Optimal PSK Signaling over Stationary Rayleigh Fading Channels

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

Rudolf Mathar
Institute for Theoretical Information Technology,
RWTH Aachen University, Germany
Email: mathar@ti.rwth-aachen.de

***Abstract*— We consider a stationary Rayleigh flat-fading channel with temporal correlation and a compactly supported power spectral density of the channel fading process. We assume that the channel state is unknown to both transmitter and receiver, while the law of the channel is presumed to be known to the receiver. Given a set of fixed signaling sequences, the optimum input distribution, with respect to the achievable rate, has the property of a constant Kullback-Leibler distance between the output distribution and a mixture of the output distributions conditioned on the different input sequences. Based on this, we determine the set of optimum input distributions for PSK signaling. In addition, we identify the special case of transmitting one pilot symbol to acquire a phase reference as being included in the set of optimum input distributions. We derive an integral expression for the capacity constrained to PSK signaling depending on the autocorrelation of the channel and the SNR. Evaluation of the asymptotic high SNR behavior shows a loss in the constrained capacity with respect to the case of perfect channel knowledge corresponding to at least one signaling dimension, i.e., the information transmitted by one symbol.**

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 transmitter and 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 enable the receiver to track the channel, in many systems training sequences are inserted into the data stream by the transmitter. These training sequences can be understood as a specific type of code [1], which has not been proven to be capacity-achieving.

The capacity of fading channels where the channel state information is unknown, sometimes referred to as noncoherent capacity, has received a lot of attention in the recent literature, see, e.g., [1], [2], [3]. Even though there exist bounds on the capacity for flat-fading channels, which are tight in a specific SNR regime, see, e.g., [1], [4], [2], neither exact expressions for the capacity, nor the capacity-achieving input distribution are known. In [5], bounds on the achievable rates for different input distribution, including discrete ones, have been derived for the case of a Gauss-Markov flat-fading channel. In [6], it has been shown that Gaussian input distributions, which are capacity-achieving in case the receiver is aware of the channel state, are in general not capacity-achieving in case the channel state is unknown to the receiver. In contrast, discrete input distributions are capacity-achieving for a variety of conditionally Gaussian channels with bounded input constraints [7]. E.g., for the case of a Rayleigh flat-fading channel without temporal correlation, it has been shown that the capacity-achieving input distribution is discrete with a finite number of mass points [8]. The scenario in the present paper falls into the class of conditionally Gaussian channels. These observations and the fact that practical realizable systems use discrete input distributions are the motivation to study the achievable rates for the given scenario with the restriction to discrete input distributions, leading to the following question: We have a bounded and closed subset $\mathcal{S} \subset \mathbb{C}^N$, where N corresponds to the length of the transmit sequence in symbols, and a maximum number M of support points $\mathbf{x}_i \in \mathcal{S}$, $i = 1, \ldots, M$, corresponding to the signaling sequences. What is then the optimum choice of the support points and what is their optimum distribution $\mathbf{p} = [p_1, \ldots, p_M]$, with p_i being the probability of transmitting the sequence $\mathbf{x}_i$, that maximizes the mutual information between channel input and output?

In this paper, we restrict to the special case where the set $\mathcal{S}$ consists of a fixed amount of predefined support points $\mathbf{x}_1, \ldots, \mathbf{x}_M$ representing the possible transmit sequences. Then, the input distribution $\mathbf{p}$ that maximizes the mutual information can be evaluated. We will refer to this input distribution as the *optimum* input distribution. Furthermore, we will name the maximum mutual information constrained to a given set of support points *constrained capacity*.

For additive noise channels, this problem has been examined in [9]. In the present paper, we extend the work in [9] to Rayleigh flat-fading channels with temporal correlation, where the receiver has no knowledge of the channel state. The channel fading process is characterized by a compactly supported power spectral density (PSD) with a normalized maximum Doppler frequency $f_d < 0.5$, i.e., it is assumed to be *nonregular* [1]. The optimum input distribution $\mathbf{p}$ is characterized by a constant Kullback-Leibler distance between the output distribution and a mixture of the output distributions conditioned on the different input sequences. For PSK signaling, we explicitly characterize the set of optimum input distributions $\mathbf{p}$. The special case of transmitting one pilot symbol, i.e., a symbol that is known to the receiver, lies within this set and thus is optimum. In addition, the asymptotic high SNR constrained capacity is degraded at least by a factor of $\frac{N-1}{N}$ with respect to the case of perfect channel state information at the receiver.

