 1 see Section V-A, for a proof of Note 2 see Section V-B, and for a proof of Note 3 see Section V-C.

A. Proof of Theorem 1

To prove Theorem 1, we first derive a lower bound on the capacity, and proceed then to analyze its asymptotic growth as the SNR tends to infinity.

1) *Capacity Lower Bound:* To derive a lower bound on the capacity we consider inputs $\{X_k, k \in \mathbb{Z}\}$ that are i.i.d., zero-mean, and circularly symmetric, and for which $|X_k|^2$ is uniformly distributed over the interval $[0, A^2]$. Our derivation is based on the lower bound

$$\frac{1}{n} I(X_1^n; Y_1^n) \geq \frac{1}{n} I(X_1^n; Y_1^n | H_1^n) - \frac{1}{n} I(H_1^n; Y_1^n | X_1^n) \quad (8)$$

which follows from the chain rule

$$\begin{aligned} I(X_1^n; Y_1^n) &= I(X_1^n, H_1^n; Y_1^n) - I(H_1^n; Y_1^n | X_1^n) \\ &= I(H_1^n; Y_1^n) + I(X_1^n; Y_1^n | H_1^n) - I(H_1^n; Y_1^n | X_1^n) \end{aligned}$$

and the nonnegativity of mutual information.

We first study the first term on the right-hand side (RHS) of (8). Making use of the stationarity of the channel and of the fact that the inputs are i.i.d. we have

$$\frac{1}{n}I(X_1^n; Y_1^n | H_1^n) = I(X_1; Y_1 | H_1). \quad (9)$$

We lower bound the RHS of (9) as follows. For any fixed $\Upsilon > 0$ we have

$$\begin{aligned} I(X_1; Y_1 | H_1) &= h(H_1 X_1 + Z_1 | H_1) - h(Z_1) \\ &= \int_{|h_1| \geq \Upsilon} h(H_1 X_1 + Z_1 | H_1 = h_1) dP_{H_1}(h_1) \\ &\quad + \int_{|h_1| < \Upsilon} h(H_1 X_1 + Z_1 | H_1 = h_1) dP_{H_1}(h_1) - h(Z_1) \\ &\geq \int_{|h_1| \geq \Upsilon} h(H_1 X_1 + Z_1 | H_1 = h_1) dP_{H_1}(h_1) \\ &\quad + \Pr[|H_1| < \Upsilon] h(Z_1) - h(Z_1) \\ &\geq \int_{|h_1| \geq \Upsilon} (\log |h_1|^2 + h(X_1)) dP_{H_1}(h_1) \\ &\quad + \Pr[|H_1| < \Upsilon] h(Z_1) - h(Z_1) \\ &\geq \Pr[|H_1| \geq \Upsilon] (\log \Upsilon^2 + h(X_1)) \\ &\quad + \Pr[|H_1| < \Upsilon] h(Z_1) - h(Z_1) \\ &= \Pr[|H_1| \geq \Upsilon] (\log \Upsilon^2 + \log \pi + h(|X_1|^2)) \\ &\quad + \Pr[|H_1| < \Upsilon] h(Z_1) - h(Z_1) \\ &= \Pr[|H_1| \geq \Upsilon] \log A^2 + \Pr[|H_1| \geq \Upsilon] \log(\pi \Upsilon^2) \\ &\quad + \Pr[|H_1| < \Upsilon] h(Z_1) - h(Z_1) \\ &= \Pr[|H_1| \geq \Upsilon] \log A^2 + \Pr[|H_1| \geq \Upsilon] \log(\pi \Upsilon^2) \\ &\quad + (\Pr[|H_1| < \Upsilon] - 1) \log(\pi e \sigma^2) \\ &= \Pr[|H_1| \geq \Upsilon] \log \text{SNR} - \Pr[|H_1| \geq \Upsilon] (1 - \log \Upsilon^2) \end{aligned} \quad (10)$$

where $P_{H_1}(\cdot)$ denotes the distribution of the fading H_1 . Here the third step follows by conditioning the entropy in the second integral on X_1 ; the fourth step follows by conditioning the entropy in the first integral on Z_1 and by the behavior of differential entropy under scaling [4, Th. 9.6.4]; the fifth step follows because over the range of integration $|h_1| \geq \Upsilon$ we have $\log |h_1|^2 \geq \log \Upsilon^2$; the sixth step follows because X_1 is circularly symmetric [6, Lemma 6.16]; the seventh step follows by computing the entropy of a random variable that is uniformly distributed over the interval $[0, A^2]$; the eighth step follows by evaluating the entropy of a zero-mean, variance- σ^2 , circularly symmetric, complex Gaussian random variable $h(Z_k) = \log(\pi e \sigma^2)$; and the last step follows from $\Pr[|H_1| \geq \Upsilon] = 1 - \Pr[|H_1| < \Upsilon]$.

