On the Gain of Joint Processing of Pilot and Data Symbols in Stationary Rayleigh Fading Channels

Meik Dörpinghaus, *Member, IEEE,* Adrian Ispas, and Heinrich Meyr, *Life Fellow, IEEE*

Abstract

In many typical mobile communication receivers the channel is estimated based on pilot symbols to allow for a coherent detection and decoding in a separate processing step. Currently much work is spent on receivers which break up this separation, e.g., by enhancing channel estimation based on reliability information on the data symbols. In the present work, we evaluate the possible gain of a joint processing of data and pilot symbols in comparison to the case of a separate processing in the context of stationary Rayleigh flat-fading channels. Therefore, we discuss the nature of the possible gain of a joint processing of pilot and data symbols. We show that the additional information that can be gained by a joint processing is captured in the temporal correlation of the channel estimation error of the solely pilot based channel estimation, which is not retrieved by the channel decoder in case of separate processing. In addition, we derive a new lower bound on the achievable rate for joint processing of pilot and data symbols.

Index Terms

Channel capacity, fading channels, information rates, joint processing, mismatched decoding, noncoherent, Rayleigh, time-selective.

I. INTRODUCTION

VIRTUALLY all practical mobile communication systems face the problem that communication takes place over a time varying fading channel whose realization is unknown to the receiver. However, for coherent detection and decoding an estimate of the channel fading process is required. For the purpose of channel estimation usually pilot symbols, i.e., symbols which are known to the receiver, are introduced into the transmit sequence. In conventional receiver design the channel is estimated based on these pilot symbols. Based on these channel estimates, in a separate step coherent detection and decoding is performed. Both processing steps are executed separately.

In recent years, much effort has been spent on the study of iterative joint channel estimation and decoding schemes, i.e., schemes, in which the channel estimation is iteratively enhanced based on reliability information on the data symbols delivered by the decoder, see, e.g., [1]–[4]. In this context, the channel estimation is not solely based on pilot symbols, but also on data symbols. This approach is an instance of a *joint processing* of data and pilot symbols in contrast to the separate processing in conventional receiver design. Obviously, this joint processing results in an increased receiver complexity. To evaluate the payoff for the increased receiver complexity, it is important to study the possible performance gain that can be achieved by a joint processing, e.g., in form of an iterative code-aided channel estimation and decoding based receiver, in comparison to a *separate processing* as it is performed in conventional synchronized detection based receivers, where the channel estimation is solely based on pilot symbols.

Therefore, in the present work we evaluate the performance of a joint processing in comparison to synchronized detection with a solely pilot based channel estimation based on the achievable rate. Regarding the channel statistics, we assume a stationary Rayleigh flat-fading channel as it is usually applied to model the fading in a mobile environment without a line of sight component. Furthermore, we assume that the power spectral density (PSD) of the channel fading process is compactly supported, and that the fading process is *non-regular* [5], which is reasonable as the maximum Doppler frequency of typical fading channels is small in comparison to the inverse of the symbol duration. Furthermore, we assume that the receiver is aware of the law of the channel, while neither the transmitter nor the receiver knows the realization of the channel fading process.

There has been a variety of publications studying the achievable rate with pilot symbols, see, e.g., [6]–[12]. Many of these works discuss the achievable rate under the assumption that a channel estimate is acquired based on pilot symbols which is then used for coherent detection, i.e., separate processing. Some of these works consider block-fading, [7], [10], and [12], while [8] and [9] specifically discuss the case of stationary fading. For the case of a stationary single-input single-output Rayleigh flat-fading channel, as we study in the present work, tight bounds on the achievable rate with synchronized detection with a solely pilot based channel estimation, i.e., separate processing, have been given in [8]. In contrast, for the case of a joint processing there is not much knowledge on the achievable rate. Very recently, in [13] the value of joint processing of pilot and data symbols has been studied in the context of a block-fading channel. To the best of our knowledge, there are no results concerning the gain of joint processing of pilot and data symbols for the case of stationary fading channels. Thus, in the present work, we study the achievable rate with a joint processing of pilot and data symbols. We identify the nature of the possible gain of a joint processing of pilot and data symbols in comparison to a separate processing. Furthermore, we derive a lower bound on the achievable rate with joint processing of pilot and data symbols, which, thus, can be seen as an extension of the work given in [13] to the case of stationary Rayleigh flat-fading. In addition, we compare the given lower bound on the achievable rate with joint processing of pilot and data symbols to bounds on the achievable rate with separate processing given in [8] and to bounds on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols given in [14], i.e., without the assumption on pilot symbols inserted into the transmit sequence.

The rest of the paper is organized as follows. In Section II the system model is introduced. Subsequently, in Section III we discuss the nature of the gain by a joint processing of pilot and data symbols, i.e., we discuss which information is discarded in case of a separate processing. Furthermore, existing bounds on the achievable rate with separate processing are briefly recalled. Afterwards, in Section IV a new lower bound on the achievable rate with a joint processing of pilot and data symbols is derived, before it is numerically evaluated and compared to the achievable rate with separate processing and to the achievable rate with i.i.d. zero-mean proper Gaussian inputs in Section V. Finally, Section VI concludes the paper with a brief summary.

II. System Model

We consider a discrete-time zero-mean jointly proper Gaussian flat-fading channel with the following input-output relation

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

with the diagonal matrices $\mathbf{H} = \mathrm{diag}(\mathbf{h})$ and $\mathbf{X} = \mathrm{diag}(\mathbf{x})$. Here the $\mathrm{diag}(\cdot)$ operator generates a diagonal matrix whose diagonal elements are given by the argument vector. The vector $\mathbf{y} = [y_1, \ldots, y_N]^T$ contains the channel output symbols in temporal order. Analogous, $\mathbf{x} = [x_1, \ldots, x_N]^T$, $\mathbf{n} = [n_1, \ldots, n_N]^T$, and $\mathbf{h} = [h_1, \ldots, h_N]^T$ contain the channel input symbols, the additive noise samples and the channel fading weights. All vectors are of length N.

The samples of the additive noise process are assumed to be i.i.d. zero-mean jointly proper Gaussian with variance σ_n^2 and, thus, $\mathbf{R}_n = \mathrm{E}\left[\mathbf{n}\mathbf{n}^H\right] = \sigma_n^2 \mathbf{I}_N$, with $\mathbf{I}_N$ being the identity matrix of size $N \times N$.

The channel fading process is zero-mean jointly proper Gaussian with the temporal correlation characterized by

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

Its variance is given by $r_h(0) = \sigma_h^2$. For mathematical reasons we assume that the autocorrelation function $r_h(l)$ is absolutely summable, i.e.,

$$\sum_{l=-\infty}^{\infty} |r_h(l)| < \infty. \tag{3}$$

The PSD of the channel fading process is defined as

$$S_h(f) = \sum_{m=-\infty}^{\infty} r_h(m)e^{-j2\pi mf}, \qquad |f| \leq 0.5. \tag{4}$$

We assume that the PSD exists, which for a jointly proper Gaussian fading process implies ergodicity. Furthermore, we assume the PSD to be compactly supported within the interval $[-f_d, f_d]$ with f_d being the maximum Doppler shift and $0 < f_d < 0.5$. This means that $S_h(f) = 0$ for $f \notin [-f_d, f_d]$. The assumption of a PSD with limited support is motivated by the fact that the velocity of the transmitter, the receiver, and of objects in the environment is limited. To ensure ergodicity, we exclude the case $f_d = 0$. In matrix-vector notation, the temporal correlation is expressed by the autocorrelation matrix $\mathbf{R}_h$ given by

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

For the following derivation we introduce the subvectors $\mathbf{x}_D$ containing all data symbols of $\mathbf{x}$ and the vector $\mathbf{x}_P$ containing all pilot symbols of $\mathbf{x}$. Correspondingly, we define the vectors $\mathbf{h}_D$, $\mathbf{h}_P$, $\mathbf{y}_D$, $\mathbf{y}_P$, $\mathbf{n}_D$, and $\mathbf{n}_P$.

The transmit symbol sequence consists of data symbols with a maximal average power σ_x^2, i.e.,

$$\frac{1}{N_D}\mathrm{E}\left[\mathbf{x}_D^H\mathbf{x}_D\right] \leq \sigma_x^2 \tag{6}$$

with N_D being the length of the vector $\mathbf{x}_D$, and periodically inserted pilot symbols with a fixed transmit power σ_x^2. Each L-th symbol is a pilot symbol. We assume that the pilot spacing is chosen such that the channel fading process is sampled at least with Nyquist rate, i.e.,

$$L < \frac{1}{2f_d}. \tag{7}$$

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

Based on the preceding definitions the average SNR ρ is given by

$$\rho = \frac{\sigma_x^2 \sigma_h^2}{\sigma_n^2}. \tag{8}$$

III. THE NATURE OF THE GAIN BY JOINT PROCESSING OF DATA AND PILOT SYMBOLS

Before we quantitatively discuss the value of a joint processing of data and pilot symbols, we discuss the nature of the possible gain of such a joint processing in comparison to a separate processing of data and pilot symbols. The mutual information between the transmitter and the receiver is given by $\mathcal{I}(\mathbf{x}_D; \mathbf{y}_D, \mathbf{y}_P, \mathbf{x}_P)$. As the pilot symbols are known to the receiver, the pilot symbol vector $\mathbf{x}_P$ is found at the RHS of the semicolon. We separate $\mathcal{I}(\mathbf{x}_D; \mathbf{y}_D, \mathbf{y}_P, \mathbf{x}_P)$ as follows

$$\mathcal{I}(\mathbf{x}_D; \mathbf{y}_D, \mathbf{y}_P, \mathbf{x}_P) \overset{(a)}{=} \mathcal{I}(\mathbf{x}_D; \mathbf{y}_D|\mathbf{y}_P, \mathbf{x}_P) + \mathcal{I}(\mathbf{x}_D; \mathbf{y}_P|\mathbf{x}_P) + \mathcal{I}(\mathbf{x}_D; \mathbf{x}_P)$$
$$\overset{(b)}{=} \mathcal{I}(\mathbf{x}_D; \mathbf{y}_D|\mathbf{y}_P, \mathbf{x}_P) \tag{9}$$

where (a) follows from the chain rule for mutual information and (b) holds due to the independency of the data and pilot symbols. The question is, which portion of $\mathcal{I}(\mathbf{x}_D; \mathbf{y}_D|\mathbf{y}_P, \mathbf{x}_P)$ can be achieved by synchronized detection with a solely pilot based channel estimation, i.e., with separate processing.

A. Separate Processing

The receiver has to find the most likely data sequence $\mathbf{x}_D$ based on the observation $\mathbf{y}$ while knowing the pilots $\mathbf{x}_P$, i.e.,

$$\hat{\mathbf{x}}_D = \arg \max_{\mathbf{x}_D \in \mathcal{C}_D} p(\mathbf{y}|\mathbf{x}) = \arg \max_{\mathbf{x}_D \in \mathcal{C}_D} p(\mathbf{y}_D|\mathbf{x}_D, \mathbf{y}_P, \mathbf{x}_P) \tag{10}$$

with the set $\mathcal{C}_D$ containing all possible data sequences $\mathbf{x}_D$. It can be shown that the probability density function (PDF) $p(\mathbf{y}_D|\mathbf{x}_D, \mathbf{y}_P, \mathbf{x}_P)$ is proper Gaussian and, thus, is completely described by the conditional mean and covariance

$$\mathrm{E}\left[\mathbf{y}_D|\mathbf{x}_D, \mathbf{y}_P, \mathbf{x}_P\right] = \mathbf{X}_D \mathrm{E}\left[\mathbf{h}_D|\mathbf{y}_P, \mathbf{x}_P\right] = \mathbf{X}_D \hat{\mathbf{h}}_{\mathrm{pil},D} \tag{11}$$

$$\mathrm{cov}[\mathbf{y}_D|\mathbf{x}_D, \mathbf{y}_P, \mathbf{x}_P] = \mathbf{X}_D \mathbf{R}_{e_{\mathrm{pil}},D} \mathbf{X}_D^H + \sigma_n^2 \mathbf{I}_{N_D} \tag{12}$$

where $\mathbf{X}_D = \mathrm{diag}(\mathbf{x}_D)$ and $\mathbf{I}_{N_D}$ is an identity matrix of size $N_D \times N_D$. The vector $\hat{\mathbf{h}}_{\mathrm{pil},D}$ is an MMSE channel estimate at the data symbol time instances based on the pilot symbols, which is denoted by the index pil. Furthermore, the corresponding channel estimation error

$$\mathbf{e}_{\mathrm{pil},D} = \mathbf{h}_D - \hat{\mathbf{h}}_{\mathrm{pil},D} \tag{13}$$

is zero-mean proper Gaussian and

$$\mathbf{R}_{e_{\mathrm{pil}},D} = \mathrm{E}\left[\mathbf{e}_{\mathrm{pil},D}\mathbf{e}_{\mathrm{pil},D}^H|\mathbf{x}_P\right] \tag{14}$$

is its correlation matrix, which is independent of $\mathbf{y}_P$ due to the principle of orthogonality.

