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

Meik Dörpinghaus, *Member, IEEE,* Adrian Ispas, *Student Member, IEEE,*
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. Finally, the results are extended to multiple-input multiple-output channels.**

***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. This setting is sometimes referred to as *noncoherent* fading. 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. With these channel estimates, in a separate step coherent detection and decoding is performed. Both processing steps are executed separately.

This work has been supported by the UMIC (Ultra High Speed Mobile Information and Communication) research centre. The material in this paper was presented in part at the 2010 International Zurich Seminar on Communications, Zurich, Switzerland, March 2010.

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

A. Ispas and H. Meyr are with the Institute for Integrated Signal Processing Systems, RWTH Aachen University, 52056 Aachen, Germany (e-mail: {ispas, meyr}@iss.rwth-aachen.de). Currently, H. Meyr is a visiting professor at École Polytechnique Fédérale de Lausanne (EPFL), Switzerland.

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 of narrowband channels in a mobile environment without a line of sight component. Furthermore, we assume 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. In addition, 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. First, we perform this study for the case of a single-input single-output (SISO) channel, before extending it to multiple-input multiple-output (MIMO) channels.

There has been a variety of publications studying the achievable rate of noncoherent fading channels. On the one hand, there is a line of work aiming to characterize the capacity of such channels, i.e., without assuming any pilot symbols, see, e.g., [6]–[10]. As the evaluation of the capacity of stationary fading channels turns out to be notoriously difficult, many authors like [6] and [7] consider a block-fading model, where the channel stays constant for a fixed amount of symbol time instances and then changes to an independent new state. Both of these contributions consider the MIMO case. On the other hand, [8] and [9] consider the case of stationary fading while focusing on the high, respectively, low SNR regime. Furthermore, [10] bridges the gap between the block-fading model and stationary fading by considering block-stationarity.

In addition, there is a row of papers discussing the achievable rate with i.i.d. Gaussian input symbols, on the one hand for the block-fading channel [11], and on the other hand for a stationary Rayleigh flat-fading channel as considered in the present work [12]–[14].

The achievable rate with pilot symbols has also been broadly studied. Concerning the block-fading model the achievable rate while using pilot sequences is discussed in [15], [16], and [7] while all of them consider MIMO channels. On the other hand, in [17] and [18] the impact of imperfect channel state information due to channel estimation errors on the achievable rate has been discussed. Furthermore, in [19], [20], and [12] the achievable rate of time-continuous stationary fading channels while using pilot sequences for channel tracking and a coherent detection based on the acquired channel estimate has been treated. In [19] and [20] tight bounds on the achievable rate with synchronized detection with a solely pilot based channel estimation, i.e., separate processing, are given for the case of a stationary Rayleigh flat-fading channel, as considered in the present work. Here [19] discusses the SISO scenario while [20] is an extension to MIMO channels. Furthermore, [21] develops a unified treatment of block-fading and continuous fading for the case of separate processing. In addition, there are several papers discussing the optimum pilot overhead and placement, see, e.g., [21]–[23].

All of the above mentioned papers consider the case of separate processing of pilot and data symbols. In contrast, for the case of a joint processing there is not much knowledge on the achievable rate. In [24] joint detection of pilot and data symbols has been examined for a MIMO block-fading channel with regard to the decoding error probability. Regarding the achievable rate, very recently in [25] 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. Here we identify that coherent detection based on the solely pilot based channel estimation corresponds to a mismatched detection yielding an information loss, see also [18]. 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 [25] 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 [19] and to bounds on the achievable rate with i.i.d. zero-mean proper Gaussian input symbols given in [12], [13], i.e., without the assumption on pilot symbols inserted into the transmit sequence. Finally, we extend the study to the case of MIMO channels. In this context, we also extend the lower bound on the achievable rate with joint processing of pilot and data symbols to the MIMO case and compare it with bounds on the achievable rate for a separate processing in the case of a MIMO channel given in [20].

To summarize, our contributions are the following:

- We identify that in the separate processing case information is discarded due to ignoring the temporal correlation of the channel estimation error when performing coherent detection, which corresponds to mismatched decoding. This amount of information is the one that can be gained while using joint processing, which is optimum, instead of separate processing.
- We derive a new lower bound on the achievable rate with joint processing of pilot and data symbols for stationary Rayleigh flat-fading channels for the case that neither the transmitter nor the receiver knows the channel state while the channel law is known to the receiver. The channel fading process is assumed to be non-regular and it is sampled by the pilot symbols at least with Nyquist frequency.
- We extend the lower bound on the achievable rate with joint processing of pilot and data symbols to the case of MIMO channels.
- To study the gain while using a joint processing of pilot and data symbols in comparison to a separate processing, we numerically compare the new lower bound on the achievable rate with joint processing to bounds on the achievable rate with synchronized detection based on a solely pilot based channel estimation given in [19], [20].