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

II. System Model

We consider a discrete-time 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.

Furthermore, we assume that the noise n_k is a sequence of i.i.d. circularly symmetric complex Gaussian random variables of zero mean and with variance σ_n^2. The channel fading process h_k is zero mean jointly proper Gaussian [10]. In addition, it is time selective and characterized by its autocorrelation function

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

where $(\cdot)^*$ indicates complex conjugation. Its variance is normalized to $r_h(0) = \sigma_h^2 = 1$. The PSD of the channel fading process is defined by

$$S(f) = \sum_{m=-\infty}^{\infty} r_h(m) e^{-j2\pi m f}, \ |f| < 0.5. \tag{3}$$

Here, we assume that the PSD exists. For a jointly proper Gaussian process this implies ergodicity [11].

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 compactly supported within the interval $[-f_d, f_d]$, with $0 < f_d < 0.5$, i.e., $S(f) = 0$ for $f \notin [-f_d, f_d]$. The parameter f_d corresponds to the maximum Doppler shift and thus indicates the dynamic of the channel. To assure ergodicity we exclude the case $f_d = 0$. As we consider a transmission with a duration of N time instances, we introduce the following vector-matrix notation of the system model:

$$\mathbf{y} = \mathbf{H} \cdot \mathbf{x} + \mathbf{n} = \mathbf{X} \cdot \mathbf{h} + \mathbf{n} \tag{4}$$

where the vectors are defined as $\mathbf{y} = [y_1, y_2, \ldots, y_N]^T$, $\mathbf{x} = [x_1, x_2, \ldots, x_N]^T$, and $\mathbf{n} = [n_1, n_2, \ldots, n_N]^T$. The matrix $\mathbf{H}$ is diagonal and defined as $\mathbf{H} = \mathrm{diag}(\mathbf{h})$ with $\mathbf{h} = [h_1, h_2, \ldots, h_N]^T$. Here, the diag(.) operator generates a diagonal matrix whose diagonal elements are given by the argument vector. The diagonal matrix $\mathbf{X}$ is given by $\mathbf{X} = \mathrm{diag}(\mathbf{x})$.

Using this vector notation, the correlation over time of the fading process is expressed by the correlation matrix

$$\mathbf{R}_h = \mathrm{E}[\mathbf{h} \cdot \mathbf{h}^H] \tag{5}$$

which has a Hermitian Toeplitz structure.

III. Optimum discrete input distributions

As the PSD of the fading process exists (3) and as the fading process is jointly proper Gaussian, the channel fading process is ergodic. Therefore, operational and information theoretic capacity coincide [11]. This allows us to base our following derivations on the concept of the ergodic capacity.

Due to the restriction to finite size transmit constellations, the input symbol x_k is a discrete random variable with Q support points. Consequently, the input vector $\mathbf{x}$ is a random variable with

$$M = Q^N. \tag{6}$$

support points $\mathbf{x}_1, \ldots, \mathbf{x}_M \in \mathbb{C}^N$.

The mutual information of the channel model given in (1) can be calculated as

$$\mathcal{I}(\mathbf{y};\mathbf{x}) = h(\mathbf{y}) - h(\mathbf{y}|\mathbf{x}) \tag{7}$$

where $h(\cdot)$ denotes the differential entropy. We will examine the capacity of the channel given by the model in (1) with the constraint on a discrete input distribution, where the input vector $\mathbf{x}$ is from the finite set $\mathcal{S}$ given by $\mathcal{S} = \{\mathbf{x}_1, \ldots, \mathbf{x}_M\}$. We are going to determine the optimum probability distribution $\mathbf{p} = [p_1, \ldots, p_M]$ with p_i the probability of transmitting the sequence $\mathbf{x}_i$. Then the constrained ergodic capacity of (1) is given by

$$C = \frac{1}{N} \max_{\mathbf{p} \in \mathcal{C}} \mathcal{I}(\mathbf{y};\mathbf{x}) \tag{8}$$

where the set $\mathcal{C}$ is convex and given by

$$\mathcal{C} = \left\{ \mathbf{p} = [p_1, \ldots, p_M] \,\middle|\, \sum_{i=1}^{M} p_i = 1, p_i \geq 0 \right\}. \tag{9}$$