We next turn to the second term on the RHS of (8). In order to upper bound it we proceed along the lines of [7], but for non-Gaussian fading. Let \mathbf{Y} , \mathbf{H} , and \mathbf{Z} be the random vectors $(Y_1, \dots, Y_n)^\top$, $(H_1, \dots, H_n)^\top$, and $(Z_1, \dots, Z_n)^\top$ (where \mathbf{A}^\top

denotes the transpose of \mathbf{A}), and let \mathbb{X} be a diagonal matrix with diagonal entries x_1, \dots, x_n . It follows from (1) that

$$\mathbf{Y} = \mathbb{X}\mathbf{H} + \mathbf{Z}.$$

The conditional covariance matrix of \mathbf{Y} , conditional on x_1, \dots, x_n , is given by

$$\mathbb{E}[(\mathbf{Y} - \mathbb{E}[\mathbf{Y}])(\mathbf{Y} - \mathbb{E}[\mathbf{Y}])^\dagger | X_1^n = x_1^n] = \mathbb{X}\mathbf{K}_{\mathbf{H}\mathbf{H}}\mathbb{X}^\dagger + \sigma^2 \mathbf{I}_n$$

where \mathbf{I}_n is the $n \times n$ identity matrix, $(\cdot)^\dagger$ denotes Hermitian conjugation, and

$$\mathbf{K}_{\mathbf{H}\mathbf{H}} \triangleq \mathbb{E}[(\mathbf{H} - \mathbb{E}[\mathbf{H}])(\mathbf{H} - \mathbb{E}[\mathbf{H}])^\dagger].$$

Using the entropy maximizing property of circularly symmetric Gaussian vectors [4, Th. 9.6.5], we have

$$\begin{aligned} \frac{1}{n}I(H_1^n; Y_1^n | X_1^n) &= \frac{1}{n}h(Y_1^n | X_1^n) - \frac{1}{n}h(Z_1^n) \\ &\leq \frac{1}{n}\mathbb{E} \left[\log \det \left(\mathbf{I}_n + \frac{1}{\sigma^2} \mathbb{X}\mathbf{K}_{\mathbf{H}\mathbf{H}}\mathbb{X}^\dagger \right) \right] \\ &= \frac{1}{n}\mathbb{E} \left[\log \det \left(\mathbf{I}_n + \frac{1}{\sigma^2} \mathbf{K}_{\mathbf{H}\mathbf{H}}\mathbb{X}^\dagger \mathbb{X} \right) \right] \\ &\leq \frac{1}{n} \log \det \left(\mathbf{I}_n + \frac{\mathbb{E}[|X_1|^2]}{\sigma^2} \mathbf{K}_{\mathbf{H}\mathbf{H}} \right) \\ &= \frac{1}{n} \sum_{k=1}^n \log \left(1 + \frac{\mathbb{E}[|X_1|^2]}{\sigma^2} \lambda_k \right) \\ &\leq \frac{1}{n} \sum_{k=1}^n \log(1 + \text{SNR} \lambda_k) \end{aligned} \quad (11)$$

where \mathbb{X} is a random diagonal matrix with diagonal entries X_1, \dots, X_n , and where $\lambda_1, \dots, \lambda_n$ denote the eigenvalues of $\mathbf{K}_{\mathbf{H}\mathbf{H}}$. Here the third step follows from the identity $\det(\mathbf{I}_n + \mathbf{A}\mathbf{B}) = \det(\mathbf{I}_n + \mathbf{B}\mathbf{A})$; the fourth step follows from Jensen's inequality and by noting that $\{X_k, k \in \mathbb{Z}\}$ is i.i.d., so $\mathbb{E}[\mathbb{X}^\dagger \mathbb{X}] = \mathbb{E}[|X_1|^2] \mathbf{I}_n$; the fifth step follows because the determinant of a matrix is given by the product of its eigenvalues; and the last step follows because, by (3), we have $\mathbb{E}[|X_1|^2] \leq A^2$.

