On the Achievable Rate of Stationary Rayleigh Flat-Fading Channels with Gaussian Inputs

Meik Dörpinghaus, *Member, IEEE*, Heinrich Meyr, *Life Fellow, IEEE*,
and Rudolf Mathar, *Member, IEEE*

***Abstract*—In this work, a discrete-time stationary Rayleigh flat-fading channel with unknown channel state information at transmitter and receiver side is studied. The law of the channel is presumed to be known to the receiver. For independent identically distributed (i.i.d.) zero-mean proper Gaussian input distributions the achievable rate is investigated. The main contribution of this paper is the derivation of two new upper bounds on the achievable rate with Gaussian input symbols. One of these bounds is based on the one-step channel prediction error variance but is not restricted to peak power constrained input symbols like known bounds. Moreover, it is shown that Gaussian inputs yield the same *pre-log* as the peak power constrained capacity. The derived bounds are compared with a known lower bound on the capacity given by Deng et al. and with bounds on the peak power constrained capacity given by Sethuraman et al.. Finally, the achievable rate with i.i.d. Gaussian input symbols is compared to the achievable rate using a coherent detection in combination with a solely pilot based channel estimation.**

***Index Terms*—Channel capacity, fading channels, Gaussian distributions, information rates, noncoherent, Rayleigh, time-selective.**

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. In order to acquire channel state information, the temporal correlation of the channel can be exploited by the system, e.g., by inserting training sequences at transmit side. While these training sequences can be understood as a specific type of code [1], we are interested in the achievable rate on this channel irrespective of the use of training sequences.

The capacity of fading channels where the channel state information is unknown, i.e., sometimes referred to as noncoherent capacity, has received a lot of attention in the recent literature. E.g., [2] considers a block fading channel model, where the channel is assumed to be constant over a block of N symbols and changes independently from block to block. This model is non-stationary and, therefore, different from the one we consider in the present work. On the other hand, in [3] and [4] the achievable rate of time-continuous fading channels has been examined under the assumption of the use of training sequences for channel tracking and a coherent detection based on the acquired channel estimate. Furthermore, in [1] the asymptotic high SNR capacity of a stationary Gaussian fading channel has been investigated, whereas in [5] an approximate behavior of the capacity for different SNR regimes has been considered. In addition, in [6] and [7] bounds on the capacity for temporally correlated Rayleigh fading channels with a peak power constraint have been derived with specific emphasis on the low SNR regime. Furthermore, the case of frequency selective stationary fading channels has been discussed, e.g., in [8] and [9].

The main goal of the present work is the investigation of the achievable rate with independent identically distributed (i.i.d.) zero-mean proper Gaussian input symbols — in the following referred to as i.i.d. Gaussian — on noncoherent stationary discrete-time Rayleigh flat-fading channels. The optimal input distribution in the noncoherent case is not Gaussian but becomes often peaky, which makes it impractical for real system designs. Although Gaussian inputs are suboptimal, they are of interest since they serve well to upper-bound the achievable rate of practical modulation and coding schemes in the coherent case [10]. In [11] the achievable rate with i.i.d. Gaussian inputs has been studied for the block fading channel. Furthermore, in [12] bounds on the mutual information with Gaussian input distributions have been derived for a Gauss-Markov fading channel, whose PSD has an unbounded support. The results in [12] indicate that at moderate SNR and/or slow fading, Gaussian inputs still work well.

In the present work we study also the achievable rate for a noncoherent discrete-time stationary Rayleigh flat-fading channel with i.i.d. Gaussian input symbols. In contrast to [12] we assume that the fading process is bandlimited, i.e., the power spectral density (PSD) of the fading process has a normalized maximum Doppler frequency $f_d < 0.5$ (*nonregular*

This work has been supported by the UMIC (Ultra High Speed Mobile Information and Communication) research centre. The work of M. Dörpinghaus was supported in part by the DFG under Grant DO1568/1-1. The material in this paper was presented in part at the International Symposium on Information Theory and its Applications (ISITA) 2008, Auckland, New Zealand, December 2008, and at the International Symposium on Information Theory and its Applications (ISITA) 2010, Taichung, Taiwan, October 2010.

M. Dörpinghaus was with the Institute for Integrated Signal Processing Systems, RWTH Aachen University, 52056 Aachen, Germany and is now with the Institute for Theoretical Information Technology, RWTH Aachen University, 52056 Aachen, Germany (e-mail: doerpinghaus@ti.rwth-aachen.de).

H. Meyr is with the École Polytechnique Fédérale de Lausanne (EPFL), Switzerland, on leave from the Institute for Integrated Signal Processing Systems, RWTH Aachen University, 52056 Aachen, Germany (e-mail: meyr@iss.rwth-aachen.de).

R. Mathar is with the Institute for Theoretical Information Technology, RWTH Aachen University, 52056 Aachen, Germany (e-mail: mathar@ti.rwth-aachen.de).

fading [13]), as it, e.g., corresponds to the widely used Jakes' model [14].

The paper is structured as follows. In Section II, we introduce the channel model. In Section III, we present the first new upper bound on the achievable rate with Gaussian input symbols. The derivation is based on a new lower bound on the conditional output entropy rate $h'(\mathbf{y}|\mathbf{x})$. The lower bound on $h'(\mathbf{y}|\mathbf{x})$ is not based on a peak power constraint and therefor applies for Gaussian input symbols. Together with a known lower bound on the achievable rate with i.i.d. Gaussian input symbols given in [15] we obtain a set of bounds which is tight in the sense that the difference is bounded by $(1+\mu(\{S_h(f) > 0\}))\gamma$ [nat/channel use] for all SNRs, where $\gamma \approx 0.57721$ is the Euler constant, $S_h(f)$ is the normalized PSD of the channel fading process, and $\mu(\cdot)$ is the Lebesgue measure on $[-\frac{1}{2}, \frac{1}{2}]$. In addition, we show that the asymptotic high SNR slope (*pre-log*) of the achievable rate for Gaussian inputs is given by $1 - \mu(\{S_h(f) > 0\})$. This is the same high SNR behavior as it has been observed for the peak power constrained capacity in [1]. Moreover, we compare the upper bound on the achievable rate with i.i.d. Gaussian input symbols to the upper bound on the peak power constrained capacity given in [6].

In Section IV, we present the other main result of this paper. We derive an upper bound on the achievable rate for i.i.d. symbols which is based on the one-step channel prediction error variance. This approach is already known for the peak power constrained capacity [1]. Our derivation makes no use of a peak power constraint and, thus, applies for Gaussian input symbols. However, due to the restriction to i.i.d. input symbols we only get an upper bound on the achievable rate and not on the capacity. We evaluate this upper bound, on the one hand, for Gaussian input symbols and, on the other hand for peak power constrained input symbols.

In Section V, we compare the achievable rate with i.i.d. Gaussian inputs to the achievable rate with synchronized detection, which is of practical interest. Here, synchronized detection means that the channel is estimated solely based on pilot symbols and then, in a second separate step, the channel estimate is used for coherent detection [3], [16, Ch. 4.3.1]. Finally, Section VI concludes the paper.

II. System Model

We consider an ergodic discrete-time jointly proper Gaussian [17] flat-fading channel, whose output at time k is given by

$$y_k = h_k \cdot x_k + n_k \tag{1}$$

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 white Gaussian noise. The processes $\{h_k\}$, $\{x_k\}$, and $\{n_k\}$ are assumed to be mutually independent.

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

$$r_h(l) = \mathrm{E}[h_{k+l} \cdot h_k^*]. \tag{2}$$

The normalized PSD of the channel fading process is defined by

$$S_h(f) = \sum_{l=-\infty}^{\infty} r_h(l) e^{-j2\pi lf}, \quad |f| < 0.5 \tag{3}$$

where $j = \sqrt{-1}$. Here, the frequency f is normalized with respect to the symbol duration. Although $S_h(f)$ is normalized, we refer to it as PSD for simplification. We assume that the PSD exists. For a jointly proper Gaussian process, the existence of the PSD implies ergodicity [18]. As the channel fading process $\{h_k\}$ is assumed to be stationary, $S_h(f)$ is real-valued. Because of the limitation of the velocity of the transmitter, the receiver, and of objects in the environment, the spread of the PSD is limited, and we assume it to be supported within the interval $[-f_d, f_d]$, with $0 < f_d < 0.5$, i.e., $S_h(f) = 0$ for $f \notin [-f_d, f_d]$. The parameter f_d corresponds to the normalized maximum Doppler shift and, thus, indicates the dynamics of the channel. To ensure ergodicity, we exclude the case $f_d = 0$. Following the definition given in [13], this fading channel is sometimes referred to as *nonregular*.

Typical fading channels, as they are observed in mobile communication environments, are characterized by relatively small normalized Doppler frequencies f_d in the regime of $f_d \ll 0.1$. Thus, the restriction to channels with $f_d < 0.5$, i.e., *nonregular* fading, in the present work is reasonable.

The average SNR is given by

$$\rho = \frac{\sigma_x^2 \sigma_h^2}{\sigma_n^2} \tag{4}$$

where σ_x^2 is the maximum average power of the input symbols.

Note that the symbol rate discrete-time model given in (1) has some limitations. With respect to the underlying continuous-time model symbol rate sampling at the receiver does not yield a sufficient statistic. The reason is that the multiplication of the input signal and the fading process in time domain yields a broadening of the PSD of the useful signal components at the receiver to a normalized bandwidth of at least $1 + 2f_d$. See [16, Ch. 4.2] for a detailed discussion regarding the generation of a sufficient statistic. However, as typical systems usually apply symbol rate sampling at the output of the matched filter the symbol rate model is of practical relevance.

Moreover, note that the ergodicity of the channel fading process assures the existence of a coding theorem such that the information theoretic capacity and the operational capacity coincide.

Further on we use the following matrix-vector notation of the system model:

$$\mathbf{y} = \mathbf{X}\mathbf{h} + \mathbf{n} \tag{5}$$

where the vector $\mathbf{h}$ is given by $\mathbf{h} = [h_1, \ldots, h_N]^T$. The vectors $\mathbf{y}$ and $\mathbf{n}$ are defined analogously. The matrix $\mathbf{X}$ is diagonal and defined as $\mathbf{X} = \mathrm{diag}(\mathbf{x})$ with $\mathbf{x} = [x_1, \ldots, x_N]^T$. Here the $\mathrm{diag}(\cdot)$ operator generates a diagonal matrix whose diagonal

elements are given by the argument vector. The quantity N is the number of considered symbols. Later on, we investigate the case of $N \to \infty$ to evaluate the achievable rate.

Using this vector notation, we express the temporal correlation of the fading process by the correlation matrix

$$\mathbf{R}_h = \mathrm{E}[\mathbf{h}\mathbf{h}^H] \tag{6}$$

which has a Hermitian Toeplitz structure.

III. The Achievable Rate with Gaussian Inputs

The main focus of the present paper is the discussion of the achievable rate with Gaussian inputs. In this regard, we derive a new upper bound on the achievable rate

$$\mathcal{I}'(\mathbf{y};\mathbf{x}) = \lim_{N\to\infty} \frac{1}{N}\mathcal{I}(\mathbf{y};\mathbf{x}) \tag{7}$$

which can be applied to arbitrary exchangeable input distributions $p(\mathbf{x})$ with a maximum average power σ_x^2, i.e., to all distributions in the set

$$\mathcal{P}_{\text{ex}} = \Big\{ p(\mathbf{x}) \Big| \mathbf{x} \in \mathbb{C}^N,\ p(x_1,\ldots,x_N) = p(x_{\pi(1)},\ldots,x_{\pi(N)}) \\ \forall \pi \in \mathcal{S}(N),\ \mathrm{E}[|x_k|^2] \le \sigma_x^2\ \forall k \Big\} \tag{8}$$

where $\mathcal{S}(N)$ is the group of permutations of $\{1,\ldots,N\}$. Later on, this bound is given for the special case of zero-mean i.i.d. Gaussian input symbols which are included in the set $\mathcal{P}_{\text{ex}}$.