Based on (11) and (12) conditioning of $\mathbf{y}_D$ on $\mathbf{x}_D, \mathbf{y}_P, \mathbf{x}_P$ is equivalent to conditioning on $\mathbf{x}_D, \hat{\mathbf{h}}_{\mathrm{pil},D}, \mathbf{x}_P$, i.e.,

$$p(\mathbf{y}_D|\mathbf{x}_D, \mathbf{y}_P, \mathbf{x}_P) = p(\mathbf{y}_D|\mathbf{x}_D, \hat{\mathbf{h}}_{\mathrm{pil},D}, \mathbf{x}_P) \tag{15}$$

as all information on $\mathbf{h}_D$ delivered by $\mathbf{y}_P$ is contained in $\hat{\mathbf{h}}_{\mathrm{pil},D}$ while conditioning on $\mathbf{x}_P$. Thus, (10) can be written as

$$\hat{\mathbf{x}}_D = \arg \max_{\mathbf{x}_D \in \mathcal{C}_D} p(\mathbf{y}_D|\mathbf{x}_D, \hat{\mathbf{h}}_{\mathrm{pil},D}, \mathbf{x}_P) = \arg \max_{\mathbf{x}_D \in \mathcal{C}_D} p(\mathbf{y}|\mathbf{x}_D, \hat{\mathbf{h}}_{\mathrm{pil}}, \mathbf{x}_P). \tag{16}$$

For ease of notation in the following we will use the metric on the RHS of (16) where $\hat{\mathbf{h}}_{\mathrm{pil}}$ corresponds to $\hat{\mathbf{h}}_{\mathrm{pil},D}$ but also contains channel estimates at the pilot symbol time instances, i.e.,

$$\hat{\mathbf{h}}_{\mathrm{pil}} = \mathrm{E}\left[\mathbf{h}|\mathbf{y}_P, \mathbf{x}_P\right]. \tag{17}$$

Based on $\hat{\mathbf{h}}_{\mathrm{pil}}$, (1) can be expressed by

$$\mathbf{y} = \mathbf{X}(\hat{\mathbf{h}}_{\mathrm{pil}} + \mathbf{e}_{\mathrm{pil}}) + \mathbf{n} \tag{18}$$

where $\mathbf{e}_{\mathrm{pil}}$ is the estimation error including the pilot symbol time instances. As the channel estimation is an interpolation, the error process is not white but temporally correlated, i.e.,

$$\mathbf{R}_{e_{\mathrm{pil}}} = \mathrm{E}\left[\mathbf{e}_{\mathrm{pil}}\mathbf{e}_{\mathrm{pil}}^H|\mathbf{x}_P\right] \tag{19}$$

is not diagonal, cf. (35). As the estimation error process is zero-mean proper Gaussian, the PDF in (16) is given by

$$p(\mathbf{y}|\mathbf{x}_D, \hat{\mathbf{h}}_{\mathrm{pil}}, \mathbf{x}_P) = \mathcal{CN}\left(\mathbf{X}\hat{\mathbf{h}}_{\mathrm{pil}}, \mathbf{X}\mathbf{R}_{e_{\mathrm{pil}}}\mathbf{X}^H + \sigma_n^2 \mathbf{I}_N\right) \tag{20}$$

where $\mathcal{CN}(\boldsymbol{\mu}, \mathbf{C})$ denotes a proper Gaussian PDF with mean $\boldsymbol{\mu}$ and covariance $\mathbf{C}$ and where $\mathbf{I}_N$ is the $N \times N$ identity matrix.[1]

---

[1] Note that for the case of data transmission only (20) becomes $p(\mathbf{y}|\mathbf{x}_D) = \mathcal{CN}(\mathbf{0}, \mathbf{X}\mathbf{R}_h\mathbf{X}^H + \sigma_n^2\mathbf{I}_N)$ as in this case $\hat{\mathbf{h}}_{\mathrm{pil}} = \mathbf{0}$ and $\mathbf{R}_{e_{\mathrm{pil}}} = \mathbf{R}_h$.

Corresponding to (15), we can also rewrite $p(\mathbf{y}_D|\mathbf{y}_P, \mathbf{x}_P)$ as follows

$$
\begin{aligned}
p(\mathbf{y}_D|\mathbf{y}_P, \mathbf{x}_P) &= \int p(\mathbf{y}_D|\mathbf{x}_D, \mathbf{y}_P, \mathbf{x}_P)p(\mathbf{x}_D|\mathbf{y}_P, \mathbf{x}_P)d\mathbf{x}_D \\
&\overset{(a)}{=} \int p(\mathbf{y}_D|\mathbf{x}_D, \hat{\mathbf{h}}_{\mathrm{pil},D}, \mathbf{x}_P)p(\mathbf{x}_D)d\mathbf{x}_D \\
&= p(\mathbf{y}_D|\hat{\mathbf{h}}_{\mathrm{pil},D}, \mathbf{x}_P)
\end{aligned}
\tag{21}
$$

where for (a) we have used (15) and the independency of $\mathbf{x}_D$ of $\mathbf{x}_P$ and $\mathbf{y}_P$.

Based on (15) and (21), we can also rewrite (9) as

$$
\mathcal{I}(\mathbf{x}_D; \mathbf{y}_D|\mathbf{y}_P, \mathbf{x}_P) = \mathcal{I}(\mathbf{x}_D; \mathbf{y}_D|\hat{\mathbf{h}}_{\mathrm{pil}}, \mathbf{x}_P) \overset{(a)}{=} \mathcal{I}(\mathbf{x}_D; \mathbf{y}_D|\hat{\mathbf{h}}_{\mathrm{pil}})
\tag{22}
$$

and where (a) holds as the pilot symbols are deterministic.

However, typical channel decoders like a Viterbi decoder are not able to exploit the temporal correlation of the channel estimation error. Therefore, the decoder performs mismatched decoding based on the assumption that the estimation error process is white, i.e., $p(\mathbf{y}|\mathbf{x}_D, \hat{\mathbf{h}}_{\mathrm{pil}}, \mathbf{x}_P)$ is approximated by

$$
p(\mathbf{y}|\mathbf{x}_D, \hat{\mathbf{h}}_{\mathrm{pil}}, \mathbf{x}_P) \approx \mathcal{CN}\left(\mathbf{X}\hat{\mathbf{h}}_{\mathrm{pil}}, \sigma_{e_{\mathrm{pil}}}^2 \mathbf{X}\mathbf{X}^H + \sigma_n^2 \mathbf{I}_N\right).
\tag{23}
$$

As it is assumed that the channel is at least sampled with Nyquist frequency, see (7), for an infinite block length $N \to \infty$ the channel estimation error variance $\sigma_{e_{\mathrm{pil}}}^2$ is independent of the symbol time instant [8] and is given by

$$
\sigma_{e_{\mathrm{pil}}}^2 = \int_{f=-\frac{1}{2}}^{\frac{1}{2}} S_{e_{\mathrm{pil}}}(f)df = \int_{f=-\frac{1}{2}}^{\frac{1}{2}} \frac{S_h(f)}{\frac{\rho}{L}\frac{S_h(f)}{\sigma_h^2} + 1}df
\tag{24}
$$

where $S_{e_{\mathrm{pil}}}(f)$ is the PSD of the channel estimation error process in case the channel estimation is solely based on pilot symbols, which is given in (102) in Appendix B. Hence, the variance of the channel estimation process, i.e., the entries of $\hat{\mathbf{h}}_{\mathrm{pil}}$, is given by $\sigma_h^2 - \sigma_{e_{\mathrm{pil}}}^2$, which follows from the principle of orthogonality in LMMSE estimation.

As the information contained in the temporal correlation of the channel estimation error is not retrieved by synchronized detection with a solely pilot based channel estimation, the mutual information in this case corresponds to the sum of the mutual information for each individual data symbol time instant. As, obviously, by this separate processing information is discarded, the following inequality for the achievable rate holds:

$$
\begin{aligned}
\lim_{N \to \infty} \frac{1}{N}\mathcal{I}(\mathbf{x}_D; \mathbf{y}_D|\hat{\mathbf{h}}_{\mathrm{pil}}) = \mathcal{I}'(\mathbf{x}_D; \mathbf{y}_D|\hat{\mathbf{h}}_{\mathrm{pil}}) \\
\geq \frac{L-1}{L}\mathcal{I}(x_{D_k}; y_{D_k}|\hat{\mathbf{h}}_{\mathrm{pil}}) \\
= \frac{L-1}{L}\mathcal{I}(x_{D_k}; y_{D_k}|\hat{h}_{\mathrm{pil},D_k}) = \mathcal{R}_{\mathrm{sep}}
\end{aligned}
\tag{25}
$$

where $\mathcal{I}'$ denotes the mutual information rate and the index D_k refers to an arbitrarily chosen data symbol, i.e., $x_{D_k} = [\mathbf{x}_D]_k$. Furthermore, $\hat{h}_{\mathrm{pil},D_k}$ is the solely pilot based channel estimate at the data symbol time instant D_k. The pre-factor $(L-1)/L$ arises from the fact that each L-th symbol is a pilot symbol. In the following, we denote the achievable rate with separate processing by $\mathcal{R}_{\mathrm{sep}}$.

As the LHS of (25) is the mutual information of the channel and as the RHS of (25) is the mutual information achievable with synchronized detection with a metric corresponding to (23) and a solely pilot based channel estimation, i.e., a separate processing, the difference of both terms upper bounds the possible gain due to joint processing of data and pilot symbols. Obviously, the additional information that can be gained by a joint processing in contrast to the separate processing is contained in the temporal correlation of the channel estimation error process.

Regarding synchronized detection in combination with a solely pilot based channel estimation, i.e., the separate processing approach, in [8] bounds on the achievable rate have been given, which for zero-mean proper Gaussian data symbols become

$$\mathcal{R}_{\text{sep}} \geq \mathcal{R}_{L,\text{sep}} = \frac{L-1}{L} \mathrm{E}_{\hat{h}_{\text{pil},D_k}} \left[\log \left(1 + \frac{\sigma_x^2 |\hat{h}_{\text{pil},D_k}|^2}{\sigma_{e_{\text{pil}}}^2 \sigma_x^2 + \sigma_n^2} \right) \right]$$

$$= \frac{L-1}{L} \int_{z=0}^{\infty} \log \left(1 + \rho \frac{1 - \frac{\sigma_{e_{\text{pil}}}^2}{\sigma_h^2}}{1 + \rho \frac{\sigma_{e_{\text{pil}}}^2}{\sigma_h^2}} z \right) e^{-z} dz \tag{26}$$

$$\mathcal{R}_{\text{sep}} \leq \mathcal{R}_{U,\text{sep}} = \mathcal{R}_{L,\text{sep}} + \frac{L-1}{L} \mathrm{E}_{x_{D_k}} \left[\log \left(\frac{\sigma_x^2 \sigma_{e_{\text{pil}}}^2 + \sigma_n^2}{|x_{D_k}|^2 \sigma_{e_{\text{pil}}}^2 + \sigma_n^2} \right) \right]$$

$$= \mathcal{R}_{L,\text{sep}} + \frac{L-1}{L} \left(\log \left(1 + \rho \frac{\sigma_{e_{\text{pil}}}^2}{\sigma_h^2} \right) - \int_{z=0}^{\infty} \log \left(1 + \rho \frac{\sigma_{e_{\text{pil}}}^2}{\sigma_h^2} z \right) e^{-z} dz \right). \tag{27}$$

Based on the lower bound in (26) 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 $\left(1 - \frac{\sigma_{e_{\text{pil}}}^2}{\sigma_h^2} \right) / \left(1 + \rho \frac{\sigma_{e_{\text{pil}}}^2}{\sigma_h^2} \right)$ due to the channel estimation error variance. The additional term in the upper bound in (27) arises from the fact that the effective noise, i.e., $e_{\text{pil},D_k} x_{D_k} + n_{D_k}$, is non-Gaussian. Here e_{D_k} is the estimation error at the data symbol time instant D_k, i.e., $e_{D_k} = [\mathbf{e}_{\text{pil},D}]_k$.