To evaluate the RHS of (11) in the limit as n tends to infinity, we apply Szegő's Theorem on the asymptotic behavior of the eigenvalues of Hermitian Toeplitz matrices [8] (see also [9, Th. 2.7.13]). We obtain

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n}I(H_1^n; Y_1^n | X_1^n) &\leq \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \log(1 + \text{SNR} \lambda_k) \\ &= \int_{-1/2}^{1/2} \log(1 + \text{SNR} F'(\lambda)) d\lambda. \end{aligned} \quad (12)$$

Combining (8), (9), (10), and (12) yields the final lower bound

$$\begin{aligned}
 C(\text{SNR}) &\geq \Pr[|H_1| \geq \Upsilon] \log \text{SNR} \\
 &\quad - \Pr[|H_1| \geq \Upsilon] (1 - \log \Upsilon^2) \\
 &\quad - \int_{-1/2}^{1/2} \log(1 + \text{SNR} F'(\lambda)) d\lambda \quad (13)
 \end{aligned}$$

which holds for any fixed $\Upsilon > 0$. Note that this lower bound applies to all mean- d , unit-variance, stationary, and ergodic fading processes $\{H_k, k \in \mathbb{Z}\}$ with spectral distribution function $F(\cdot)$, irrespective of whether $\Pr[H_k = 0]$ is zero or not.

2) *Asymptotic Analysis:* In the following, we prove Theorem 1 by computing the limiting ratio of the lower bound (13) to $\log \text{SNR}$ as SNR tends to infinity.

We first show that

$$\lim_{\text{SNR} \rightarrow \infty} \int_{-1/2}^{1/2} \frac{\log(1 + \text{SNR} F'(\lambda))}{\log \text{SNR}} d\lambda = \mu(\{\lambda: F'(\lambda) > 0\}). \quad (14)$$

To this end, we divide the integral into three parts, depending on whether λ takes part in the set \mathcal{S}_1 , \mathcal{S}_2 , or \mathcal{S}_3 , where

$$\begin{aligned}
 \mathcal{S}_1 &\triangleq \{\lambda \in [-1/2, 1/2]: F'(\lambda) = 0\} \\
 \mathcal{S}_2 &\triangleq \{\lambda \in [-1/2, 1/2]: F'(\lambda) \geq 1\} \\
 \mathcal{S}_3 &\triangleq \{\lambda \in [-1/2, 1/2]: 0 < F'(\lambda) < 1\}.
 \end{aligned}$$

For $\lambda \in \mathcal{S}_1$ the integrand is zero and, hence

$$\lim_{\text{SNR} \rightarrow \infty} \int_{\mathcal{S}_1} \frac{\log(1 + \text{SNR} F'(\lambda))}{\log \text{SNR}} d\lambda = 0. \quad (15)$$

For $\lambda \in \mathcal{S}_2$, i.e., when $F'(\lambda) \geq 1$, we note that, for $\text{SNR} > 1$, the function

$$\text{SNR} \mapsto \frac{\log(1 + \text{SNR} F'(\lambda))}{\log \text{SNR}}$$

is monotonically decreasing in SNR . Therefore, applying the Monotone Convergence Theorem [10, Th. 1.26], we have

$$\begin{aligned}
 &\lim_{\text{SNR} \rightarrow \infty} \int_{\mathcal{S}_2} \frac{\log(1 + \text{SNR} F'(\lambda))}{\log \text{SNR}} d\lambda \\
 &= \int_{\mathcal{S}_2} \lim_{\text{SNR} \rightarrow \infty} \frac{\log(1 + \text{SNR} F'(\lambda))}{\log \text{SNR}} d\lambda \\
 &= \mu(\mathcal{S}_2) \\
 &= \mu(\{\lambda: F'(\lambda) \geq 1\}). \quad (16)
 \end{aligned}$$

For $\lambda \in \mathcal{S}_3$, i.e., when $0 < F'(\lambda) < 1$, we have for $\text{SNR} \geq e$

$$0 < \frac{\log(1 + \text{SNR} F'(\lambda))}{\log \text{SNR}} < \frac{\log(1 + \text{SNR})}{\log \text{SNR}} \leq \log(1 + e)$$

where the last step follows because, for $\text{SNR} > 1$, the function

$$\text{SNR} \mapsto \frac{\log(1 + \text{SNR})}{\log \text{SNR}}$$

is monotonically decreasing in SNR . Since $\log(1 + e)$ is integrable over \mathcal{S}_3 , we can apply the Dominated Convergence Theorem [10, Th. 1.34] to obtain

$$\begin{aligned}
 &\lim_{\text{SNR} \rightarrow \infty} \int_{\mathcal{S}_3} \frac{\log(1 + \text{SNR} F'(\lambda))}{\log \text{SNR}} d\lambda \\
 &= \int_{\mathcal{S}_3} \lim_{\text{SNR} \rightarrow \infty} \frac{\log(1 + \text{SNR} F'(\lambda))}{\log \text{SNR}} d\lambda \\
 &= \mu(\mathcal{S}_3) \\
 &= \mu(\{\lambda: 0 < F'(\lambda) < 1\}). \quad (17)
 \end{aligned}$$

Adding (15), (16), and (17) yields (14).

