The Achievable Rate of Stationary Rayleigh Flat-Fading Channels with IID Input Symbols

Meik Dörpinghaus
Institute for Theoretical Information Technology
RWTH Aachen University, 52056 Aachen, Germany
Email: Doerpinghaus@ti.rwth-aachen.de

Heinrich Meyr and Gerd Ascheid
Institute for Integrated Signal Processing Systems
RWTH Aachen University, 52056 Aachen, Germany
Email: {Meyr, Ascheid}@iss.rwth-aachen.de

***Abstract*—In this work, we derive a new upper bound on the achievable rate of stationary Rayleigh flat-fading channels with i.i.d. input symbols. The novelty lies in the fact that this bound is not restricted to peak power constrained input symbols like known bounds, e.g., in [1] or [2]. Therefore, the derived upper bound can also be used to evaluate the achievable rate with i.i.d. proper Gaussian input symbols, which are capacity achieving in the coherent case. The derivation of the upper bound is based on the prediction error variance of the one-step channel predictor.**

I. Introduction

In this paper, we consider a stationary Rayleigh flat-fading channel with temporal correlation. We assume that the channel state information is unknown to the transmitter and the receiver, while the receiver is aware of the channel law. The capacity of this scenario is particularly important, as it applies to many realistic mobile communication systems.

The capacity of fading channels where the channel state information is unknown, i.e., sometimes referred to as non-coherent capacity, has been studied in several publications, see, e.g., [1]–[5]. Most of the existing work on the capacity of/achievable rate on stationary fading channels is restricted to peak power constrained input symbols. On the one hand, this restriction is reasonable as any realistic transmitter has a peak limited transmit power. Furthermore, this approach seems to simplify the mathematical derivation of bounds on the capacity. On the other hand, in the coherent case independent identically distributed (i.i.d.) zero-mean proper Gaussian input symbols are capacity achieving, which are obviously not peak power limited. Furthermore, in many cases the capacity achieving input distribution becomes peaky and, thus, impractical for real system design. In contrast, i.i.d. zero-mean proper Gaussian input distributions serve well to upper-bound the achievable rate with practical modulation and coding schemes, see also [6]. Therefore, we are interested in bounds on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols for a stationary Rayleigh flat-fading channel. Furthermore, the achievable rate with i.i.d. zero-mean proper Gaussian input symbols will converge to the coherent capacity for asymptotically small channel dynamics. However, if we want to evaluate the achievable rate with proper Gaussian input symbols, this requires the derivation of bounds on the achievable rate, which also hold in case of non-peak power constrained input symbols. In [6] the achievable rate with i.i.d. proper Gaussian input symbols has been computed for a Rayleigh block-fading channel. Concerning the case of a stationary Rayleigh flat-fading channel, in [7] we have already given bounds on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols. These bounds rely on a purely mathematical derivation and do not give any link to a physical interpretation like the channel prediction error variance as it has been used in [1]. In the present work, we give a new upper bound on the achievable rate which is also based on the channel prediction error variance and is not restricted to peak power constrained input symbols. In contrast, for the derivation of the channel prediction based capacity bounds in [1], the peak power constraint has been required for technical reasons. We do not give an upper bound on the capacity but only on the achievable rate, as we must restrict to i.i.d. input symbols for mathematical reasons, which are not capacity achieving in general [2]. In conclusion, the contribution of the present work is the derivation of a new upper bound on the achievable rate of a single-antenna discrete-time Rayleigh flat-fading channel with i.i.d. input symbols. We consider a stationary zero-mean jointly proper Gaussian [8] fading process. Its realization is unknown to both, the transmitter and the receiver, while the receiver is aware of the channel law. In addition, we assume that the power spectral density (PSD) of the fading process has compact support. And in contrast to the upper bound on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols given in [7], which holds only for a rectangular PSD of the fading process, the upper bound given in the present work holds for an arbitrary PSD with compact support.

Finally, we evaluate the new upper bound on the achievable rate with i.i.d. input symbols, on the one hand, for peak power constrained input symbols and, on the other hand, for zero-mean proper Gaussian data symbols. For the case of a peak power constraint, we compare the new upper bound to capacity bounds given in [2], and for the case of proper Gaussian input symbols we compare the new upper bound to the bounds on the achievable rate for the same scenario given in [7].

II. System Model

We consider a discrete-time zero-mean jointly proper Gaussian flat-fading channel with the following I/O-relation

$$\mathbf{y} = \mathbf{X}\mathbf{h} + \mathbf{n} \tag{1}$$

with the diagonal matrix $\mathbf{X} = \operatorname{diag}(\mathbf{x})$. Here the $\operatorname{diag}(\cdot)$ operator generates a diagonal matrix whose diagonal is given

During the course of this work Meik Dörpinghaus was with the Institute for Integrated Signal Processing Systems, RWTH Aachen University, 52056 Aachen, Germany.

This work has been supported by the UMIC (Ultra High Speed Mobile Information and Communication) research centre.

by the argument vector. The vector $\mathbf{y} = [y_1, \ldots, y_N]^T$ contains the output symbols in temporal order. Analogously, $\mathbf{x}$, $\mathbf{n}$, and $\mathbf{h}$ contain the channel input symbols, the additive noise samples, and the channel fading weights. All vectors are of length N.

The additive noise process is white and zero-mean jointly proper Gaussian with variance σ_n^2. The fading process is zero-mean jointly proper Gaussian with the temporal correlation $r_h(l) = \mathrm{E}[h_{k+l}h_k^*]$ and variance $r_h(0) = \sigma_h^2$. The PSD of the channel fading process is defined as

$$S_h(f) = \sum_{m=-\infty}^{\infty} r_h(m) e^{-j2\pi mf}, \qquad |f| \leq 0.5. \tag{2}$$

We assume that the PSD exists, which for a jointly proper Gaussian fading process implies ergodicity. Furthermore, we presume the PSD to be compactly supported within the interval $[-f_d, f_d]$ with f_d being the maximum normalized Doppler shift and $0 < f_d < 0.5$. The assumption of a PSD with limited support is motivated by the fact that the velocity of the transmitter, the receiver, and of objects in the environment is limited. To ensure ergodicity, the case $f_d = 0$ is excluded.

For the transmit symbols, which are contained in $\mathbf{x}$, we currently only make the assumption that they are i.i.d. and have a maximum average power of σ_x^2. We name the set of input probability density functions $p(\mathbf{x})$ fulfilling these properties $\mathcal{P}_{\text{i.i.d.}}$.

