Convergence of the Conditional Per Symbol Entropy for Stationary Gaussian Fading Channels for Almost All Input Sequences

Meik Dörpinghaus*, Norbert Gaffke†, Lorens A. Imhof‡, and Rudolf Mathar*

*Inst. for Theoretical Information Technology, RWTH Aachen University, 52056 Aachen, Germany, {doerpinghaus, mathar}@ti.rwth-aachen.de

†Faculty of Mathematics, University of Magdeburg 39016 Magdeburg, Germany, norbert.gaffke@mathematik.uni-magdeburg.de

‡Department of Statistics and the Hausdorff Center for Mathematics, Bonn University, 53113 Bonn, Germany, limhof@uni-bonn.de

***Abstract*—In the present work, a discrete-time stationary Rayleigh flat-fading channel with unknown channel state information at the transmitter and the receiver is considered. The law of the channel is assumed to be known at the receiver and the fading process is supposed to be a stationary Gaussian process with an absolutely summable autocorrelation function. The conditional per symbol entropy of the output given the input is shown to converge to a constant for almost every realization of i.i.d. input variables. This implies the existence of the corresponding conditional entropy rate. Moreover, a novel inequality yielding a lower bound for the rate is derived.**

I. Motivation and Setup

We consider a stationary Rayleigh flat-fading channel where the channel state information is unknown at the transmitter and the receiver. Moreover, the law of the channel is assumed to be known at the receiver. Often, this channel is referred to as *noncoherent fading channel*. As this scenario corresponds to the basic model of nearly all realistic mobile communication systems it is particularly important. Nevertheless, determining the capacity of this channel turns out to be notoriously difficult and the problem is still open in general.

There have been already several attempts to approximate the capacity of noncoherent fading channels by bounds, see, e.g., [1], [2]. For the case of i.i.d. zero-mean proper Gaussian input symbols, which are capacity-achieving in the coherent setup, in [3] bounds on the achievable rate have been derived. One of the hardest problems when studying the capacity or achievable rate of stationary Rayleigh fading channels is the evaluation of the conditional entropy rate of the channel fading process. This is the main topic of the present paper. The proofs and extended material are contained in technical report [4].

A. Channel Model

We consider an ergodic discrete-time jointly proper Gaussian [5] flat-fading channel, whose output at time k is given by

$$y_k = h_k \cdot x_k + n_k$$

This work was partially supported by the Deutsche Forschungsgemeinschaft under Grant DO1568/1-1 and by UMIC at RWTH Aachen University.

where $x_k \in \mathbb{C}$ is the complex-valued channel input, $h_k \in \mathbb{C}$ represents the channel fading coefficient, and $n_k \in \mathbb{C}$ is additive Gaussian noise. The discrete-time processes $\{h_k\}$, $\{x_k\}$, and $\{n_k\}$ are assumed to be jointly independent.

We assume that the noise $\{n_k\}$ is a sequence of i.i.d. proper Gaussian random variables of zero-mean and variance $\sigma_n^2 > 0$. The stationary channel fading process $\{h_k\}$ is zero-mean jointly proper Gaussian with variance σ_h^2 and covariance function

$$r_\ell = E[h_{k+\ell} \cdot h_k^*].$$

For technical reasons we confine ourselves to absolutely summable covariance functions, i.e., $\sum_{\ell=-\infty}^{\infty} |r_\ell| < \infty$. The normalized PSD of the channel fading process is defined by

$$S_h(f) = \sum_{\ell=-\infty}^{\infty} r_\ell e^{-j2\pi\ell f}, \quad |f| < \frac{1}{2},$$

where $j = \sqrt{-1}$. Here, the frequency f is normalized with respect to the symbol duration.

We use the following matrix-vector notation of the system model:

$$\mathbf{y}_N = \mathbf{X}_N \mathbf{h}_N + \mathbf{n}_N$$

where the vector $\mathbf{h}_N$ is given by $\mathbf{h}_N = [h_1, \ldots, h_N]^T$. The vectors $\mathbf{y}_N$ and $\mathbf{n}_N$ are defined analogously. The matrix $\mathbf{X}_N$ is diagonal and defined as $\mathbf{X}_N = \text{diag}(\mathbf{x}_N)$ with $\mathbf{x}_N = [x_1, \ldots, x_N]^T$. Here, $\text{diag}(\cdot)$ denotes the diagonal matrix whose entries are given by the components of the argument vector. The quantity N is the number of considered symbols. In this paper, we will investigate the limit as $N \to \infty$.