To continue with the asymptotic analysis of (13) we note that by (14)

$$\begin{aligned}
 \Pi &\triangleq \overline{\lim}_{\text{SNR} \rightarrow \infty} \frac{C(\text{SNR})}{\log \text{SNR}} \\
 &\geq \Pr[|H_1| \geq \Upsilon] - \mu(\{\lambda: F'(\lambda) > 0\}) \\
 &= \mu(\{\lambda: F'(\lambda) = 0\}) - \Pr[|H_1| < \Upsilon] \quad (18)
 \end{aligned}$$

for any $\Upsilon > 0$. If the law of the fading process has no mass point at zero, then

$$\lim_{\Upsilon \downarrow 0} \Pr[|H_1| < \Upsilon] = 0$$

and Theorem 1 follows, thus, from (18) by letting Υ tend to zero from above.

B. Proof of Note 2

We prove Note 2 by demonstrating that there exists a stationary and ergodic fading process of some spectral distribution function $F(\cdot)$ for which

$$\Pi < \mu(\{\lambda: F'(\lambda) = 0\}).$$

By Theorem 1, the law of such a process must have a mass point at zero, i.e.

$$\Pr[H_k = 0] > 0, \quad k \in \mathbb{Z}.$$

We first show that the capacity pre-log is upper bounded by

$$\Pi \leq \Pr[|H_1| > 0]. \quad (19)$$

Indeed, the capacity $C(\text{SNR})$ does not decrease when the receiver additionally knows the realization of $\{H_k, k \in \mathbb{Z}\}$, and when the inputs have to satisfy an average-power constraint rather than a peak-power constraint, i.e.

$$C(\text{SNR}) \leq \lim_{n \rightarrow \infty} \frac{1}{n} \sup I(X_1^n; Y_1^n | H_1^n) \quad (20)$$

where the maximization is over all input distributions on X_1, \dots, X_n satisfying the average-power constraint

$$\frac{1}{n} \sum_{k=1}^n \frac{\mathbb{E}[|X_k|^2]}{\sigma^2} \leq \text{SNR}. \quad (21)$$

(This follows because the availability of additional information cannot decrease the capacity, and because any distribution on

the inputs satisfying the peak-power constraint (3) satisfies also (21.) It is well known that the expression on the RHS of (20) is equal to

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sup I(X_1^n; Y_1^n | H_1^n) = \mathbb{E}[\log(1 + |H_1|^2 \text{SNR})]$$

(see, e.g., [11, eq. (3.3.10)]), which can be further upper bounded by

$$\begin{aligned} & \mathbb{E}[\log(1 + |H_1|^2 \text{SNR})] \\ &= \Pr[|H_1| > 0] \mathbb{E}[\log(1 + |H_1|^2 \text{SNR}) | |H_1| > 0] \\ &\leq \Pr[|H_1| > 0] \log(1 + \mathbb{E}[|H_1|^2 | |H_1| > 0] \text{SNR}) \\ &= \Pr[|H_1| > 0] \log\left(1 + \frac{\text{SNR}}{\Pr[|H_1| > 0]}\right). \end{aligned} \quad (22)$$

Here the first step follows by writing the expectation as

$$\begin{aligned} & \mathbb{E}[\log(1 + |H_1|^2 \text{SNR})] \\ &= \Pr[|H_1| = 0] \mathbb{E}[\log(1 + |H_1|^2 \text{SNR}) | |H_1| = 0] \\ &\quad + \Pr[|H_1| > 0] \mathbb{E}[\log(1 + |H_1|^2 \text{SNR}) | |H_1| > 0] \end{aligned}$$

and by noting that $\mathbb{E}[\log(1 + |H_1|^2 \text{SNR}) | |H_1| = 0] = 0$; the second step follows from Jensen's inequality; and the last step follows because $\mathbb{E}[|H_1|^2] = 1$, which implies that

$$\mathbb{E}[|H_1|^2 | |H_1| > 0] = \frac{1}{\Pr[|H_1| > 0]}.$$

Dividing the RHS of (22) by $\log \text{SNR}$ and computing the limit as SNR tends to infinity yields (19).