The processes $\{x_k\}$, $\{h_k\}$, and $\{n_k\}$ are assumed to be mutually independent. With the preceding definitions, the nominal mean SNR[1] is given by $\rho = \frac{\sigma_x^2 \sigma_h^2}{\sigma_n^2}$.

III. The Achievable Rate

Using differential entropies the mutual information equals

$$\mathcal{I}(\mathbf{y};\mathbf{x}) = h(\mathbf{y}) - h(\mathbf{y}|\mathbf{x}). \tag{3}$$

As we study the achievable rate, we consider an infinite transmission length and evaluate the mutual information rate

$$\mathcal{I}'(\mathbf{y};\mathbf{x}) = \lim_{N\to\infty} \frac{1}{N}\mathcal{I}(\mathbf{y};\mathbf{x}) = h'(\mathbf{y}) - h'(\mathbf{y}|\mathbf{x}) \tag{4}$$

where $h'(\cdot)$ is the differential entropy rate.

We construct an upper bound on the achievable rate based on channel prediction. As the fading process is stationary and ergodic, and as we assume i.i.d. input symbols, it holds that

$$\begin{aligned} h'(\mathbf{y}|\mathbf{x}) &= \lim_{N\to\infty} \frac{1}{N} h(\mathbf{y}|\mathbf{x}) \stackrel{(a)}{=} \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} h(y_k|\mathbf{x}, \mathbf{y}_1^{k-1}) \\ &\stackrel{(b)}{=} \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} h(y_k|\mathbf{x}_1^k, \mathbf{y}_1^{k-1}) \stackrel{(c)}{=} \lim_{N\to\infty} h(y_N|\mathbf{x}_1^N, \mathbf{y}_1^{N-1}) \end{aligned} \tag{5}$$

where, e.g., the vector $\mathbf{y}_1^{N-1}$ contains all channel output symbols from the time instant 1 to the time instant $N-1$. Here, for (a) we have used the chain rule for differential entropy, (b) is based on the fact that y_k conditioned on $\mathbf{y}_1^{k-1}$ and $\mathbf{x}_1^k$ is independent of the symbols $\mathbf{x}_{k+1}^N$ due to the independency of the transmit symbols. Equality (c) follows from the ergodicity and stationarity of all processes, see [9, Chapter 4.2].

[1] We use the term *nominal mean SNR* as in case of a peak power constraint it is in general not optimal to use the maximum average transmit power. The nominal mean SNR is the actual mean SNR if the average Tx power is σ_x^2.

Correspondingly, $h'(\mathbf{y})$ can be written as [9, Chapter 4.2]

$$h'(\mathbf{y}) = \lim_{N\to\infty} h(y_N|\mathbf{y}_1^{N-1}). \tag{6}$$

Thus, based on (5) and (6), the achievable rate is given by

$$\mathcal{I}'(\mathbf{y};\mathbf{x}) = \lim_{N\to\infty} \left\{ h(y_N|\mathbf{y}_1^{N-1}) - h(y_N|\mathbf{x}_1^N, \mathbf{y}_1^{N-1}) \right\}. \tag{7}$$

A. An Upper Bound based on Channel Prediction

Now, we upper-bound the achievable rate based on (7).

1) Upper Bound on $h'(\mathbf{y})$: An upper bound on $h'(\mathbf{y})$ is given by the following derivation. As conditioning reduces entropy, we can upper-bound $h(y_N|\mathbf{y}_1^{N-1})$ in (6) by

$$h(y_N|\mathbf{y}_1^{N-1}) \leq h(y_N). \tag{8}$$

Using (6), (8), ergodicity, and stationarity, we get

$$h'(\mathbf{y}) \leq h(y_N) \stackrel{(a)}{\leq} \log\left(\pi e \left(\alpha\sigma_x^2\sigma_h^2 + \sigma_n^2\right)\right) = h'_U(\mathbf{y}) \tag{9}$$

where for (a) we used the fact that proper Gaussian distributions maximize entropy, see [8], and that the average transmit power is given by $\alpha\sigma_x^2$ with $\alpha \in [0,1]$. Using an average transmit power of $\alpha\sigma_x^2$ still enables to choose average transmit powers smaller than the maximum average transmit power σ_x^2.

2) The Entropy Rate $h'(\mathbf{y}|\mathbf{x})$: To lower-bound $h'(\mathbf{y}|\mathbf{x})$, we express $h(y_N|\mathbf{x}_1^N, \mathbf{y}_1^{N-1})$ at the RHS of (5) based on the one-step channel prediction error variance. As the following argumentation will show, the channel output y_N conditioned on $\mathbf{x}_1^N, \mathbf{y}_1^{N-1}$ is proper Gaussian and, thus, fully characterized by its conditional mean and conditional variance. The conditional mean is given by

$$\mathrm{E}\left[y_N|\mathbf{x}_1^N, \mathbf{y}_1^{N-1}\right] = x_N \mathrm{E}\left[h_N|\mathbf{x}_1^{N-1}, \mathbf{y}_1^{N-1}\right] = x_N \hat{h}_N \tag{10}$$

where $\hat{h}_N$ is the MMSE estimate of h_N based on the channel output observations at all previous time instances and the channel input symbols at these time instances. Based on $\hat{h}_N$ the channel output y_N can be written as

$$y_N = x_N h_N + n_N = x_N\left(\hat{h}_N + e_N\right) + n_N \tag{11}$$

with the prediction error $e_N = h_N - \hat{h}_N$.

As both, the noise as well as the fading process, are jointly proper Gaussian, the MMSE estimate is equivalent to the linear minimum mean squared error (LMMSE). Since $\hat{h}_N$ and h_N are jointly proper Gaussian and zero-mean, it follows that the estimation error e_N is zero-mean proper Gaussian.