The temporal correlation of the fading process is now expressed by the covariance matrix

$$\mathbf{R}_N = E[\mathbf{h}_N \mathbf{h}_N^H]$$

which is an $N \times N$ nonnegative definite Toeplitz matrix.

We assume that the input variables $x_1, x_2, \ldots$ are i.i.d. and that $E[|x_1|^4] < \infty$. This implies that the x_k satisfy an average power constraint of the form

$$E\left[|x_k|^2\right] \le \sigma_x^2, \quad k = 1, 2, \ldots. \tag{1}$$

B. Mutual Information Rate

The mutual information rate between the channel input and its output, i.e., the achievable rate for a given input distribution is given by

$$\begin{aligned}\mathcal{I}'(\mathbf{y};\mathbf{x}) &= \lim_{N\to\infty} \frac{1}{N}\mathcal{I}(\mathbf{y}_N;\mathbf{x}_N)\\ &= \lim_{N\to\infty} \frac{1}{N}\left\{h(\mathbf{y}_N) - h(\mathbf{y}_N|\mathbf{x}_N)\right\}\\ &= h'(\mathbf{y}) - h'(\mathbf{y}|\mathbf{x}),\end{aligned} \tag{2}$$

provided the limits exist. Here $h(\cdot)$ and $h'(\cdot)$ are the differential entropy and the differential entropy rate, respectively. Moreover, $\mathbf{x} = \{x_k\}_{k\in\mathbb{N}}$ denotes the stochastic process of input variables. The process $\mathbf{y}$ is defined analogously. Stationarity of $\mathbf{y}$ ensures that the limit $h'(\mathbf{y})$ exists.

The focus of the present paper is on the second term of the RHS of (2), i.e., on $h'(\mathbf{y}|\mathbf{x})$. Consider a sequence of deterministic input symbols $\xi_1, \xi_2, \ldots \in \mathbb{C}$. Let $\boldsymbol{\xi}_N = [\xi_1, \ldots, \xi_N]^T$ and $\boldsymbol{\Xi}_N = \operatorname{diag}(\boldsymbol{\xi}_N)$. Conditional on $\mathbf{x}_N = \boldsymbol{\xi}_N$, $\mathbf{y}_N$ is proper Gaussian with zero mean and covariance matrix

$$E\left[\mathbf{y}_N\mathbf{y}_N^H|\mathbf{x}_N = \boldsymbol{\xi}_N\right] = \boldsymbol{\Xi}_N\mathbf{R}_N\boldsymbol{\Xi}_N^H + \sigma_n^2\mathbf{I}_N$$

where $\mathbf{I}_N$ is the identity matrix of size $N\times N$. The conditional entropy of $\mathbf{y}_N$ given $\mathbf{x}_N = \boldsymbol{\xi}_N$ is therefore given by

$$\begin{aligned}h(\mathbf{y}_N|\mathbf{x}_N = \boldsymbol{\xi}_N) &= \log\det\left(\pi e\left(\boldsymbol{\Xi}_N\mathbf{R}_N\boldsymbol{\Xi}_N^H + \sigma_n^2\mathbf{I}_N\right)\right)\\ &= \log\det\left(\pi e\left(\boldsymbol{\Xi}_N^H\boldsymbol{\Xi}_N\mathbf{R}_N + \sigma_n^2\mathbf{I}_N\right)\right).\end{aligned} \tag{3}$$

Hence, the conditional entropy of $\mathbf{y}_N$ given $\mathbf{x}_N$ is

$$h(\mathbf{y}_N|\mathbf{x}_N) = E\left[\log\det\left(\pi e\left(\mathbf{Z}_N\mathbf{R}_N + \sigma_n^2\mathbf{I}_N\right)\right)\right],$$

where $\mathbf{Z}_N = \mathbf{X}_N^H\mathbf{X}_N$ is the diagonal matrix with entries $z_k = |x_k|^2$, $k = 1, \ldots, N$. By definition, the conditional entropy rate $h'(\mathbf{y}|\mathbf{x})$ reads as

$$\begin{aligned}h'(\mathbf{y}|\mathbf{x}) &= \lim_{N\to\infty}\frac{1}{N}h(\mathbf{y}_N|\mathbf{x}_N)\\ &= \lim_{N\to\infty}\frac{1}{N}E\left[\log\det\left(\pi e\left(\mathbf{Z}_N\mathbf{R}_N + \sigma_n^2\mathbf{I}_N\right)\right)\right]\end{aligned} \tag{4}$$