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. In Section V, the study is extended to the MIMO case by introducing an extended system model, briefly recalling known bounds on the achievable rate with separate processing for the case of MIMO channels, and extending the lower bound on the achievable rate with joint processing of pilot and data symbols to the MIMO scenario. Afterwards, the bounds on the achievable rate with joint processing are 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 VI. Finally, Section VII concludes the paper with a brief summary.

II. System Model

First, we consider a discrete-time zero-mean jointly proper Gaussian SISO 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. Analogously, $\mathbf{x} = [x_1, \ldots, x_N]^T$, $\mathbf{n} = [n_1, \ldots, n_N]^T$, 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[1], i.e.,

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

The power spectral density (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 supported within the interval $[-f_d, f_d]$ with f_d being the maximum (normalized) 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.,

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

[1]Note that the assumption on an absolutely summable autocorrelation function in (3) is not very restrictive. In this regard, consider that an absolutely summable autocorrelation function corresponds to a smooth PSD in the sense that it is Riemann integrable [26]. Furthermore, PSDs of practically relevant channel fading processes can be tightly approximated by PSDs that are smooth in this sense. E.g., the often assumed rectangular PSD ($S_h(f) = \sigma_h^2/(2f_d)$ for $|f| \leq f_d$ and 0 otherwise), whose autocorrelation function $r_h(l) = \sigma_h^2\mathrm{sinc}(2f_dl)$ is not absolutely summable, can be arbitrarily closely approximated by a PSD with raised cosine shape, whose corresponding correlation function is absolutely summable [27, Ch. 9.2.1].

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

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

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

$$\begin{aligned}\mathrm{E}\left[\mathbf{y}_D|\mathbf{x}_D, \mathbf{y}_P, \mathbf{x}_P\right] &= \mathbf{X}_D\mathrm{E}\left[\mathbf{h}_D|\mathbf{y}_P, \mathbf{x}_P\right] \\ &= \mathbf{X}_D\hat{\mathbf{h}}_{\mathrm{pil},D} \qquad (11)\end{aligned}$$

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

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 a minimum mean square error (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 $\mathbf{x}_D, \hat{\mathbf{h}}_{\text{pil},D}, \mathbf{x}_P$, i.e.,

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

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

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

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

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

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

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

where $\mathbf{e}_{\text{pil}}$ 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.,

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

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

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

where $\mathcal{CN}(\boldsymbol{\mu}, \mathbf{C})$ denotes a proper Gaussian PDF with mean $\boldsymbol{\mu}$ and covariance $\mathbf{C}$.[2]

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

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

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

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

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

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

However, at this point it is important to remark that for practical channel decoders like the Viterbi or the BCJR-decoder the computational complexity per decoded symbol must be independent of the sequence length. This implies that the metric calculated within the decoder is a sum of metric increments for each symbol, i.e., the metric increments have to be statistically independent. This corresponds to $\mathbf{R}_{e_{\text{pil}}}$ in (20) being diagonal. Clearly, $\mathbf{R}_{e_{\text{pil}}}$ is in general non-diagonal due to the temporal correlation of the channel estimation error. By neglecting this correlation, practical decoders, which are using symbol-wise demapping, therefore perform mismatched decoding based on the assumption that the estimation error process is white, i.e., $p(\mathbf{y}|\mathbf{x}_D, \hat{\mathbf{h}}_{\text{pil}}, \mathbf{x}_P)$ is approximated by

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

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 $\sigma_{e_{\text{pil}}}^2$ is independent of the symbol time instant [19] and is given by

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

where $S_{e_{\text{pil}}}(f)$ is the PSD of the channel estimation error process in case the channel estimation is solely based on pilot symbols, which is given in (149) in Appendix C. Hence, the variance of the channel estimation process, i.e., the entries of $\hat{\mathbf{h}}_{\text{pil}}$, is given by $\sigma_h^2 - \sigma_{e_{\text{pil}}}^2$, which follows from the principle of orthogonality in linear minimum mean square error (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:

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

where $\mathcal{I}'$ denotes the mutual information rate and the index D_k refers to an arbitrarily chosen data symbol time instant, i.e., $x_{D_k} = [\mathbf{x}_D]_k$. Furthermore, $\hat{h}_{\text{pil},D_k}$ 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 $\mathcal{R}_{\text{sep}}$.