A. Coherent Mutual Information Rate

The mutual information rate in case the channel is known at the receiver $\mathcal{I}'(\mathbf{y};\mathbf{x}|\mathbf{h})$ — the coherent case — is an upper bound on $\mathcal{I}'(\mathbf{y};\mathbf{x})$ as [19]

$$\mathcal{I}'(\mathbf{y};\mathbf{x}) = \mathcal{I}'(\mathbf{y};\mathbf{x}|\mathbf{h}) - \mathcal{I}'(\mathbf{x};\mathbf{h}|\mathbf{y}). \tag{9}$$

Here, $\mathcal{I}'(\mathbf{x};\mathbf{h}|\mathbf{y})$ is the penalty due to the channel uncertainty. For (9) the independency of $\mathbf{x}$ and $\mathbf{h}$ has been used.

The coherent mutual information rate $\mathcal{I}'(\mathbf{y};\mathbf{x}|\mathbf{h})$ is maximized by i.i.d. zero-mean proper Gaussian inputs yielding[1]

$$\begin{aligned}\mathcal{I}'(\mathbf{y};\mathbf{x}|\mathbf{h}) &= \mathrm{E}_{h_k}\left[\log\left(1+\rho\frac{|h_k|^2}{\sigma_h^2}\right)\right] \\ &= \int_0^\infty \log(1+\rho z)\, e^{-z} dz \\ &= -e^{1/\rho}\mathrm{Ei}(-1/\rho)\end{aligned} \tag{10}$$

with the exponential integral $\mathrm{Ei}(x) = \int_{-\infty}^{x} \frac{e^t}{t} dt$. Note, the capacity of the coherent channel is independent of the temporal correlation of the channel.

B. Upper Bound on the Achievable Rate for Input Symbols with an Exchangeable Distribution

The mutual information rate can be expressed by

$$\mathcal{I}'(\mathbf{y};\mathbf{x}) = h'(\mathbf{y}) - h'(\mathbf{y}|\mathbf{x}) \tag{11}$$

with the differential entropy rate $h'(\cdot) = \lim_{N\to\infty} \frac{1}{N} h(\cdot)$ where $h(\cdot)$ is the differential entropy.

[1]All logarithms in this paper are to the base e and, unless stated otherwise, all rates are in *nat*.

1) Upper Bound on $h'(\mathbf{y})$: As the entropy $h(\mathbf{y})$ of a zero-mean complex random vector is upper-bounded by the entropy of a proper Gaussian random vector with the same covariance matrix [17], it follows that

$$h'(\mathbf{y}) \le \log\left(\pi e\left(\sigma_x^2\sigma_h^2 + \sigma_n^2\right)\right) \tag{12}$$

where we have additionally used Hadamard's inequality.

2) Lower Bound on $h'(\mathbf{y}|\mathbf{x})$: As the probability density of $\mathbf{y}$ conditioned on $\mathbf{x}$ is zero-mean proper Gaussian, the entropy $h(\mathbf{y}|\mathbf{x})$ is given by

$$h(\mathbf{y}|\mathbf{x}) = \mathrm{E}_{\mathbf{x}}\left[\log\left((\pi e)^N \det(\mathbf{R}_{y|x})\right)\right] \tag{13}$$

with

$$\begin{aligned}\mathbf{R}_{y|x} &= \mathrm{E}_{\mathbf{h},\mathbf{n}}\left[\mathbf{y}\mathbf{y}^H\middle|\mathbf{x}\right] \\ &= \mathbf{X}\mathbf{R}_h\mathbf{X}^H + \sigma_n^2\mathbf{I}_N\end{aligned} \tag{14}$$

where $\mathbf{I}_N$ is the identity matrix of size $N \times N$. Substituting (14) into (13) yields

$$\begin{aligned}h(\mathbf{y}|\mathbf{x}) &= \mathrm{E}_{\mathbf{x}}\left[\log\det\left(\pi e\left(\mathbf{X}^H\mathbf{X}\mathbf{R}_h + \sigma_n^2\mathbf{I}_N\right)\right)\right] \\ &= \mathrm{E}_{\mathbf{z}}\left[\log\det\left(\pi e\left(\mathbf{Z}\mathbf{U}^H\mathbf{\Lambda}_h\mathbf{U} + \sigma_n^2\mathbf{I}_N\right)\right)\right] \\ &\ge \mathrm{E}_{\mathbf{z}}\left[\log\det\left(\pi e\left(\mathbf{\Lambda}_h\mathbf{Z} + \sigma_n^2\mathbf{I}_N\right)\right)\right]\end{aligned} \tag{15}$$

where we several times apply the relation $\det(\mathbf{A}\mathbf{B}+\mathbf{I}) = \det(\mathbf{B}\mathbf{A}+\mathbf{I})$ for $\mathbf{A}$ and $\mathbf{B}$ being matrices of appropriate size [20, Theorem 1.3.20]. Moreover, $\mathbf{Z} = \mathbf{X}^H\mathbf{X}$ where $\mathbf{Z} = \mathrm{diag}(\mathbf{z})$ is a diagonal matrix with the elements $[\mathbf{z}]_k = z_k = |x_k|^2$. In addition, we have substituted $\mathbf{R}_h$ by its spectral decomposition

$$\mathbf{R}_h = \mathbf{U}^H\mathbf{\Lambda}_h\mathbf{U} \tag{16}$$

where $\mathbf{U}$ is unitary and $\mathbf{\Lambda}_h = \mathrm{diag}([\lambda_1^{(N)},\ldots,\lambda_N^{(N)}])$ is diagonal containing the eigenvalues $\lambda_k^{(N)}$ of $\mathbf{R}_h$. Here the superscript (N) corresponds to the size of the corresponding matrix $\mathbf{R}_h$.

It remains to proof the inequality in (15) which is equivalent to

$$\begin{aligned}&\mathrm{E}_{\mathbf{z}}\left[\log\det\left(\mathbf{U}\mathbf{Z}\mathbf{U}^H + \sigma_n^2\mathbf{\Lambda}_h^{-1}\right)\right] \\ &\quad\ge \mathrm{E}_{\mathbf{z}}\left[\log\det\left(\mathbf{Z} + \sigma_n^2\mathbf{\Lambda}_h^{-1}\right)\right]\end{aligned} \tag{17}$$

Let

$$F(\mathbf{z}) = -\log\det\left(\mathbf{U}\mathbf{Z}\mathbf{U}^H + \sigma_n^2\mathbf{\Lambda}_h^{-1}\right) \tag{18}$$

with $\mathbf{z} \in [0,\infty)^N$. First, we proof that $F(\mathbf{z})$ is L-superadditive.

A function $F : [0,\infty)^N \to \mathbb{R}$ is called *L-superadditive* if

$$\begin{aligned}&F(\mathbf{z}+h\mathbf{e}_i+k\mathbf{e}_j) - F(\mathbf{z}+h\mathbf{e}_i) - F(\mathbf{z}+k\mathbf{e}_j) + F(\mathbf{z}) \ge 0 \\ &\qquad \forall \mathbf{z} \in [0,\infty)^N,\ \forall 1 \le i < j \le N,\ \forall h,k \in (0,\infty)\end{aligned} \tag{19}$$

with $\mathbf{e}_i$ and $\mathbf{e}_j$ being the i-th and j-th unit vector in $\mathbb{R}^N$, see [21, p. 213 Def. C.2], [22, p. 58].

An equivalent condition for L-superadditivity is the following. For all $\mathbf{z} \in [0,\infty)^N, 1 \le i < j \le N$ and $h \in (0,\infty)$ the function

$$[0,\infty) \ni k \to F(\mathbf{z}+h\mathbf{e}_i+k\mathbf{e}_j) - F(\mathbf{z}+k\mathbf{e}_j) \tag{20}$$

is monotonically increasing.

Let

$$\mathbf{A}(k) = \mathbf{U}\mathbf{Z}\mathbf{U}^H + \sigma_n^2 \mathbf{\Lambda}_h^{-1} + k\mathbf{u}_j\mathbf{u}_j^H \tag{21}$$

with $\mathbf{u}_j$ the j-th column vector of $\mathbf{U}$ such that

$$\mathbf{U}\mathbf{Z}\mathbf{U}^H = \sum_{l=1}^{N} z_l \mathbf{u}_l \mathbf{u}_l^H. \tag{22}$$

For $k \geq 0$ it holds that

$$\begin{aligned} &F(\mathbf{z} + h\mathbf{e}_i + k\mathbf{e}_j) - F(\mathbf{z} + k\mathbf{e}_j) \\ &\quad = -\log\det\left(\mathbf{A}(k) + h\mathbf{u}_i\mathbf{u}_i^H\right) + \log\det\left(\mathbf{A}(k)\right) \\ &\quad = -\log\left(1 + h\mathbf{u}_i^H\mathbf{A}^{-1}(k)\mathbf{u}_i\right) \end{aligned} \tag{23}$$

where for (23) we have used

$$\det\left(\mathbf{A}(k) + h\mathbf{u}_i\mathbf{u}_i^H\right) = \det\left(\mathbf{A}(k)\right)\left(1 + h\mathbf{u}_i^H\mathbf{A}^{-1}(k)\mathbf{u}_i\right). \tag{24}$$

Obviously, $[0,\infty) \ni k \to \mathbf{A}(k)$ is monotonically increasing in the sense of the Loewner partial ordering of Hermitian matrices. As $\mathbf{A}(k)$ is positive definite for all $k \geq 0$ this implies that $[0,\infty) \ni k \to \mathbf{A}^{-1}(k)$ is monotonically decreasing in the sense of the Loewner partial ordering [21, p. 672 E.3.b.]. Hence,

$$[0,\infty) \ni k \to F(\mathbf{z} + h\mathbf{e}_i + k\mathbf{e}_j) - F(\mathbf{z} + k\mathbf{e}_j) \tag{25}$$

is monotonically increasing.

Using [22, Corollary 3, Part A] yields the following inequality

$$\begin{aligned} &\mathrm{E}_{\mathbf{z}}\left[\log\det\left(\mathbf{U}\mathbf{Z}\mathbf{U}^H + \sigma_n^2\mathbf{\Lambda}_h^{-1}\right)\right] \\ &\quad \geq \mathrm{E}_{\mathbf{z}}\left[\log\det\left(\mathbf{U}z_1\mathbf{I}_N\mathbf{U}^H + \sigma_n^2\mathbf{\Lambda}_h^{-1}\right)\right] \\ &\quad = \mathrm{E}_{\mathbf{z}}\left[\log\det\left(z_1\mathbf{I}_N + \sigma_n^2\mathbf{\Lambda}_h^{-1}\right)\right] \\ &\quad = \sum_{k=1}^{N}\mathrm{E}_{\mathbf{z}}\left[\log\left(z_1 + \frac{\sigma_n^2}{\lambda_k^{(N)}}\right)\right] \\ &\quad = \mathrm{E}_{\mathbf{z}}\left[\log\det\left(\mathbf{Z} + \sigma_n^2\mathbf{\Lambda}_h^{-1}\right)\right] \end{aligned} \tag{26}$$

where (26) follows from the fact that the diagonal elements of $\mathbf{Z}$ are exchangeable due to the exchangeability of the elements of $\mathbf{x}$, see (8). Thus, (17) and, hence, the inequality in (15) is proven.