As e_N is proper Gaussian, it can be easily seen by (11) that y_N conditioned on $\mathbf{x}_1^N, \mathbf{y}_1^{N-1}$ is also proper Gaussian. Thus, for the evaluation of $h(y_N|\mathbf{x}_1^N, \mathbf{y}_1^{N-1})$, we calculate the conditional variance of the channel output y_N which is given by

$$\begin{aligned} \mathrm{var}\left[y_N|\mathbf{x}_1^N, \mathbf{y}_1^{N-1}\right] &= \mathrm{E}\left[\left|y_N - \mathrm{E}\left[y_N|\mathbf{x}_1^N, \mathbf{y}_1^{N-1}\right]\right|^2 \middle| \mathbf{x}_1^N, \mathbf{y}_1^{N-1}\right] \\ &= |x_N|^2 \mathrm{E}\left[|e_N|^2 \middle| \mathbf{x}_1^{N-1}, \mathbf{y}_1^{N-1}\right] + \sigma_n^2 = |x_N|^2 \sigma_{e_{\text{pred}}}^2(\mathbf{x}_1^{N-1}) + \sigma_n^2 \end{aligned} \tag{12}$$

where

$$\sigma_{e_{\text{pred}}}^2(\mathbf{x}_1^{N-1}) \stackrel{(a)}{=} \mathrm{E}\left[|e_N|^2 \middle| \mathbf{x}_1^{N-1}\right] \tag{13}$$

is the prediction error variance of the MMSE estimator for $\hat{h}_N$. For (a) we have used the fact that the zero-mean estimation error e_N is orthogonal to and, thus, independent of the observations $\mathbf{y}_1^{N-1}$. The prediction error variance depends on the input symbols $\mathbf{x}_1^{N-1}$, which is indicated by writing $\sigma_{e_{\text{pred}}}^2(\mathbf{x}_1^{N-1})$.

Based on $\sigma^2_{e_{\text{pred}}}(\mathbf{x}_1^{N-1})$, we can write $h(y_N|\mathbf{x}_1^N,\mathbf{y}_1^{N-1})$ as

$$h(y_N|\mathbf{x}_1^N,\mathbf{y}_1^{N-1})=\mathrm{E}_{\mathbf{x}}\Big[\log\Big(\pi e\Big(\sigma_n^2+\sigma^2_{e_{\text{pred}}}(\mathbf{x}_1^{N-1})|x_N|^2\Big)\Big)\Big]. \quad (14)$$

With (5) and (14), we get for i.i.d. input symbols

$$h'(\mathbf{y}|\mathbf{x})=\mathrm{E}_{x_k}\left[\mathrm{E}_{\mathbf{x}_{-\infty}^{k-1}}\left[\log\Big(\pi e\left(\sigma_n^2+\sigma^2_{e_{\text{pred},\infty}}(\mathbf{x}_{-\infty}^{k-1})|x_k|^2\right)\Big)\right]\right] \quad (15)$$

where $\sigma^2_{e_{\text{pred},\infty}}(\mathbf{x}_{-\infty}^{k-1})$ is the prediction error variance in (13) for an infinite number of channel observations in the past, i.e.,

$$\sigma^2_{e_{\text{pred},\infty}}(\mathbf{x}_{-\infty}^{k-1})=\lim_{N\to\infty}\sigma^2_{e_{\text{pred}}}(\mathbf{x}_1^{N-1}) \quad (16)$$

which is indicated by writing $\sigma^2_{e_{\text{pred},\infty}}(\mathbf{x}_{-\infty}^{k-1})$. Note that we have switched the notation and now predict at the time instant k instead of predicting at the time instant N. This is possible, as the channel fading process is stationary, the input symbols are assumed to be i.i.d., and as we consider an infinitely long past.

3) Upper Bound on the Achievable Rate: With (4), (9), and (15), we can give the following upper bound on the achievable rate with i.i.d. input symbols

$$\mathcal{I}'(\mathbf{y};\mathbf{x})\le\log(\alpha\rho+1)-\mathrm{E}_{x_k}\left[\mathrm{E}_{\mathbf{x}_{-\infty}^{k-1}}\left[\log\left(1+\frac{\sigma^2_{e_{\text{pred},\infty}}(\mathbf{x}_{-\infty}^{k-1})}{\sigma_n^2}|x_k|^2\right)\right]\right]. \quad (17)$$

Obviously, the upper bound in (17) still depends on the channel prediction error variance $\sigma^2_{e_{\text{pred},\infty}}(\mathbf{x}_{-\infty}^{k-1})$ given in (16), which itself depends on the distribution of the input symbols in the past. Effectively $\sigma^2_{e_{\text{pred},\infty}}(\mathbf{x}_{-\infty}^{k-1})$ is itself a random quantity. For infinite transmission lengths, i.e., $N\to\infty$, its distribution is independent of the time instant k, as the channel fading process is stationary and as the transmit symbols are i.i.d..

4) The Prediction Error Variance $\sigma^2_{e_{\text{pred},\infty}}(\mathbf{x}_{-\infty}^{k-1})$: The prediction error variance $\sigma^2_{e_{\text{pred},\infty}}(\mathbf{x}_{-\infty}^{k-1})$ in (16) depends on the distribution of the input symbols $\mathbf{x}_{-\infty}^{k-1}$. To construct an upper bound on the RHS of (17) we need to find a distribution of the transmit symbols in the past, i.e., $\mathbf{x}_{-\infty}^{k-1}$, which leads to a distribution of $\sigma^2_{e_{\text{pred},\infty}}(\mathbf{x}_{-\infty}^{k-1})$ that maximizes the RHS of (17). Therefore, we have to express the channel prediction error variance $\sigma^2_{e_{\text{pred},\infty}}(\mathbf{x}_{-\infty}^{k-1})$ as a function of the transmit symbols in the past, i.e., $\mathbf{x}_{-\infty}^{k-1}$. In a first step, we give such an expression for the case of a finite past time horizon, i.e., for $\sigma^2_{e_{\text{pred}}}(\mathbf{x}_1^{N-1})$ in (13) which can be expressed by

$$\sigma^2_{e_{\text{pred}}}(\mathbf{x}_1^{N-1})=\sigma_h^2-\mathbf{r}^H_{\mathbf{y}_1^{N-1}h_N|\mathbf{x}_1^{N-1}}\mathbf{R}^{-1}_{\mathbf{y}_1^{N-1}|\mathbf{x}_1^{N-1}}\mathbf{r}_{\mathbf{y}_1^{N-1}h_N|\mathbf{x}_1^{N-1}} \quad (18)$$

where $\mathbf{R}_{\mathbf{y}_1^{N-1}|\mathbf{x}_1^{N-1}}$ is the correlation matrix of the observations $\mathbf{y}_1^{N-1}$ while the past transmit symbols $\mathbf{x}_1^{N-1}$ are known