Evidently, the difference between the mutual information rate achievable using the mismatched PDF in (23) discarding the temporal correlation, i.e., $\mathcal{R}_{\text{sep}}$, and the mutual information rate of the channel, i.e., the LHS of (25), corresponding to the achievable rate with joint processing of data and pilot symbols using the true PDF in (20) is the possible gain which can be realized by employing joint processing in comparison to separate processing. 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.

Note that the loss of mutual information in (25) is not caused by using the solely pilot based channel estimation, but due to the approximation of the decoding in (16) by ignoring the temporal correlation of the channel estimation

[2]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 $\hat{\mathbf{h}}_{\text{pil}} = \mathbf{0}$ and $\mathbf{R}_{e_{\text{pil}}} = \mathbf{R}_h$.

error when using (23), i.e., by assuming $\mathbf{R}_{e_{\text{pil}}}$ to be diagonal. If one were to use the true $\mathbf{R}_{e_{\text{pil}}}$ as in (20), then one would be able to retrieve the complete mutual information between the transmitter and the receiver. However, this would require a maximum likelihood sequence detection by exhaustive search — just as in the case of directly evaluating (10) based on $p(\mathbf{y}|\mathbf{x})$, which is not based on any channel estimation at all — which is computationally infeasible as the complexity increases exponentially with the sequence length. Furthermore, note that (10) as well as (16) correspond to joint processing of pilot and data symbols, which is optimal. We finally remark that joint processing is approximately performed in receivers based on iterative joint channel estimation and decoding. It can be shown that such an iterative receiver solves a set of fixed point equations defined by the optimization problem given by (16) and (20). However, in the derivation of this set of fixed point equations one makes the approximation that the symbols after deinterleaving and, thus, at the input to the demapper/decoder are statistically independent. This has the effect that the correlation matrix of the channel estimation error is assumed to be diagonal. Note that in this case the error correlation matrix is different to $\mathbf{R}_{e_{\text{pil}}}$ in (19) as the channel estimator uses reliability information on the data symbols in addition to the pilot symbols. For a more detailed discussion on iterative joint channel estimation and decoding, we refer the reader to [4]. However, as the achievable rate with any iterative joint channel estimation and decoding scheme is upper-bounded by the achievable rate with a joint processing of pilot and data symbols, which is the optimum, we study the achievable rate with joint processing in the following section.

Regarding synchronized detection in combination with a solely pilot based channel estimation, i.e., the separate processing approach, in [19] bounds on the achievable rate have been given, which for zero-mean proper Gaussian data symbols become

$$\begin{aligned}\mathcal{R}_{\text{sep}} &\geq \mathcal{R}_{L,\text{sep}} = \frac{L-1}{L}\mathrm{E}_{\hat{h}_{\text{pil},D_k}}\left[\log\left(1+\frac{\sigma_x^2|\hat{h}_{\text{pil},D_k}|^2}{\sigma_{e_{\text{pil}}}^2\sigma_x^2+\sigma_n^2}\right)\right] \\ &= \frac{L-1}{L}\int_0^\infty \log\left(1+\rho\frac{1-\frac{\sigma_{e_{\text{pil}}}^2}{\sigma_h^2}}{1+\rho\frac{\sigma_{e_{\text{pil}}}^2}{\sigma_h^2}}z\right)e^{-z}dz \qquad (26)\\ \mathcal{R}_{\text{sep}} &\leq \mathcal{R}_{U,\text{sep}} \\ &= \mathcal{R}_{L,\text{sep}} + \frac{L-1}{L}\mathrm{E}_{x_{D_k}}\left[\log\left(\frac{\sigma_x^2\sigma_{e_{\text{pil}}}^2+\sigma_n^2}{|x_{D_k}|^2\sigma_{e_{\text{pil}}}^2+\sigma_n^2}\right)\right] \\ &= \mathcal{R}_{L,\text{sep}} + \frac{L-1}{L}\left(\log\left(1+\rho\frac{\sigma_{e_{\text{pil}}}^2}{\sigma_h^2}\right)\right. \\ &\qquad \left. - \int_0^\infty \log\left(1+\rho\frac{\sigma_{e_{\text{pil}}}^2}{\sigma_h^2}z\right)e^{-z}dz\right). \qquad (27)\end{aligned}$$

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 $\left(1-\sigma_{e_{\text{pil}}}^2/\sigma_h^2\right)/\left(1+\rho\sigma_{e_{\text{pil}}}^2/\sigma_h^2\right)$ 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_{pil,D_k} is the estimation error at the data symbol time instant D_k, i.e., $e_{\text{pil},D_k} = [\mathbf{e}_{\text{pil},D}]_k$.