With (15) the differential entropy rate $h'(\mathbf{y}|\mathbf{x})$ is lower-bounded by

$$\begin{aligned} h'(\mathbf{y}|\mathbf{x}) &\geq \lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}\mathrm{E}_{\mathbf{z}}\log\left(\pi e\left(\lambda_k^{(N)}z_k + \sigma_n^2\right)\right) \\ &= \lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}\mathrm{E}_{z_1}\log\left(\pi e\left(\lambda_k^{(N)}z_1 + \sigma_n^2\right)\right) \quad (27) \\ &= \int_{-\frac{1}{2}}^{\frac{1}{2}}\mathrm{E}\left[\log\left(\pi e\left(|x_1|^2 S_h(f) + \sigma_n^2\right)\right)\right]df \quad (28) \end{aligned}$$

where (27) results as $z_1, \ldots, z_N$ are exchangeable. Moreover, (28) follows by Szegö's theorem on the asymptotic eigenvalue distribution of Hermitian Toeplitz matrices [23, pp. 64-65], [24]. This may be seen as follows. Let

$$g:\ \lambda_k^{(N)} \to \mathrm{E}_{z_1}\left[\log\left(\pi e\left(\lambda_k^{(N)}z_1 + \sigma_n^2\right)\right)\right]. \tag{29}$$

The function g is continuous on the finite interval $m \leq \lambda_k^{(N)} \leq M$, where m and M are the essential lower and upper bound of $S_h(f)$, so that Szegö's theorem is applicable. The continuity of g follows from the fact that $\mathrm{E}_{z_1}[\log(\pi e(\lambda_k^{(N)}z_1 + \sigma_n^2))]$ is a continuous function of the argument $\lambda_k^{(N)}$. This may be seen by applying Lebesgue's dominated convergence theorem [25, Theorem 1.34] as

$$\left|\log\left(\pi e\left(\lambda_k^{(N)}z_1 + \sigma_n^2\right)\right)p(z_1)\right| \leq u(z_1) \tag{30}$$

for the integrable function $u(z_1) = |\log(\pi e\sigma_n^2)|p(z_1) + z_1\frac{M}{\sigma_n^2}p(z_1)$.

3) Upper Bound on the Achievable Rate: Using (11), (12), and (28) we are now able to state the following theorem:

Theorem 1. *For all exchangeable input distributions contained in the set $\mathcal{P}_{ex}$ in (8) the achievable rate is upper-bounded by*

$$\mathcal{I}'(\mathbf{y};\mathbf{x}) \leq \log(\rho+1) - \int_{-\frac{1}{2}}^{\frac{1}{2}}\mathrm{E}\left[\log\left(\rho\frac{|x_1|^2}{\sigma_x^2}\frac{S_h(f)}{\sigma_h^2} + 1\right)\right]df \tag{31}$$

where ρ is the average SNR defined in (4).

In particular, for i.i.d. zero-mean proper Gaussian input symbols we obtain:

Lemma 1. *For i.i.d. zero-mean proper Gaussian input symbols and SNR ρ the achievable rate is upper-bounded by*

$$\begin{aligned} &\mathcal{I}'(\mathbf{y};\mathbf{x}) \\ &\leq \log(\rho+1) - \int_{f=-\frac{1}{2}}^{\frac{1}{2}}\int_{z=0}^{\infty}\log\left(\rho z\frac{S_h(f)}{\sigma_h^2} + 1\right)e^{-z}dzdf \\ &= \mathcal{I}'_U(\mathbf{y};\mathbf{x}). \end{aligned} \tag{32}$$

As the mutual information rate in case of perfect channel state information at the receiver $\mathcal{I}'(\mathbf{y};\mathbf{x}|\mathbf{h})$ in (10) always upper-bounds the mutual information rate in the absence of channel state information, we can modify the upper bound in (32) as follows:

$$\mathcal{I}'_{U_{mod}}(\mathbf{y};\mathbf{x}) = \min\{\mathcal{I}'_U(\mathbf{y};\mathbf{x}), \mathcal{I}'(\mathbf{y};\mathbf{x}|\mathbf{h})\}. \tag{33}$$

C. Lower Bound on the Achievable Rate with I.I.D. Gaussian Input Symbols [15]

In [15] the following lower bound on the capacity is given, which is achievable with i.i.d. Gaussian input symbols

$$\begin{aligned} &\mathcal{I}'(\mathbf{y};\mathbf{x}) \\ &\geq \mathcal{I}'(\mathbf{y};\mathbf{x}|\mathbf{h}) - \int_{-\frac{1}{2}}^{\frac{1}{2}}\log\left(\rho\frac{S_h(f)}{\sigma_h^2} + 1\right)df \\ &= \int_0^{\infty}\log(\rho z + 1)\,e^{-z}dz - \int_{-\frac{1}{2}}^{\frac{1}{2}}\log\left(\rho\frac{S_h(f)}{\sigma_h^2} + 1\right)df \\ &= \mathcal{I}'_L(\mathbf{y};\mathbf{x}). \end{aligned} \tag{34}$$

The lower bound corresponds to the mutual information rate in case of perfect channel knowledge at the receiver $\mathcal{I}'(\mathbf{y};\mathbf{x}|\mathbf{h})$ reduced by a term which accounts for the channel uncertainty.

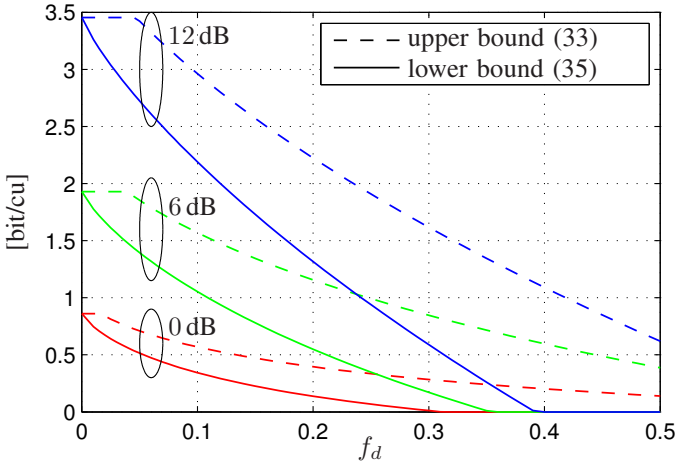

Fig. 1. Upper and lower bound on the achievable rate with i.i.d. Gaussian input symbols on a Rayleigh flat-fading channel with a rectangular $S_h(f)$ in bits per channel use (cu).

For i.i.d. Gaussian input symbols $\mathcal{I}'(\mathbf{y};\mathbf{x}|\mathbf{h})$ corresponds to the coherent capacity given in (10).

As the mutual information rate is nonnegative, we modify the lower bound in (34) as follows:

$$\mathcal{I}'_{L_{mod}}(\mathbf{y};\mathbf{x}) = \max\{\mathcal{I}'_L(\mathbf{y};\mathbf{x}), 0\}. \tag{35}$$

D. Numerical Evaluation

Fig. 1 shows the upper bound (33) and the lower bound (35) on the achievable rate with i.i.d. Gaussian input symbols for a rectangular PSD $S_h(f)$ given by

$$S_h(f)\big|_{\text{Rect}} = \begin{cases} \frac{\sigma_h^2}{2f_d} & \text{for } |f| \le f_d \\ 0 & \text{for } f_d < |f| \le 0.5 \end{cases}. \tag{36}$$

The achievable rate strongly decreases with increasing channel dynamics f_d. Moreover, the gap between the upper and the lower bound increases with the SNR.

E. Tightness of Bounds

The difference between the upper bound and the lower bound on the achievable rate with Gaussian input symbols in (32) and (34) is given by

$$\begin{aligned}\Delta_{\mathcal{I}'(\mathbf{y};\mathbf{x})} &= \mathcal{I}'_U(\mathbf{y};\mathbf{x}) - \mathcal{I}'_L(\mathbf{y};\mathbf{x}) \\ &= \log(\rho+1) - \int_0^\infty \log(\rho z+1)e^{-z}dz \\ &\quad + \int_{f=-\frac{1}{2}}^{\frac{1}{2}} \left[\log\left(\rho\frac{S_h(f)}{\sigma_h^2}+1\right) \right. \\ &\quad \left. - \int_{z=0}^\infty \log\left(\rho\frac{S_h(f)}{\sigma_h^2}z+1\right)e^{-z}dz \right] df. \end{aligned} \tag{37}$$

We first evaluate the difference of the first two terms in (37) given by

$$\Delta_1 = \log(\rho+1) - \int_0^\infty \log(\rho z+1)e^{-z}dz. \tag{38}$$

For $\rho \to 0$ the difference Δ_1 converges to zero, while $\lim_{\rho\to\infty}\Delta_1 = \gamma \approx 0.57721$ [nat/cu], see [26], where γ is the Euler constant. Moreover, Δ_1 monotonically increases with ρ as[2]

$$\begin{aligned}\frac{\partial\Delta_1}{\partial\rho} &= \frac{1}{1+\rho} - \int_0^\infty \frac{z}{1+\rho z}e^{-z}dz && (39)\\ &\ge 0 && (40)\end{aligned}$$

where for (40) we have used that $\frac{z}{1+\rho z}$ is concave in z and, thus, we can apply Jensen's inequality. Thus, Δ_1 is bounded by $0 \le \Delta_1 \le \gamma$.

Lets name the term with the integral over f in (37) Δ_2. As its integrand is structurally similar to Δ_1 (substitute $\rho\frac{S_h(f)}{\sigma_h^2}$ by ρ), it can be shown analogously to Δ_1 that whenever $S_h(f) > 0$ this integrand is bounded between 0 and γ. For $S_h(f) = 0$ the integrand is zero. Thus, it follows that $0 \le \Delta_2 \le \gamma\mu(\{S_h(f) > 0\})$ [nat/cu] where $\mu(\cdot)$ is the Lebesgue measure on the interval $[-\frac{1}{2}, \frac{1}{2}]$. Note that we use here the Lebesgue measure to allow for an arbitrary PSD $S_h(f)$ as defined in (3). In case $S_h(f)$ is supported on $[-f_d, f_d]$, i.e., $S_h(f) > 0\ \forall |f| \le f_d$, it holds that $\mu(\{S_h(f) > 0\}) = 2f_d$.

Hence, $\Delta_{\mathcal{I}'(\mathbf{y};\mathbf{x})}$ is bounded by

$$0 \le \Delta_{\mathcal{I}'(\mathbf{y};\mathbf{x})} \le \gamma(1+\mu(\{S_h(f) > 0\})) \quad \text{[nat/cu]} \tag{41}$$

where the lower limit is achieved for $\rho \to 0$ and

$$\lim_{\rho\to\infty} \Delta_{\mathcal{I}'(\mathbf{y};\mathbf{x})} = \gamma(1+\mu(\{S_h(f) > 0\})). \tag{42}$$

For asymptotically small channel dynamics the lower bound in (34) converges to the mutual information rate in case of perfect channel knowledge (10). This corresponds to the physical interpretation that an arbitrarily slowly changing channel can be estimated arbitrarily well and, thus, the penalty term $\mathcal{I}'(\mathbf{x};\mathbf{h}|\mathbf{y})$ in (9) approaches zero.

F. The Asymptotic High SNR Behavior

The lower bound on the achievable rate with i.i.d. Gaussian input symbols given in (34) is characterized by the following high SNR slope[3], which is often named *pre-log*

$$\begin{aligned}&\lim_{\rho\to\infty} \frac{\partial\mathcal{I}'_L(\mathbf{y};\mathbf{x})}{\partial\log(\rho)} \\ &= \lim_{\rho\to\infty}\left[\int_0^\infty \frac{\rho z}{\rho z+1}e^{-z}dz - \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{\frac{S_h(f)}{\sigma_h^2}\rho}{\frac{S_h(f)}{\sigma_h^2}\rho+1}df\right] && (43)\\ &= 1-\mu(\{S_h(f) > 0\}) && (44)\\ &= \lim_{\rho\to\infty} \frac{\partial\mathcal{I}'(\mathbf{y};\mathbf{x})}{\partial\log(\rho)} && (45)\end{aligned}$$

[2]Note that the change of the order of differentiation and integration required in (39) is possible as can be shown using Lebesgue's dominated convergence theorem [25, Theorem 1.34].