IV. JOINT PROCESSING OF DATA AND PILOT SYMBOLS

Now, we give a new lower bound on the achievable rate for a joint processing of data and pilot symbols. The following approach can be seen as an extension of the work in [13] for the case of a block-fading channel to the stationary Rayleigh flat-fading scenario discussed in the present work. Therefore, analogously to [13] we decompose and lower-bound the mutual information between the transmitter and the receiver $\mathcal{I}(\mathbf{x}_D; \mathbf{y}_D, \mathbf{y}_P, \mathbf{x}_P)$ as follows

$$\mathcal{I}(\mathbf{x}_D; \mathbf{y}_D, \mathbf{y}_P, \mathbf{x}_P) \overset{(a)}{=} \mathcal{I}(\mathbf{x}_D; \mathbf{y}_D, \mathbf{y}_P, \mathbf{x}_P, \mathbf{h}) - \mathcal{I}(\mathbf{x}_D; \mathbf{h} | \mathbf{y}_D, \mathbf{y}_P, \mathbf{x}_P)$$

$$\overset{(b)}{=} \mathcal{I}(\mathbf{x}_D; \mathbf{y}_D, \mathbf{h}) - h(\mathbf{h} | \mathbf{y}_D, \mathbf{y}_P, \mathbf{x}_P) + h(\mathbf{h} | \mathbf{x}_D, \mathbf{y}_D, \mathbf{y}_P, \mathbf{x}_P)$$

$$\overset{(c)}{\geq} \mathcal{I}(\mathbf{x}_D; \mathbf{y}_D, \mathbf{h}) - h(\mathbf{h} | \mathbf{y}_P, \mathbf{x}_P) + h(\mathbf{h} | \mathbf{x}_D, \mathbf{y}_D, \mathbf{y}_P, \mathbf{x}_P) \tag{28}$$

where (a) follows from the chain rule for mutual information. For the first term in (b) we have used the fact that due to the knowledge on $\mathbf{h}$, the knowledge on $\mathbf{y}_P$ and $\mathbf{x}_P$ does not increase the mutual information between $\mathbf{x}_D$ and $\mathbf{y}_D$. Finally, (c) is due to the fact that conditioning reduces entropy. Note, the first term on the RHS of (28) is the mutual information in case of perfect channel knowledge.

In the following we deviate from the derivation given in [13]. Now, we calculate both differential entropy terms at the RHS of (28). Therefore, we rewrite the RHS of (28) as follows

$$\mathcal{I}(\mathbf{x}_D; \mathbf{y}_D, \mathbf{y}_P, \mathbf{x}_P) \geq \mathcal{I}(\mathbf{x}_D; \mathbf{y}_D, \mathbf{h}) - h(\mathbf{h} | \mathbf{y}_P, \mathbf{x}_P) + h(\mathbf{h} | \mathbf{x}_D, \mathbf{y}_D, \mathbf{y}_P, \mathbf{x}_P)$$

$$\overset{(a)}{=} \mathcal{I}(\mathbf{x}_D; \mathbf{y}_D, \mathbf{h}) - h(\mathbf{h} | \hat{\mathbf{h}}_{\text{pil}}, \mathbf{x}_P) + h(\mathbf{h} | \hat{\mathbf{h}}_{\text{joint}}, \mathbf{x}_D, \mathbf{x}_P)$$

$$\overset{(b)}{=} \mathcal{I}(\mathbf{x}_D; \mathbf{y}_D, \mathbf{h}) - h(\hat{\mathbf{h}}_{\text{pil}} + \mathbf{e}_{\text{pil}} | \hat{\mathbf{h}}_{\text{pil}}, \mathbf{x}_P) + h(\hat{\mathbf{h}}_{\text{joint}} + \mathbf{e}_{\text{joint}} | \hat{\mathbf{h}}_{\text{joint}}, \mathbf{x}_D, \mathbf{x}_P)$$

$$\overset{(c)}{=} \mathcal{I}(\mathbf{x}_D; \mathbf{y}_D, \mathbf{h}) - h(\mathbf{e}_{\text{pil}} | \mathbf{x}_P) + h(\mathbf{e}_{\text{joint}} | \mathbf{x}_D, \mathbf{x}_P)$$

$$\overset{(d)}{=} \mathcal{I}(\mathbf{x}_D; \mathbf{y}_D, \mathbf{h}) - \mathrm{E}_{\mathbf{x}_P} \left[\log \det \left(\pi e \mathbf{R}_{e_{\text{pil}}} \right) \right] + \mathrm{E}_{\mathbf{x}_P, \mathbf{x}_D} \left[\log \det \left(\pi e \mathbf{R}_{e_{\text{joint}}} \right) \right]$$

$$\overset{(e)}{=} \mathcal{I}(\mathbf{x}_D; \mathbf{y}_D, \mathbf{h}) - \log \det \left(\mathbf{R}_{e_{\text{pil}}} \right) + \mathrm{E}_{\mathbf{x}_D} \left[\log \det \left(\mathbf{R}_{e_{\text{joint}}} \right) \right] \tag{29}$$

where for the second term in (a) we have substituted the condition on $\mathbf{y}_P$ by $\hat{\mathbf{h}}_{\text{pil}}$, which is possible as the estimate $\hat{\mathbf{h}}_{\text{pil}}$ contains the same information on $\mathbf{h}$ as $\mathbf{y}_P$ while conditioning on $\mathbf{x}_P$. Corresponding to the solely pilot based channel estimate $\hat{\mathbf{h}}_{\text{pil}}$, based on $\mathbf{x}_D$, $\mathbf{x}_P$, $\mathbf{y}_D$, and $\mathbf{y}_P$, we can calculate the estimate $\hat{\mathbf{h}}_{\text{joint}}$, which is based on data and pilot symbols. Like $\hat{\mathbf{h}}_{\text{pil}}$ this estimate is a MAP estimate, which, due to the jointly Gaussian nature of the problem, is an MMSE estimate, i.e.,

$$\hat{\mathbf{h}}_{\text{joint}} = \mathrm{E}\left[\mathbf{h}|\mathbf{y}_P, \mathbf{x}_P, \mathbf{y}_D, \mathbf{x}_D\right]. \tag{30}$$

Thus, for (a) we have substituted the conditioning on $\mathbf{y}_D$ and $\mathbf{y}_P$ by conditioning on $\hat{\mathbf{h}}_{\text{joint}}$ in the third term, as $\hat{\mathbf{h}}_{\text{joint}}$ contains all information on $\mathbf{h}$ that is contained in $\mathbf{y}_D$ and $\mathbf{y}_P$ while $\mathbf{x}_D$ and $\mathbf{x}_P$ are known. For equality (b) we have used for the second term that $\mathbf{h}$ can be expressed as a sum of its estimate $\hat{\mathbf{h}}_{\text{pil}}$ and the estimation error $\mathbf{e}_{\text{pil}}$, cf. (18). Analogously, for the third term we used the separation of $\mathbf{h}$ into the estimate $\hat{\mathbf{h}}_{\text{joint}}$ and the corresponding estimation error $\mathbf{e}_{\text{joint}}$, i.e.,

$$\mathbf{e}_{\text{joint}} = \mathbf{h} - \hat{\mathbf{h}}_{\text{joint}}. \tag{31}$$

Equality (c) is due to the fact that the addition of a constant does not change differential entropy and that the estimation error $\mathbf{e}_{\text{pil}}$ is independent of the estimate $\hat{\mathbf{h}}_{\text{pil}}$ and analogously $\mathbf{e}_{\text{joint}}$, which depends on $\mathbf{x}_P$ and $\mathbf{x}_D$, is independent of $\hat{\mathbf{h}}_{\text{joint}}$ due to the orthogonality principle in LMMSE estimation. Finally, (d) follows from the fact that the estimation error processes are zero-mean jointly proper Gaussian. Here the error correlation matrices are given by (19) and by

$$\mathbf{R}_{e_{\text{joint}}} = \mathrm{E}\left[\mathbf{e}_{\text{joint}}\mathbf{e}_{\text{joint}}^H|\mathbf{x}_D, \mathbf{x}_P\right]. \tag{32}$$

For (e) we have used that the pilot symbols are deterministic. Therefore, the expectation over $\mathbf{x}_P$ in the second and third term can be removed. However, the channel estimation error $\mathbf{e}_{\text{joint}}$ depends on the distribution of the data symbols $\mathbf{x}_D$. Concerning the third term on the RHS of (29), it can be shown that the differential entropy rate $h'(\mathbf{e}_{\text{joint}}|\mathbf{x}_D, \mathbf{x}_P)$, i.e.,

$$h'(\mathbf{e}_{\text{joint}}|\mathbf{x}_D, \mathbf{x}_P) = \lim_{N\to\infty} \frac{1}{N} h(\mathbf{e}_{\text{joint}}|\mathbf{x}_D, \mathbf{x}_P) \tag{33}$$

is minimized for a given average transmit power σ_x^2 if the data symbols are constant modulus (CM) symbols with power σ_x^2, see Appendix A. Within this proof the restriction to an absolutely summable autocorrelation function $r_h(l)$, see (3), is required.

Thus, based on (29) a lower bound for the achievable rate with joint processing of data and pilot symbols is given by

$$\begin{aligned}
\mathcal{I}'(\mathbf{x}_D; \mathbf{y}_D, \mathbf{y}_P, \mathbf{x}_P) &= \lim_{N\to\infty} \frac{1}{N}\mathcal{I}(\mathbf{x}_D; \mathbf{y}_D, \mathbf{y}_P, \mathbf{x}_P) \\
&\geq \lim_{N\to\infty} \frac{1}{N}\left\{\mathcal{I}(\mathbf{x}_D; \mathbf{y}_D, \mathbf{h}) - \log\det\left(\mathbf{R}_{e_{\text{pil}}}\right) + \log\det\left(\mathbf{R}_{e_{\text{joint,CM}}}\right)\right\} \\
&\stackrel{(a)}{=} \lim_{N\to\infty} \frac{1}{N}\mathcal{I}(\mathbf{x}_D; \mathbf{y}_D, \mathbf{h}) - \int_{-\frac{1}{2}}^{\frac{1}{2}} \log\left(\frac{S_{e_{\text{pil}}}(f)}{S_{e_{\text{joint,CM}}}(f)}\right) df
\end{aligned} \tag{34}$$

with $\mathbf{R}_{e_{\text{joint,CM}}}$ corresponding to (32), but under the assumption of CM data symbols with transmit power σ_x^2. As $\mathbf{R}_{e_{\text{joint,CM}}}$ only depends on the distribution of the magnitude of the data symbols contained in $\mathbf{x}_D$, which is constant and deterministic, we can remove the expectation operation with respect to $\mathbf{x}_D$. Note that the CM assumption has only been used to lower-bound the third term at the RHS of (29), and not the whole expression at the RHS of (29). For (a) in (34) we have used Szegö's theorem on the asymptotic eigenvalue distribution of Hermitian Toeplitz matrices [15]. $S_{e_{\text{pil}}}(f)$ and $S_{e_{\text{joint,CM}}}(f)$ are the PSDs of the channel estimation error processes, on the one hand, if the estimation is solely based on pilot symbols,

and on the other hand, if the estimation is based on data and pilot symbols, assuming CM data symbols. They are given by

$$S_{e_{\text{pil}}}(f) = \frac{S_h(f)}{\frac{\rho}{L}\frac{S_h(f)}{\sigma_h^2} + 1} \tag{35}$$

$$S_{e_{\text{joint,CM}}}(f) = \frac{S_h(f)}{\rho\frac{S_h(f)}{\sigma_h^2} + 1}. \tag{36}$$

The derivation of these PSDs is given in Appendix B.