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 [25] for the case of a block-fading channel to the stationary Rayleigh flat-fading scenario discussed in the present work. Therefore, analogously to [25] 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

$$\begin{aligned}&\mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{y}_P,\mathbf{x}_P) \\ &\overset{(a)}{=} \mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{y}_P,\mathbf{x}_P,\mathbf{h}) - \mathcal{I}(\mathbf{x}_D;\mathbf{h}|\mathbf{y}_D,\mathbf{y}_P,\mathbf{x}_P) \\ &\overset{(b)}{=} \mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{h}) - h(\mathbf{h}|\mathbf{y}_D,\mathbf{y}_P,\mathbf{x}_P) + h(\mathbf{h}|\mathbf{x}_D,\mathbf{y}_D,\mathbf{y}_P,\mathbf{x}_P) \\ &\overset{(c)}{\geq} \mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{h}) - h(\mathbf{h}|\mathbf{y}_P,\mathbf{x}_P) + h(\mathbf{h}|\mathbf{x}_D,\mathbf{y}_D,\mathbf{y}_P,\mathbf{x}_P) \qquad (28)\end{aligned}$$

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 [25]. Now, we calculate both differential entropy terms at the RHS of (28). Therefore, we rewrite the RHS of (28) as follows

$$\begin{aligned}&\mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{y}_P,\mathbf{x}_P) \\ &\geq \mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{h}) - h(\mathbf{h}|\mathbf{y}_P,\mathbf{x}_P) + h(\mathbf{h}|\mathbf{x}_D,\mathbf{y}_D,\mathbf{y}_P,\mathbf{x}_P) \\ &\overset{(a)}{=} \mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{h}) - h(\mathbf{h}|\hat{\mathbf{h}}_{\text{pil}},\mathbf{x}_P) + h(\mathbf{h}|\hat{\mathbf{h}}_{\text{joint}},\mathbf{x}_D,\mathbf{x}_P) \\ &\overset{(b)}{=} \mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{h}) - h(\hat{\mathbf{h}}_{\text{pil}}+\mathbf{e}_{\text{pil}}|\hat{\mathbf{h}}_{\text{pil}},\mathbf{x}_P) \\ &\quad + h(\hat{\mathbf{h}}_{\text{joint}}+\mathbf{e}_{\text{joint}}|\hat{\mathbf{h}}_{\text{joint}},\mathbf{x}_D,\mathbf{x}_P) \\ &\overset{(c)}{=} \mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{h}) - h(\mathbf{e}_{\text{pil}}|\mathbf{x}_P) + h(\mathbf{e}_{\text{joint}}|\mathbf{x}_D,\mathbf{x}_P) \\ &\overset{(d)}{=} \mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{h}) - \mathrm{E}_{\mathbf{x}_P}\left[\log\det\left(\pi e\mathbf{R}_{e_{\text{pil}}}\right)\right] \\ &\quad + \mathrm{E}_{\mathbf{x}_P,\mathbf{x}_D}\left[\log\det\left(\pi e\mathbf{R}_{e_{\text{joint}}}\right)\right] \\ &\overset{(e)}{=} \mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{h}) - \log\det\left(\mathbf{R}_{e_{\text{pil}}}\right) + \mathrm{E}_{\mathbf{x}_D}\left[\log\det\left(\mathbf{R}_{e_{\text{joint}}}\right)\right] \qquad (29)\end{aligned}$$

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.,

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

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 have 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.,[3]

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

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

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

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 $\mathbf{x}_D$, see (139) in Appendix A, 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). Equality (a) in (34) is based on Szegö's theorem on the asymptotic eigenvalue distribution of Hermitian Toeplitz matrices [28], see Appendix B for a detailed derivation. $S_{e_{\text{pil}}}(f)$ 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 C.