[3]When using the term *high SNR slope* we refer to the high SNR limit of the derivative of the achievable rate (bound) with respect to the logarithm of the SNR.

where $\mu(\cdot)$ again is the Lebesgue measure on $[-\frac{1}{2},\frac{1}{2}]$.[4] As the gap between the upper and the lower bound $\Delta_{\mathcal{I}'(\mathbf{y};\mathbf{x})}$ in (37) converges to a constant for high SNR, cf. (42), the slope of the actual achievable rate $\mathcal{I}'(\mathbf{y};\mathbf{x})$ is also given by $1-\mu(\{S_h(f)>0\})$ yielding (45). This corresponds to the pre-log of the peak power constrained capacity for *nonregular* fading [1]. In case $S_h(f)$ is supported on the interval $[-f_d, f_d]$ the pre-log of the achievable rate is, thus, given by $1-2f_d$, clearly showing the degradation with increasing channel dynamics.

G. Comparison to the Upper Bound on the Peak Power Constrained Capacity in [6]

In [6, Prop. 2.2] the following upper bound for the peak power constrained capacity with $\frac{1}{N}\mathrm{E}\left[\mathbf{x}^H\mathbf{x}\right] \leq \sigma_x^2$ and $|x_k|^2 \leq P_{\text{peak}}$ is given

$$C_{\text{peak}} \leq \log\left(\alpha_{\text{opt}}\rho+1\right) - \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{\alpha_{\text{opt}}}{\beta}\log\left(\frac{\rho\beta}{\sigma_h^2}S_h(f)+1\right)df \tag{46}$$

with

$$\alpha_{\text{opt}} = \min\left\{1, \left(\frac{1}{\beta}\int_{-\frac{1}{2}}^{\frac{1}{2}}\log\left(\frac{\rho\beta}{\sigma_h^2}S_h(f)+1\right)df\right)^{-1} - \frac{1}{\rho}\right\} \tag{47}$$

and with the *nominal peak-to-average power ratio* $\beta = \frac{P_{\text{peak}}}{\sigma_x^2}$. Note that in case of a peak power constraint it is not necessarily optimal to use the maximum average transmit power σ_x^2. Thus, in case of a peak power constraint the SNR ρ is actually a *nominal average SNR*.

Applying the peak power constraint to Theorem 1 yields the same upper bound as in (46) in terms of the expression. However, the derivation of Theorem 1 requires the restriction to input symbols with an exchangeable distribution which is not required in [6, Prop. 2.2].

The high SNR pre-log of (46) is given by $1-\mu(\{S_h(f)>0\})/\beta$, which is higher than the capacity pre-log of $1-\mu(\{S_h(f)>0\})$, showing that for large peak-to-average power ratios β and high SNRs the upper bound on the peak power constrained capacity becomes loose. This intuitively demonstrates the relevance of the upper bound on the achievable rate with i.i.d. Gaussian input symbols.

IV. Alternative Upper Bound on the Achievable Rate with I.I.D. Input Symbols Based on the One-Step Channel Prediction Error Variance

In the present section, we give further a new upper bound on the achievable rate with i.i.d. input symbols which is based on the channel prediction error variance and is like the upper bound given in Section III-B not restricted to peak power constrained input symbols. Differently, for the derivation of the channel prediction based capacity bounds in [1] the peak power constraint has been required for technical reasons.

In the first part of the following derivation, we only restrict to i.i.d. input symbols with a maximum average power constraint, i.e., the input distribution is contained in the set

$$\mathcal{P}_{\text{i.i.d.}} = \Bigg\{p(\mathbf{x})\Bigg|\mathbf{x}\in\mathbb{C}^N,\ p(\mathbf{x}) = \prod_{i=1}^{N} p(x_i), \\ p(x_i) = p(x_j)\ \forall i,j,\ \mathrm{E}[|x_k|^2] \leq \sigma_x^2 \quad \forall k\Bigg\}. \tag{48}$$

A. An Upper Bound based on the Channel Prediction Error Variance

Corresponding to Section III-B, we express $\mathcal{I}'(\mathbf{y};\mathbf{x})$ based on the separation in (11).

1) Upper Bound on $h'(\mathbf{y})$: For the term $h'(\mathbf{y})$ we apply the upper bound given in (12), i.e.

$$h'(\mathbf{y}) \leq \log\left(\pi e\left(\alpha\sigma_x^2\sigma_h^2+\sigma_n^2\right)\right) \tag{49}$$

where we additionally introduce the factor $\alpha\in[0,1]$, enabling to choose average transmit powers smaller than the maximum average transmit power σ_x^2.

2) The Entropy Rate $h'(\mathbf{y}|\mathbf{x})$: Our aim is to express $h'(\mathbf{y}|\mathbf{x})$ based on the channel prediction error variance. As the fading channel is stationary and ergodic we get for i.i.d. input symbols

$$\begin{aligned} h'(\mathbf{y}|\mathbf{x}) &= \lim_{N\to\infty}\frac{1}{N}h(\mathbf{y}|\mathbf{x}) \\ &= \lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}h(y_k|\mathbf{x},\mathbf{y}_1^{k-1}) &(50)\\ &= \lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}h(y_k|\mathbf{x}_1^{k},\mathbf{y}_1^{k-1}) &(51)\\ &= \lim_{N\to\infty}h(y_N|\mathbf{x}_1^{N},\mathbf{y}_1^{N-1}) &(52)\end{aligned}$$

where 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 (50) we have used the chain rule for differential entropy, (51) 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. For equality (52) we have used the Cesáro mean based on the convergence of $h(y_k|\mathbf{x}_1^k,\mathbf{y}_1^{k-1})$ to a fixed value for $k\to\infty$ due to the stationarity of the channel fading process and the assumption on independent transmit symbols, see [27, Theorem 4.2.3].

As $p(\mathbf{y}|\mathbf{x})$ is proper Gaussian, 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 its conditional variance given by

$$\begin{aligned}\mathrm{E}\left[y_N|\mathbf{x}_1^N,\mathbf{y}_1^{N-1}\right] &= \mathrm{E}\left[x_Nh_N+n_N|\mathbf{x}_1^N,\mathbf{y}_1^{N-1}\right] \\ &= x_N\mathrm{E}\left[h_N|\mathbf{x}_1^{N-1},\mathbf{y}_1^{N-1}\right] \\ &= x_N\hat{h}_N &(53)\end{aligned}$$

[4] Using Lebesgue's dominated convergence theorem [25, Theorem 1.34] one can show that the change of the order of differentiation and integration required to get equality (43) and the change of the order of the limit and the integration required to derive equality (44) is allowed.

$$\begin{aligned}
&\operatorname{var}\left[y_N|\mathbf{x}_1^N,\mathbf{y}_1^{N-1}\right]\\
&\quad=\mathrm{E}\left[\left|y_N-\mathrm{E}\left[y_N|\mathbf{x}_1^N,\mathbf{y}_1^{N-1}\right]\right|^2\left|\mathbf{x}_1^N,\mathbf{y}_1^{N-1}\right.\right]\\
&\quad=\mathrm{E}\left[\left|x_N(h_N-\hat{h}_N)+n_N\right|^2\left|\mathbf{x}_1^N,\mathbf{y}_1^{N-1}\right.\right]\\
&\quad=|x_N|^2\sigma^2_{e_{\text{pred}}}(\mathbf{x}_1^{N-1})+\sigma_n^2 \qquad (54)
\end{aligned}$$

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. 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). Furthermore, $\sigma^2_{e_{\text{pred}}}(\mathbf{x}_1^{N-1})$ is the prediction error variance of the MMSE estimate $\hat{h}_N$ given by

$$\begin{aligned}
\sigma^2_{e_{\text{pred}}}(\mathbf{x}_1^{N-1})&=\mathrm{E}\left[\left|h_N-\hat{h}_N\right|^2\left|\mathbf{x}_1^{N-1},\mathbf{y}_1^{N-1}\right.\right]\\
&=\mathrm{E}\left[|e_N|^2\left|\mathbf{x}_1^{N-1},\mathbf{y}_1^{N-1}\right.\right]\\
&=\mathrm{E}\left[|e_N|^2\left|\mathbf{x}_1^{N-1}\right.\right] \qquad (55)
\end{aligned}$$

with the prediction error $e_N = h_N - \hat{h}_N$. For (55) we have used the fact that the zero-mean estimation error e_N is orthogonal to and, thus, independent of the observations $\mathbf{y}_1^{N-1}$ as both are proper Gaussian. However, the prediction error variance depends on the input symbols $\mathbf{x}_1^{N-1}$, which is indicated by writing $\sigma^2_{e_{\text{pred}}}(\mathbf{x}_1^{N-1})$.

Based on the channel prediction error variance, we can rewrite the entropy $h(y_N|\mathbf{x}_1^N,\mathbf{y}_1^{N-1})$ as

$$\begin{aligned}
h(y_N|\mathbf{x}_1^N,\mathbf{y}_1^{N-1})&=\mathrm{E}_{\mathbf{x}}\left[\log\left(\pi e\ \operatorname{var}\left[y_N|\mathbf{x}_1^N,\mathbf{y}_1^{N-1}\right]\right)\right]\\
&=\mathrm{E}_{\mathbf{x}}\left[\log\left(\pi e\left(\sigma_n^2+\sigma^2_{e_{\text{pred}}}(\mathbf{x}_1^{N-1})|x_N|^2\right)\right)\right]. \qquad (56)
\end{aligned}$$

With (52) and (56), we get for i.i.d. input symbols

$$h'(\mathbf{y}|\mathbf{x})=\mathrm{E}_{x_k}\left[\mathrm{E}_{\mathbf{x}_{-\infty}^{k-1}}\left[\log\left(\pi e\left(\sigma_n^2+\sigma^2_{e_{\text{pred},\infty}}(\mathbf{x}_{-\infty}^{k-1})|x_k|^2\right)\right)\right]\right] \qquad (57)$$

where $\sigma^2_{e_{\text{pred},\infty}}(\mathbf{x}_{-\infty}^{k-1})$ denotes the prediction error variance in (55) for an infinite number of channel observations in the past. 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 (11), (49), and (57), we can give the following upper bound on the achievable rate with i.i.d. input symbols:

$$\begin{aligned}
\mathcal{I}'(\mathbf{y};\mathbf{x})\leq\ &\log(\alpha\rho+1)\\
&-\mathrm{E}_{x_k}\left[\mathrm{E}_{\mathbf{x}_{-\infty}^{k-1}}\left[\log\left(1+\frac{\sigma^2_{e_{\text{pred},\infty}}(\mathbf{x}_{-\infty}^{k-1})}{\sigma_n^2}|x_k|^2\right)\right]\right] \qquad (58)
\end{aligned}$$

where ρ is the average SNR, see (4). Obviously, the upper bound in (58) still depends on the channel prediction error variance $\sigma^2_{e_{\text{pred},\infty}}(\mathbf{x}_{-\infty}^{k-1})$, 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 $\sigma^2_{e_{pred,\infty}}(\mathbf{x}_{-\infty}^{k-1})$: The prediction error variance $\sigma^2_{e_{\text{pred},\infty}}(\mathbf{x}_{-\infty}^{k-1})$ depends on the distribution of the input symbols $\mathbf{x}_{-\infty}^{k-1}$. To construct an upper bound on the RHS of (58), 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 (58). 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, for the case of a finite past time horizon it is given by

$$\sigma^2_{e_{\text{pred}}}(\mathbf{x}_1^{N-1})=\sigma_h^2-\mathbf{r}^H_{\mathbf{y}_1^{N-1}h_N|\mathbf{x}_1^{N-1}}\mathbf{R}^{-1}_{\mathbf{y}_1^{N-1}|\mathbf{x}_1^{N-1}}\mathbf{r}_{\mathbf{y}_1^{N-1}h_N|\mathbf{x}_1^{N-1}} \qquad (59)$$