$$\mathbf{R}_{\mathbf{y}_1^{N-1}|\mathbf{x}_1^{N-1}}=\mathrm{E}\left[\mathbf{y}_1^{N-1}(\mathbf{y}_1^{N-1})^H\middle|\mathbf{x}_1^{N-1}\right]=\mathbf{X}_{N-1}\mathbf{R}_h\mathbf{X}^H_{N-1}+\sigma_n^2\mathbf{I}_{N-1} \quad (19)$$

with $\mathbf{X}_{N-1}$ being a diagonal matrix containing the past transmit symbols such that $\mathbf{X}_{N-1}=\operatorname{diag}\left(\mathbf{x}_1^{N-1}\right)$ and $\mathbf{I}_{N-1}$ being the identity matrix of size $(N-1)\times(N-1)$. In addition, $\mathbf{R}_h$ is the autocorrelation matrix of the channel fading process

$$\mathbf{R}_h=\mathrm{E}\left[\mathbf{h}_1^{N-1}(\mathbf{h}_1^{N-1})^H\right] \quad (20)$$

where $\mathbf{h}_1^{N-1}$ contains the fading weights from time instant 1 to $N-1$. The cross correlation vector $\mathbf{r}_{\mathbf{y}_1^{N-1}h_N|\mathbf{x}_1^{N-1}}$ between the observation vector $\mathbf{y}_1^{N-1}$ and the fading weight h_N while knowing the past transmit symbols $\mathbf{x}_1^{N-1}$ is given by

$$\mathbf{r}_{\mathbf{y}_1^{N-1}h_N|\mathbf{x}_1^{N-1}}=\mathrm{E}\left[\mathbf{y}_1^{N-1}h_N^*\middle|\mathbf{x}_1^{N-1}\right]=\mathbf{X}_{N-1}\mathbf{r}_{h,\text{pred}} \quad (21)$$

with $\mathbf{r}_{h,\text{pred}}=\left[r_h(-(N-1))\dots r_h(-1)\right]^T$ where $r_h(l)$ is the autocorrelation function as defined in Section II.

Substituting (19) and (21) into (18) yields

$$\begin{aligned}\sigma^2_{e_{\text{pred}}}(\mathbf{x}_1^{N-1})&=\sigma_h^2-\mathbf{r}^H_{h,\text{pred}}\Big(\mathbf{R}_h+\sigma_n^2\left(\mathbf{X}^H_{N-1}\mathbf{X}_{N-1}\right)^{-1}\Big)^{-1}\mathbf{r}_{h,\text{pred}}\\&\overset{(a)}{=}\sigma_h^2-\mathbf{r}^H_{h,\text{pred}}\left(\mathbf{R}_h+\sigma_n^2\mathbf{Z}^{-1}\right)^{-1}\mathbf{r}_{h,\text{pred}}\end{aligned} \quad (22)$$

where for (a) we have used $\mathbf{Z}=\mathbf{X}^H_{N-1}\mathbf{X}_{N-1}$ i.e., $\mathbf{Z}$ is a diagonal matrix containing the powers of the individual transmit symbols in the past from time instant 1 to $N-1$. For ease of notation, we omit the index $N-1$.

Remember that we want to derive an upper bound on the achievable rate with i.i.d. input symbols by maximizing the RHS of (17) over all i.i.d. distributions of the transmit symbols in the past with an average power $\alpha\sigma_x^2$. Obviously, the distribution of the phase of the past transmit symbols $\mathbf{x}_1^{N-1}$ has no influence on the channel prediction error variance $\sigma^2_{e_{\text{pred}}}(\mathbf{x}_1^{N-1})$. Thus, it rests to evaluate, for which distribution of the power of the past transmit symbols the RHS of (17) is maximized. In the following, we will show that the RHS of (17) is maximized in case the past transmit symbols have a constant power $\alpha\sigma_x^2$. I.e., calculation of the prediction error variance under the assumption that the past transmit symbols are constant modulus symbols with transmit power $|x_k|^2=\alpha\sigma_x^2$ maximizes the RHS of (17) over all i.i.d. input distributions for the given average power constraint of $\alpha\sigma_x^2$.

To prove this statement, we use the fact that the expression in the expectation operation at the RHS of (17) (but here for the case of a finite past time horizon) with (22), i.e.,

$$\log\left(1+\frac{|x_N|^2}{\sigma_n^2}\Big(\sigma_h^2-\mathbf{r}^H_{h,\text{pred}}\left(\mathbf{R}_h+\sigma_n^2\mathbf{Z}^{-1}\right)^{-1}\mathbf{r}_{h,\text{pred}}\Big)\right) \quad (23)$$

is convex with respect to each individual element of the diagonal of $\mathbf{Z}$, which we name $\mathbf{z}$, see Appendix A for a proof. As the transmit symbols are i.i.d., using convexity and Jensen's inequality, we get

$$\begin{aligned}&\mathrm{E}_{\mathbf{z}}\left[\log\left(1+\frac{|x_N|^2}{\sigma_n^2}\left(\sigma_h^2-\mathbf{r}^H_{h,\text{pred}}\left(\mathbf{R}_h+\sigma_n^2\mathbf{Z}^{-1}\right)^{-1}\mathbf{r}_{h,\text{pred}}\right)\right)\right]\\&\ge\log\left(1+\frac{|x_N|^2}{\sigma_n^2}\left(\sigma_h^2-\mathbf{r}^H_{h,\text{pred}}\Big(\mathbf{R}_h+\sigma_n^2\left(\mathrm{E}_{\mathbf{z}}[\mathbf{Z}]\right)^{-1}\Big)^{-1}\mathbf{r}_{h,\text{pred}}\right)\right)\\&=\log\left(1+\frac{|x_N|^2}{\sigma_n^2}\left(\sigma_h^2-\mathbf{r}^H_{h,\text{pred}}\left(\mathbf{R}_h+\frac{\sigma_n^2}{\alpha\sigma_x^2}\mathbf{I}_{N-1}\right)^{-1}\mathbf{r}_{h,\text{pred}}\right)\right)\\&=\log\left(1+\frac{|x_N|^2}{\sigma_n^2}\sigma^2_{e_{\text{pred,CM}}}\right)\end{aligned} \quad (24)$$

where $\sigma^2_{e_{\text{pred,CM}}}$ is the channel prediction error variance in case all past transmit symbols are constant modulus symbols with power $\alpha\sigma_x^2$. Here, the index CM denotes constant modulus.