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

$$\mathcal{R}_{L,\text{joint}} = \frac{L-1}{L} C_{\text{perf}}(\rho) - \int_{-\frac{1}{2}}^{\frac{1}{2}} \log\left(\frac{\frac{\rho}{\sigma_h^2} S_h(f) + 1}{\frac{\rho}{L\sigma_h^2} S_h(f) + 1}\right) df \tag{37}$$

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

$$\begin{aligned}
C_{\text{perf}}(\rho) &= \mathrm{E}_{h_k}\left[\log\left(1 + \rho\frac{|h_k|^2}{\sigma_h^2}\right)\right] \\
&= \int_0^\infty \log\left(1+\rho z\right) e^{-z} dz
\end{aligned} \tag{38}$$

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

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

Substituting (38) into (37) 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[4] of the channel fading process, i.e.,

$$S_h(f) = \begin{cases} \frac{\sigma_h^2}{2f_d} & \text{for } |f| \leq f_d \\ 0 & \text{otherwise} \end{cases} \tag{39}$$

the lower bound in (37) becomes

$$\begin{aligned}
&\mathcal{R}_{L,\text{joint}}\big|_{\text{rect.}S_h(f)} \\
&\quad = \frac{L-1}{L} \int_0^\infty \log\left(1+\rho z\right) e^{-z} dz - 2f_d \log\left(\frac{\frac{\rho}{2f_d} + 1}{\frac{\rho}{L2f_d} + 1}\right).
\end{aligned} \tag{40}$$

[3]Note that the differential entropy rate $h'(\mathbf{e}_{\text{joint}}|\mathbf{x}_D, \mathbf{x}_P)$ does not exist, i.e., the limit when calculating this entropy diverges to $-\infty$. This is a result of the fact that $S_h(f)$ is bandlimited. However, finally only the difference $h'(\mathbf{e}_{\text{pil}}|\mathbf{x}_P) - h'(\mathbf{e}_{\text{joint}}|\mathbf{x}_D, \mathbf{x}_P)$, which exists, see (141), is used. Thus, the non-existence of $h'(\mathbf{e}_{\text{joint}}|\mathbf{x}_D, \mathbf{x}_P)$ and $h'(\mathbf{e}_{\text{pil}}|\mathbf{x}_P)$ does not pose any problems and could also be avoided while deriving the lower bound on the achievable rate with joint processing given in (37). Nevertheless, we choose to use the separation into $h'(\mathbf{e}_{\text{joint}}|\mathbf{x}_D, \mathbf{x}_P)$ and $h'(\mathbf{e}_{\text{pil}}|\mathbf{x}_P)$ as these terms give valuable physical insight by representing the randomness of the estimation error processes and, thus, reflecting the estimation quality. The same holds for the differential entropy terms of the estimation error processes in the derivation of the lower bound on the achievable rate for the MIMO case in Section V.

[4]Note that a rectangular PSD $S_h(f)$ corresponds to $r_h(l) = \sigma_h^2 \mathrm{sinc}(2f_d l)$, which is not absolutely summable. However, recall that the rectangular PSD can be arbitrarily closely approximated by a PSD with a raised cosine shape, whose corresponding correlation function is absolutely summable [27, Ch. 9.2.1].

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

Obviously, the lower bound in (40) 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 (40) 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 (40) 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 in general. For the special case of PSK signaling, it is shown in [29] 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 (40) is maximized for

$$L_{\text{opt}} = \left\lfloor \frac{1}{2f_d} \right\rfloor. \tag{41}$$

To prove this statement we differentiate the RHS of (40) with respect to L and set the result equal to zero, which yields that the RHS of (40) has a unique local extremum at

$$\tilde{L}_{\text{opt}} = \frac{1}{2f_d} \frac{C_{\text{perf}}(\rho)\rho}{\rho - C_{\text{perf}}(\rho)}. \tag{42}$$

Numerical evaluation shows that the factor $\frac{C_{\text{perf}}(\rho)\rho}{\rho - C_{\text{perf}}(\rho)}$ is larger than one. As (42) is the only local extremum of the RHS of (40), 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 (41).

Substituting L in (40) by L_{opt} in (41) yields a lower bound on the achievable rate with joint processing in case the pilot spacing can be arbitrarily chosen while fulfilling (7).

V. Extension to MIMO Channels

Now, we extend the results given previously for SISO channels to MIMO channels.