In view of (19), it suffices to demonstrate that there exists a fading process of some spectral distribution function $F(\cdot)$ that satisfies

$$\Pr[|H_1| > 0] < \mu(\{\lambda : F'(\lambda) = 0\}). \quad (23)$$

A first attempt of defining such a process (which, alas, does not work) is

$$\dots, H_{-1}, H_0, H_1, \dots = \begin{cases} \dots, 0, 0, 0, \dots & \text{w.p. } \delta \\ \dots, B_{-1}, B_0, B_1, \dots & \text{w.p. } 1 - \delta \end{cases}$$

("w.p." stands for "with probability"), where $\{B_k, k \in \mathbb{Z}\}$ is a zero-mean, circularly symmetric, stationary and ergodic, complex Gaussian process of variance $1/(1 - \delta)$ and of spectral distribution function $G(\cdot)$; and δ and $G(\cdot)$ are chosen so that

$$1 - \delta < \mu(\{\lambda : G'(\lambda) = 0\}).$$

This process satisfies (23) because $\Pr[|H_1| > 0] = 1 - \delta$, and because

$$\begin{aligned} \mathbb{E}[(H_{k+m} - d)(H_k - d)^*] &= (1 - \delta) \mathbb{E}[B_{k+m} B_k^*], \\ &\quad (k \in \mathbb{Z}, m \in \mathbb{Z}) \end{aligned}$$

which implies that $F(\lambda) = (1 - \delta)G(\lambda)$ almost everywhere, so

$$\mu(\{\lambda : F'(\lambda) = 0\}) = \mu(\{\lambda : G'(\lambda) = 0\}).$$

Alas, the above fading process is stationary but not ergodic.

In the following, we exhibit a fading process that is stationary and ergodic and that satisfies (23). Let

$$\dots, A_{-1}, A_0, A_1, A_2, \dots = \begin{cases} \dots, 0, 1, 0, 1, \dots & \text{w.p. } \frac{1}{2} \\ \dots, 1, 0, 1, 0, \dots & \text{w.p. } \frac{1}{2} \end{cases}$$

and let $\{B_k, k \in \mathbb{Z}\}$ be a zero-mean, variance-2, circularly symmetric, stationary, ergodic, complex Gaussian process of spectral distribution function $G(\cdot)$. Furthermore, let $\{A_k, k \in \mathbb{Z}\}$ and $\{B_k, k \in \mathbb{Z}\}$ be independent of each other. Finally, let the fading process be given by

$$H_k = A_k B_k, \quad k \in \mathbb{Z}.$$

Note that $\{H_k, k \in \mathbb{Z}\}$ is of mean zero, and its law satisfies

$$\Pr[|H_k| > 0] = \Pr[A_k = 1] = \frac{1}{2}, \quad k \in \mathbb{Z}. \quad (24)$$

We first argue that $\{H_k, k \in \mathbb{Z}\}$ is stationary and ergodic. Indeed, $\{A_k, k \in \mathbb{Z}\}$ is stationary and ergodic. And since a Gaussian process is ergodic if, and only if, it is weakly mixing (see, e.g., [12, Sec. II]), we have that $\{B_k, k \in \mathbb{Z}\}$ is stationary and weakly mixing. (See [13, Sec. 2.6] for a definition of weakly mixing stochastic processes.) It thus follows from [14, Prop. 1.6] that the process $\{(A_k, B_k), k \in \mathbb{Z}\}$ is jointly stationary and ergodic, which implies that

$$\{H_k, k \in \mathbb{Z}\} = \{A_k B_k, k \in \mathbb{Z}\}$$

is stationary and ergodic.

We next demonstrate that $G(\cdot)$ can be chosen so that $\{H_k, k \in \mathbb{Z}\}$ satisfies (23). We choose

$$G'(\lambda) = \begin{cases} \frac{1}{W}, & \text{if } |\lambda| \leq W \\ 0, & \text{otherwise} \end{cases}$$

for some $W \in (0, 1/8)$, which corresponds to the autocovariance function

$$\mathbb{E}[B_{k+m} B_k^*] = 2\text{sinc}(2Wm), \quad (k \in \mathbb{Z}, m \in \mathbb{Z}).$$

Here $\text{sinc}(\cdot)$ denotes the sinc-function, i.e., $\text{sinc}(0) = 1$ and $\text{sinc}(x) = \sin(\pi x)/(\pi x)$ for $|x| > 0$. Using that

$$\mathbb{E}[A_{k+m} A_k^*] = \frac{1}{2} \mathbb{I}\{m \text{ is even}\}, \quad (k \in \mathbb{Z}, m \in \mathbb{Z})$$

(where $\mathbb{I}\{\text{statement}\}$ is 1 if the statement is true and 0 otherwise), we have

$$\begin{aligned} & \mathbb{E}[H_{k+m} H_k^*] \\ &= \mathbb{E}[A_{k+m} B_{k+m} A_k^* B_k^*] \\ &= \mathbb{E}[A_{k+m} A_k^*] \mathbb{E}[B_{k+m} B_k^*] \\ &= \mathbb{I}\{m \text{ is even}\} \text{sinc}(2Wm), \quad (k \in \mathbb{Z}, m \in \mathbb{Z}) \end{aligned}$$

and the corresponding spectrum is given by

$$F'(\lambda) = \begin{cases} \frac{1}{4W}, & \text{if } (|\lambda| \leq W) \text{ or } (\frac{1}{2} - W \leq |\lambda| \leq \frac{1}{2}) \\ 0, & \text{otherwise.} \end{cases}$$