For the calculation of the channel output entropy conditioned on the channel input $h(\mathbf{y}|\mathbf{x})$, the conditional probability density function $p(\mathbf{y}|\mathbf{x})$ is required. It is given as

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{\pi^N \det\left(\mathbf{R}_{y|x}\right)} \exp\left(-\mathbf{y}^H \mathbf{R}_{y|x}^{-1} \mathbf{y}\right) \tag{10}$$

with

$$\begin{aligned} \mathbf{R}_{y|x} &= E_{\mathbf{h},\mathbf{n}}[\mathbf{y}\mathbf{y}^H] = E_{\mathbf{h}}[\mathbf{X}\mathbf{h}\mathbf{h}^H\mathbf{X}^H] + \sigma_n^2 \mathbf{I}_N \\ &= \mathbf{X}\mathbf{R}_h\mathbf{X}^H + \sigma_n^2 \mathbf{I}_N. \end{aligned} \tag{11}$$

Here $\mathbf{I}_N$ indicates the identity matrix of dimension $N \times N$. The distribution of the channel output is given by

$$p(\mathbf{y}) = \sum_{i=1}^{M} p_i p(\mathbf{y}|\mathbf{x}_i). \tag{12}$$

As the entropies in (7) are given by

$$\begin{aligned} h(\mathbf{y}) &= -\int p(\mathbf{y}) \log p(\mathbf{y}) d\mathbf{y} \\ &= -\int \sum_{i=1}^{M} p_i p(\mathbf{y}|\mathbf{x}_i) \log \left(\sum_{j=1}^{M} p_j p(\mathbf{y}|\mathbf{x}_j) \right) d\mathbf{y} \end{aligned} \tag{13}$$

$$\begin{aligned} h(\mathbf{y}|\mathbf{x}) &= -\int \int p(\mathbf{y},\mathbf{x}) \log p(\mathbf{y}|\mathbf{x}) d\mathbf{x}\, d\mathbf{y} \\ &= -\int \sum_{i=1}^{M} p_i p(\mathbf{y}|\mathbf{x}_i) \log p(\mathbf{y}|\mathbf{x}_i) d\mathbf{y} \end{aligned} \tag{14}$$

we get by interchanging of summation and integration for the mutual information

$$\begin{aligned}\mathcal{I}(\mathbf{y};\mathbf{x}) &= -\sum_{i=1}^{M} p_i \int p(\mathbf{y}|\mathbf{x}_i)\log\left(\sum_{j=1}^{M} p_j p(\mathbf{y}|\mathbf{x}_j)\right) d\mathbf{y} \\ &\quad + \sum_{i=1}^{M} p_i \int p(\mathbf{y}|\mathbf{x}_i)\log p(\mathbf{y}|\mathbf{x}_i) d\mathbf{y} \\ &= \sum_{i=1}^{M} p_i \int p(\mathbf{y}|\mathbf{x}_i)\log\left(\frac{p(\mathbf{y}|\mathbf{x}_i)}{\sum_{j=1}^{M} p_j p(\mathbf{y}|\mathbf{x}_j)}\right) d\mathbf{y} \\ &= \sum_{i=1}^{M} p_i D\left(p(\mathbf{y}|\mathbf{x}_i)\middle|\middle|\sum_{j=1}^{M} p_j p(\mathbf{y}|\mathbf{x}_j)\right) \end{aligned} \quad (15)$$

where $D(g||h) = \int g \log\left(\frac{g}{h}\right)$ is the Kullback-Leibler distance, or the relative entropy between the densities g and h, see also [12].

We now seek to find the distribution $\mathbf{p} = [p_1, ..., p_M]$ that maximizes (15) by calculating the gradient of $f = \mathcal{I}(\mathbf{y};\mathbf{x})$ with respect to $\mathbf{p}$ and using directional derivatives. However, we would like to point out that there also exists an alternative approach as described at the end of this section.