if the limit exists. It is well-known that, in general, the entropy rate may fail to exist, see [6, pp. 74-75] for a simple example. For the unconditional entropy $h(\mathbf{y}_N) = -E[\log p_N(\mathbf{y}_N)]$, where p_N is the density of $\mathbf{y}_N$, the existence of the limit of $\frac{1}{N}h(\mathbf{y}_N)$ is an immediate consequence of the stationarity of $\mathbf{y}$. However, the fact that for the ergodic process $\mathbf{y}$, $-\frac{1}{N}\log p_N(\mathbf{y}_N)$ converges almost surely to $h'(\mathbf{y})$ is a deep result [7], [8]. The study of the limit of $\frac{1}{N}h(\mathbf{y}_N|\mathbf{x}_N = \boldsymbol{\xi}_N)$ in (3) and the existence of $h'(\mathbf{y}|\mathbf{x})$ in (4) are the main focus of the present paper. Moreover, we present bounds for $h'(\mathbf{y}|\mathbf{x})$.

II. Existence of the Conditional Entropy Rate

Let $z_1, z_2, \ldots$ be nonnegative i.i.d. random variables and $\mathbf{Z}_N = \operatorname{diag}(z_1, \ldots, z_N)$. Let $r_j \in \mathbb{C}$, $j \in \mathbb{Z}$, and let $r_{-j} = r_j^*$ for all j. Let $\mathbf{R}_N = (r_{j-k})_{j,k=1}^N$.

Theorem 1. *Suppose that $E[z_1^2] < \infty$. Suppose $\mathbf{R}_N$ is nonnegative definite for all N and $\sum_{j=0}^{\infty}|r_j| < \infty$. Then there is a constant $c \in [0,\infty)$ such that*

$$\lim_{N\to\infty}\frac{1}{N}\log\det(\mathbf{Z}_N\mathbf{R}_N + \mathbf{I}_N) = c \qquad \textit{a.s.} \tag{5}$$

Outline of the proof: Denoting the eigenvalues of $\mathbf{Z}_N\mathbf{R}_N$ by $\lambda_j^{(N)}$ we show that $\frac{1}{N}\sum_{j=1}^N f(\lambda_j^{(N)})$ converges a.s. to a constant when $f(x) = \log(x+1)$. We first consider the band Toeplitz matrices obtained from the $\mathbf{R}_N$ by replacing r_j with zero for all j with $|j| > \ell$, where ℓ is some fixed integer. For these matrices, we prove the corresponding convergence result for the functions $f(x) = x^s$, $s \in \mathbb{N}$. Letting ℓ become large, we then extend this result to Toeplitz matrices $\mathbf{R}_N$ with $\sum_{j=0}^{\infty}|r_j| < \infty$ under the assumption that the z_j are bounded. This convergence result for the functions x^s can be extended to $\log(x+1)$ by a standard approximation argument. Finally, we apply a truncation argument to weaken the boundedness assumption to the assumption that the z_j have a finite second moment. ∎

From Theorem 1 it follows that there is a constant c' such that $\frac{1}{N}h(\mathbf{y}_N|\mathbf{x}_N = \boldsymbol{\xi}_N)$ converges to c' for $P_{\mathbf{x}}$-almost every input sequence $\{\xi_k\} \in \mathbb{C}^\infty$, where $P_{\mathbf{x}}$ denotes the law of the input process $\mathbf{x}$. Since the sequence given by the argument of the limit in (5) is uniformly integrable, the expectation converges to c' as well, see, e.g., [9, Theorems 16.14 (i), 25.12]. Thus, the conditional entropy rate $h'(\mathbf{y}|\mathbf{x}) = \lim_{N\to\infty}\frac{1}{N}h(\mathbf{y}_N|\mathbf{x}_N)$ exists and is equal to the constant c' yielding the following corollary.

Corollary 1. *If $E[z_1^2] < \infty$, $\mathbf{R}_N$ is nonnegative definite for all N and $\sum_{j=0}^{\infty}|r_j| < \infty$, then there is a constant $c \in [0,\infty)$ such that*

$$\lim_{N\to\infty}\frac{1}{N}E[\log\det(\mathbf{Z}_N\mathbf{R}_N + \mathbf{I}_N)] = c.$$

In particular, the conditional entropy rate $h'(\mathbf{y}|\mathbf{x})$ exists in $\mathbb{R}$ and for $P_{\mathbf{x}}$-almost every input sequence $\{\xi_k\} \in \mathbb{C}^\infty$,

$$\lim_{N\to\infty}\frac{1}{N}h(\mathbf{y}_N|\mathbf{x}_N = \boldsymbol{\xi}_n) = h'(\mathbf{y}|\mathbf{x}).$$

III. A New Inequality for Bounding the Conditional Entropy Rate

In this section, we apply a rearrangement argument to establish a new determinantal inequality, which implies a lower bound for $h(\mathbf{y}_N|\mathbf{x}_N)$. A simple convexity argument yields an upper bound for $h(\mathbf{y}_N|\mathbf{x}_N)$, and these bounds together with Szegö's theorem on the asymptotic eigenvalue distribution of Hermitian Toeplitz matrices then lead to bounds for the rate $h'(\mathbf{y}|\mathbf{x})$ in terms of the PSD $S_h(f)$.