However, the application of Szegö's theorem for (a) in (34) requires several steps, which we discuss in the following. The limit over the second and the third term on the LHS of (a) in (34) can be transformed as follows

$$\lim_{N \to \infty} \frac{1}{N} \left\{ \log \det \left(\mathbf{R}_{e_{\text{pil}}} \right) - \log \det \left(\mathbf{R}_{e_{\text{joint,CM}}} \right) \right\}$$

$$\stackrel{(a)}{=} \lim_{N \to \infty} \frac{1}{N} \left\{ \log \det \left(\mathbf{C}_{e_{\text{pil}}} \right) - \log \det \left(\mathbf{C}_{e_{\text{joint,CM}}} \right) \right\}$$

$$\stackrel{(b)}{=} \lim_{N \to \infty} \frac{1}{N} \left\{ \log \det \left(\mathbf{F}\mathbf{\Lambda}_{e_{\text{pil}}}\mathbf{F}^H \right) - \log \det \left(\mathbf{F}\mathbf{\Lambda}_{e_{\text{joint,CM}}}\mathbf{F}^H \right) \right\}$$

$$= \lim_{N \to \infty} \frac{1}{N} \left\{ \log \det \left(\mathbf{F}\mathbf{\Lambda}_{e_{\text{pil}}}\mathbf{\Lambda}_{e_{\text{joint,CM}}}^{-1}\mathbf{F}^H \right) \right\}$$

$$\stackrel{(c)}{=} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log \left(\frac{S_{e_{\text{pil}}}(f)}{S_{e_{\text{joint,CM}}}(f)} \right) df$$

$$\stackrel{(d)}{=} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log \left(\frac{\rho\frac{S_h(f)}{\sigma_h^2} + 1}{\frac{\rho}{L}\frac{S_h(f)}{\sigma_h^2} + 1} \right) df \tag{37}$$

where for (a) we have substituted the Toeplitz matrices $\mathbf{R}_{e_{\text{pil}}}$ and $\mathbf{R}_{e_{\text{joint,CM}}}$ by their asymptotic equivalent circulant matrices $\mathbf{C}_{e_{\text{pil}}}$ and $\mathbf{C}_{e_{\text{joint,CM}}}$, see [16]. Furthermore, for (b) we have used the spectral decompositions of the circulant matrices given by

$$\mathbf{C}_{e_{\text{pil}}} = \mathbf{F}\mathbf{\Lambda}_{e_{\text{pil}}}\mathbf{F}^H \tag{38}$$

$$\mathbf{C}_{e_{\text{joint,CM}}} = \mathbf{F}\mathbf{\Lambda}_{e_{\text{joint,CM}}}\mathbf{F}^H \tag{39}$$

where $\mathbf{\Lambda}_{e_{\text{pil}}}$ and $\mathbf{\Lambda}_{e_{\text{joint,CM}}}$ are diagonal matrices containing the eigenvalues of $\mathbf{C}_{e_{\text{pil}}}$ and $\mathbf{C}_{e_{\text{joint,CM}}}$, and the matrix $\mathbf{F}$ is a unitary DFT-matrix whose elements are given by

$$[F]_{k,l} = \frac{1}{\sqrt{N}}e^{j2\pi\frac{(k-1)(l-1)}{N}}. \tag{40}$$

For (c) in (37) we have then used Szegö's theorem on the asymptotic eigenvalue distribution of Hermitian Toeplitz matrices [15]. Therefore, first consider that the matrix $\mathbf{F}\mathbf{\Lambda}_{e_{\text{pil}}}\mathbf{\Lambda}_{e_{\text{joint,CM}}}^{-1}\mathbf{F}^H$ on the LHS of (c) is again a circulant matrix and that there exists an asymptotically equivalent Toeplitz matrix. Furthermore, the eigenvalues of $\mathbf{C}_{e_{\text{pil}}}$ are samples of the PSD $S_{e_{\text{pil}}}(f)$ and the eigenvalues of $\mathbf{C}_{e_{\text{joint,CM}}}$ are samples of the PSD $S_{e_{\text{joint,CM}}}(f)$. Here we assume a construction of the circulant matrices as described in [16, (4.32)], see also in Appendix A from (72) to (76). Furthermore, the application of Szegö's theorem requires that the log-function is continuous on the support of the eigenvalues of the matrix $\mathbf{\Lambda}_{e_{\text{pil}}}\mathbf{\Lambda}_{e_{\text{joint,CM}}}^{-1}$. This means that we have to show that the eigenvalues of $\mathbf{\Lambda}_{e_{\text{pil}}}\mathbf{\Lambda}_{e_{\text{joint,CM}}}^{-1}$ are bounded away from zero and from infinity. That this is indeed the case will become obvious after introducing $S_{e_{\text{pil}}}(f)$ and $S_{e_{\text{joint,CM}}}(f)$ given in (35) and (36) as it has been done in (d). Obviously, the argument of the log at the RHS of (37) is larger than zero and smaller than infinity on the interval $f \in [-0.5, 0.5]$. Therefore, the integral on the RHS of (37) exists, implying that also the LHS of (c) in (37) is bounded and, thus, that the eigenvalues of $\mathbf{\Lambda}_{e_{\text{pil}}}\mathbf{\Lambda}_{e_{\text{joint,CM}}}^{-1}$

are bounded away from zero and from infinity. Thus, in conclusion we have shown that Szegö's theorem is applicable and that (a) in (34) holds.

The first term on the RHS of (34) is the mutual information rate in case of perfect channel state information, which for an average power constraint is maximized with i.i.d. zero-mean proper Gaussian data symbols. Thus, we get the following lower bound on the achievable rate with joint processing

$$\mathcal{R}_{L,\text{joint}} = \frac{L-1}{L}C_{\text{perf}}(\rho) - \int_{-\frac{1}{2}}^{\frac{1}{2}} \log\left(\frac{\frac{\rho}{\sigma_h^2}S_h(f)+1}{\frac{\rho}{L\sigma_h^2}S_h(f)+1}\right)df \tag{41}$$

where $C_{\text{perf}}(\rho)$ corresponds to the coherent capacity with

$$C_{\text{perf}}(\rho) = \mathrm{E}_{h_k}\left[\log\left(1+\rho\frac{|h_k|^2}{\sigma_h^2}\right)\right] = \int_{z=0}^{\infty} \log\left(1+\rho z\right)e^{-z}dz \tag{42}$$

and the factor $(L-1)/L$ arises as each L-th symbol is a pilot symbol.

A. Lower Bound on the Achievable Rate for a Joint Processing of Data and Pilot Symbols and a Fixed Pilot Spacing

Substituting (42) into (41) we have found a lower bound on the achievable rate with joint processing of data and pilot symbols, for a given pilot spacing L and stationary Rayleigh flat-fading.

For the special case of a rectangular PSD[2] of the channel fading process, i.e.,

$$S_h(f) = \begin{cases} \frac{\sigma_h^2}{2f_d} & \text{for } |f| \leq f_d \\ 0 & \text{otherwise} \end{cases} \tag{43}$$

the lower bound in (41) becomes

$$\mathcal{R}_{L,\text{joint}}\Big|_{\text{rect.}S_h(f)} = \frac{L-1}{L}\int_{z=0}^{\infty} \log\left(1+\rho z\right)e^{-z}dz - 2f_d\log\left(\frac{\frac{\rho}{2f_d}+1}{\frac{\rho}{L2f_d}+1}\right). \tag{44}$$

B. Lower Bound on the Achievable Rate for a Joint Processing of Data and Pilot Symbols and an Optimal Pilot Spacing

Obviously, the lower bound in (44) still depends on the pilot spacing L. In case the pilot spacing is not fixed, we can further enhance it by calculating the supremum of (44) with respect to L. In this regard, it has to be considered that the pilot spacing L is an integer value. Furthermore, we have to take into account that the derivation of the lower bound in (44) is based on the assumption that the pilot spacing is chosen such that the channel fading process is at least sampled with Nyquist rate, i.e., (7) has to be fulfilled. In case the pilot spacing L is chosen larger than the Nyquist rate, the estimation error process is no longer stationary, which is required for our derivation. At this point it is also important to remark that periodically inserted pilot symbols do not maximize the achievable rate. For the special case of PSK signaling, it is shown in [17] that the use of a single pilot symbol, i.e., not periodically inserted pilot symbols, is optimal in the sense that it maximizes the achievable rate. However, in the present work we restrict to the assumption of periodically inserted pilot symbols with a pilot spacing fulfilling (7), which is customary and reasonable as this enables detection and decoding with manageable complexity.

For these conditions, i.e., positive integer values for L fulfilling (7), it can be shown that the lower bound $\mathcal{R}_{L,\text{joint}}\Big|_{\text{rect.}S_h(f)}$ in (44) is maximized for

$$L_{\text{opt}} = \left\lfloor\frac{1}{2f_d}\right\rfloor. \tag{45}$$

---

[2]Note that a rectangular PSD $S_h(f)$ corresponds to $r_h(l) = \sigma_h^2\text{sinc}(2f_d l)$ which is not absolutely summable. However, the rectangular PSD can be arbitrarily closely approximated by a PSD with a raised cosine shape, whose corresponding correlation function is absolutely summable.

To prove this statement we differentiate the RHS of (44) with respect to L and set the result equal to zero, which yields that the RHS of (44) has a unique local extremum at

$$\tilde{L}_{\text{opt}} = \frac{1}{2f_d} \frac{C_{\text{perf}}(\rho)\rho}{\rho - C_{\text{perf}}(\rho)}. \tag{46}$$

Numerical evaluation shows that the factor $\frac{C_{\text{perf}}(\rho)\rho}{\rho - C_{\text{perf}}(\rho)}$ is larger than one. As (46) is the only local extremum of the RHS of (44), and with the constraints on L given by (7) and the fact that L is an integer value, and considering that $\mathcal{R}_{L,\text{joint}}\big|_{\text{rect.}S_h(f)}$ monotonically increases with L for $L < \tilde{L}_{\text{opt}}$ we can conclude that the lower bound is maximized by L_{opt} in (45).

Substituting L in (44) by L_{opt} in (45) yields a lower bound on the achievable rate with joint processing in case the pilot spacing can be arbitrarily chosen while fulfilling (7).

V. NUMERICAL EVALUATION

Fig. 1 shows a comparison of the bounds on the achievable rate for separate and joint processing of data and pilot symbols.

On the one hand, the lower bound on the achievable rate for joint processing in (44) is compared to bounds on the achievable rate with separate processing of data and pilot symbols, i.e, (26) and (27), for a fixed pilot spacing. As the upper and lower bound on the achievable rate with separate processing are relatively tight, we choose the pilot spacing such that the lower bound on the achievable rate for separate processing in (26) is maximized. It can be seen that except for very high channel dynamics, i.e., very large f_d the lower bound on the achievable rate for joint processing is larger than the bounds on the achievable rate with separate processing. This indicates the possible gain while using joint processing of data and pilot symbols for a given pilot spacing. Note, the observation that the lower bound for joint processing for large f_d is smaller than the bounds on the achievable rate with separate processing is a result of the lower bounding, i.e., it indicates that the lower bound is not tight for these parameters.

On the other hand, also the lower bound on the achievable rate with joint processing and a pilot spacing that maximizes this lower bound, i.e., (44) in combination with (45), is shown. In this case the pilot spacing is always chosen such that the channel fading process is sampled by the pilot symbols with Nyquist rate. Obviously, this lower bound is larger than or equal to the lower bound for joint processing while choosing the pilot spacing as it is optimal for separate processing of data and pilot symbols. This behavior arises from the effect that for separate processing in case of small f_d a pilot rate is chosen that is higher than the Nyquist rate of the channel fading process to enhance the channel estimation quality. In case of a joint processing all symbols are used for channel estimation anyway. Therefore, a pilot rate higher than Nyquist rate always leads to an increased loss in the achievable rate as less symbols can be used for data transmission.

Fig. 2 shows the lower bound on the achievable rate for joint processing of data and pilot symbols when choosing L as given in (45), which maximizes the lower bound in (44). This lower bound is compared to the following bounds on the achievable rate with i.i.d. zero-mean proper Gaussian (PG) input symbols for a rectangular PSD of the channel fading process, see (43), which have been given in [14]

$$\mathcal{I}'_L(\mathbf{y};\mathbf{x})\big|_{\text{PG}} = \max\left\{ C_{\text{perf}}(\rho) - 2f_d \log\left(1 + \frac{\rho}{2f_d}\right), 0 \right\} \tag{47}$$

$$\mathcal{I}'_U(\mathbf{y};\mathbf{x})\big|_{\text{PG}} = \min\left\{ \log(1+\rho) - 2f_d \int_{z=0}^{\infty} \log\left(1 + \frac{\rho}{2f_d}z\right) e^{-z} dz, C_{\text{perf}}(\rho) \right\}. \tag{48}$$

with $C_{\text{perf}}(\rho)$ being the coherent capacity of a Rayleigh flat-fading channel given in (42).