As this lower bounding of the LHS of (24) can be performed for an arbitrary N, i.e., for an arbitrary long past, we can also conclude that the RHS of (17) is upper bounded by

$$\mathcal{I}'(\mathbf{y};\mathbf{x}) \le \log(\alpha\rho+1) - \mathrm{E}_{x_k}\left[\log\left(1+\frac{\sigma^2_{e_{\text{pred,CM},\infty}}}{\sigma_n^2}|x_k|^2\right)\right] \quad (25)$$

where $\sigma^2_{e_{\text{pred,CM},\infty}}$ is the channel prediction error variance in case all past transmit symbols are constant modulus symbols with a power $\alpha\sigma_x^2$ and an infinitely long past observation horizon. In this case, the prediction error variance is no longer a random quantity but is constant for all time instances k.

Constant modulus symbols are in general not the capacity maximizing input distribution. However, we only use them to find a distribution of $\sigma^2_{e_{\text{pred},\infty}}(\mathbf{x}_{-\infty}^{k-1})$ that maximizes (17).

For constant modulus input symbols and an infinitely long past, the prediction error variance is given by, cf. [1]

$$\sigma^2_{e_{\text{pred,CM},\infty}} = \frac{\sigma_n^2}{\alpha\sigma_x^2}\left\{\exp\left(\int_{-\frac{1}{2}}^{\frac{1}{2}}\log\left(1+\frac{\alpha\sigma_x^2}{\sigma_n^2}S_h(f)\right)df\right)-1\right\} \quad (26)$$

5) Effect of Constraints on the Input Distribution: We evaluate the upper bound given in (25) for different constraints on the input distribution. First, we consider the case of a peak power constrained to P_{peak} in addition to the average power constraint. With the nominal peak-to-average power ratio[2] $\beta = P_{\text{peak}}/\sigma_x^2$, we get the following upper bound on the achievable rate with i.i.d. input symbols

$$\begin{aligned}
&\sup_{\mathcal{P}^{\text{peak}}_{\text{i.i.d.}}} \mathcal{I}'(\mathbf{y};\mathbf{x}) \\
&\le \sup_{\alpha\in[0,1]} \sup_{\mathcal{P}^{\text{peak}}_{\text{i.i.d.}}|\alpha} \left\{\log(\alpha\rho+1) - \mathrm{E}_{x_k}\left[\log\left(1+\frac{\sigma^2_{e_{\text{pred,CM},\infty}}}{\sigma_n^2}|x_k|^2\right)\right]\right\} \\
&\overset{(a)}{\le} \sup_{\alpha\in[0,1]}\left\{\log(\alpha\rho+1) - \frac{\alpha}{\beta}\log\left(1+\frac{\sigma^2_{e_{\text{pred,CM},\infty}}}{\sigma_h^2}\rho\beta\right)\right\} \quad (27)
\end{aligned}$$

where $\mathcal{P}^{\text{peak}}_{\text{i.i.d.}}$ corresponds to $\mathcal{P}_{\text{i.i.d.}}$ but with an additional peak power constraint $|x_k|^2 \le \beta\sigma_x^2$. $\mathcal{P}^{\text{peak}}_{\text{i.i.d.}}|\alpha$ corresponds to $\mathcal{P}^{\text{peak}}_{\text{i.i.d.}}$ but with the average transmit power fixed to $\alpha\sigma_x^2$. Inequality (a) can be shown by calculating a lower bound on the infimum of $\mathrm{E}_{x_k}\left[\log\left(1+\frac{\sigma^2_{e_{\text{pred,CM},\infty}}}{\sigma_n^2}|x_k|^2\right)\right]$ over $\mathcal{P}^{\text{peak}}_{\text{i.i.d.}}|\alpha$. Note that the prediction error variance $\sigma^2_{e_{\text{pred,CM},\infty}}$ depends on α. Now, we would have to calculate the supremum of the RHS of (27) with respect to α which turns out to be difficult due to the dependency of $\sigma^2_{e_{\text{pred,CM},\infty}}$ on α. However, $\sigma^2_{e_{\text{pred,CM},\infty}}$ monotonically decreases with an increasing α. Furthermore, the RHS of (27) monotonically increases with a decreasing $\sigma^2_{e_{\text{pred,CM},\infty}}$. Thus, we can upper-bound (27) by setting $\alpha = 1$ within $\sigma^2_{e_{\text{pred,CM},\infty}}$ in (26), i.e., $\sigma^2_{e_{\text{pred,CM},\infty}}\big|_{\alpha=1}$, and obtain

$$\begin{aligned}
\sup_{\mathcal{P}^{\text{peak}}_{\text{i.i.d.}}}\mathcal{I}'(\mathbf{y};\mathbf{x}) &\le \sup_{\alpha\in[0,1]}\left\{\log(\alpha\rho+1) - \frac{\alpha}{\beta}\log\left(1+\frac{\sigma^2_{e_{\text{pred,CM},\infty}}\big|_{\alpha=1}}{\sigma_h^2}\rho\beta\right)\right\} \\
&= \log(\alpha_{\text{opt}}\rho+1) - \frac{\alpha_{\text{opt}}}{\beta}\log\left(1+\frac{\sigma^2_{e_{\text{pred,CM},\infty}}\big|_{\alpha=1}}{\sigma_h^2}\rho\beta\right) \\
&= \mathcal{I}'_U(\mathbf{y};\mathbf{x})\big|_{\text{pred},P_{\text{peak}}} \quad (28)
\end{aligned}$$

with

$$\alpha_{\text{opt}} = \min\left\{1, \left(\frac{1}{\beta}\log\left(1+\frac{\sigma^2_{e_{\text{pred,CM},\infty}}\big|_{\alpha=1}}{\sigma_h^2}\rho\beta\right)\right)^{-1} - \frac{1}{\rho}\right\}. \quad (29)$$

As the bound in (28) becomes loose for $\beta \to \infty$, we also give an upper bound on the achievable rate with i.i.d. zero-mean proper Gaussian (PG) input symbols which is given by

$$\mathcal{I}'_U(\mathbf{y};\mathbf{x})\big|_{\text{pred,PG}} = \log(\rho+1) - \int_0^\infty \log\left(1+\frac{\sigma^2_{e_{\text{pred,CM},\infty}}\big|_{\alpha=1}}{\sigma_h^2}\rho z\right)e^{-z}dz \quad (30)$$

where we set $\alpha = 1$, as in the non-peak power constrained case the upper bound is maximized for the maximum average transmit power σ_x^2.