In the following derivation, we closely follow the lines in [9]. Therefore we calculate the gradient $\nabla f = \left(\frac{df}{dp_i}\right)_{i=1,\ldots,M}$

$$\frac{df}{dp_i} = \frac{d}{dp_i}\sum_{i=1}^{M} p_i \int p(\mathbf{y}|\mathbf{x}_i)\left[\log p(\mathbf{y}|\mathbf{x}_i) - \log\left(\sum_{j=1}^{M} p_j p(\mathbf{y}|\mathbf{x}_j)\right)\right] d\mathbf{y}$$

$$= \int p(\mathbf{y}|\mathbf{x}_i)\log p(\mathbf{y}|\mathbf{x}_i) d\mathbf{y} - \int p(\mathbf{y}|\mathbf{x}_i)\log\left(\sum_{j=1}^{M} p_j p(\mathbf{y}|\mathbf{x}_i)\right) d\mathbf{y} - 1.$$

The directional derivative of f at $\hat{\mathbf{p}} = [\hat{p}_1, \ldots, \hat{p}_M]$ in the direction of $\mathbf{p}$ is given by

$$\langle \nabla f(\hat{\mathbf{p}}), \mathbf{p} - \hat{\mathbf{p}} \rangle = \sum_{i=1}^{M} (p_i - \hat{p}_i)(c_i - b_i(\hat{\mathbf{p}})) \quad (16)$$

where

$$b_i(\hat{\mathbf{p}}) = \int p(\mathbf{y}|\mathbf{x}_i) \cdot \log\left(\sum_{j=1}^{M} \hat{p}_j p(\mathbf{y}|\mathbf{x}_j)\right) d\mathbf{y} \quad (17)$$

$$c_i = \int p(\mathbf{y}|\mathbf{x}_i) \cdot \log p(\mathbf{y}|\mathbf{x}_i) d\mathbf{y}. \quad (18)$$

Due to the concavity of the log function, $b_i(\hat{\mathbf{p}})$ is concave in $\mathcal{C}$ for $i = 1, \ldots, M$. Thus, the maximum of f is attained at $\hat{\mathbf{p}} \in \mathcal{C}$ iff the directional derivatives at $\hat{\mathbf{p}}$ in any direction $\mathbf{p} \in \mathcal{C}$ is non-positive, i.e.,

$$\sum_{i=1}^{M} (p_i - \hat{p}_i)(c_i - b_i(\hat{\mathbf{p}})) \leq 0. \quad (19)$$

Hence, $\hat{\mathbf{p}}$ is an optimum point iff

$$\begin{aligned}\sum_{i=1}^{M} \hat{p}_i (c_i - b_i(\hat{\mathbf{p}})) &= \max_{\mathbf{p}\in\mathcal{C}} \sum_{i=1}^{M} p_i (c_i - b_i(\hat{\mathbf{p}})) \\ &= \max_{i=1,\ldots,M} (c_i - b_i(\hat{\mathbf{p}})). \end{aligned} \quad (20)$$

Equality is only achieved iff $c_i - b_i(\hat{\mathbf{p}})$ equals some constant χ for all i with $\hat{p}_i > 0$. Recognizing that

$$c_i - b_i(\hat{\mathbf{p}}) = D\left(p(\mathbf{y}|\mathbf{x}_i)\middle|\middle|\sum_{j=1}^{M} \hat{p}_j p(\mathbf{y}|\mathbf{x}_j)\right) \quad (21)$$

we state the following proposition.

Proposition 1: Given the signaling vectors $\mathbf{x}_1, \ldots, \mathbf{x}_M \in \mathbb{C}^N$ for the input variable $\mathbf{x}$ in channel model (1), the distribution $\hat{\mathbf{p}}$ is optimum, i.e., achieves the constrained capacity, if and only if

$$D\left(p(\mathbf{y}|\mathbf{x}_i)\middle|\middle|\sum_{j=1}^{M} \hat{p}_j p(\mathbf{y}|\mathbf{x}_j)\right) = \chi \quad (22)$$

for some $\chi \in \mathbb{R}$ and all indices i with $\hat{p}_i > 0$.