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, i.e.,

$$\begin{aligned}
\mathbf{R}_{\mathbf{y}_1^{N-1}|\mathbf{x}_1^{N-1}}&=\mathrm{E}\left[\mathbf{y}_1^{N-1}(\mathbf{y}_1^{N-1})^H\left|\mathbf{x}_1^{N-1}\right.\right]\\
&=\mathbf{X}_{N-1}\mathbf{R}_h\mathbf{X}_{N-1}^H+\sigma_n^2\mathbf{I}_{N-1} \qquad (60)
\end{aligned}$$

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

$$\mathbf{R}_h=\mathrm{E}\left[\mathbf{h}_1^{N-1}(\mathbf{h}_1^{N-1})^H\right] \qquad (61)$$

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

$$\begin{aligned}
\mathbf{r}_{\mathbf{y}_1^{N-1}h_N|\mathbf{x}_1^{N-1}}&=\mathrm{E}\left[\mathbf{y}_1^{N-1}h_N^*\left|\mathbf{x}_1^{N-1}\right.\right]\\
&=\mathbf{X}_{N-1}\mathbf{r}_{h,\text{pred}} \qquad (62)
\end{aligned}$$

where $\mathbf{r}_{h,\text{pred}}=[r_h(-(N-1))\dots r_h(-1)]^T$ with the autocorrelation function $r_h(l)$ defined in (2).

Substituting (60) and (62) into (59) yields

$$\begin{aligned}
&\sigma^2_{e_{\text{pred}}}(\mathbf{x}_1^{N-1})\\
&=\sigma_h^2\\
&\quad-\mathbf{r}^H_{h,\text{pred}}\mathbf{X}^H_{N-1}\left(\mathbf{X}_{N-1}\mathbf{R}_h\mathbf{X}^H_{N-1}+\sigma_n^2\mathbf{I}_{N-1}\right)^{-1}\mathbf{X}_{N-1}\mathbf{r}_{h,\text{pred}}\\
&=\sigma_h^2-\mathbf{r}^H_{h,\text{pred}}\left(\mathbf{R}_h+\sigma_n^2\left(\mathbf{X}^H_{N-1}\mathbf{X}_{N-1}\right)^{-1}\right)^{-1}\mathbf{r}_{h,\text{pred}}\\
&=\sigma_h^2-\mathbf{r}^H_{h,\text{pred}}\left(\mathbf{R}_h+\sigma_n^2\mathbf{Z}^{-1}\right)^{-1}\mathbf{r}_{h,\text{pred}} \qquad (63)
\end{aligned}$$

where for (63) we have used $\mathbf{Z}=\mathbf{X}^H_{N-1}\mathbf{X}_{N-1}$. For ease of notation, we omit the index $N-1$.[5]

Recall that we want to derive an upper bound on the achievable rate with i.i.d. input symbols by maximizing the RHS of (58) over all i.i.d. distributions of the transmit symbols in the past with an average power $\alpha\sigma_x^2$. Obviously, the distribution

[5]Note that the inverse of $\mathbf{Z}$ in (63) does not exist if a diagonal element z_i of the diagonal matrix $\mathbf{Z}$ is zero, i.e., one transmit symbol has zero power. However, as the prediction error variance is continuous in $z_i=0$ for all i this does not lead to problems in the following derivation.

of the phases of the past transmit symbols $\mathbf{x}_1^{N-1}$ has no influence on the 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 (58) is maximized. In the following, we will show that the RHS of (58) 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 (58) 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 (58) (but here for the case of a finite past time horizon) with (63), i.e.,

$$\log\left(1+\frac{|x_N|^2}{\sigma_n^2}\left(\sigma_h^2-\mathbf{r}_{h,\text{pred}}^H\left(\mathbf{R}_h+\sigma_n^2\mathbf{Z}^{-1}\right)^{-1}\mathbf{r}_{h,\text{pred}}\right)\right) \tag{64}$$

is convex with respect to each individual element z_i of the diagonal of $\mathbf{Z}$, which we name $\mathbf{z}$. See Appendix A for the proof. As the z_i are i.i.d. we can apply Jensen's inequality for each individual z_i yielding

$$\begin{aligned}
&\mathrm{E}_{\mathbf{z}}\left[\log\left(1+\frac{|x_N|^2}{\sigma_n^2}\left(\sigma_h^2-\mathbf{r}_{h,\text{pred}}^H\left(\mathbf{R}_h+\sigma_n^2\mathbf{Z}^{-1}\right)^{-1}\mathbf{r}_{h,\text{pred}}\right)\right)\right]\\
&\geq \log\left(1+\frac{|x_N|^2}{\sigma_n^2}\right.\\
&\qquad\left.\times\left(\sigma_h^2-\mathbf{r}_{h,\text{pred}}^H\left(\mathbf{R}_h+\sigma_n^2\left(\mathrm{E}_{\mathbf{z}}\left[\mathbf{Z}\right]\right)^{-1}\right)^{-1}\mathbf{r}_{h,\text{pred}}\right)\right)\\
&=\log\left(1+\frac{|x_N|^2}{\sigma_n^2}\right.\\
&\qquad\left.\times\left(\sigma_h^2-\mathbf{r}_{h,\text{pred}}^H\left(\mathbf{R}_h+\frac{\sigma_n^2}{\alpha\sigma_x^2}\mathbf{I}_{N-1}\right)^{-1}\mathbf{r}_{h,\text{pred}}\right)\right)\\
&=\log\left(1+\frac{|x_N|^2}{\sigma_n^2}\sigma^2_{e_{\text{pred,CM}}}\right)
\end{aligned} \tag{65}$$

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 $\alpha\sigma_x^2$, cf. (63). Here, the index CM denotes constant modulus.

For an arbitrarily long past we can conclude that the RHS of (58) is upper-bounded by

$$\mathcal{I}'(\mathbf{y};\mathbf{x}) \leq \log(\alpha\rho+1) - \mathrm{E}_{x_k}\left[\log\left(1+\frac{\sigma^2_{e_{\text{pred,CM},\infty}}}{\sigma_n^2}|x_k|^2\right)\right] \tag{66}$$

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 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. It is given by [1]

$$\sigma^2_{e_{\text{pred,CM},\infty}} = \frac{\sigma_n^2}{\alpha\sigma_x^2}\left\{\exp\left(\int_{-\frac{1}{2}}^{\frac{1}{2}}\log\left(1+\frac{\alpha\sigma_x^2}{\sigma_n^2}S_h(f)\right)df\right)-1\right\}. \tag{67}$$

As far as we know, the upper bound on the achievable rate in (66) 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. We have shown that the achievable rate with i.i.d. input symbols 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. Gaussian input symbols.

5) Gaussian Input Symbols: For i.i.d. proper Gaussian (PG) input symbols with (66) we get

$$\begin{aligned}
&\mathcal{I}'(\mathbf{y};\mathbf{x})\big|_{\text{PG}}\\
&\leq \sup_{\alpha\in[0,1]}\left\{\log(\alpha\rho+1)\right.\\
&\qquad\left.-\int_0^\infty\log\left(1+\frac{\sigma^2_{e_{\text{pred,CM},\infty}}}{\sigma_h^2}\alpha\rho z\right)e^{-z}dz\right\} \qquad (68)\\
&\leq \sup_{\alpha\in[0,1]}\left\{\log(\alpha\rho+1)\right.\\
&\qquad\left.-\int_0^\infty\log\left(1+\frac{\sigma^2_{e_{\text{pred,CM},\infty}}\big|_{\alpha=1}}{\sigma_h^2}\alpha\rho z\right)e^{-z}dz\right\} \qquad (69)\\
&=\log(\rho+1)-\int_0^\infty\log\left(1+\frac{\sigma^2_{e_{\text{pred,CM},\infty}}\big|_{\alpha=1}}{\sigma_h^2}\rho z\right)e^{-z}dz \qquad (70)
\end{aligned}$$

where (69) is based on the fact that $\sigma^2_{e_{\text{pred,CM},\infty}}$ monotonically decreases with an increasing α, and, thus, that the term on the RHS of (68) is maximized if the prediction error variance is calculated for $\alpha = 1$, which is denoted by writing $\sigma^2_{e_{\text{pred,CM},\infty}}\big|_{\alpha=1}$. Furthermore, (70) follows from the monotonicity of the argument of the supremum in (69) in α, which can be shown analogously to the monotonicity of Δ_1 in (40). In conclusion, the upper bound for Gaussian input symbols is maximized for the maximum average transmit power σ_x^2. Based on this and in combination with the coherent upper bound, cf. (33), we are able to state the following theorem:

Theorem 2. *For i.i.d. zero-mean proper Gaussian input symbols and SNR ρ the achievable rate is upper-bounded by*

$$\begin{aligned}
\mathcal{I}'(\mathbf{y};\mathbf{x}) \leq \min\Bigg\{&\mathcal{I}'(\mathbf{x};\mathbf{y}|\mathbf{h}),\ \log(\rho+1)\\
&-\int_0^\infty\log\left(1+\frac{\sigma^2_{e_{\text{pred,CM},\infty}}\big|_{\alpha=1}}{\sigma_h^2}\rho z\right)e^{-z}dz\Bigg\}
\end{aligned} \tag{71}$$

with $\mathcal{I}'(\mathbf{x};\mathbf{y}|\mathbf{h})$ given in (10) and $\sigma^2_{e_{\text{pred,CM},\infty}}\big|_{\alpha=1}$ given by (67) with $\alpha = 1$.

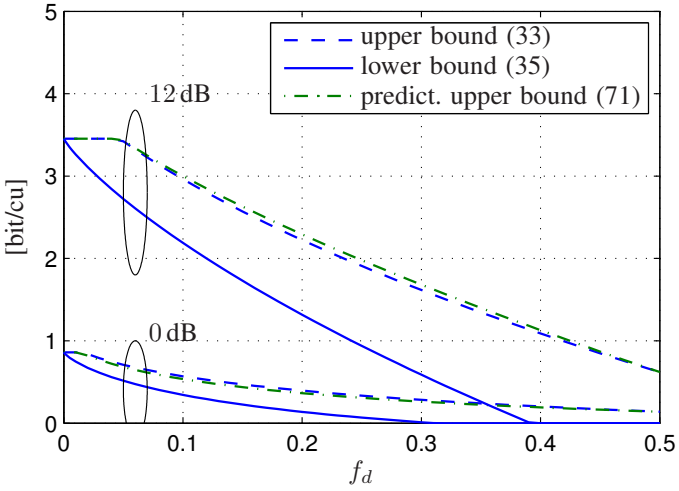

Fig. 2. Comparison of upper bound on the achievable rate with i.i.d. Gaussian inputs based on channel prediction (71) with upper bound (33) and lower bound (35) on the achievable rate with i.i.d. Gaussian inputs; rectangular $S_h(f)$.

Fig. 2 shows the prediction based upper bound on the achievable rate given in (71) in comparison to the upper and lower bound given by (33) and (35). A comparison of the prediction based upper bound (71) and the bound given in (33) shows, that it depends on the channel parameters which one is tighter. It can be shown that for $f_d \to 0$ and for $f_d = 0.5$ both bounds, i.e., (32) and (70), are equal. For other f_d it depends on the SNR ρ which bound is tighter. An analytical comparison turns out to be difficult due to the different ways of lower-bounding $h'(\mathbf{y}|\mathbf{x})$.