A. Expectation Inequalities for a Determinant

Theorem 2. *Let $\mathbf{Q}$ be a nonnegative definite $N\times N$ matrix with eigenvalues $\gamma_1, \ldots, \gamma_N$ and let $\boldsymbol{\Gamma} = \operatorname{diag}(\gamma_1, \ldots, \gamma_N)$. Let $\mathbf{W}$ be an $N\times N$ diagonal matrix whose diagonal entries*

$w_1, \dots, w_N$ *are identically distributed nonnegative random variables. Then*

$$E[\log\det(\mathbf{WQ}+\mathbf{I})] \ge E[\log\det(\mathbf{W\Gamma}+\mathbf{I})], \qquad (6)$$

and if the entries of $\mathbf{W}$ *are integrable, then*

$$E[\log\det(\mathbf{WQ}+\mathbf{I})] \le \log\det(E[\mathbf{W}]\mathbf{\Gamma}+\mathbf{I}). \qquad (7)$$

Inequality (7) is a direct consequence of Jensen's inequality. The proof of (6) rests on an inequality for rearrangements by Lorentz', see [10].

B. A Muirhead- and a Rado-type Inequality

Two special cases of (6) yield deterministic inequalities of interest in themselves.

Corollary 2. *Let* $\mathbf{C}$ *be a nonnegative definite* $N \times N$ *diagonal matrix, let* $\mathbf{D}$ *be a positive definite* $N \times N$ *diagonal matrix, and let* $\mathbf{U}$ *be a unitary* $N \times N$ *matrix. Let* $\mathcal{P}$ *be either the set of all* $N \times N$ *permutation matrices or the set of the* N *circulant permutation matrices* $(p_{jk}^{(n)})_{j,k=1}^N$, $n = 0, \dots, N-1$, *where* $p_{jk}^{(n)} = 1$ *if* $k - j = n$ *modulo* N. *Then*

$$\sum_{\mathbf{\Pi}\in\mathcal{P}} \log\det(\mathbf{U\Pi C\Pi}^H\mathbf{U}^H + \mathbf{D}) \ge \sum_{\mathbf{\Pi}\in\mathcal{P}} \log\det(\mathbf{\Pi C\Pi}^H + \mathbf{D}).$$

C. Bounds for $h(\mathbf{y}_N|\mathbf{x}_N)$ and $h'(\mathbf{y}|\mathbf{x})$

The almost sure convergence in Section II does not reveal the value of the limit $h'(\mathbf{y}|\mathbf{x})$. However, Theorem 2 yields bounds for $h(\mathbf{y}_N|\mathbf{x}_N) = E[\log\det(\pi e(\mathbf{Z}_N\mathbf{R}_N + \sigma_n^2\mathbf{I}_N))]$. Denote the eigenvalues of $\mathbf{R}_N$ by $\gamma_1^{(N)}, \dots, \gamma_N^{(N)}$. Then

$$\frac{1}{N}\sum_{j=1}^N E[\log(\pi e(z_1\gamma_j^{(N)} + \sigma_n^2))] \le \frac{1}{N}h(\mathbf{y}_N|\mathbf{x}_N) \le \frac{1}{N}\sum_{j=1}^N \log(\pi e(E[z_1]\gamma_j^{(N)} + \sigma_n^2)). \quad (8)$$

By Corollary 1, $\frac{1}{N}h(\mathbf{y}_N|\mathbf{x}_N) \to h'(\mathbf{y}|\mathbf{x})$ as $N \to \infty$. To compute the limits of the bounds in (8) we apply Szegö's theorem on the asymptotic eigenvalue distribution of Toeplitz matrices [11, pp. 64-65] yielding

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} E[\log(\pi e(|x_1|^2 S_h(f) + \sigma_n^2))]\,df \le h'(\mathbf{y}|\mathbf{x}) \le \int_{-\frac{1}{2}}^{\frac{1}{2}} \log(\pi e(E[|x_1|^2]S_h(f) + \sigma_n^2))\,df. \quad (9)$$

The lower bound has also been derived in [3] by a different approach and the upper bound is known from [12].