Obviously, for some parameters the lower bound on the achievable rate for joint processing of data and pilot symbols is larger than the lower bound on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols, i.e., without the assumption of any pilot symbols. However, this observation does not allow
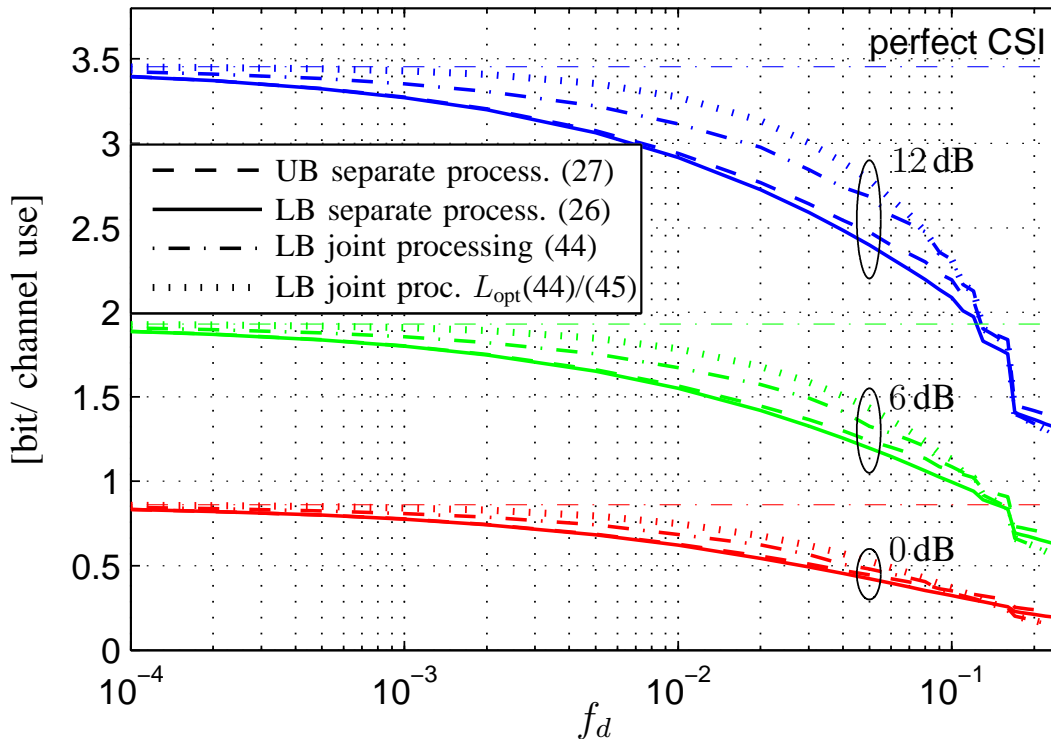
Fig. 1. Comparison of bounds on the achievable rate with separate processing of data and pilot symbols to lower bounds on the achievable rate with joint processing of data and pilot symbols; except of *LB joint proc. L_{opt}* the pilot spacing L is chosen such that the lower bound for separate processing (26) is maximized; the PSD $S_h(f)$ is assumed to be rectangular, see (43)

to argue that in these cases the use of pilot symbols is better than i.i.d. symbols, as we only compare lower bounds.

VI. SUMMARY

In the present work, we have studied the achievable rate with a joint processing of pilot and data symbols in the context of stationary Rayleigh flat-fading channels. We have discussed the nature of the possible gain when using joint processing of data and pilot symbols in contrast to separate processing. We have shown that the additional information that can be retrieved by joint processing is contained in the temporal correlation of the channel estimation error process when using a solely pilot based channel estimation, which cannot be captured by standard decoders as they are used in conventional synchronized detection based receivers with a solely pilot based channel estimation. In addition, and this is the main novelty of the present work, we have derived a lower bound on the achievable rate for joint processing of data and pilot symbols on a stationary Rayleigh flat-fading channel, giving an indication on the possible gain in terms of the achievable rate when using a joint processing of pilot and data symbols in comparison to the typically used separate processing.

APPENDIX A
MINIMIZATION OF $h'(\mathbf{e}_{\text{joint}}|\mathbf{x}_D, \mathbf{x}_P)$ BY CM MODULATION

In this appendix we will show that the differential entropy rate $h'(\mathbf{e}_{\text{joint}}|\mathbf{x}_D, \mathbf{x}_P)$ in (33), which depends on the distribution of the data symbols contained in $\mathbf{x}_D$, is minimized for constant modulus input symbols among all distributions of the data symbols with an maximum average power of σ_x^2.
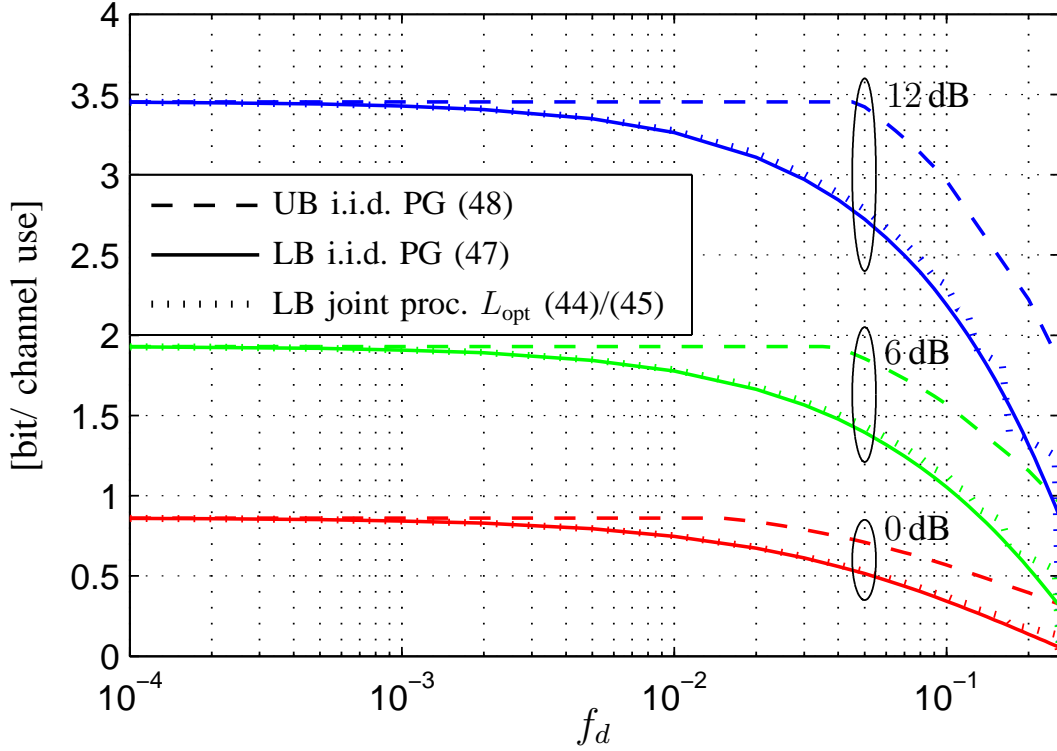
Fig. 2. Lower bound on the achievable rate with joint processing of data and pilot symbols and a pilot spacing L_{opt} that maximizes this bound, i.e., (44) in combination with (45); for comparison bounds on the achievable rate with i.i.d. zero-mean proper Gaussian (PG) input symbols are shown; rectangular PSD $S_h(f)$, see (43)

The MAP channel estimate based on pilot and perfectly known data symbols is given by

$$
\begin{aligned}
\hat{\mathbf{h}}_{\text{joint}} &= \arg\max_{\mathbf{h}} p(\mathbf{h}|\mathbf{y}, \mathbf{x}) \\
&= \arg\max_{\mathbf{h}} p(\mathbf{y}|\mathbf{h}, \mathbf{x}) p(\mathbf{h}) \\
&= \arg\max_{\mathbf{h}} \left\{ \log(p(\mathbf{y}|\mathbf{h}, \mathbf{x})) + \log(p(\mathbf{h})) \right\}
\end{aligned}
\tag{49}
$$

with

$$
p(\mathbf{y}|\mathbf{h}, \mathbf{x}) = \frac{1}{\pi^N \sigma_n^{2N}} \exp\left(-\frac{|\mathbf{y} - \mathbf{X}\mathbf{h}|^2}{\sigma_n^2} \right)
\tag{50}
$$

$$
p(\mathbf{h}) = \frac{1}{\pi^N \det(\mathbf{R}_h)} \exp\left(-\mathbf{h}^H \mathbf{R}_h^{-1} \mathbf{h} \right).
\tag{51}
$$

Thus, (49) becomes

$$
\hat{\mathbf{h}}_{\text{joint}} = \arg\max_{\mathbf{h}} \left\{ -\frac{1}{\sigma_n^2} |\mathbf{y} - \mathbf{X}\mathbf{h}|^2 - \mathbf{h}^H \mathbf{R}_h^{-1} \mathbf{h} \right\}.
\tag{52}
$$

Differentiating the argument of the maximum operation at the RHS of (52) with respect to $\mathbf{h}$ and setting the result equal to zero yields

$$
-\frac{1}{\sigma_n^2} \left\{ -\mathbf{X}^H \mathbf{y} + \mathbf{X}^H \mathbf{X} \mathbf{h} \right\} - \mathbf{R}_h^{-1} \mathbf{h} = \mathbf{0}
\tag{53}
$$

and, thus,[3]

$$\hat{\mathbf{h}}_{\text{joint}} = \mathbf{R}_h \left(\mathbf{R}_h + \sigma_n^2 \mathbf{X}^{-1} \mathbf{X}^{H^{-1}} \right)^{-1} \mathbf{X}^{-1} \mathbf{y}. \tag{55}$$

With (55) the channel estimation error correlation matrix $\mathbf{R}_{e_{\text{joint}}}$ is given by

$$\begin{aligned} \mathbf{R}_{e_{\text{joint}}} &= \mathrm{E}\left[\left(\mathbf{h} - \hat{\mathbf{h}}_{\text{joint}} \right) \left(\mathbf{h} - \hat{\mathbf{h}}_{\text{joint}} \right)^H \middle| \mathbf{x} \right] \\ &= \mathbf{R}_h - \mathbf{R}_h \left(\mathbf{R}_h + \sigma_n^2 (\mathbf{X}^H \mathbf{X})^{-1} \right)^{-1} \mathbf{R}_h. \end{aligned} \tag{56}$$

Thus, the differential entropy $h(\mathbf{e}_{\text{joint}}|\mathbf{x}_D, \mathbf{x}_P)$ becomes

$$\begin{aligned} h(\mathbf{e}_{\text{joint}}|\mathbf{x}_D, \mathbf{x}_P) &= \mathrm{E}_{\mathbf{x}}\left[\log \det \left(\pi e \mathbf{R}_{e_{\text{joint}}} \right) \right] \\ &= \log \left((\pi e)^N \det(\mathbf{R}_h) \right) + \mathrm{E}_{\mathbf{x}}\left[\log \det \left(\mathbf{I}_N - \left(\mathbf{R}_h + \sigma_n^2 (\mathbf{X}^H \mathbf{X})^{-1} \right)^{-1} \mathbf{R}_h \right) \right]. \end{aligned} \tag{57}$$

The argument of the expectation operation in the last summand on the RHS of (57) can be rewritten as

$$\begin{aligned} &\log \det \left(\mathbf{I}_N - \left(\mathbf{R}_h + \sigma_n^2 (\mathbf{X}^H \mathbf{X})^{-1} \right)^{-1} \mathbf{R}_h \right) \\ &= \log \det \left(\mathbf{I}_N - \left(\mathbf{I}_N + \mathbf{R}_h^{-1} \sigma_n^2 (\mathbf{X}^H \mathbf{X})^{-1} \right)^{-1} \right) \\ &\overset{(a)}{=} \log \det \left(\mathbf{I}_N - \left[\mathbf{I}_N - \left(\frac{1}{\sigma_n^2} \mathbf{R}_h + (\mathbf{X}^H \mathbf{X})^{-1} \right)^{-1} (\mathbf{X}^H \mathbf{X})^{-1} \right] \right) \\ &= -\log \det \left(\frac{1}{\sigma_n^2} \mathbf{R}_h \mathbf{X}^H \mathbf{X} + \mathbf{I}_N \right) \end{aligned} \tag{58}$$

where (a) follows from the matrix inversion lemma. Inserting (58) into (57) yields

$$h(\mathbf{e}_{\text{joint}}|\mathbf{x}_D, \mathbf{x}_P) = \log \left((\pi e)^N \det(\mathbf{R}_h) \right) - \mathrm{E}_{\mathbf{x}}\left[\log \det \left(\frac{1}{\sigma_n^2} \mathbf{R}_h \mathbf{X}^H \mathbf{X} + \mathbf{I}_N \right) \right]. \tag{59}$$

As the matrix $\mathbf{X} = \text{diag}(\mathbf{x})$ is diagonal, the product $\mathbf{X}\mathbf{X}^H$ is also diagonal and its diagonal elements are the powers of the individual transmit symbols. In the following we substitute this product by

$$\mathbf{Z} = \mathbf{X}\mathbf{X}^H \tag{60}$$

and $\mathbf{z} = \text{diag}(\mathbf{Z})$ contains the diagonal elements of $\mathbf{Z}$.

The aim of this appendix is to show that the entropy rate $h'(\mathbf{e}_{\text{joint}}|\mathbf{x}_D, \mathbf{x}_P)$ corresponding to the entropy in (59) is minimized by constant modulus data symbols with the power σ_x^2 among all input distributions fulfilling the maximum average power constraint in (6), i.e.,

$$\mathrm{E}\left[\mathbf{x}^H \mathbf{x} \right] = \mathrm{E}\left[\sum_{k=1}^{N} z_k \right] \le N \sigma_x^2 \tag{61}$$

where the z_k with $k = 1 \ldots N$ are the elements of $\mathbf{z}$. Therefor, in a first step, we study the entropy in (59), i.e., a finite transmission length N. Let the set $\mathcal{P}$ be the set containing all input distributions fulfilling the maximum average power constraint in (61). Note that this set $\mathcal{P}$ includes the case of having pilot symbols. However, when using pilot symbols, the transmit power of each L-th symbol is fixed to σ_x^2. For

---

[3]Note that the inverse of $\mathbf{X}$ in (55) does not exist, if a diagonal element of the diagonal matrix $\mathbf{X}$ is zero, i.e., one transmit symbol has zero power. However, as the channel estimates can be rewritten as

$$\hat{\mathbf{h}}_{\text{joint}} = \mathbf{R}_h \mathbf{X}^H \left(\mathbf{X} \mathbf{R}_h \mathbf{X}^H + \sigma_n^2 \mathbf{I}_N \right)^{-1} \mathbf{y} \tag{54}$$

it is obvious that the elements of $\hat{\mathbf{h}}_{\text{joint}}$ are continuous in x_k for all k, and, thus, this does not lead to problems in the following derivation.

the moment, we allow all input distributions contained in $\mathcal{P}$. Later on, we will come back to the special case of using pilot symbols.