A. System Model

For a discrete-time zero-mean proper Gaussian MIMO flat-fading channel with n_T transmit and n_R receive antennas the input-output relation at the time instant k is given by

$$\begin{aligned}\tilde{\mathbf{y}}(k) &= \tilde{\mathbf{H}}(k)\tilde{\mathbf{x}}(k) + \tilde{\mathbf{n}}(k) \\ &= \left(\mathbf{I}_{n_R} \otimes (\tilde{\mathbf{x}}(k))^T\right) \tilde{\mathbf{h}}(k) + \tilde{\mathbf{n}}(k)\end{aligned} \tag{43}$$

where $\tilde{\mathbf{x}}(k) = [x_1(k), \ldots, x_{n_T}(k)]^T$ is the channel input vector at the time instant k containing the input symbols at the individual transmit antennas. In addition, the elements of the fading matrix $[\tilde{\mathbf{H}}(k)]_{ij} = h_{ij}(k)$ with $1 \leq i \leq n_R$ and $1 \leq j \leq n_T$ are the fading weights of the channel from transmit antenna j to receive antenna i. $\tilde{\mathbf{y}}(k)$ contains the received signals at the individual receive antennas at the time instant k. Finally, $\tilde{\mathbf{n}}(k)$ is the additive white Gaussian noise vector. In the second representation $\otimes$ denotes the Kronecker product, $\mathbf{I}_{n_R}$ is an identity matrix of size $n_R \times n_R$, and

$$\begin{aligned}\tilde{\mathbf{h}}(k) = &[h_{11}(k), \ldots, h_{1n_T}(k), \ldots, h_{n_R1}(k), \ldots, h_{n_Rn_T}(k)]^T \\ &\in \mathbb{C}^{n_T \cdot n_R \times 1}.\end{aligned} \tag{44}$$

As in the SISO case in (1), we use the following concatenation of N individual time instances

$$\mathbf{y} = \mathbf{X}\mathbf{h} + \mathbf{n} \tag{45}$$

where $\mathbf{h}$ is now given by

$$\mathbf{h} = \left[\bar{\mathbf{h}}_{11}^T, \ldots, \bar{\mathbf{h}}_{1n_T}^T, \ldots, \bar{\mathbf{h}}_{n_R1}^T, \ldots, \bar{\mathbf{h}}_{n_Rn_T}^T\right]^T \in \mathbb{C}^{N \cdot n_T \cdot n_R \times 1} \tag{46}$$

with $\bar{\mathbf{h}}_{ij} = [h_{ij}(1), \ldots, h_{ij}(N)]^T$, $1 \leq i \leq n_R, 1 \leq j \leq n_T$, i.e., vectors containing the fading weights of the individual subchannels in temporal order.

The matrix $\mathbf{X}$ contains the transmit symbols and is given by

$$\mathbf{X} = \mathbf{I}_{n_R} \otimes \left[\bar{\mathbf{X}}_1 \ldots \bar{\mathbf{X}}_{n_T}\right] \tag{47}$$

with the diagonal matrices $\bar{\mathbf{X}}_j = \text{diag}(\bar{\mathbf{x}}_j)$. The vectors $\bar{\mathbf{x}}_j$ contain the symbols transmitted from antenna j in temporal order, i.e., $\bar{\mathbf{x}}_j = [x_j(1), \ldots, x_j(N)]^T$, $1 \leq j \leq n_T$. Finally the output is given by

$$\mathbf{y} = \left[\bar{\mathbf{y}}_1^T, \ldots, \bar{\mathbf{y}}_{n_R}^T\right]^T \in \mathbb{C}^{N \cdot n_R \times 1} \tag{48}$$

with $\bar{\mathbf{y}}_i = [y_i(1), \ldots, y_i(N)]^T, 1 \leq i \leq n_R$. Analogously, we define $\mathbf{n}$, $\bar{\mathbf{n}}_i, 1 \leq i \leq n_R$, and $\mathbf{x} = \left[\bar{\mathbf{x}}_1^T, \ldots, \bar{\mathbf{x}}_{n_T}^T\right]^T$.