As far as we know, the upper bound on the achievable rate in (25) is new. The innovation in the derivation of this bound lies in the fact that we separate the input symbols into the one at the time instant x_k and the previous input symbols contained in $\mathbf{x}_{-\infty}^{k-1}$. The latter ones are only relevant to calculate the prediction error variance, which itself is a random variable depending on the distribution of the past transmit symbols. To derive an upper bound on the achievable rate with i.i.d. input distributions, we have shown that the achievable rate is upper-bounded if the prediction error variance is calculated under the assumption that all past transmit symbols are constant modulus input symbols. As the assumption on constant modulus symbols is only used in the context of the prediction error variance, the upper bound on the achievable rate still holds for any i.i.d. input distribution with the given average power constraint. This allows us to evaluate this bound also for the case of i.i.d. zero-mean proper Gaussian input symbols.

Note that all preceding upper bounds can be enhanced, as $\mathcal{I}'(\mathbf{y};\mathbf{x})$ is upper bounded by the coherent mutual information rate $\mathcal{I}'(\mathbf{y};\mathbf{x}|\mathbf{h})$. The coherent channel capacity is known and achieved by i.i.d. zero-mean proper Gaussian input symbols. Thus, we can enhance the bounds in (28) and (30) as follows

$$\mathcal{I}'(\mathbf{y};\mathbf{x}) \le \min\left\{\mathcal{I}'_U(\mathbf{y};\mathbf{x}), \sup_{\mathcal{P}_{\text{i.i.d.}}}\mathcal{I}'(\mathbf{y};\mathbf{x}|\mathbf{h})\right\} \quad (31)$$

where $\sup_{\mathcal{P}_{\text{i.i.d.}}}\mathcal{I}'(\mathbf{y};\mathbf{x}|\mathbf{h})$ is the coherent capacity given by

$$\sup_{\mathcal{P}_{\text{i.i.d.}}}\mathcal{I}'(\mathbf{y};\mathbf{x}|\mathbf{h}) = \int_0^\infty \log(1+\rho\cdot z)\,e^{-z}dz. \quad (32)$$

IV. Numerical Evaluation

In the following, we evaluate the new upper bound on the achievable rate with i.i.d. input symbols, on the one hand, for the case of a peak power constraint, i.e., (28) and, on the other hand, for zero-mean proper Gaussian input symbols, i.e., (30), both in combination with (31). Furthermore, we compare these bounds to the upper and lower bounds on the peak power constraint capacity given in [2], and respectively with the upper and lower bound on the achievable rate with

[2]The nominal peak-to-average power ratio corresponds to the actual peak-to-average power ratio if the actual average transmit power is equal to σ_x^2.

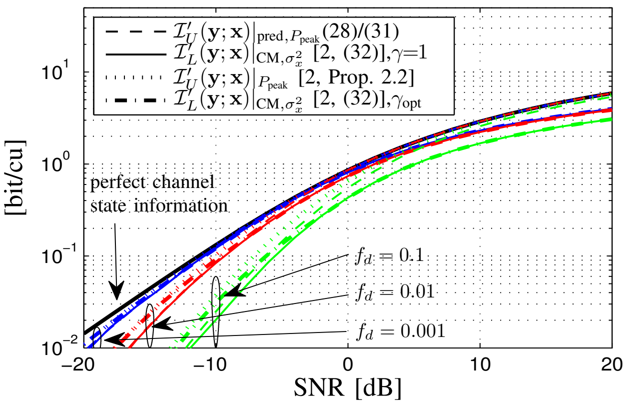

Fig. 1. Comparison of the upper bound on the achievable rate with i.i.d. symbols and a peak power constraint given in (28)/(31) based on channel prediction to the upper bound on capacity given in [2, Proposition 2.2] for $\beta = 2$; in addition the lower bound on the peak power constrained capacity [2, (32)] is shown for a constant modulus (CM) input distribution with 100 signaling points without and with time sharing, i.e., for $\gamma = 1$ and for γ_{opt}

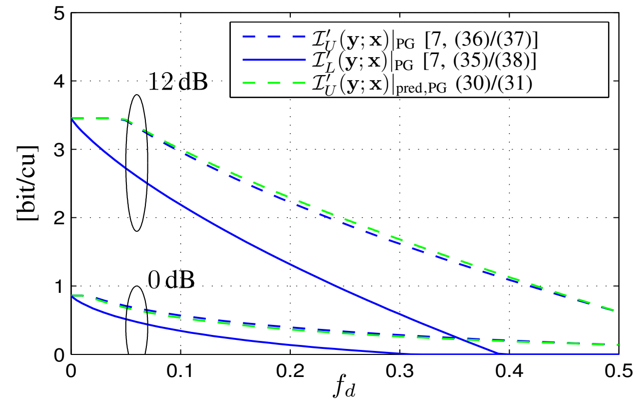

Fig. 2. Comparison of the upper bound on the achievable rate with i.i.d. zero-mean proper Gaussian (PG) inputs based on channel prediction (30)/(31) with the upper bound given in [7, (36)/(37)]; in addition the lower bound on the achievable rate with i.i.d. zero-mean proper Gaussian inputs [7, (35)/(38)] is shown

i.i.d. zero-mean proper Gaussian input symbols given in [7]. For the following evaluations, we assume in all cases that the PSD of the channel fading process is rectangular, i.e., $S_h(f) = \sigma_h^2/(2f_d)$ for $|f| \leq f_d$ and zero otherwise.