We, thus, have

$$\mu(\{\lambda: F'(\lambda) = 0\}) = 1 - 4W \quad (25)$$

and it follows from (24) and (25) that

$$\Pr[|H_k| > 0] < \mu(\{\lambda: F'(\lambda) = 0\}), \quad \text{for } W < \frac{1}{8}.$$

Thus, there exist stationary and ergodic fading processes of some spectral distribution function that give rise to a capacity pre-log that is strictly smaller than the pre-log of a Gaussian fading channel of equal spectral distribution function.

C. Proof of Note 3

To prove Note 3, we first notice that, since the phase noise is memoryless, the derivative of the spectral distribution function is

$$F'(\lambda) = 1, \quad -\frac{1}{2} \leq \lambda \leq \frac{1}{2}.$$

Hence the capacity pre-log of the Gaussian fading channel of spectral distribution function $F(\cdot)$ equals

$$\Pi_G = \mu(\{\lambda: F'(\lambda) = 0\}) = 0.$$

It remains to show that the pre-log of the phase-noise channel with memoryless phase noise is equal to

$$\Pi = \frac{1}{2}. \quad (26)$$

In [15] it was shown that at high SNR the capacity of the phase-noise channel under the average-power constraint (21) is given by

$$C_{\text{Avg}}(\text{SNR}) = \frac{1}{2} \log \left(1 + \frac{\text{SNR}}{2} \right) + o(1) \quad (27)$$

where $o(1)$ tends to zero as SNR tends to infinity. (The subscript ‘‘Avg’’ indicates that the inputs satisfy an average-power constraint and not a peak-power constraint.) Since any distribution on the inputs satisfying the peak-power constraint (3) satisfies also the average-power constraint (21), it follows that $C(\text{SNR}) \leq C_{\text{Avg}}(\text{SNR})$ and hence

$$\Pi \leq \frac{1}{2}.$$

To prove (26) it thus suffices to show that $\Pi \geq \frac{1}{2}$. To this end, we first note that, since the phase noise is memoryless, we have

$$C(\text{SNR}) = \sup I(X_1; Y_1) \quad (28)$$

where the maximization is over all distributions on X_1 satisfying with probability one

$$|X_1| \leq A.$$

We derive a lower bound on $C(\text{SNR})$ by evaluating the RHS of (28) for X_1 being a zero-mean, circularly symmetric, complex random variable with $|X_1|^2$ uniformly distributed over the interval $[0, A^2]$. We have

$$\begin{aligned} I(X_1; Y_1) &\geq I(X_1; |Y_1|^2) \\ &= h(|Y_1|^2) - h(|Y_1|^2 | X_1) \\ &\geq h(|X_1|^2) - h(|Y_1|^2 | X_1) \end{aligned} \quad (29)$$

where the first step follows from the data processing inequality [4, Th. 2.8.1]; and the last step follows by the circular symmetry of X_1 [15, p. 3, after (20)].

Computing the differential entropy of a uniformly distributed random variable, the first term on the RHS of (29) becomes

$$h(|X_1|^2) = \log A^2. \quad (30)$$

As to the second term, we note that, for a given $X_1 = x_1$, the random variable $2/\sigma^2 |Y_1|^2$ has a noncentral chi-square distribution with noncentrality parameter $2/\sigma^2 |x_1|^2$ and two degrees of freedom. Its differential entropy can be upper bounded by [15, (8)]

$$\begin{aligned} h(|Y_1|^2 | X_1) &\leq \frac{1}{2} \mathbb{E} \left[\log \left(4\pi e \left(2 + 2 \frac{2}{\sigma^2} |X_1|^2 \right) \right) \right] - \log \frac{2}{\sigma^2} \\ &\leq \frac{1}{2} \log \left(4\pi e \left(2 + 2 \frac{2}{\sigma^2} A^2 \right) \right) - \log \frac{2}{\sigma^2} \end{aligned} \quad (31)$$

where the last step follows because $|X_1| \leq A$ with probability one. Combining (30) and (31) with (29) yields thus

$$I(X_1; Y_1) \geq \frac{1}{2} \log \text{SNR} + O(1) \quad (32)$$

where

$$\lim_{\text{SNR} \rightarrow \infty} |O(1)| < \infty.$$

We finally obtain the lower bound

$$\Pi \geq \frac{1}{2}$$

upon dividing the RHS of (32) by $\log \text{SNR}$ and letting SNR tend to infinity.