If we impose, for some fixed $\sigma_x \in [0, \infty)$, the average power constraint (1), then

$$\max_{\mathbf{x}} h'(\mathbf{y}|\mathbf{x}) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \log(\pi e(\sigma_x^2 S_h(f) + \sigma_n^2))\,df,$$

where the maximum is taken over all sequences of i.i.d. input variables x_k subject to (1) and $E[|x_1|^4] < \infty$. The maximum is attained when $P(|x_1| = \sigma_x) = 1$. In this case, all inequalities in (8) and (9) are equalities, showing that these bounds are sharp.

IV. Bounds on the Mutual Information Rate

In order to bound the mutual information rate we use (9) in combination with the following lower and upper bound on $h'(\mathbf{y})$, see [13],

$$\mathcal{I}'(\mathbf{y};\mathbf{x}|\mathbf{h}) + h'(\mathbf{y}|\mathbf{x},\mathbf{h}) \le h'(\mathbf{y}) \le \log\left(\pi e\left(\sigma_x^2\sigma_h^2 + \sigma_n^2\right)\right).$$

Since $h'(\mathbf{y}|\mathbf{x},\mathbf{h}) = \log(\pi e\sigma_n^2)$ we get with (2) the following upper and lower bound on the mutual information rate, see [3] for the upper bound and [12] for the lower bound,

$$\mathcal{I}'(\mathbf{y};\mathbf{x}) \ge \mathcal{I}'(\mathbf{y};\mathbf{x}|\mathbf{h}) - \int_{-\frac{1}{2}}^{\frac{1}{2}} \log\left(\rho\frac{E[|x_1|^2]}{\sigma_x^2}\frac{S_h(f)}{\sigma_h^2} + 1\right)df$$

$$\mathcal{I}'(\mathbf{y};\mathbf{x}) \le \log(\rho + 1) - \int_{-\frac{1}{2}}^{\frac{1}{2}} \mathrm{E}\left[\log\left(\rho\frac{|x_1|^2}{\sigma_x^2}\frac{S_h(f)}{\sigma_h^2} + 1\right)\right]df$$

where $\rho = \sigma_x^2\sigma_h^2/\sigma_n^2$ is the maximum mean SNR. For a discussion of the tightness of these bounds in case of Gaussian inputs see [3].

References

[1] V. Sethuraman, L. Wang, B. Hajek, and A. Lapidoth, "Low-SNR capacity of noncoherent fading channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 4, pp. 1555–1574, Apr. 2009.

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

[3] M. Dörpinghaus, H. Meyr, and R. Mathar, "On the achievable rate of stationary Rayleigh flat-fading channels with Gaussian inputs," *IEEE Trans. Inf. Theory*, accepted [Online]. Available: http://dx.doi.org/10.1109/TIT.2012.2234206.

[4] M. Dörpinghaus, N. Gaffke, L. A. Imhof, and R. Mathar, "Convergence of the conditional per symbol entropy for stationary Gaussian fading channels for almost all input sequences," *Technical Report, Institute for Theoretical Information Technology, RWTH Aachen University, to be submitted*, 2013.

[5] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1293–1302, Jul. 1993.

[6] T. M. Cover and J. A. Thomas, *Elements of Information Theory, 2nd edition*. New York: Wiley & Sons, 2006.

[7] A. R. Barron, "The strong ergodic theorem for densities: generalized Shannon-McMillan-Breiman theorem," *Ann. Probab.*, vol. 13, no. 4, pp. 1292–1303, Nov. 1985.

[8] S. Orey, "On the Shannon-Perez-Moy theorem," *Contemp. Math.*, vol. 41, pp. 319–327, 1985.

[9] P. Billingsley, *Probability and Measure, 3rd edition*. New York: John Wiley & Sons, 1995.

[10] G. G. Lorentz, "An inequality for rearrangements," *Amer. Math. Monthly*, vol. 60, no. 3, pp. 176–179, Mar. 1953.

[11] U. Grenander and G. Szegö, *Toeplitz Forms and Their Applications*. Berkeley, CA, U.S.A.: Univ. Calif. Press, 1958.

[12] X. Deng and A. M. Haimovich, "Information rates of time varying Rayleigh fading channels," in *Proc. IEEE International Conference on Communications (ICC)*, vol. 1, Paris, France, Jun. 2004, pp. 573–577.

[13] M. Dörpinghaus, M. Senst, G. Ascheid, and H. Meyr, "On the achievable rate of stationary Rayleigh flat-fading channels with Gaussian input distribution," in *Proc. International Symposium on Information Theory and its Applications (ISITA)*, Auckland, New Zealand, Dec. 2008.