Fig. 1 shows the upper bound on the achievable rate with i.i.d. input symbols and a peak power constraint based on the channel prediction error variance in (28)/(31) in comparison to the upper bound on the peak power constrained capacity given in [2, Prop. 2.2] with $\beta = 2$ for both. For comparison we use the lower bound on the peak power constrained capacity given in [2, (32)] based on a constant modulus input distribution with 100 discrete signaling points with a uniform angular spacing. This approximates the case of a uniformly distributed phase. This lower bound is shown without time sharing ($\gamma = 1$) and with time sharing (γ_{opt}). Time sharing means, that the transmitter uses the channel only a $1/\gamma$ part of the time. Obviously, time sharing is not in accordance with the assumption on i.i.d. input symbols. Therefore, the lower bound with $\gamma = 1$ matches the new upper bound on the achievable rate with i.i.d. input symbols in (28)/(31), while the lower bound with time sharing (γ_{opt}) only matches the capacity upper bound in [2, Prop. 2.2]. From Fig. 1 it can be seen that the upper bound on the achievable rate with i.i.d. input symbols in (28)/(31) is lower or equal than the capacity upper bound in [2, Prop. 2.2]. However, (28)/(31) is only an upper bound on the achievable rate with i.i.d. input symbols and not on the capacity, as i.i.d. input symbols are in general not capacity achieving [2]. This can also be seen, as the lower bound on the achievable rate with time sharing is larger than the upper bound on the achievable rate with i.i.d. input symbols (28)/(31) for very low SNRs. Furthermore, it is worth mentioning that for the case of a nominal peak-to-average power ratio $\beta = 1$, the upper bound in (28)/(31) and the one given in [2, Prop. 2.2] coincide. In addition, the prediction based upper bound on the achievable rate in (28)/(31) as well as the capacity upper bound in [2, Prop. 2.2] become loose for large β.

Fig. 2 shows the prediction based upper bound on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols given in (30)/(31) in comparison to the upper and lower bound on the achievable rate with i.i.d. zero-mean proper Gaussian inputs given [7]. Both upper bounds are shown in combination with the coherent upper bound, see (31). A comparison of the prediction based upper bound (30)/(31) and the bound given in [7, (36)/(37)] shows, that it depends on the channel parameters, which one is tighter. It can easily be shown that for $f_d \to 0$ and for $f_d = 0.5$ both bounds are equal. For other f_d it depends on the SNR ρ which bound is tighter. An analytical comparison turns out to be difficult as in both cases we use a different way of lower bounding $h'(\mathbf{y}|\mathbf{x})$.

V. Summary

In the present paper, we have derived a new upper bound on the achievable rate with i.i.d. input symbols based on the prediction separation of the mutual information rate in (7). Based on this separation, the conditional channel output entropy rate $h'(\mathbf{y}|\mathbf{x})$ can be expressed by the one-step channel prediction error variance, which is a well known result, see, e.g., [1]. We show that for i.i.d. input symbols the prediction error variance $\sigma^2_{e_{\text{pred},\infty}}(\mathbf{x}_{-\infty}^{k-1})$ calculated under the assumption of constant modulus symbols yields an upper bound on the achievable rate. As the constant modulus assumption is only used in the context of $\sigma^2_{e_{\text{pred},\infty}}(\mathbf{x}_{-\infty}^{k-1})$, we can still give upper bounds on the achievable rate for general i.i.d. input symbol distributions, even for the case without a peak power constraint.

For a peak power constraint, we have observed that for nominal peak-to-average power ratios of $\beta = 2$ and $\beta = 1$ the upper bound on the achievable rate with i.i.d. input symbols is lower than or equal to the capacity upper bound in [2, Prop. 2.2]. But, it is not an upper bound on the capacity due to the restriction to i.i.d. input symbols. In case of i.i.d. proper Gaussian input symbols, it depends on the channel parameters if (30)/(31) or the upper bound in [7, (36)/(37)] is tighter. Concerning the case of proper Gaussian input symbols, the new bound given in (30)/(31) is more general than the upper bound in [7, (36)/(37)] as it holds for arbitrary PSDs of the fading process with compact support and is not limited to rectangular PSDs as the one in [7, (36)/(37)].

APPENDIX A

Convexity of (23): To prove that (23) is convex with respect to the individual diagonal elements of $\mathbf{Z}$ we rewrite the prediction error variance $\sigma^2_{e_{\text{pred}}}(\mathbf{x}_1^{N-1}) = \sigma^2_{e_{\text{pred}}}(\mathbf{z})$ as follows

$$\begin{aligned}\sigma^2_{e_{\text{pred}}}(\mathbf{z}) &= \sigma_h^2 - \mathbf{r}^H_{h,\text{pred}} \left(\mathbf{R}_h + \sigma_n^2 \mathbf{Z}^{-1}\right)^{-1} \mathbf{r}_{h,\text{pred}} \\ &\overset{(a)}{=} \sigma^2_{e_{\text{pred}}}(\mathbf{z}_{\setminus i}) - \frac{z_i \cdot a}{1 + z_i \lambda_{\max}} \end{aligned} \tag{33}$$

where for (a) we have used the matrix inversion lemma several times, and we have separated the diagonal matrix $\mathbf{Z}$ as follows

$$\mathbf{Z} = \mathbf{Z}_{\setminus i} + z_i \mathbf{V}_i \tag{34}$$

where $\mathbf{Z}_{\setminus i}$ corresponds to $\mathbf{Z}$ except that the i-th diagonal element is set to 0, $\mathbf{V}_i$ is a matrix with all elements zero except of the i-th diagonal element being equal to 1, and z_i is the i-th diagonal element of the matrix $\mathbf{Z}$. In addition, $\lambda_{\max}$ is the non-zero eigenvalue of the rank one matrix

$$\mathbf{B} = \left(\frac{1}{\sigma_n^2}\mathbf{Z}_{\setminus i} + \mathbf{R}_h^{-1}\right)^{-1} \frac{1}{\sigma_n^2}\mathbf{V}_i. \tag{35}$$

Furthermore, $\sigma^2_{e_{\text{pred}}}(\mathbf{z}_{\setminus i})$ is given by

$$\sigma^2_{e_{\text{pred}}}(\mathbf{z}_{\setminus i}) = \sigma_h^2 - \mathbf{r}^H_{h,\text{pred}} \left(\mathbf{R}_h^{-1} - \mathbf{R}_h^{-1}\left(\frac{\mathbf{Z}_{\setminus i}}{\sigma_n^2} + \mathbf{R}_h^{-1}\right)^{-1} \mathbf{R}_h^{-1}\right) \mathbf{r}_{h,\text{pred}}$$

which is the prediction error variance if the observation at the i-th time instant is not used for the channel prediction. Finally, for (a) in (33) we have used the substitution

$$\begin{aligned} a &= \mathbf{r}^H_{h,\text{pred}} \mathbf{R}_h^{-1}\left(\frac{\mathbf{Z}_{\setminus i}}{\sigma_n^2} + \mathbf{R}_h^{-1}\right)^{-1} \frac{\mathbf{V}_i}{\sigma_n^2}\left(\frac{\mathbf{Z}_{\setminus i}}{\sigma_n^2} + \mathbf{R}_h^{-1}\right)^{-1} \mathbf{R}_h^{-1}\mathbf{r}_{h,\text{pred}} \\ &\geq 0 \end{aligned} \tag{36}$$

where the nonnegativity follows as $\mathbf{V}_i$ is positive semidefinite.