6) Peak Power Constrained Input Distributions: For peak power constrained input symbols with the nominal peak-to-average power ratio $\beta = P_{\text{peak}}/\sigma_x^2$, with (66) we get

$$\begin{aligned}
&\sup_{\mathcal{P}_{\text{i.i.d.}}^{\text{peak}}} \mathcal{I}'(\mathbf{y};\mathbf{x}) \\
&\leq \sup_{\alpha\in[0,1]} \sup_{\mathcal{P}_{\text{i.i.d.}}^{\text{peak}}|\alpha} \Bigg\{ \log(\alpha\rho+1) \\
&\qquad\qquad - \mathrm{E}_{x_k}\left[\log\left(1+\frac{\sigma^2_{e_{\text{pred,CM},\infty}}}{\sigma_n^2}|x_k|^2\right)\right]\Bigg\} \\
&= \sup_{\alpha\in[0,1]} \Bigg\{ \log(\alpha\rho+1) \\
&\qquad\qquad - \inf_{\mathcal{P}_{\text{i.i.d.}}^{\text{peak}}|\alpha} \mathrm{E}_{x_k}\left[\log\left(1+\frac{\sigma^2_{e_{\text{pred,CM},\infty}}}{\sigma_n^2}|x_k|^2\right)\right]\Bigg\} \\
&= \sup_{\alpha\in[0,1]} \left\{ \log(\alpha\rho+1) - \frac{\alpha}{\beta}\log\left(1+\frac{\sigma^2_{e_{\text{pred,CM},\infty}}}{\sigma_h^2}\rho\beta\right)\right\}
\end{aligned} \tag{72}$$

where $\mathcal{P}_{\text{i.i.d.}}^{\text{peak}}$ corresponds to $\mathcal{P}_{\text{i.i.d.}}$ in (48) but with the additional peak power constraint $|x_k|^2 \leq \beta\sigma_x^2$. Furthermore, $\mathcal{P}_{\text{i.i.d.}}^{\text{peak}}|\alpha$ corresponds to $\mathcal{P}_{\text{i.i.d.}}^{\text{peak}}$ but with the average transmit power fixed to $\alpha\sigma_x^2$, i.e., $\mathrm{E}[|x_k|^2] = \alpha\sigma_x^2$. Equality (72) follows from

$$\begin{aligned}
&\inf_{\mathcal{P}_{\text{i.i.d.}}^{\text{peak}}|\alpha} \mathrm{E}_{x_k}\left[\log\left(1+\frac{\sigma^2_{e_{\text{pred,CM},\infty}}}{\sigma_n^2}|x_k|^2\right)\right] \\
&= \inf_{\mathcal{P}_{\text{i.i.d.}}^{\text{peak}}|\alpha} \int_0^{\sqrt{P_{\text{peak}}}} \frac{\log\left(1+\frac{\sigma^2_{e_{\text{pred,CM},\infty}}}{\sigma_n^2}z^2\right)}{z^2} z^2 p_{|x_k|}(z)dz
\end{aligned} \tag{73}$$

$$\geq \frac{\log\left(1+\frac{\sigma^2_{e_{\text{pred,CM},\infty}}}{\sigma_n^2}P_{\text{peak}}\right)}{P_{\text{peak}}} \inf_{\mathcal{P}_{\text{i.i.d.}}^{\text{peak}}|\alpha} \int_0^{\sqrt{P_{\text{peak}}}} z^2 p_{|x_k|}(z)dz \tag{74}$$

$$= \frac{\log\left(1+\frac{\sigma^2_{e_{\text{pred,CM},\infty}}}{\sigma_n^2}P_{\text{peak}}\right)}{P_{\text{peak}}}\alpha\sigma_x^2 \tag{75}$$

where for (74) we have used that all factors of the integrand are positive and that the term

$$\frac{1}{|x_k|^2}\log\left(1+\frac{\sigma^2_{e_{\text{pred,CM},\infty}}}{\sigma_n^2}|x_k|^2\right) = \frac{1}{z}\log(1+cz) \tag{76}$$

with $c = \frac{\sigma^2_{e_{\text{pred,CM},\infty}}}{\sigma_n^2}$ and $z = |x_k|^2$ is monotonically decreasing in z as

$$\begin{aligned}
\frac{\partial}{\partial z}\left\{\frac{1}{z}\log(1+cz)\right\} &= \frac{c}{(1+cz)z} - \frac{\log(1+cz)}{z^2} < 0 \\
&\Leftrightarrow \frac{cz}{1+cz} < \log(1+cz)
\end{aligned} \tag{77}$$

which holds for $cz > -1$. Thus, the term in (76) is minimized for $z = |x_k|^2 = P_{\text{peak}}$. The RHS of (75) is achievable with on-off keying and, thus, (74) holds with equality, yielding (72). An approach similar to the one used to calculate the infimum in (73) has been applied in [28] for an analogous problem.

Note that the prediction error variance $\sigma^2_{e_{\text{pred,CM},\infty}}$ in (72) depends on α. Now, we would have to calculate the supremum in (72) 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, as $\sigma^2_{e_{\text{pred,CM},\infty}}$ monotonically decreases with an increasing α, and as (72) monotonically increases with a decreasing $\sigma^2_{e_{\text{pred,CM},\infty}}$, we can upper-bound (72) by setting $\alpha = 1$ within $\sigma^2_{e_{\text{pred,CM},\infty}}$ in (67), i.e.,

$$\begin{aligned}
&\sup_{\mathcal{P}_{\text{i.i.d.}}^{\text{peak}}} \mathcal{I}'(\mathbf{y};\mathbf{x}) \\
&\leq \sup_{\alpha\in[0,1]} \left\{ \log(\alpha\rho+1) - \frac{\alpha}{\beta}\log\left(1+\frac{\sigma^2_{e_{\text{pred,CM},\infty}}\big|_{\alpha=1}}{\sigma_h^2}\rho\beta\right)\right\}.
\end{aligned} \tag{78}$$

The argument of the supremum in (78) is concave in α, and, thus, there exists a unique maximum. This allows us to state the following theorem:

Theorem 3. *The achievable rate with peak power constrained i.i.d. input symbols is bounded by*

$$\mathcal{I}'(\mathbf{y};\mathbf{x}) \leq \log(\alpha_{opt}\rho+1) - \frac{\alpha_{opt}}{\beta}\log\left(1+\frac{\sigma^2_{e_{pred,CM,\infty}}\big|_{\alpha=1}}{\sigma_h^2}\rho\beta\right) \tag{79}$$

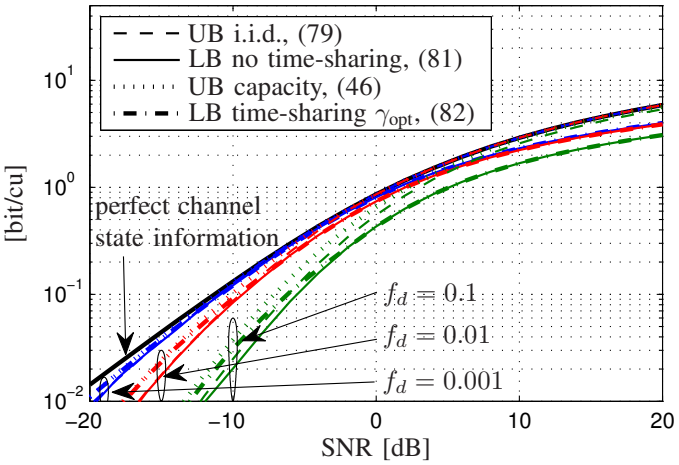

Fig. 3. Comparison of upper bound on the achievable rate with i.i.d. symbols and a peak power constraint based on channel prediction (79) to upper bound on the peak power constrained capacity (46); $\beta = 2$; all upper bounds plotted in combination with coherent upper bound (10), cf. (33); lower bound on the peak power constrained capacity for a constant modulus (CM) input distribution with 100 signaling points, without (81) and with time-sharing (82) (γ_{opt}); rectangular $S_h(f)$.

with

$$\alpha_{opt} = \min\left\{1, \left(\frac{1}{\beta}\log\left(1 + \frac{\sigma^2_{e_{pred,CM,\infty}}\big|_{\alpha=1}}{\sigma_h^2}\rho\beta\right)\right)^{-1} - \frac{1}{\rho}\right\}. \tag{80}$$

a) Comparison to Capacity Bounds in [6] and [7]: In the following, we compare the upper bound on the achievable rate with peak power constrained i.i.d. input symbols in (79) with the upper bound on the peak power constrained capacity given in [6, Prop. 2.2], see (46), and with the lower bound on the peak power constrained capacity given in [7, (35)]:

$$C_{l1}(\rho) = h(y_k|\hat{h}_k) - \int_{-\frac{1}{2}}^{\frac{1}{2}} \log\left(\pi e \sigma_n^2\left(1 + \rho\frac{S_h(f)}{\sigma_h^2}\right)\right) df \tag{81}$$

where k is an arbitrary chosen time instant with an infinitely long past and $h(y_k|\hat{h}_k)$ is the differential output entropy while conditioning on the channel estimate $\hat{h}_k$, being given by the MMSE estimate $\mathrm{E}\left[h_k\middle|\mathbf{x}_{-\infty}^{k-1}, \mathbf{y}_{-\infty}^{k-1}\right]$. Based on a time-sharing argumentation an enhanced lower bound on the peak power constrained capacity is given by [7, (29)/(35)]

$$C \geq \max_{\gamma\in[1,\beta]} \frac{1}{\gamma} C_{l1}(\rho\gamma). \tag{82}$$

b) Numerical Evaluation: Fig. 3 shows the upper bound on the achievable rate with i.i.d. input symbols and a peak power constraint in (79) in comparison to the upper bound on the peak power constrained capacity given in [6, Prop. 2.2], i.e., (46). Although not explicitly denoted, both upper bounds are combined with the coherent upper bound $\mathcal{I}'(\mathbf{y};\mathbf{x}|\mathbf{h})$ in (10), cf. (33). For comparison the lower bounds on the peak power constrained capacity without time-sharing (81) and with time-sharing (γ_{opt}) (82) are shown. The lower bound in (81) is achievable with constant modulus input symbols with a uniformly distributed phase. 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 without time-sharing matches the new upper bound on the achievable rate with i.i.d. input symbols in (79), while the lower bound with time-sharing (γ_{opt}) only matches the capacity upper bound in (46). From Fig. 3 it can be seen that the upper bound on the achievable rate with i.i.d. input symbols in (79) is lower or equal than the capacity upper bound in (46). However, (79) is only an upper bound on the achievable rate with i.i.d. input symbols and not an upper bound on the capacity, as i.i.d. input symbols are in general not capacity-achieving [6]. 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 (79) 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 (79) and the one given in [6, Prop. 2.2], i.e., (46), coincide.

V. Comparison to Synchronized Detection with Pilot Based Channel Estimation

In typical mobile communication systems periodical pilot symbols are introduced into the transmit data sequence. The pilot symbol spacing L is chosen such that the channel fading process is sampled at least with Nyquist frequency, i.e., $L < \lfloor 1/(2f_d)\rfloor$. Based on these pilot symbols the channel is estimated, allowing for a coherent detection (synchronized detection) [16, Ch. 4.3.1]. As coherent detection enables for a decoding complexity per symbol that is independent of the sequence length for long sequences, it is implemented in almost all practical systems.