VI. EXTENSION TO MISO FADING CHANNELS

Theorem 1 can be extended to multiple-input single-output (MISO) fading channels with memory, when the fading processes corresponding to the different transmit antennas are independent. For such channels, the channel output $Y_k \in \mathbb{C}$ at time $k \in \mathbb{Z}$ corresponding to the channel input $\mathbf{x}_k \in \mathbb{C}^{n_T}$ (where n_T stands for the number of antennas at the transmitter) is given by

$$Y_k = \mathbf{H}_k^T \mathbf{x}_k + Z_k, \quad k \in \mathbb{Z} \quad (33)$$

where $\mathbf{H}_k = (H_k^{(1)}, \dots, H_k^{(n_T)})^\top$, and where the processes

$$\{H_k^{(1)}, k \in \mathbb{Z}\}, \{H_k^{(2)}, k \in \mathbb{Z}\}, \dots, \{H_k^{(n_T)}, k \in \mathbb{Z}\}$$

are jointly stationary ergodic and independent. We assume that for each $t = 1, \dots, n_T$ the process $\{H_k^{(t)}, k \in \mathbb{Z}\}$ is of mean d_t , of unit variance, and of spectral distribution function $F_t(\cdot)$. We further assume that for $k \in \mathbb{Z}$

$$\begin{aligned} \Pr [H_k^{(1)} = 0] &= \Pr [H_k^{(2)} = 0] = \dots \\ &= \Pr [H_k^{(n_T)} = 0] = 0. \end{aligned}$$

The additive noise $\{Z_k, k \in \mathbb{Z}\}$ is defined as in Section II.

The capacity of this channel is given by (2), but with X_1^n replaced by \mathbf{X}_1^n , and with the peak-power constraint (3) altered accordingly

$$\|\mathbf{X}_k\| \leq A \quad \text{with probability one, } k \in \mathbb{Z}$$

where $\|\mathbf{a}\|$ denotes the Euclidean norm of the vector \mathbf{a} , i.e.

$$\|\mathbf{a}\| = \sqrt{\sum_{\ell=1}^L |a_\ell|^2}, \quad \mathbf{a} = (a_1, \dots, a_L)^\top.$$

The pre-log of MISO fading channels is defined as in the single-antenna case (4). For Gaussian fading, i.e., when $\{H^{(t)} - d_t, k \in \mathbb{Z}\}$, $1 \leq t \leq n_T$ are circularly symmetric, complex Gaussian processes, the pre-log is given by [16, Cor. 13]

$$\Pi_G = \max_{1 \leq t \leq n_T} \mu(\{\lambda: F_t'(\lambda) = 0\}). \quad (34)$$

(See [17, Sec. 4.6] for a proof of this results.)

Proving that the capacity pre-log Π of MISO fading channels is lower bounded by the pre-log of the MISO Gaussian fading channel of equal spectral distribution functions—namely $F_1(\cdot), \dots, F_{n_T}(\cdot)$ —is straightforward. Let Π_t , $1 \leq t \leq n_T$ denote the capacity pre-log of a single-antenna fading channel with fading process $\{H_k^{(t)}, k \in \mathbb{Z}\}$, and let

$$t_* = \arg \max_{1 \leq t \leq n_T} \Pi_t.$$

By signaling only from antenna t_* while keeping the other antennas silent, we achieve the pre-log Π_{t_*} , so

$$\Pi \geq \max_{1 \leq t \leq n_T} \Pi_t. \quad (35)$$

Theorem 1 yields then

$$\Pi_t \geq \mu(\{\lambda: F_t'(\lambda) = 0\}), \quad 1 \leq t \leq n_T$$

which together with (35) proves the claim

$$\Pi \geq \max_{1 \leq t \leq n_T} \mu(\{\lambda: F_t'(\lambda) = 0\}). \quad (36)$$

VII. SUMMARY AND DISCUSSION

We showed that, among all stationary and ergodic fading processes of a given spectral distribution function and whose law has no mass point at zero, the Gaussian process gives rise to the smallest capacity pre-log. We further showed that if the fading law is allowed to have a mass point at zero, then the above statement is not necessarily true anymore. Roughly speaking, we can say that for a large class of fading processes the Gaussian process is the worst. This demonstrates the robustness of the Gaussian assumption in the analysis of fading channels at high SNR.