Thus, with (33) we have found a separation of the channel prediction error variance $\sigma^2_{e_{\text{pred}}}(\mathbf{z})$ into the term $\sigma^2_{e_{\text{pred}}}(\mathbf{z}_{\setminus i})$ being independent of z_i, and an additional term, which depends on z_i. Note that a and $\lambda_{\max}$ in the second term on the RHS of (33) are independent of z_i and that the element i is an arbitrarily chosen element. I.e., we can use this separation for each diagonal element of the matrix $\mathbf{Z}$.

By substituting the RHS of (33) into (23) we get

$$\log\left(1 + \frac{|x_N|^2}{\sigma_n^2}\left(\sigma^2_{e_{\text{pred}}}(\mathbf{z}_{\setminus i}) - \frac{z_i \cdot a}{1 + z_i\lambda_{\max}}\right)\right) = K. \tag{37}$$

Recall that we want to show the convexity of (37) with respect to the element z_i. Therefore, we calculate its second derivative with respect to z_i which is given by

$$\frac{\partial^2 K}{(\partial z_i)^2} = \frac{\frac{|x_N|^2}{\sigma_n^2}\frac{a 2\lambda_{\max}(1+z_i\lambda_{\max})}{(1+z_i\lambda_{\max})^4}\left\{1 + \frac{|x_N|^2}{\sigma_n^2}\left(\sigma^2_{e_{\text{pred}}}(\mathbf{z}_{\setminus i}) - \frac{a\left(z_i + \frac{1}{2\lambda_{\max}}\right)}{1+z_i\lambda_{\max}}\right)\right\}}{\left(1 + \frac{|x_N|^2}{\sigma_n^2}\left(\sigma^2_{e_{\text{pred}}}(\mathbf{z}_{\setminus i}) - \frac{a z_i}{1+z_i\lambda_{\max}}\right)\right)^2}$$

and will show that it is nonnegative, i.e.,

$$\frac{\partial^2 K}{(\partial z_i)^2} \geq 0. \tag{38}$$

Therefore, first we show that $\lambda_{\max}$ is nonnegative. This can be done based on the definition of the eigenvalues of the matrix $\mathbf{B}$

$$\mathbf{B}\mathbf{u} = \left(\frac{1}{\sigma_n^2}\mathbf{Z}_{\setminus i} + \mathbf{R}_h^{-1}\right)^{-1} \frac{1}{\sigma_n^2}\mathbf{V}_i \mathbf{u} = \lambda_{\max}\mathbf{u}$$

$$\Rightarrow \frac{1}{\sigma_n^2}\mathbf{u}^H\mathbf{V}_i\mathbf{u} = \lambda_{\max}\mathbf{u}^H\left(\frac{1}{\sigma_n^2}\mathbf{Z}_{\setminus i} + \mathbf{R}_h^{-1}\right)\mathbf{u} \overset{(a)}{\Rightarrow} \lambda_{\max} \geq 0$$

where (a) follows from the fact that the eigenvalues of $\left(\frac{1}{\sigma_n^2}\mathbf{Z}_{\setminus i} + \mathbf{R}_h^{-1}\right)$ are nonnegative, as $\mathbf{R}_h$ is positive definite and the diagonal entries of the diagonal matrix $\mathbf{Z}_{\setminus i}$ are also nonnegative. In addition, $\mathbf{V}_i$ is also positive semidefinite.

With $\lambda_{\max}$, z_i, and a being nonnegative, for the proof of (38), it rests to show that

$$\sigma^2_{e_{\text{pred}}}(\mathbf{z}_{\setminus i}) - \frac{a}{1 + z_i\lambda_{\max}}\left(z_i + \frac{1}{2\lambda_{\max}}\right) \geq 0. \tag{39}$$

To prove this inequality, we calculate the derivative of the LHS of (39) with respect to z_i, which is given by

$$\frac{\partial}{\partial z_i}\left\{\sigma^2_{e_{\text{pred}}}(\mathbf{z}_{\setminus i}) - \frac{a\left(z_i + \frac{1}{2\lambda_{\max}}\right)}{1+z_i\lambda_{\max}}\right\} = \frac{-a}{2(1+z_i\lambda_{\max})^2} \leq 0 \tag{40}$$

where for the last inequality we have used (36). I.e., the LHS of (39) monotonically decreases in z_i. Furthermore, for $z_i \to \infty$ the LHS of (39) becomes

$$\lim_{z_i\to\infty}\left\{\sigma^2_{e_{\text{pred}}}(\mathbf{z}_{\setminus i}) - \frac{a\left(z_i + \frac{1}{2\lambda_{\max}}\right)}{1 + z_i\lambda_{\max}}\right\} \overset{(a)}{=} \lim_{z_i\to\infty}\sigma^2_{e_{\text{pred}}}(\mathbf{z}) \overset{(b)}{\geq} 0$$

where (a) follows due to (33), and where (b) holds as the prediction error variance must be nonnegative. As the LHS of (39) is monotonically decreasing in z_i and as its limit for $z_i \to \infty$ is nonnegative, (39) must hold.

With (39), (38) holds and, thus, (37) is convex in z_i. As the element i has been chosen arbitrarily, in conclusion, we have shown that (23) is convex in each z_i for $i = 1, \ldots, N-1$.

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] 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.

[3] M. Médard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 933–946, May 2000.

[4] J. Baltersee, G. Fock, and H. Meyr, "An information theoretic foundation of synchronized detection," *IEEE Trans. Commun.*, vol. 49, no. 12, pp. 2115–2123, Dec. 2001.

[5] G. Durisi, U. G. Schuster, H. Bölcskei, and S. Shamai (Shitz), "Noncoherent capacity of underspread fading channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 367–395, Jan. 2010.

[6] F. Rusek, A. Lozano, and N. Jindal, "Mutual information of IID complex Gaussian signals on block Rayleigh-faded channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Austin, TX, U.S.A., Jun. 2010, pp. 300–304.

[7] 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. Int. Symp. Inf. Theory and its Applications (ISITA)*, Auckland, New Zealand, Dec. 2008.

[8] 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.

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