In conventional receivers, the channel estimation and the detection/decoding are two separate steps, such that the channel is estimated solely based on the pilot symbols. The resulting channel estimation error process is temporally correlated. However, performing coherent detection, the information contained in this temporal correlation is discarded. For a detailed discussion on this, we refer to [29]. Bounds on the achievable rate for this separate processing have been given in [3]. For i.i.d. Gaussian data symbols these bounds become

$$\mathcal{R}_{\text{sep}} \geq \mathcal{R}_{L,\text{sep}} = \frac{L-1}{L}\int_0^\infty \log\left(1 + \rho\frac{1 - \frac{\sigma^2_{e_{\text{pil}}}}{\sigma_h^2}}{1 + \rho\frac{\sigma^2_{e_{\text{pil}}}}{\sigma_h^2}}z\right)e^{-z}dz \tag{83}$$

$$\mathcal{R}_{\text{sep}} \leq \mathcal{R}_{U,\text{sep}} = \mathcal{R}_{L,\text{sep}} + \frac{L-1}{L}\left(\log\left(1 + \rho\frac{\sigma^2_{e_{\text{pil}}}}{\sigma_h^2}\right) - \int_0^\infty \log\left(1 + \rho\frac{\sigma^2_{e_{\text{pil}}}}{\sigma_h^2}z\right)e^{-z}dz\right) \tag{84}$$

where $\sigma^2_{e_{\text{pil}}}$ is the channel estimation error variance when estimating the channel solely based on pilot symbols which is given by

$$\sigma^2_{e_{\text{pil}}} = \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{S_h(f)}{\frac{\rho}{L}\frac{S_h(f)}{\sigma_h^2} + 1} df. \tag{85}$$

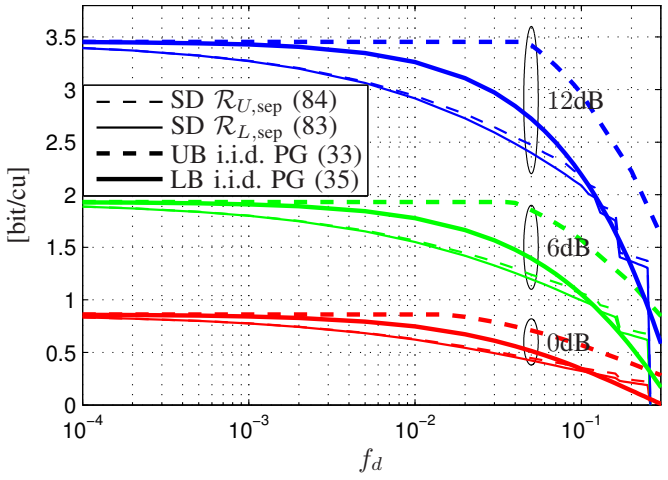

Fig. 4. Comparison of bounds on the achievable rate with synchronized detection and a solely pilot based channel estimation (SD) and bounds on the achievable rate with i.i.d. zero-mean proper Gaussian (PG) input symbols; rectangular $S_h(f)$.

With (83) it can easily be seen that the achievable rate is decreased in comparison to perfect channel knowledge by two factors. First, symbol time instances that are used for pilot symbols are lost for data symbols leading to the pre-log factor $\frac{L-1}{L}$, and secondly, the average SNR is decreased by the factor $(1-\sigma^2_{e_{\text{pil}}}/\sigma^2_h)/(1+\rho\sigma^2_{e_{\text{pil}}}/\sigma^2_h)$ due to the channel estimation error variance.

Fig. 4 shows a comparison of the bounds on the achievable rate with synchronized detection based on a solely pilot based channel estimation in (83) and (84) with the bounds on the achievable rate with i.i.d. Gaussian input symbols. For synchronized detection with a solely pilot based channel estimation the pilot spacing has been chosen such that the lower bound on the achievable rate in (83) is maximized. As this lower bound is relatively tight, the chosen pilot spacing should be close to the one that maximizes the achievable rate with synchronized detection using a solely pilot based channel estimation. Obviously, for the practically important range of small channel dynamics, i.e., $f_d \ll 0.1$, the achievable rate with synchronized detection using a solely pilot based channel estimation stays below the achievable rate with i.i.d. Gaussian input symbols, indicating the possible gain when using enhanced receiver structures. Even in case of using pilot symbols, the receiver performance can be enhanced by using a joint processing of pilot and data symbols instead of a separate processing. For a more detailed discussion on the difference between separate and joint processing we refer to [29]. One possibility of such a joint processing is to use an iterative code-aided channel estimation, where the channel estimation is enhanced based on reliability information on the data symbols delivered by the decoder. Based on this enhanced channel estimation detection and decoding is performed again, see e.g., [30] and [31].

VI. Concluding Remarks

The main focus of the present paper is the study of the achievable rate with i.i.d. Gaussian inputs. We have derived two new upper bounds on the achievable rate for this class of input distributions which are functions of the SNR and the PSD of the channel fading process. Among other things, we have shown that the achievable rate with i.i.d. Gaussian inputs has the same high SNR slope as the peak power constrained capacity, meaning that this kind of input distribution is not severely suboptimal.

Gaussian inputs are of practical interest. They serve well to upper-bound the achievable rate of practical modulation and coding schemes in the coherent case [10]. We have been particularly interested to compare the bounds on the achievable rate with i.i.d. Gaussian input symbols to those for synchronized detection employing solely pilot based channel estimation, a method which today is almost universally used in practice. The difference of the bounds is a useful indication of how much gain can be achieved using advanced receiver algorithms, for example iterative code-aided channel estimation. The numerical results show that the potential gain for such receivers is moderate for channel dynamics of practical relevance ($f_d < 10^{-3}$).

Appendix A
Convexity of (64)

To prove that (64) 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:

$$\sigma^2_{e_{\text{pred}}}(\mathbf{z}) = \sigma^2_h - \mathbf{r}^H_{h,\text{pred}}\left(\mathbf{R}_h + \sigma^2_n\mathbf{Z}^{-1}\right)^{-1}\mathbf{r}_{h,\text{pred}}$$
$$= \sigma^2_h - \mathbf{r}^H_{h,\text{pred}}\left(\mathbf{R}_h^{-1} - \mathbf{R}_h^{-1}\left(\frac{\mathbf{Z}}{\sigma^2_n} + \mathbf{R}_h^{-1}\right)^{-1}\mathbf{R}_h^{-1}\right)\mathbf{r}_{h,\text{pred}} \tag{86}$$

$$= \sigma^2_h - \mathbf{r}^H_{h,\text{pred}} \times \left(\mathbf{R}_h^{-1} - \mathbf{R}_h^{-1}\left(\frac{z_i\mathbf{V}_i + \mathbf{Z}_{\setminus i}}{\sigma^2_n} + \mathbf{R}_h^{-1}\right)^{-1}\mathbf{R}_h^{-1}\right)\mathbf{r}_{h,\text{pred}} \tag{87}$$

$$= \sigma^2_h - \mathbf{r}^H_{h,\text{pred}}\left(\mathbf{R}_h^{-1} - \mathbf{R}_h^{-1}\left[\left(\frac{1}{\sigma^2_n}\mathbf{Z}_{\setminus i} + \mathbf{R}_h^{-1}\right) \times \left\{\left(\frac{1}{\sigma^2_n}\mathbf{Z}_{\setminus i} + \mathbf{R}_h^{-1}\right)^{-1}\frac{z_i}{\sigma^2_n}\mathbf{V}_i + \mathbf{I}\right\}\right]^{-1}\mathbf{R}_h^{-1}\right)\mathbf{r}_{h,\text{pred}}$$
$$= \sigma^2_h - \mathbf{r}^H_{h,\text{pred}}\left(\mathbf{R}_h^{-1} - \mathbf{R}_h^{-1}\left\{\left(\frac{\mathbf{Z}_{\setminus i}}{\sigma^2_n} + \mathbf{R}_h^{-1}\right)^{-1}\frac{z_i}{\sigma^2_n}\mathbf{V}_i + \mathbf{I}\right\}^{-1} \times \left(\frac{1}{\sigma^2_n}\mathbf{Z}_{\setminus i} + \mathbf{R}_h^{-1}\right)^{-1}\mathbf{R}_h^{-1}\right)\mathbf{r}_{h,\text{pred}}$$
$$= \sigma^2_h - \mathbf{r}^H_{h,\text{pred}}\left(\mathbf{R}_h^{-1} - \mathbf{R}_h^{-1}\left\{\mathbf{I} - \frac{z_i}{1+z_i\lambda_{\max}}\left(\frac{\mathbf{Z}_{\setminus i}}{\sigma^2_n} + \mathbf{R}_h^{-1}\right)^{-1} \times \frac{\mathbf{V}_i}{\sigma^2_n}\right\}\left(\frac{\mathbf{Z}_{\setminus i}}{\sigma^2_n} + \mathbf{R}_h^{-1}\right)^{-1}\mathbf{R}_h^{-1}\right)\mathbf{r}_{h,\text{pred}} \tag{88}$$

$$= \sigma_h^2 - \mathbf{r}_{h,\text{pred}}^H \left(\mathbf{R}_h^{-1} - \mathbf{R}_h^{-1} \left(\frac{\mathbf{Z}_{\setminus i}}{\sigma_n^2} + \mathbf{R}_h^{-1} \right)^{-1} \mathbf{R}_h^{-1} \right) \mathbf{r}_{h,\text{pred}}$$
$$- \frac{z_i \mathbf{r}_{h,\text{pred}}^H \mathbf{R}_h^{-1} \left(\frac{\mathbf{Z}_{\setminus i}}{\sigma_n^2} + \mathbf{R}_h^{-1} \right)^{-1} \frac{\mathbf{V}_i}{\sigma_n^2} \left(\frac{\mathbf{Z}_{\setminus i}}{\sigma_n^2} + \mathbf{R}_h^{-1} \right)^{-1} \mathbf{R}_h^{-1} \mathbf{r}_{h,\text{pred}}}{1 + z_i \lambda_{\max}}$$
$$= \sigma_{e_{\text{pred}}}^2(\mathbf{z}_{\setminus i}) - \frac{z_i \cdot a}{1 + z_i \lambda_{\max}} \tag{89}$$

where for (86) we have used the matrix inversion lemma, and for (87) we have separated the diagonal matrix $\mathbf{Z}$ as follows:

$$\mathbf{Z} = \mathbf{Z}_{\setminus i} + z_i \mathbf{V}_i \tag{90}$$

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, for (88) we have used the Sherman-Morrison formula yielding

$$\left(\mathbf{I} + \mathbf{a}\mathbf{b}^T \right)^{-1} = \mathbf{I} - \frac{1}{1 + \mathbf{b}^T \mathbf{a}} \mathbf{a}\mathbf{b}^T \tag{91}$$

for arbitrary vectors $\mathbf{a}, \mathbf{b}$ of equal length and $\mathbf{b}^T\mathbf{a} \neq -1$. To apply (91) we identify the rank one matrix

$$\mathbf{B} = \left(\frac{1}{\sigma_n^2} \mathbf{Z}_{\setminus i} + \mathbf{R}_h^{-1} \right)^{-1} \frac{1}{\sigma_n^2} \mathbf{V}_i. \tag{92}$$

with $\frac{\mathbf{a}\mathbf{b}^T}{z_i}$. Hence $\lambda_{\max}$ in (88) is the non-zero eigenvalue of the rank one matrix $\mathbf{B}$. Furthermore, for (89) we substituted $\sigma_{e_{\text{pred}}}^2(\mathbf{z}_{\setminus i})$ for

$$\sigma_{e_{\text{pred}}}^2(\mathbf{z}_{\setminus i})$$
$$= \sigma_h^2 - \mathbf{r}_{h,\text{pred}}^H \left(\mathbf{R}_h^{-1} - \mathbf{R}_h^{-1} \left(\frac{\mathbf{Z}_{\setminus i}}{\sigma_n^2} + \mathbf{R}_h^{-1} \right)^{-1} \mathbf{R}_h^{-1} \right) \mathbf{r}_{h,\text{pred}} \tag{93}$$

which is the prediction error variance if the observation at the i-th time instant is not used for channel prediction. Additionally, for (89) we have also used the substitution

$$a = \mathbf{r}_{h,\text{pred}}^H \mathbf{R}_h^{-1} \left(\frac{\mathbf{Z}_{\setminus i}}{\sigma_n^2} + \mathbf{R}_h^{-1} \right)^{-1} \frac{\mathbf{V}_i}{\sigma_n^2} \left(\frac{\mathbf{Z}_{\setminus i}}{\sigma_n^2} + \mathbf{R}_h^{-1} \right)^{-1} \mathbf{R}_h^{-1} \mathbf{r}_{h,\text{pred}}$$
$$\geq 0 \tag{94}$$

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

Thus, with (89) we have found a separation of the channel prediction error variance $\sigma_{e_{\text{pred}}}^2(\mathbf{z})$ into the term $\sigma_{e_{\text{pred}}}^2(\mathbf{z}_{\setminus i})$ being independent of z_i, and an additional term, which depends on z_i. Note that a and $\lambda_{\max}$ in the second term in (89) are independent of z_i and that the element i is arbitrarily chosen. I.e., we can use this separation for each diagonal element of $\mathbf{Z}$.