Thus, for the optimum $\hat{\mathbf{p}}$ the Kullback-Leibler distance $D\left(p(\mathbf{y}|\mathbf{x}_i)\middle|\middle|\sum_{j=1}^{M} \hat{p}_j p(\mathbf{y}|\mathbf{x}_j)\right)$ is constant for all i with positive $\hat{p}_i$. With (15), the constrained ergodic capacity amounts to

$$C = \frac{1}{N}\max_{\mathbf{p}\in\mathcal{C}} \mathcal{I}(\mathbf{y};\mathbf{x}) = \frac{1}{N}\chi. \quad (23)$$

Alternatively, we can use the analogy of our problem to the problem of finding the capacity-achieving input distribution of the discrete memoryless channel (DMC). Due to the restriction to a finite set of input sequences with finite length, the Rayleigh fading channel is similar to the DMC, except that its output is continuous. For the DMC a characterization of the capacity-achieving distribution is given in [13, Theorem 4.5.1.]. The same methodology, essentially the Karush-Kuhn-Tucker conditions, may be used in our context, yielding proposition 1.

IV. CONSTANT MODULUS INPUT DISTRIBUTIONS

In this section, we will give an explicit characterization of the optimum input distribution $\mathbf{p}$ for the special case of constant modulus input distributions, i.e., PSK type signaling with

$$x_k = \sigma_x^2 \exp\left(j2\pi\frac{i}{Q}\right), \quad i = 0, \ldots, Q-1 \quad (24)$$

where $\sigma_x^2 = |x_k|^2, \forall k$. Here, j is the imaginary unit, i.e., $j = \sqrt{-1}$. Without loss of generality, we assume $\sigma_x^2 = 1$ for the rest of this work.

A. Distinguishable Transmit Sequences

For input signals given by (24) the probability density function of the output conditioned on the input sequence (10) can be simplified to

$$p(\mathbf{y}|\mathbf{x}_i) = \frac{\exp\left(-\mathbf{y}^H\mathbf{X}_i\left(\mathbf{R}_h + \sigma_n^2\mathbf{I}_N\right)^{-1}\mathbf{X}_i^H\mathbf{y}\right)}{\pi^N \det\left(\mathbf{R}_h + \sigma_n^2\mathbf{I}_N\right)}. \quad (25)$$

It can be shown that the density function conditioned on the two input sequences $\mathbf{x}_m$ and $\mathbf{x}_n$ is equal, i.e.,

$$p(\mathbf{y}|\mathbf{x}_n) = p(\mathbf{y}|\mathbf{x}_m) \quad (26)$$

if and only if

$$\mathbf{x}_n = \mathbf{x}_m \exp(j\phi) \text{ for some } \phi \in [0, 2\pi). \tag{27}$$

Thus, transmit sequences fulfilling the constraint given in (27) can not be distinguished by the receiver.

We select a subset $\mathcal{S}_0 \subseteq \mathcal{S}$ of maximal cardinality such that the elements of $\mathcal{S}_0$ are pairwise distinguishable, i.e.,

$$p(\mathbf{y}|\mathbf{x}_n) \neq p(\mathbf{y}|\mathbf{x}_m) \text{ for any } \mathbf{x}_n \neq \mathbf{x}_m \in \mathcal{S}_0. \tag{28}$$

It is easy to see that $|\mathcal{S}_0| = \frac{M}{Q} = Q^{N-1}$.

B. Optimum Input Distribution

Based on $\mathcal{S}_0$, the Kullback-Leibler distance in (22) can be transformed to

$$\begin{aligned} &D\left(p(\mathbf{y}|\mathbf{x}_i)\middle|\middle| \sum_{j=1}^{M} \hat{p}_j p(\mathbf{y}|\mathbf{x}_j)\right) \\ &= \int p(\mathbf{y}|\mathbf{x}_i) \log\left(\frac{p(\mathbf{y}|\mathbf{x}_i)}{\sum_{j=1}^{M} \hat{p}_j p(\mathbf{y}|\mathbf{x}_j)}\right) d\mathbf{y} \\ &= \int p(\mathbf{y}|\mathbf{x}_i) \log\left(\frac{p(\mathbf{y}|\mathbf{x}_i)}{\sum_{\mathbf{x}_j \in \mathcal{S}_0} p(\mathbf{y}|\mathbf{x}_j) \sum_{\{k|\mathbf{x}_k = \mathbf{x}_j e^{j\phi}\}} \hat{p}_k}\right) d\mathbf{y} \end{aligned} \tag{29}$$

In the last step, we use (26) and (27).