B. Stochastic Characterization

We assume a spatially uncorrelated MIMO channel, yielding the channel correlation matrix

$$\mathbf{R}_{\mathbf{h}} = \text{E}\left[\mathbf{h}\mathbf{h}^H\right] = \mathbf{I}_{n_R} \otimes \mathbf{I}_{n_T} \otimes \mathbf{R}_{\bar{\mathbf{h}}} \tag{49}$$

with the temporal correlation matrix of each subchannel

$$\mathbf{R}_{\bar{\mathbf{h}}} = \text{E}\left[\bar{\mathbf{h}}_{ij}\bar{\mathbf{h}}_{ij}^H\right], \quad 1 \leq i \leq n_R, 1 \leq j \leq n_T. \tag{50}$$

The individual subchannels have the same stochastic properties as in the SISO case, i.e., they are zero-mean jointly proper Gaussian and their temporal correlation is given by the autocorrelation function $r_h(l)$ in (2) or the PSD $S_h(f)$ in (4). The fading processes on all subchannels are characterized by the same PSD $S_h(f)$.

The noise $\mathbf{n}$ is zero-mean proper Gaussian and i.i.d. in the temporal and the spatial domain such that $\mathbf{R}_{\mathbf{n}} = \text{E}[\mathbf{n}\mathbf{n}^H] = \sigma_n^2 \mathbf{I}_{Nn_R}$.

The transmit symbol sequence $\{\tilde{\mathbf{x}}(k)\}$ consists of data symbols and periodically inserted pilot symbols. The average power of the data symbols is limited to P, i.e.,

$$\mathrm{E}[(\tilde{\mathbf{x}}_D(k))^H \tilde{\mathbf{x}}_D(k)] \leq P \tag{51}$$

where the index D indicates a data symbol. To estimate the individual subchannels based on the pilot symbols, we assume that the pilot sequences transmitted from the individual transmit antennas are mutually orthogonal. This requires pilot sequences of length n_T at minimum, where we choose the pilot sequence length to be n_T. These pilot sequences are periodically introduced into the transmit symbol sequence with a period length of L. This pilot spacing L is chosen such that the fading processes are sampled by the pilot symbols at least with Nyquist rate, i.e.,

$$n_T \leq L < 1/(2f_d). \tag{52}$$

Without loss of generality, we assume that the transmission length N is an integer multiple of the pilot spacing L. Due to the orthogonality of the pilot sequences transmitted from different antennas it holds that

$$(\bar{\mathbf{x}}_{i,P})^H \bar{\mathbf{x}}_{j,P} = 0 \quad \text{for } i \neq j \tag{53}$$

where $\bar{\mathbf{x}}_{i,P}$ corresponds to $\bar{\mathbf{x}}_i$ but contains only the time instances used for pilot symbols. Analogous to the SISO case, we assume that the pilot symbols have constant power:

$$(\tilde{\mathbf{x}}_P(k))^H \tilde{\mathbf{x}}_P(k) = P \tag{54}$$

where the index P indicates a pilot symbol.

Like in the SISO case we use $\mathbf{x}_P$, $\mathbf{x}_D$, $\mathbf{y}_P$, $\mathbf{y}_D$, $\mathbf{h}_P$, $\mathbf{h}_D$, $\mathbf{n}_P$, and $\mathbf{n}_D$ as subvectors of $\mathbf{x}$, $\mathbf{y}$, $\mathbf{h}$, and $\mathbf{n}$ containing all time slots used to transmit pilot symbols and data symbols respectively.

All fading processes $\{h_{ij}(k)\}$, the additive noise processes $\{n_i(k)\}$ and the transmit symbol sequences $\{x_j(k)\}$ are mutually independent. Finally, the average SNR is now given by

$$\rho = \frac{P\sigma_h^2}{\sigma_n^2}. \tag{55}$$

C. Separate Processing of Pilot and Data Symbols

Analogous to the SISO case discussed in Section III, it can be shown that the mutual information between the transmitter and the receiver can be expressed as, cf. (9) to (22)

$$\mathcal{I}(\mathbf{x}_D; \mathbf{y}_D, \mathbf{y}_P, \mathbf{x}_P) = \mathcal{I}(\mathbf{x}_D; \mathbf{y}_D | \hat{\mathbf{h}}_{\text{pil}}) \tag{56}$$

with the solely pilot based channel estimate

$$\hat{\mathbf{h}}_{\text{pil}} = \mathrm{E}\left[\mathbf{h} | \mathbf{y}_P, \mathbf{x}_P\right] \tag{57}$$

and the input-output relation at the data symbol time instances given by

$$\mathbf{y}_D = \mathbf{X}_D \mathbf{h}_D + \mathbf{n}_D = \mathbf{X}_D \left(\hat{\mathbf{h}}_{\text{pil},D} + \mathbf{e}_{\text{pil},D}\right) + \mathbf{n}_D \tag{58}$$

with $\hat{\mathbf{h}}_{\text{pil},D} = \mathrm{E}\left[\mathbf{h}_D | \mathbf{y}_P, \mathbf{x}_P\right]$ and the estimation error $\mathbf{e}_{\text{pil},D} = \mathbf{h}_D - \hat{\mathbf{h}}_{\text{pil},D}$.