By substituting (89) into (64) we get

$$\log \left(1 + \frac{|x_N|^2}{\sigma_n^2} \left(\sigma_{e_{\text{pred}}}^2(\mathbf{z}_{\setminus i}) - \frac{z_i \cdot a}{1 + z_i \lambda_{\max}} \right) \right) = K. \tag{95}$$

Recall that we want to show the convexity of (95) with respect to the element z_i. Therefore, we calculate its second derivative with respect to z_i which is given by

$$\frac{\partial^2 K}{(\partial z_i)^2} = \frac{\frac{|x_N|^2}{\sigma_n^2} \frac{a 2\lambda_{\max}(1 + z_i \lambda_{\max})}{(1 + z_i \lambda_{\max})^4}}{\left(1 + \frac{|x_N|^2}{\sigma_n^2} \left(\sigma_{e_{\text{pred}}}^2(\mathbf{z}_{\setminus i}) - \frac{a z_i}{1 + z_i \lambda_{\max}} \right) \right)^2}$$
$$\times \left\{ 1 + \frac{|x_N|^2}{\sigma_n^2} \left(\sigma_{e_{\text{pred}}}^2(\mathbf{z}_{\setminus i}) - \frac{a \left(z_i + \frac{1}{2\lambda_{\max}} \right)}{1 + z_i \lambda_{\max}} \right) \right\} \tag{96}$$

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

$$\frac{\partial^2 K}{(\partial z_i)^2} \geq 0. \tag{97}$$

Therefore, first we show that $\lambda_{\max}$ is nonnegative based on the definition of the eigenvalues of $\mathbf{B}$

$$\mathbf{B}\mathbf{u} = \left(\frac{1}{\sigma_n^2} \mathbf{Z}_{\setminus i} + \mathbf{R}_h^{-1} \right)^{-1} \frac{1}{\sigma_n^2} \mathbf{V}_i \mathbf{u} = \lambda_{\max} \mathbf{u}$$
$$\Rightarrow \frac{1}{\sigma_n^2} \mathbf{u}^H \mathbf{V}_i \mathbf{u} = \lambda_{\max} \mathbf{u}^H \left(\frac{1}{\sigma_n^2} \mathbf{Z}_{\setminus i} + \mathbf{R}_h^{-1} \right) \mathbf{u}$$
$$\Rightarrow \lambda_{\max} \geq 0 \tag{98}$$

where (98) 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 (97), it rests to show that

$$\sigma_{e_{\text{pred}}}^2(\mathbf{z}_{\setminus i}) - \frac{a}{1 + z_i \lambda_{\max}} \left(z_i + \frac{1}{2\lambda_{\max}} \right) \geq 0. \tag{99}$$

To prove (99), we calculate the derivative of its LHS with respect to z_i, which is given by

$$\frac{\partial}{\partial z_i} \left\{ \sigma_{e_{\text{pred}}}^2(\mathbf{z}_{\setminus i}) - \frac{a \left(z_i + \frac{1}{2\lambda_{\max}} \right)}{1 + z_i \lambda_{\max}} \right\} = - \frac{a}{2(1 + z_i \lambda_{\max})^2}$$
$$\leq 0 \tag{100}$$

where for the last inequality we have used (94). I.e., the LHS of (99) monotonically decreases in z_i. Furthermore, for $z_i \to \infty$ the LHS of (99) becomes

$$\lim_{z_i \to \infty} \left\{ \sigma_{e_{\text{pred}}}^2(\mathbf{z}_{\setminus i}) - \frac{a \left(z_i + \frac{1}{2\lambda_{\max}} \right)}{1 + z_i \lambda_{\max}} \right\} = \lim_{z_i \to \infty} \sigma_{e_{\text{pred}}}^2(\mathbf{z}) \tag{101}$$
$$\geq 0 \tag{102}$$

where (101) follows due to (89), and where (102) holds as the prediction error variance must be nonnegative. As the LHS of (99) is monotonically decreasing in z_i and as its limit for $z_i \to \infty$ is nonnegative, (99) must hold.

Hence, with (99) inequality (97) holds and, thus, (95) is convex in z_i. As the element i has been chosen arbitrarily, in conclusion, we have shown that (64) is convex in each z_i for $i = 1, \ldots, N-1$.

ACKNOWLEDGMENT

The authors are very grateful to Norbert Gaffke for contributing the proof of inequality (17), which is a crucial point for the paper and is elaborated in (18)-(26).

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] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 139–157, Jan. 1999.

[3] 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.

[4] 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.

[5] R. H. Etkin and D. N. C. Tse, "Degrees of freedom in some underspread MIMO fading channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1576–1608, Apr. 2006.

[6] 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.

[7] V. Sethuraman, B. Hajek, and K. Narayanan, "Capacity bounds for noncoherent fading channels with a peak constraint," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Adelaide, Australia, Sep. 2005, pp. 515–519.

[8] 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.

[9] T. Koch and A. Lapidoth, "On multipath fading channels at high SNR," *IEEE Trans. Inf. Theory*, vol. 56, no. 12, pp. 5945–5957, Dec. 2010.

[10] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inf. Theory*, vol. 28, no. 1, pp. 55–67, Jan. 1982.

[11] 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.

[12] R.-R. Chen, B. Hajek, R. Koetter, and U. Madhow, "On fixed input distributions for noncoherent communication over high-SNR Rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3390–3396, Dec. 2004.

[13] J. Doob, *Stochastic Processes*. New York: Wiley, 1990.

[14] W. C. Jakes, Ed., *Microwave Mobile Communications*. New York, NY, U.S.A.: John Wiley & Sons Inc, 1975.

[15] 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.

[16] H. Meyr, M. Moeneclaey, and S. Fechtel, *Digital Communication Receivers: Synchronization, Channel Estimation and Signal Processing, 1st ed.* New York, NY, USA: John Wiley & Sons, 1998.

[17] 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.

[18] 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.

[19] 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.

[20] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

[21] A. Marshall, I. Olkin, and B. Arnold, *Inequalities: Theory of Majorization and Its Applications*, ser. Springer Series in Statistics. Springer, 2010.

[22] L. Rüschendorf, "Solution of a statistical optimization problem by rearrangement methods," *Metrika*, vol. 30, pp. 55–61, 1983.

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

[24] R. M. Gray, "Toeplitz and circulant matrices: A review," *Foundations and Trends in Communications and Information Theory*, vol. 2, no. 3, pp. 155–239, 2006.

[25] W. Rudin, *Real and complex analysis*. New York: McGraw-Hill, 1987.

[26] A. Lapidoth and S. Shamai (Shitz), "Fading channels: How perfect need "perfect side information" be?" *IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 1118–1134, May 2002.

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

[28] S. Verdú, "On channel capacity per unit cost," *IEEE Trans. Inf. Theory*, vol. 36, no. 5, pp. 1019–1030, Sep. 1990.

[29] M. Dörpinghaus, A. Ispas, and H. Meyr, "On the gain of joint processing of pilot and data symbols in stationary Rayleigh fading channels," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2963–2982, May 2012.

[30] M. C. Valenti and B. D. Woerner, "Iterative channel estimation and decoding of pilot symbol assisted Turbo codes over flat-fading channels," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 9, pp. 1697–1705, Sep. 2001.

[31] L. Schmitt, H. Meyr, and D. Zhang, "Systematic design of iterative ML receivers for flat fading channels," *IEEE Trans. Commun.*, vol. 58, no. 7, pp. 1897–1901, Jul. 2010.

Meik Dörpinghaus (S'04–M'10) received the Dipl.-Ing. degree in electrical engineering and information technology (with distinction) and the Dr.-Ing. degree in electrical engineering and information technology (*summa cum laude*) both from RWTH Aachen University, Aachen, Germany, in 2003 and 2010, respectively. Currently, he is a postdoctoral researcher at RWTH Aachen University. In 2007, he was a visiting researcher at ETH Zurich, Switzerland. His research interests are in the areas of communication and information theory.

Dr. Dörpinghaus received the Friedrich-Wilhelm Preis of RWTH Aachen in 2011 for an outstanding PhD thesis, and the Friedrich-Wilhelm Preis of RWTH Aachen in 2004 and the Siemens Preis in 2004 for an excellent diploma thesis.

Heinrich Meyr (M'75–SM'83–F'86–LF'10) received his M.Sc. and Ph.D. from ETH Zurich, Switzerland. He spent over 12 years in various research and management positions in industry before accepting a professorship in Electrical Engineering at Aachen University of Technology (RWTH Aachen) where he founded the Institute for Integrated Signal Processing Systems. In 2007 he has assumed the rank of emeritus. Presently he is a visiting professor at the LSI lab of the EPFL directed by Professor Giovanni de Micheli.

During the last thirty years, Dr. Meyr has worked extensively in the areas of communication theory, digital signal processing and CAD tools for system-level design. He has published numerous IEEE papers and holds many patents. He is a Life-Fellow of the IEEE and has received three IEEE best paper awards. In 2000, Dr. Meyr was the recipient of the prestigious Vodafone prize for his outstanding contribution in the area of wireless communications.

Dr. Meyr has a dual career as entrepreneur. He has co-founded a number of successful companies. The last company he has co-founded was LISATek Inc., a company with breakthrough technology to design application specific processors. The company merged with CoWare in February 2003 and was acquired by Synopsys in spring 2010. From 2003 until 2010 Dr. Meyr held the position of Chief Scientific Officer of CoWare.

Rudolf Mathar (M'98) received his Diploma and Ph.D. degree in 1978 and 1981 from RWTH Aachen University. Previous positions include lecturer positions at Augsburg University and at the European Business School. In 1989, he joined the Faculty of Natural Sciences at RWTH Aachen University. He has held the International IBM Chair in Computer Science at Brussels Free University in 1999. In 2004 he was appointed head of the Institute for Theoretical Information Technology in the Faculty of Electrical Engineering and Information Technology at RWTH Aachen University. His research interests include mobile communication systems, particularly optimization, resource allocation and information theory. He has published extensively in IEEE journals and conferences. In 2002, he was the recipient of the prestigious Vodafone Innovation Award, and in 2010 he was elected member of the NRW Academy of Sciences and Arts. He is co-founder of two spin-off enterprises. Since 2011 he is serving as the dean of the Faculty of Electrical and Engineering and Information Technology.