Proposition 2: The distribution

$$\sum_{\{k|\mathbf{x}_k = \mathbf{x}_l e^{j\phi}\}} \hat{p}_k = \psi = \frac{Q}{M} = \frac{1}{Q^{N-1}}, \quad \forall l : \mathbf{x}_l \in \mathcal{S}_0 \tag{30}$$

is optimum, i.e., achieves the constrained capacity.

Intuitively, the optimum input distribution corresponds to a uniform distribution over the space of distinguishable transmit sequences.

Proof: We have to show that for the input distribution given in (30), the Kullback-Leibler distance (29) is independent of the index i. With (21) the Kullback-Leibler distance can be expressed by $c_i - b_i(\hat{\mathbf{p}})$, with c_i and $b_i(\hat{\mathbf{p}})$ given in (18) and (17).

We will first show that the term c_i given in (18) is independent of the index i for constant modulus input distributions. All signaling sequences $\mathbf{x}_i$ can be generated as

$$\mathbf{x}_i = \mathbf{U}_i \mathbf{x}_1 \tag{31}$$

where the matrix $\mathbf{U}_i$ is diagonal, orthonormal and, thus, unitary.

The conditional density $p(\mathbf{y}|\mathbf{x}_i)$, see (25), obeys the following property

$$p(\mathbf{y}|\mathbf{x}_i) = p(\mathbf{y}|\mathbf{U}_i\mathbf{x}_1) = p(\mathbf{U}_i^H\mathbf{y}|\mathbf{x}_1). \tag{32}$$

With (32), we get

$$\begin{aligned} c_i &= \int p(\mathbf{y}|\mathbf{x}_i) \log p(\mathbf{y}|\mathbf{x}_i) d\mathbf{y} = \int p(\mathbf{U}_i^H\mathbf{y}|\mathbf{x}_1) \log p(\mathbf{U}_i^H\mathbf{y}|\mathbf{x}_1) d\mathbf{y} \\ &= \int p(\mathbf{y}|\mathbf{x}_1) \log p(\mathbf{y}|\mathbf{x}_1) d\mathbf{y} \end{aligned} \tag{33}$$

as $\mathbf{U}_i$ is unitary. Thus, c_i is independent of the index i.

For $b_i(\hat{\mathbf{p}})$ we get with (29) and (30)

$$\begin{aligned} b_i(\hat{\mathbf{p}}) &= \int p(\mathbf{y}|\mathbf{x}_i) \log\left(\psi \sum_{\mathbf{x}_j \in \mathcal{S}_0} p(\mathbf{y}|\mathbf{x}_j)\right) d\mathbf{y} \\ &= \int p(\mathbf{U}_i^H\mathbf{y}|\mathbf{x}_1) \log\left(\psi \sum_{\mathbf{x}_j \in \mathcal{S}_0} p(\mathbf{y}|\mathbf{x}_j)\right) d\mathbf{y} \\ &= \int p(\mathbf{y}|\mathbf{x}_1) \log\left(\psi \sum_{\mathbf{x}_j \in \mathcal{S}_0} p(\mathbf{y}|\mathbf{U}_i^H\mathbf{x}_j)\right) d\mathbf{y} \\ &= \int p(\mathbf{y}|\mathbf{x}_1) \log\left(\psi \sum_{\mathbf{x}_j \in \mathcal{S}_0} p(\mathbf{y}|\mathbf{x}_j)\right) d\mathbf{y} \end{aligned} \tag{34}$$

where we used (32) and for the last equality that $\sum_{\mathbf{x}_j \in \mathcal{S}_0} p(\mathbf{y}|\mathbf{U}_i^H\mathbf{x}_j) = \sum_{\mathbf{x}_j \in \mathcal{S}_0} p(\mathbf{y}|\mathbf{x}_j)$. Therefore, $b_i(\hat{\mathbf{p}})$ is independent of the index i for the distribution (30). Finally, we have shown that for the distribution in (30), the Kullback-Leibler distance $D\left(p(\mathbf{y}|\mathbf{x}_i)\middle|\middle| \sum_{j=1}^{M} \hat{p}_j p(\mathbf{y}|\mathbf{x}_j)\right)$ is constant, and, thus, (30) is optimum. ■