We want to find the input vector $\mathbf{z}$ that minimizes (59) provided that the average power constraint is fulfilled. Therefor, we first show that the argument of the expectation operation on the RHS of (59), i.e.,

$$g(\mathbf{Z}) = \log \det \left(\frac{1}{\sigma_n^2} \mathbf{R}_h \mathbf{Z} + \mathbf{I}_N \right) \tag{62}$$

is concave in $\mathbf{Z}$. To verify the concavity of $g(\mathbf{Z})$, we follow along the lines of [18, Chapter 3.1.5] and consider an arbitrary line $\mathbf{Z} = \bar{\mathbf{Z}} + t\boldsymbol{\Delta}$. Based on this, we define $g(t)$ as

$$
\begin{aligned}
g(t) &= \log \det \left(\frac{1}{\sigma_n^2} \mathbf{R}_h \left(\bar{\mathbf{Z}} + t\boldsymbol{\Delta} \right) + \mathbf{I}_N \right) \\
&= \log \det \left(\frac{1}{\sigma_n^2} \mathbf{R}_h \right) + \log \det \left(\bar{\mathbf{Z}} + \sigma_n^2 \mathbf{R}_h^{-1} + t\boldsymbol{\Delta} \right) \\
&\overset{(a)}{=} \log \det \left(\frac{1}{\sigma_n^2} \mathbf{R}_h \right) + \log \det \left(\mathbf{Q} + t\boldsymbol{\Delta} \right) \\
&= \log \det \left(\frac{\mathbf{R}_h}{\sigma_n^2} \right) + \log \det \left(\mathbf{Q}^{\frac{H}{2}} \left(\mathbf{I}_N + t\mathbf{Q}^{-\frac{H}{2}} \boldsymbol{\Delta} \mathbf{Q}^{-\frac{1}{2}} \right) \mathbf{Q}^{\frac{1}{2}} \right) \\
&= \log \det \left(\frac{\mathbf{R}_h}{\sigma_n^2} \right) + \log \det \left(\mathbf{Q} \right) + \log \det \left(\mathbf{I}_N + t\mathbf{Q}^{-\frac{H}{2}} \boldsymbol{\Delta} \mathbf{Q}^{-\frac{1}{2}} \right) \\
&= \log \det \left(\frac{1}{\sigma_n^2} \mathbf{R}_h \bar{\mathbf{Z}} + \mathbf{I}_N \right) + \log \det \left(\mathbf{I}_N + t \left(\bar{\mathbf{Z}} + \sigma_n^2 \mathbf{R}_h^{-1} \right)^{-\frac{H}{2}} \boldsymbol{\Delta} \left(\bar{\mathbf{Z}} + \sigma_n^2 \mathbf{R}_h^{-1} \right)^{-\frac{1}{2}} \right) \\
&\overset{(b)}{=} \log \det \left(\frac{1}{\sigma_n^2} \mathbf{R}_h \bar{\mathbf{Z}} + \mathbf{I}_N \right) + \sum_{k=1}^{N} \log \left(1 + t\lambda_k \right)
\end{aligned}
\tag{63}
$$

where for (a) we have used the substitution $\mathbf{Q} \overset{\triangle}{=} \bar{\mathbf{Z}} + \sigma_n^2 \mathbf{R}_h^{-1}$ to simplify notation. Furthermore, the λ_k in (b) are the eigenvalues of $\left(\bar{\mathbf{Z}} + \sigma_n^2 \mathbf{R}_h^{-1} \right)^{-\frac{H}{2}} \boldsymbol{\Delta} \left(\bar{\mathbf{Z}} + \sigma_n^2 \mathbf{R}_h^{-1} \right)^{-\frac{1}{2}}$.

Based on (63) the derivatives of $g(t)$ with respect to t are given by

$$\frac{dg(t)}{dt} = \sum_{k=1}^{N} \frac{\lambda_k}{1 + t\lambda_k} \tag{64}$$

$$\frac{d^2 g(t)}{dt^2} = -\sum_{k=1}^{N} \frac{\lambda_k^2}{(1 + t\lambda_k)^2}. \tag{65}$$

As the second derivative $\frac{d^2 g(t)}{dt^2}$ is always negative, $g(\mathbf{Z})$ is concave on the set of diagonal matrices $\mathbf{Z}$ with non-negative diagonal entries.

Based on the concavity of $g(\mathbf{Z})$ with respect to $\mathbf{Z}$ we can lower-bound $h(\mathbf{e}_{joint}|\mathbf{x}_D, \mathbf{x}_P)$ in (59) by using Jensen's inequality as follows, cf. (62):

$$
\begin{aligned}
h(\mathbf{e}_{\text{joint}}|\mathbf{x}_D, \mathbf{x}_P) &= \log \det \left((\pi e)^N \det(\mathbf{R}_h) \right) - \mathrm{E}_{\mathbf{z}} \left[g(\mathbf{Z}) \right] \\
&\geq \log \det \left((\pi e)^N \det(\mathbf{R}_h) \right) - \log \det \left(\frac{1}{\sigma_n^2} \mathbf{R}_h \mathrm{E}\left[\mathbf{Z} \right] + \mathbf{I}_N \right).
\end{aligned}
\tag{66}
$$

Recall, that we want to show that constant modulus data symbols with the power σ_x^2 minimize the entropy rate $h'(\mathbf{e}_{\text{joint}}|\mathbf{x}_D, \mathbf{x}_P)$. Therefore, from here on we consider the entropy rate which is given by

$$
\begin{aligned}
h'(\mathbf{e}_{\text{joint}}|\mathbf{x}_D, \mathbf{x}_P) &= \lim_{N \to \infty} \frac{1}{N} h(\mathbf{e}_{\text{joint}}|\mathbf{x}_D, \mathbf{x}_P) \\
&= \lim_{N \to \infty} \frac{1}{N} \left[\log \det \left((\pi e)^N \det(\mathbf{R}_h) \right) - \log \det \left(\frac{1}{\sigma_n^2} \mathbf{R}_h \mathrm{E}\left[\mathbf{Z} \right] + \mathbf{I}_N \right) \right].
\end{aligned}
\tag{67}
$$

In the next step, we show for which kind of distribution of $\mathbf{z}$ fulfilling the maximum average power constraint in (61) the RHS of (67) is minimized. I.e., we have to find

$$\lim_{N \to \infty} \frac{1}{N} \sup_{\mathcal{P}} \log \det \left(\frac{1}{\sigma_n^2} \mathbf{R}_h \mathrm{E}\left[\mathbf{Z}\right] + \mathbf{I}_N \right) \tag{68}$$

where the set $\mathcal{P}$ contains all input distributions fulfilling the maximum average power constraint in (61).

For the evaluation of (68) we substitute the Toeplitz matrix $\mathbf{R}_h$ by an asymptotic equivalent circulant matrix $\mathbf{C}_h$, which is possible, as we are finally interested in the supremum in (68) for the case of an infinite transmission length, i.e., $N \to \infty$. In the following, we will formalize the construction of $\mathbf{C}_h$ and show that the following holds

$$\lim_{N \to \infty} \frac{1}{N} \sup_{\mathcal{P}} \log \det \left(\frac{1}{\sigma_n^2} \mathbf{R}_h \mathrm{E}\left[\mathbf{Z}\right] + \mathbf{I}_N \right) = \lim_{N \to \infty} \frac{1}{N} \sup_{\mathcal{P}} \log \det \left(\frac{1}{\sigma_n^2} \mathbf{C}_h \mathrm{E}\left[\mathbf{Z}\right] + \mathbf{I}_N \right) \tag{69}$$

Therefore, we express the channel correlation matrix $\mathbf{R}_h$ by its spectral decomposition

$$\mathbf{R}_h = \mathbf{R}_h^{(N)} = \mathbf{U}^{(N)} \mathbf{\Lambda}_h^{(N)} \left(\mathbf{U}^{(N)} \right)^H \tag{70}$$

where we introduced the superscript (N) to indicate the size of the matrices. Furthermore, the matrix $\mathbf{U}^{(N)}$ is unitary and $\mathbf{\Lambda}_h^{(N)} = \mathrm{diag}(\lambda_1^{(N)}, \ldots, \lambda_N^{(N)})$ is diagonal and contains the eigenvalues $\lambda_k^{(N)}$ of $\mathbf{R}_h^{(N)}$.

We construct the circulant matrix $\mathbf{C}_h^{(N)}$ which is asymptotically equivalent to the Toeplitz matrix $\mathbf{R}_h^{(N)}$ following along the lines of [16, Section 4.4, Eq. (4.32)]. The first column of the circulant matrix $\mathbf{C}_h^{(N)}$ is given by $(c_0^{(N)}, c_1^{(N)}, \ldots, c_{N-1}^{(N)})^T$ with the elements

$$c_k^{(N)} = \frac{1}{N} \sum_{l=0}^{N-1} \tilde{S}_h \left(\frac{l}{N} \right) e^{j2\pi \frac{lk}{N}}. \tag{71}$$

Here $\tilde{S}_h(f)$ is the periodic continuation of $S_h(f)$ given in (4), i.e.,

$$\tilde{S}_h(f) = \sum_{k=-\infty}^{\infty} \delta(f - k) \star S_h(f) \tag{72}$$

and $S_h(f)$ being zero outside the interval $|f| \leq 0.5$ for which it is defined in (4). The asterisk $\star$ in (72) denotes convolution.

As we assume that the autocorrelation function of the channel fading process is absolutely summable, see (3), the PSD of the channel fading process $\tilde{S}_h(f)$ is Riemann integrable, and it holds that

$$\begin{aligned}
\lim_{N \to \infty} c_k^{(N)} &= \lim_{N \to \infty} \frac{1}{N} \sum_{l=0}^{N-1} \tilde{S}_h \left(\frac{l}{N} \right) e^{j2\pi \frac{lk}{N}} \\
&= \int_0^1 \tilde{S}_h(f) e^{j2\pi k f} df \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}} S_h(f) e^{j2\pi k f} df = r_h(k)
\end{aligned} \tag{73}$$

with $r_h(k)$ defined in (2).

As the eigenvectors of a circulant matrix are given by a discrete Fourier transform (DFT), the eigenvalues $\breve{\lambda}_k^{(N)}$ with $k = 1, \ldots, N$ of the circulant matrix $\mathbf{C}_h^{(N)}$ are given by

$$
\begin{aligned}
\breve{\lambda}_k^{(N)} &= \sum_{l=0}^{N-1} c_l^{(N)} e^{-j2\pi \frac{(k-1)l}{N}} \\
&= \sum_{l=0}^{N-1} \left(\frac{1}{N} \sum_{m=0}^{N-1} \tilde{S}_h\left(\frac{m}{N}\right) e^{j2\pi \frac{ml}{N}} \right) e^{-j2\pi \frac{l(k-1)}{N}} \\
&= \sum_{m=0}^{N-1} \tilde{S}_h\left(\frac{m}{N}\right) \left\{ \frac{1}{N} \sum_{l=0}^{N-1} e^{j2\pi \frac{l(m-(k-1))}{N}} \right\} \\
&= \tilde{S}_h\left(\frac{k-1}{N}\right).
\end{aligned}
\tag{74}
$$

Consequently, the spectral decomposition of the circulant matrix $\mathbf{C}_h^{(N)}$ is given by

$$
\mathbf{C}_h^{(N)} = \mathbf{F}^{(N)} \breve{\mathbf{\Lambda}}_h^{(N)} \left(\mathbf{F}^{(N)} \right)^H
\tag{75}
$$

where the matrix $\mathbf{F}^{(N)}$ is a unitary DFT matrix, i.e., its elements are given by

$$
\left[\mathbf{F}^{(N)} \right]_{k,l} = \frac{1}{\sqrt{N}} e^{j2\pi \frac{(k-1)(l-1)}{N}}.
\tag{76}
$$

Furthermore, the matrix $\breve{\mathbf{\Lambda}}_h^{(N)}$ is diagonal with the elements $\breve{\lambda}_k^{(N)}$ given in (74).

By this construction the circulant matrix $\mathbf{C}_h^{(N)}$ is asymptotically equivalent to the Toeplitz matrix $\mathbf{R}_h^{(N)}$, see [16, Lemma 4.6], if the autocorrelation function $r_h(k)$ is absolutely summable, which is assumed to be fulfilled, see (3).