To give an intuition why Gaussian processes give rise to the smallest pre-log, we recall that for Gaussian fading [1, (33) and (47)]

$$C(\text{SNR}) = \log \frac{1}{\epsilon_{\text{pred}}^2(1/\text{SNR})} + o(\log \text{SNR})$$

where

$$\lim_{\text{SNR} \rightarrow \infty} \frac{o(\log \text{SNR})}{\log \text{SNR}} = 0$$

and where $\epsilon_{\text{pred}}^2(\delta)$ denotes the mean-square error in predicting the present fading H_0 from a variance- δ noisy observation of its past $H_{-1} + W_{-1}, H_{-2} + W_{-2}, \dots$ (with $\{W_k, k \in \mathbb{Z}\}$ being a sequence of i.i.d., zero-mean, variance- δ , circularly symmetric, complex Gaussian random variables). Thus, for Gaussian fading the capacity pre-log is determined by $\epsilon_{\text{pred}}^2(1/\text{SNR})$, and it is plausible that also the pre-log of non-Gaussian fading channels is connected with the ability of predicting the present fading from a noisy observation of its past. Since, among all stationary and ergodic processes of a given spectral distribution function, the Gaussian process is hardest to predict, it is, therefore, plausible that the Gaussian process gives rise to the smallest pre-log.

ACKNOWLEDGMENT

The authors wish to thank E. Telatar for his valuable comments.

REFERENCES

- [1] A. Lapidoth, "On the asymptotic capacity of stationary Gaussian fading channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 437–446, Feb. 2005.
- [2] A. Lapidoth, "On the high SNR capacity of stationary Gaussian fading channels," in *Proc. 41st Allerton Conf. Commun., Contr. Comput.*, Monticello, IL, Oct. 1–3, 2003.
- [3] J. Doob, *Stochastic Processes*. New York: Wiley, 1990.
- [4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 1st ed. New York: Wiley, 1991.
- [5] Y.-H. Kim, "A coding theorem for a class of stationary channels with feedback," *IEEE Trans. Inf. Theory*, vol. 54, no. 4, pp. 1488–1499, Apr. 2008.
- [6] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat fading channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2426–2467, Oct. 2003.
- [7] X. Deng and A. M. Haimovich, "Information rates of time varying Rayleigh fading channels," in *Proc. ICC*, Paris, France, Jun. 20–24, 2004.
- [8] U. Grenander and G. Szegő, *Toeplitz Forms and Their Applications*. Los Angeles: Univ. California Press, 1958.

- [9] B. Simon, *Orthogonal Polynomials on the Unit Circle. Part 1: Classical Theory*. New York: Amer. Math. Soc., 2005.
- [10] W. Rudin, *Real and Complex Analysis*, 3rd ed. New York: McGraw-Hill, 1987.
- [11] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.
- [12] V. Sethuraman and B. Hajek, "Capacity per unit energy of fading channels with a peak constraint," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3102–3120, Sep. 2005.
- [13] K. Petersen, *Ergodic Theory*, ser. Cambridge Studies in Advanced Mathematics 2. Cambridge, U.K.: Cambridge Univ. Press, 1983.
- [14] J. R. Brown, *Ergodic Theory and Topological Dynamics*. New York: Academic, 1976.
- [15] A. Lapidoth, "On phase noise channels at high SNR," in *Proc. Inf. Theory Workshop (ITW)*, Bangalore, India, Oct. 20–25, 2002.
- [16] 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. 4–9, 2005.
- [17] T. Koch, "On heating up and fading in communication channels," Ph.D. dissertation, Signal and Inf. Proc. Lab., ETH Zurich, Zurich, Switzerland, Jul. 2009.

Tobias Koch (S'02–M'09) received the M.Sc. degree in electrical engineering (with distinction) and the Ph.D. degree in electrical engineering from ETH Zurich, Switzerland, in 2004 and 2009, respectively.

His research interests include digital communication theory and information theory with particular focus on wireless communication and communication in electronic circuits.

Amos Lapidoth (S'89–M'95–SM'00–F'04) received the B.A. degree in mathematics (*summa cum laude*) in 1986, the B.Sc. degree in electrical engineering (*summa cum laude*) in 1986, and the M.Sc. degree in electrical engineering, in 1990, all from the Technion—Israel Institute of Technology. He received the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1995.

During 1995–1999, he was an Assistant and Associate Professor with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, and was the KDD Career Development Associate Professor in Communications and Technology. He is now a Professor of Information Theory at ETH Zurich, Switzerland. He is the author of the book *A Foundation in Digital Communication* (Cambridge, U.K.: Cambridge University Press, 2009). His research interests are in digital communications and information theory.

Dr. Lapidoth served during 2003–2004 and 2009 as an Associate Editor for Shannon Theory for the IEEE TRANSACTIONS ON INFORMATION THEORY.