With (22), (23), (29), and (30) the constrained capacity is given by

$$\begin{aligned} C &= \frac{1}{N} \int p(\mathbf{y}|\mathbf{x}_i) \log\left(\frac{p(\mathbf{y}|\mathbf{x}_i)}{\frac{1}{Q^{N-1}} \sum_{\mathbf{x}_j \in \mathcal{S}_0} p(\mathbf{y}|\mathbf{x}_j)}\right) d\mathbf{y} \\ &= \frac{N-1}{N} \log(Q) \\ &\quad - \frac{1}{N} \int p(\mathbf{y}|\mathbf{x}_i) \log\left(1 + \frac{\sum_{\mathbf{x}_j \in \mathcal{S}_0 \setminus \mathbf{x}_i} p(\mathbf{y}|\mathbf{x}_j)}{p(\mathbf{y}|\mathbf{x}_i)}\right) d\mathbf{y}. \end{aligned} \tag{35}$$

C. Asymptotic SNR Behavior

As we assume the PSD of the channel fading process to be compactly supported and characterized by a maximum normalized Doppler frequency $f_d < 0.5$, there are eigenvalues of the channel covariance matrix $\mathbf{R}_h$ which are close to zero if f_d is not close to 0.5 and if the blocklength N is sufficiently large. Thus, in this case $\mathbf{R}_h$ is close to singular. As in addition, the sequences constituting the set $\mathcal{S}_0$ are distinguishable, we conjecture that the integral in (35) is close to zero and hence,

$$\lim_{\sigma_n^2 \to 0} C \approx \frac{N-1}{N} \log(Q) \tag{36}$$

for N sufficiently large and f_d sufficiently small. This behavior can already be observed for the parameters $N = 6$ and $f_d = 0.2$ used in the numerical evaluation in Section IV-E.

D. Interpretation

The optimum input distribution given in (30) intuitively corresponds to a uniform distribution over the space of distinguishable transmit sequences. One specific solution, being included in the set of optimum input distributions is to use only distinguishable transmit sequences, i.e., sequences taken from one set $\mathcal{S}_0$, thus fulfilling (28), i.e., given by the distribution

$$\hat{p}_i = \begin{cases} \frac{Q}{M} & \text{for } \mathbf{x}_i \in \mathcal{S}_0 \\ 0 & \text{for } \mathbf{x}_i \notin \mathcal{S}_0 \end{cases} . \quad (37)$$

As the cardinality of a subset $\mathcal{S}_0$ is Q^{N-1}, the constrained capacity is limited to $\frac{N-1}{N}\log(Q)$, independent of f_d, corresponding to (36).

In case the set $\mathcal{S}_0$ is constructed such that all used transmit sequences are characterized by having a fixed symbol at a predetermined time instant, this solution corresponds to the use of a pilot symbol. This intuitively explains why at least one signaling dimension, i.e. the information transmitted by one symbol, is lost for requiring a phase reference at the receiver.

The above result should not be understood in the way that it is not possible to use all the transmit sequences of the set $\mathcal{S}$. In this case it has to be assured, that the information that is mapped to non-distinguishable sequences $\mathbf{x}$ is equivalent, as the differentiation between these sequences is impossible.

Following the argumentation in Section IV-C, we conjecture that for infinitely long transmission intervals, i.e., $N \to \infty$, and the infinite SNR case we get $\lim_{N\to\infty} \lim_{\sigma_n^2 \to 0} C = \log(Q)$, which corresponds to the case where the receiver knows the channel fading process.

E. Numerical Results

Fig. 1 shows the result of a Monte Carlo simulation of (35) for $Q = 2$, i.e., BPSK, and for $Q = 4$, i.e., QPSK. The temporal correlation of the channel fading process is determined by the Jakes' spectrum with maximum Doppler frequency f_d. Thus the corresponding covariance matrix is given by (5) and the autocorrelation function $r_h(l) = J_0(2\pi f_d l)$, where J_0 is the zeroth-order Bessel function of the first kind. For comparison also the mutual information in case of perfect channel state information (CSI) is shown. For $SNR \to \infty$, the curves converge to the expression given in (36). In addition, we see that for a given SNR and a given sequence length N the constrained capacity decreases with increasing f_d. Furthermore, it should be noted that the length of the transmission sequence N influences the constrained capacity. The smaller the blocklength N is, the smaller is the constrained capacity. Due to space constraints, this is not shown here.