In the context of proving [16, Lemma 4.6], it is shown that the weak norm of the difference of $\mathbf{R}_h^{(N)}$ and $\mathbf{C}_h^{(N)}$ converges to zero as $N \to \infty$, i.e.,

$$
\lim_{N \to \infty} \left| \mathbf{R}_h^{(N)} - \mathbf{C}_h^{(N)} \right| = 0
\tag{77}
$$

where the weak norm of a matrix $\mathbf{B}$ is defined as

$$
|\mathbf{B}| = \left(\frac{1}{N} \mathrm{Tr} \left[\mathbf{B}^H \mathbf{B} \right] \right)^{\frac{1}{2}}.
\tag{78}
$$

This fact will be used later on.

To exploit the asymptotic equivalence of $\mathbf{R}_h^{(N)}$ and $\mathbf{C}_h^{(N)}$ for the current problem, we have to show that the matrices in the argument of the $\log \det$ operation on the LHS and the RHS of (69), i.e.,

$$
\mathbf{K}_1^{(N)} = \frac{1}{\sigma_n^2} \mathbf{R}_h^{(N)} \mathrm{E}\left[\mathbf{Z}\right] + \mathbf{I}_N
\tag{79}
$$

$$
\mathbf{K}_2^{(N)} = \frac{1}{\sigma_n^2} \mathbf{C}_h^{(N)} \mathrm{E}\left[\mathbf{Z}\right] + \mathbf{I}_N
\tag{80}
$$

are asymptotically equivalent.

In this context, we have to show that both matrices are bounded in the strong norm, and the weak norm of their difference converges to zero for $N \to \infty$ [16, Section 2.3].

Concerning the condition with respect to the strong norm we have to show that

$$
\left\| \mathbf{K}_1^{(N)} \right\| = \left\| \frac{1}{\sigma_n^2} \mathbf{R}_h^{(N)} \mathrm{E}\left[\mathbf{Z}\right] + \mathbf{I}_N \right\| < \infty
\tag{81}
$$

$$
\left\| \mathbf{K}_2^{(N)} \right\| = \left\| \frac{1}{\sigma_n^2} \mathbf{C}_h^{(N)} \mathrm{E}\left[\mathbf{Z}\right] + \mathbf{I}_N \right\| < \infty
\tag{82}
$$

with the strong norm of the matrix $\mathbf{B}$ defined by

$$\|\mathbf{B}\|^2 = \max_k \gamma_k \tag{83}$$

where γ_k are the eigenvalues of the Hermitian nonnegative definite matrix $\mathbf{BB}^H$. The diagonal matrix $\mathrm{E}[\mathbf{Z}]$ contains the average transmit powers of the individual transmit symbols on its diagonal. Thus, its entries are bounded. In addition, as the strong norms of $\mathbf{R}_h^{(N)}$ and $\mathbf{C}_h^{(N)}$ are bounded, too, the strong norms of $\mathbf{K}_1^{(N)}$ and $\mathbf{K}_2^{(N)}$ are bounded. Concerning the boundedness of the eigenvalues of the Hermitian Toeplitz matrix $\mathbf{R}_h^{(N)}$ see [16, Lemma 4.1].

Furthermore, the weak norm of the difference $\mathbf{K}_1^{(N)} - \mathbf{K}_2^{(N)}$ converges to zero for $N \to \infty$ as

$$\left| \mathbf{K}_1^{(N)} - \mathbf{K}_2^{(N)} \right| = \left| \frac{1}{\sigma_n^2} \mathbf{R}_h^{(N)} \mathrm{E}[\mathbf{Z}] + \mathbf{I}_N - \frac{1}{\sigma_n^2} \mathbf{C}_h^{(N)} \mathrm{E}[\mathbf{Z}] - \mathbf{I}_N \right|$$

$$= \left| \frac{1}{\sigma_n^2} \left(\mathbf{R}_h^{(N)} - \mathbf{C}_h^{(N)} \right) \mathrm{E}[\mathbf{Z}] \right|$$

$$\stackrel{(a)}{\leq} \frac{1}{\sigma_n^2} \left| \mathbf{R}_h^{(N)} - \mathbf{C}_h^{(N)} \right| \|\mathrm{E}[\mathbf{Z}]\| \tag{84}$$

where for (a) we have used [16, Lemma 2.3]. As $\|\mathrm{E}[\mathbf{Z}]\|$ is bounded, we get for $N \to \infty$

$$\lim_{N \to \infty} \left| \mathbf{K}_1^{(N)} - \mathbf{K}_2^{(N)} \right| \leq \lim_{N \to \infty} \frac{1}{\sigma_n^2} \left| \mathbf{R}_h^{(N)} - \mathbf{C}_h^{(N)} \right| \|\mathrm{E}[\mathbf{Z}]\| = 0 \tag{85}$$

due to (77). Thus we have shown the asymptotic equivalence of $\mathbf{K}_1^{(N)}$ and $\mathbf{K}_2^{(N)}$.

As $\mathbf{K}_1^{(N)}$ and $\mathbf{K}_2^{(N)}$ are asymptotically equivalent, with [16, Theorem 2.4] the equality in (69) holds. For ease of notation, in the following we omit the use of the superscript (N) for all matrices and eigenvalues.

Based on (69) the evaluation of the supremum in (68) can be substituted by

$$\lim_{N \to \infty} \frac{1}{N} \sup_{\mathcal{P}} \log \det \left(\frac{1}{\sigma_n^2} \mathbf{C}_h \mathrm{E}[\mathbf{Z}] + \mathbf{I}_N \right) \stackrel{(a)}{=} \lim_{N \to \infty} \frac{1}{N} \sup_{\mathcal{P}} \log \det \left(\frac{1}{\sigma_n^2} \mathbf{F} \breve{\mathbf{\Lambda}}_h \mathbf{F}^H \mathrm{E}[\mathbf{Z}] + \mathbf{I}_N \right)$$

$$\stackrel{(b)}{=} \lim_{N \to \infty} \frac{1}{N} \sup_{\mathcal{P}} \log \det \left(\frac{1}{\sigma_n^2} \breve{\mathbf{\Lambda}}_h \mathbf{F}^H \mathrm{E}[\mathbf{Z}] \mathbf{F} + \mathbf{I}_N \right) \tag{86}$$

where for (a) we have used (75) and (b) is based on the following relation

$$\det(\mathbf{AB} + \mathbf{I}) = \det(\mathbf{BA} + \mathbf{I}) \tag{87}$$

which holds as $\mathbf{AB}$ has the same eigenvalues as $\mathbf{BA}$ for $\mathbf{A}$ and $\mathbf{B}$ being square matrices [19, Theorem 1.3.20].

As the matrix $\frac{1}{\sigma_n^2} \breve{\mathbf{\Lambda}}_h \mathbf{F}^H \mathrm{E}[\mathbf{Z}] \mathbf{F} + \mathbf{I}_N$ in the argument of the logarithm on the RHS of (86) is positive definite, using Hadamard's inequality we can upper-bound the argument of the supremum on the RHS of (86) as follows

$$\log \det \left(\frac{1}{\sigma_n^2} \breve{\mathbf{\Lambda}}_h \mathbf{F}^H \mathrm{E}[\mathbf{Z}] \mathbf{F} + \mathbf{I}_N \right) \leq \sum_{k=1}^{N} \log \left(\frac{1}{\sigma_n^2} \breve{\lambda}_k \left[\mathbf{F}^H \mathrm{E}[\mathbf{Z}] \mathbf{F} \right]_{k,k} + 1 \right) \tag{88}$$

where $\left[\mathbf{F}^H \mathrm{E}[\mathbf{Z}] \mathbf{F} \right]_{k,k}$ are the diagonal entries of the matrix $\mathbf{F}^H \mathrm{E}[\mathbf{Z}] \mathbf{F}$. Note, this means that distributions of the input sequences $\mathbf{z}$ which lead to the case that the matrix $\mathbf{F}^H \mathrm{E}[\mathbf{Z}] \mathbf{F}$ is diagonal maximize the RHS

of (86). Using (88), the RHS of (86) is given by

$$\lim_{N\to\infty} \frac{1}{N} \sup_{\mathcal{P}} \log\det\left(\frac{1}{\sigma_n^2}\breve{\mathbf{\Lambda}}_h \mathbf{F}^H \mathrm{E}\left[\mathbf{Z}\right]\mathbf{F} + \mathbf{I}_N\right)$$

$$= \lim_{N\to\infty} \frac{1}{N} \sup_{\mathcal{P}} \sum_{k=1}^{N} \log\left(\frac{1}{\sigma_n^2}\breve{\lambda}_k\left(\frac{1}{N}\sum_{l=1}^{N}\mathrm{E}\left[z_l\right]\right) + 1\right)$$

$$= \lim_{N\to\infty} \frac{1}{N} \sup_{\mathcal{P}} \sum_{k=1}^{N} \log\left(\frac{1}{\sigma_n^2}\breve{\lambda}_k\left(\mathrm{E}\left[\frac{1}{N}\sum_{l=1}^{N}z_l\right]\right) + 1\right). \tag{89}$$

It rests to evaluate the supremum on the RHS of (89). However, as the logarithm is a monotonically increasing function with the maximum average power constraint in (61) the supremum in (89) is given by

$$\lim_{N\to\infty} \frac{1}{N} \sup_{\mathcal{P}} \sum_{k=1}^{N} \log\left(\frac{1}{\sigma_n^2}\breve{\lambda}_k\left(\mathrm{E}\left[\frac{1}{N}\sum_{l=1}^{N}z_l\right]\right) + 1\right) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \log\left(\frac{\sigma_x^2}{\sigma_n^2}\breve{\lambda}_k + 1\right)$$

$$\overset{(a)}{=} \lim_{N\to\infty} \frac{1}{N} \log\det\left(\frac{\sigma_x^2}{\sigma_n^2}\mathbf{C}_h + 1\right)$$

$$\overset{(b)}{=} \lim_{N\to\infty} \frac{1}{N} \log\det\left(\frac{\sigma_x^2}{\sigma_n^2}\mathbf{R}_h + 1\right) \tag{90}$$

where (a) is based on (75) and for (b) we have used the asymptotic equivalence of the circulant matrix $\mathbf{C}_h$ and the Toeplitz matrix $\mathbf{R}_h$.

Now, using (69), (86), (89), and (90) the supremum in (68) is given by

$$\lim_{N\to\infty} \frac{1}{N} \sup_{\mathcal{P}} \log\det\left(\frac{1}{\sigma_n^2}\mathbf{R}_h\mathrm{E}\left[\mathbf{Z}\right] + \mathbf{I}_N\right) = \lim_{N\to\infty} \frac{1}{N} \log\det\left(\frac{\sigma_x^2}{\sigma_n^2}\mathbf{R}_h + \mathbf{I}_N\right). \tag{91}$$

However, this means that the entropy rate $h'(\mathbf{e}_{\mathrm{joint}}|\mathbf{x}_D, \mathbf{x}_P)$ in (67) is lower-bounded by

$$h'(\mathbf{e}_{\mathrm{joint}}|\mathbf{x}_D, \mathbf{x}_P) = \lim_{N\to\infty} \frac{1}{N}\left[\log\det\left((\pi e)^N \det(\mathbf{R}_h)\right) - \log\det\left(\frac{1}{\sigma_n^2}\mathbf{R}_h\mathrm{E}\left[\mathbf{Z}\right] + \mathbf{I}_N\right)\right]$$

$$\geq \lim_{N\to\infty} \frac{1}{N}\left[\log\det\left((\pi e)^N \det(\mathbf{R}_h)\right) - \log\det\left(\frac{\sigma_x^2}{\sigma_n^2}\mathbf{R}_h + \mathbf{I}_N\right)\right]$$

$$\overset{(a)}{=} \lim_{N\to\infty} \frac{1}{N} \log\det\left(\pi e \mathbf{R}_{e_{\mathrm{joint,CM}}}\right) \tag{92}$$

where for (a) we have used (57) and (58), and where $\mathbf{R}_{e_{\mathrm{joint,CM}}}$ is the estimation error correlation matrix in case all input symbols have a constant modulus with power σ_x^2, cf. (56)

$$\mathbf{R}_{e_{\mathrm{joint,CM}}} = \mathbf{R}_h - \mathbf{R}_h\left(\mathbf{R}_h + \frac{\sigma_n^2}{\sigma_x^2}\mathbf{I}_N\right)^{-1}\mathbf{R}_h. \tag{93}$$

This mean, that the entropy rate $h'(\mathbf{e}_{\mathrm{joint}}|\mathbf{x}_D, \mathbf{x}_P)$ is minimized for the given maximum average power constraint in (6) when all input symbols are constant modulus input symbols with power σ_x^2. Note that this includes the case that each L-th symbol is a pilot symbol with power σ_x^2 and all other symbols are constant modulus data symbols with power σ_x^2.