As the numerical evaluation is based on a Monte Carlo simulation, the calculation time increases strongly with N and Q, as the number of possible transmit sequences increases.

V. Conclusion

In this paper, we have considered a Rayleigh flat-fading channel with temporal correlation, where the channel state is unknown to the transmitter and receiver, while the receiver is aware of the channel law. For PSK signaling sequences, we have derived an explicit expression for the optimum input distribution achieving the constrained capacity. Furthermore, we have identified the strategy of transmitting one pilot symbol as being included in the set of optimum input distributions. For asymptotic high SNR, the constrained capacity is at least degraded by a factor of $(N-1)/N$ compared to the case of perfect channel state information at receiver side.

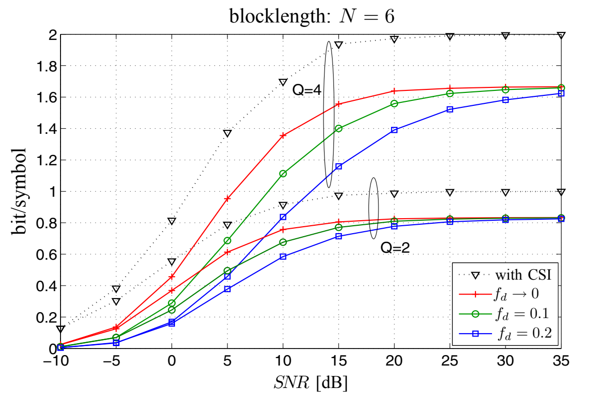

Figure 1. Effect of Q and f_d on the constr. capacity (35), num. evaluation

The extension of this work to signaling sets making use of the amplitude component, e.g., QAM signal constellations is subject to further study. Furthermore, a general optimization over the input distribution, including the choice of the optimal signaling constellation, is of high relevance.

Acknowledgment

The authors would like to thank an anonymous reviewer for pointing out the relation of proposition 1 to [13, Theorem 4.5.1.] and M. Senst for helpful comments on this paper.

References

[1] A. Lapidoth, "On the asymptotic capacity of stationary Gaussian fading channels," *IEEE Trans. on 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," *arXiv:0712.2872v1 [cs.IT]*, Dec. 2007.
[3] G. Durisi, H. Bölcskei, and S. Shamai, "Capacity of underspread noncoherent WSSUS fading channels under peak signal constraints," in *IEEE Int. Symp. on Information Theory (ISIT)*, Jun. 2007, pp. 156–160.
[4] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels," *IEEE Trans. on Inf. Theory*, vol. 49, no. 10, pp. 2426–2467, Oct. 2003.
[5] 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. on Inf. Theory*, vol. 50, no. 12, pp. 3390–3396, Dec. 2004.
[6] A. Lapidoth and S. Shamai, "Fading channels: how perfect need "perfect side information" be?" *IEEE Trans. on Inf. Theory*, vol. 48, no. 5, pp. 1118–1134, May 2002.
[7] T. H. Chan, S. Hranilovic, and F. R. Kschischang, "Capacity-achieving probability measure for conditionally Gaussian channels with bounded inputs," *IEEE Trans. on Inf. Theory*, vol. 51, no. 6, pp. 2073–2088, Jun. 2005.
[8] I. C. Abou-Faycal, M. D. Trott, and S. Shamai, "The capacity of discrete-time memoryless Rayleigh-fading channels," *IEEE Trans. on Inf. Theory*, vol. 47, no. 4, pp. 1290–1301, May 2001.
[9] A. Feiten and R. Mathar, "Capacity-achieving discrete signaling over additive noise channels," in *Proc. IEEE International Conf. on Communications (ICC), Glasgow, UK*, Jun. 2007, pp. 5401–5405.
[10] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. on Inf. Theory*, vol. 39, no. 4, pp. 1293–1302, Jul. 1993.
[11] V. Sethuraman and B. Hajek, "Capacity per unit energy of fading channels with a peak constraint," *IEEE Trans. on Inf. Theory*, vol. 51, no. 9, pp. 3102–3120, Sep. 2005.
[12] T. Cover and J. Thomas, *Elements of Information Theory*. New York: John Wiley, 1991.
[13] R. G. Gallager, *Information Theory and Reliable Communication*. Wiley, 1968.