In conclusion, we have shown that the differential entropy rate $h'(\mathbf{e}_{\mathrm{joint}}|\mathbf{x}_D, \mathbf{x}_P)$ is minimized for constant modulus data symbols with power σ_x^2, i.e.,

$$h'(\mathbf{e}_{\mathrm{joint}}|\mathbf{x}_D, \mathbf{x}_P) \geq h'(\mathbf{e}_{\mathrm{joint}}|\mathbf{x}_D, \mathbf{x}_P)\big|_{\mathrm{CM}}$$

$$= \lim_{N\to\infty} \frac{1}{N} \log\det\left(\pi e \mathbf{R}_{e_{\mathrm{joint,CM}}}\right). \tag{94}$$

APPENDIX B

ESTIMATION ERROR SPECTRA $S_{e_{\mathrm{pil}}}(f)$ AND $S_{e_{\mathrm{joint,CM}}}(f)$

First, we calculate the PSD $S_{e_{\mathrm{pil}}}(f)$ of the channel estimation error in case of a solely pilot based channel estimation. The channel estimation error in the frequency domain is given by

$$E_N(e^{j2\pi f}) = \sum_{k=1}^{N} e_{\mathrm{pil},k} \cdot e^{-j2\pi fk} \tag{95}$$

where $e_{\mathrm{pil},k}$ are the elements of the vector $\mathbf{e}_{\mathrm{pil}}$. In the following we are interested in the case $N \to \infty$. As in this case the sum in (95) does not exist, in the following we discuss $\lim_{N\to\infty} \frac{1}{N} E_N(e^{j2\pi f})$, which can be expressed as follows

$$
\begin{aligned}
\lim_{N\to\infty} \frac{1}{N} E_N(e^{j2\pi f}) &\overset{(a)}{=} \lim_{N\to\infty} \frac{1}{N} \sum_{l=1}^{L} E_{N,l}(e^{j2\pi Lf}) e^{-j2\pi lf} \\
&\overset{(b)}{=} \lim_{N\to\infty} \frac{1}{N} \sum_{l=1}^{L} \left[H_{N,l}(e^{j2\pi Lf}) - W_l(e^{j2\pi Lf}) \frac{Y_{N,P}(e^{j2\pi Lf})}{\sigma_x} \right] e^{-j2\pi lf} \\
&\overset{(c)}{=} \lim_{N\to\infty} \frac{1}{N} \left[H_N(e^{j2\pi f}) - \sum_{l=1}^{L} W(e^{j2\pi Lf}) e^{j2\pi lf} \frac{Y_{N,P}(e^{j2\pi Lf})}{\sigma_x} e^{-j2\pi lf} \right] \\
&= \lim_{N\to\infty} \frac{1}{N} \left[H_N(e^{j2\pi f}) - L \cdot W(e^{j2\pi Lf}) \frac{Y_{N,P}(e^{j2\pi Lf})}{\sigma_x} \right] \\
&\overset{(d)}{=} \lim_{N\to\infty} \frac{1}{N} \left[H_N(e^{j2\pi f}) - L \cdot W(e^{j2\pi Lf}) \left[H_{N,P}(e^{j2\pi Lf}) + \frac{N_{N,P}(e^{j2\pi Lf})}{\sigma_x} \right] \right]. \tag{96}
\end{aligned}
$$

For (a) we have used that the estimation error in frequency domain is the sum of the interpolation errors at the individual symbols time instances between the pilot symbols, where the temporal shift yields the phase shift of $2\pi lf$. Here $E_{N,l}(e^{j2\pi Lf})$ is the frequency transform of the estimation error at the symbol position with the distance l to the next pilot symbols, i.e.,

$$E_{N,l}(e^{j2\pi Lf}) = \sum_{k=1}^{\frac{N}{L}} e_{\mathrm{pil},(k-1)L+1+l} \cdot e^{-j2\pi fkL}, \quad \text{for } l = 0, \dots, L-1 \tag{97}$$

where without loss of generality we assume that N is an integer multiple of L and that the transmit sequence starts with a pilot symbol. Equality (b) results from expressing $E_{N,l}(e^{j2\pi Lf})$ by the difference between the actual channel realization and the estimated channel realization at the different interpolation positions in time domain transferred to frequency domain. Here, without loss of generality, we assume that the pilot symbols are given by σ_x. Furthermore, $W_l(e^{j2\pi Lf})$ is the transfer function of the interpolation filter for the symbols at distance l from the previous pilot symbol. Furthermore, $Y_{N,P}(e^{j2\pi Lf})$ is the channel output at the pilot symbols time instance transferred to frequency domain. For (c) we have used that the sum of the phase shifted channel realizations in frequency domain at sampling rate $1/L$ corresponds to the frequency domain representation of the fading process at symbol rate. In addition, we have used that for $N \to \infty$ the interpolation filter transfer function $W_l(e^{j2\pi Lf})$, which is an MMSE interpolation filter, can be expressed as

$$W_l(e^{j2\pi Lf}) = W(e^{j2\pi Lf}) e^{j2\pi lf} \tag{98}$$

i.e., the interpolation filter transfer functions for the individual time shifts are equal except of a phase shift. Finally, for (d) we have expressed $Y_{N,P}(e^{j2\pi Lf})$ as the sum of the frequency domain representations of the fading process and the additive noise process.

Based on (96) the PSD $S_{e_{\text{pil}}}(f)$ is given by

$$
\begin{aligned}
S_{e_{\text{pil}}}(f) &= \lim_{N \to \infty} \frac{1}{N} \mathrm{E}\left[|E_N(e^{j2\pi f})|^2\right] \\
&= \lim_{N \to \infty} \frac{1}{N} \mathrm{E}\Bigg[|H_N(e^{j2\pi f})|^2 - L \cdot H_N(e^{j2\pi f})W^*(e^{j2\pi Lf})H_{N,P}^*(e^{j2\pi Lf}) \\
&\qquad\qquad - L \cdot H_N^*(e^{j2\pi f})W(e^{j2\pi Lf})H_{N,P}(e^{j2\pi Lf}) \\
&\qquad\qquad + L^2|W(e^{j2\pi Lf})|^2\left[|H_{N,P}(e^{j2\pi Lf})|^2 + \left|\frac{N_{N,P}(e^{j2\pi Lf})}{\sigma_x}\right|^2\right]\Bigg] \\
&= S_h(e^{j2\pi f}) - \lim_{N \to \infty} \frac{1}{N} \mathrm{E}\Bigg[L \cdot W^*(e^{j2\pi Lf})\sum_{l=1}^{L}H_{N,l}(e^{j2\pi Lf})e^{-j2\pi lf}H_{N,P}^*(e^{j2\pi Lf}) \\
&\qquad\qquad + L \cdot W(e^{j2\pi Lf})\sum_{l=1}^{L}H_{N,l}^*(e^{j2\pi Lf})e^{j2\pi lf}H_{N,P}(e^{j2\pi Lf})\Bigg] \\
&\qquad + L^2|W(e^{j2\pi Lf})|^2\left[\frac{1}{L}S_h(e^{j2\pi Lf}) + \frac{1}{L}\frac{\sigma_n^2}{\sigma_x^2}\right] \\
&= S_h(e^{j2\pi f}) - L \cdot W^*(e^{j2\pi Lf})\sum_{l=1}^{L}\frac{1}{L}\cdot S_h(e^{j2\pi Lf}) - L \cdot W(e^{j2\pi Lf})\sum_{l=1}^{L}\frac{1}{L}\cdot S_h^*(e^{j2\pi Lf}) \\
&\qquad + L^2|W(e^{j2\pi Lf})|^2\left[\frac{1}{L}S_h(e^{j2\pi Lf}) + \frac{1}{L}\frac{\sigma_n^2}{\sigma_x^2}\right] \\
&\overset{(a)}{=} S_h(e^{j2\pi f}) - 2L \cdot W(e^{j2\pi Lf})S_h(e^{j2\pi Lf}) + L|W(e^{j2\pi Lf})|^2\left[S_h(e^{j2\pi Lf}) + \frac{\sigma_n^2}{\sigma_x^2}\right] \quad (99)
\end{aligned}
$$

where for (a) we have used that $S_h(f)$ is real and, thus, the MMSE filter $W(e^{j2\pi Lf})$ is also real, see below.

The MMSE filter transfer function $W(e^{j2\pi Lf})$ is given by

$$
W(e^{j2\pi Lf}) = \frac{S_h(e^{j2\pi Lf})}{S_h(e^{j2\pi Lf}) + \frac{\sigma_n^2}{\sigma_x^2}} = \frac{\frac{1}{L}S_h(e^{j2\pi f})}{\frac{1}{L}S_h(e^{j2\pi f}) + \frac{\sigma_n^2}{\sigma_x^2}} \quad (100)
$$

where we have used that

$$
S_h(e^{j2\pi Lf}) = \frac{1}{L}S_h(e^{j2\pi f}). \quad (101)
$$

Inserting (100) into (99) yields

$$
\begin{aligned}
S_{e_{\text{pil}}}(f) &= S_h(e^{j2\pi f}) - 2L\frac{S_h(e^{j2\pi f})}{S_h(e^{j2\pi f}) + L\frac{\sigma_n^2}{\sigma_x^2}}S_h(e^{j2\pi Lf}) + L\left|\frac{S_h(e^{j2\pi f})}{S_h(e^{j2\pi f}) + L\frac{\sigma_n^2}{\sigma_x^2}}\right|^2\left[S_h(e^{j2\pi Lf}) + \frac{\sigma_n^2}{\sigma_x^2}\right] \\
&\overset{(a)}{=} S_h(e^{j2\pi f}) - 2\cdot\frac{S_h(e^{j2\pi f})}{S_h(e^{j2\pi f}) + L\frac{\sigma_n^2}{\sigma_x^2}}S_h(e^{j2\pi f}) + \frac{\left|S_h(e^{j2\pi f})\right|^2}{S_h(e^{j2\pi f}) + L\frac{\sigma_n^2}{\sigma_x^2}} \\
&= \frac{S_h(e^{j2\pi f})L\frac{\sigma_n^2}{\sigma_x^2}}{S_h(e^{j2\pi f}) + L\frac{\sigma_n^2}{\sigma_x^2}}
\end{aligned}
$$

$$= \frac{S_h(e^{j2\pi f})}{\frac{\rho}{L}\frac{S_h(e^{j2\pi f})}{\sigma_h^2} + 1}$$

$$\overset{(b)}{=} \frac{S_h(f)}{\frac{\rho}{L}\frac{S_h(f)}{\sigma_h^2} + 1} \tag{102}$$

where (a) results from (101) and for (b) we simplified the notation and substituted $e^{j2\pi f}$ by f to get a consistent notation with (4).

The PSD $S_{e_{\text{joint,CM}}}(f)$ is then obviously given by setting $L = 1$ in (102), i.e.,

$$S_{e_{\text{joint,CM}}}(f) = \frac{S_h(f)}{\rho\frac{S_h(f)}{\sigma_h^2} + 1} \tag{103}$$

as all data symbols are assumed to be known and of constant modulus with power σ_x^2, cf. (34).

REFERENCES

[1] 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.

[2] N. Noels, V. Lottici, A. Dejonghe, H. Steendam, M. Moeneclaey, M. Luise, and L. Vandendorpe, "A theoretical framework for soft-information-based synchronization in iterative (Turbo) receivers," *EURASIP J. Wirel. Commun. Netw.*, vol. 2005, no. 2, pp. 117–129, 2005.

[3] S. Godtmann, N. Hadaschik, A. Pollok, G. Ascheid, and H. Meyr, "Iterative code-aided phase noise synchronization based on the LMMSE criterion," in *Proc. IEEE 8th Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Helsinki, Finland*, Jun. 2007.

[4] 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.

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

[6] 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.

[7] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[8] 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.

[9] ——, "Achievable rate of MIMO channels with data-aided channel estimation and perfect interleaving," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 12, pp. 2358–2368, Dec. 2001.

[10] S. Furrer and D. Dahlhaus, "Multiple-antenna signaling over fading channels with estimated channel state information: Performance analysis," *IEEE Trans. Inf. Theory*, vol. 53, no. 6, pp. 2010–2027, Jun. 2007.

[11] 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.

[12] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.

[13] N. Jindal, A. Lozano, and T. Marzetta, "What is the value of joint processing of pilots and data in block-fading channels?" in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Seoul, Korea, Jun. 2009.

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

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

[16] 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.

[17] M. Dörpinghaus, G. Ascheid, H. Meyr, and R. Mathar, "Optimal PSK signaling over stationary Rayleigh fading channels," in *Proc. International Symposium on Information Theory (ISIT)*, Toronto, Canada, Jul. 2008, pp. 126–130.

[18] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

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