As already discussed in the context of the SISO case the temporal correlation of the channel estimation error $\mathbf{e}_{\text{pil},D}$ cannot be exploited by a coherent detection/demapping and, thus, the corresponding information is discarded when using synchronized detection, i.e., separate processing, yielding the following inequality:

$$\begin{aligned}\lim_{N\to\infty} \frac{1}{N} \mathcal{I}(\mathbf{x}_D; \mathbf{y}_D | \hat{\mathbf{h}}_{\text{pil}}) &= \mathcal{I}'(\mathbf{x}_D; \mathbf{y}_D | \hat{\mathbf{h}}_{\text{pil}}) \\ &\geq \frac{L-n_T}{L} \mathcal{I}\left(\tilde{\mathbf{x}}_D(k); \tilde{\mathbf{y}}_D(k) \middle| \hat{\tilde{\mathbf{h}}}_{\text{pil},D}(k)\right) = \mathcal{R}_{\text{sep}}\end{aligned} \tag{59}$$

where $\hat{\tilde{\mathbf{h}}}_{\text{pil},D}(k)$ contains the elements of $\hat{\mathbf{h}}_{\text{pil}}$ at the data symbol time instant k. The achievable rate with separate processing, i.e., the RHS of (59), is the mutual information at an arbitrarily chosen data symbol time instant multiplied by the factor $(L-n_T)/L$, as in each pilot interval of length L, n_T symbols are used for pilot symbols and, thus, are not used for data transmission. In [20] bounds on the achievable rate with separate processing have been derived, which for zero-mean proper Gaussian data symbols that are i.i.d. in time and space are given by

$$\begin{aligned}\mathcal{R}_{\text{sep}} &\geq \mathcal{R}_{L,\text{sep}} \\ &= \frac{L-n_T}{L} \mathrm{E}_{\tilde{\mathbf{H}}(k)} \log\det\left(\mathbf{I}_{n_R} + \frac{\rho}{n_T} \frac{1-\frac{\sigma^2_{e_{\text{pil}}}}{\sigma_h^2}}{1+\rho\frac{\sigma^2_{e_{\text{pil}}}}{\sigma_h^2}} \frac{\tilde{\mathbf{H}}(k)(\tilde{\mathbf{H}}(k))^H}{\sigma_h^2}\right) \\ &= \frac{L-n_T}{L} \int_0^\infty \log\left(1 + \frac{\rho}{n_T} \frac{1-\frac{\sigma^2_{e_{\text{pil}}}}{\sigma_h^2}}{1+\rho\frac{\sigma^2_{e_{\text{pil}}}}{\sigma_h^2}} z\right) \\ &\qquad \times \sum_{k=0}^{m-1} \frac{k![L_k^{n-m}(z)]^2}{(k+n-m)!} z^{n-m} e^{-z} dz\end{aligned} \tag{60}$$

$$\begin{aligned}\mathcal{R}_{\text{sep}} &\leq \mathcal{R}_{U,\text{sep}} \\ &= \mathcal{R}_{L,\text{sep}} + \frac{L-n_T}{L} n_R \mathrm{E}_{\tilde{\mathbf{x}}_D(k)} \log\left(\frac{P\sigma^2_{e_{\text{pil}}} + \sigma_n^2}{(\tilde{\mathbf{x}}_D(k))^H \tilde{\mathbf{x}}_D(k) \sigma^2_{e_{\text{pil}}} + \sigma_n^2}\right) \\ &= \mathcal{R}_{L,\text{sep}} + \frac{L-n_T}{L} n_R \int_0^\infty \log\left(\frac{P\sigma^2_{e_{\text{pil}}} + \sigma_n^2}{\frac{P}{n_T}\sigma^2_{e_{\text{pil}}} z + \sigma_n^2}\right) \frac{z^{n_T-1} e^{-z}}{\Gamma(n_T)} dz.\end{aligned} \tag{61}$$

The integral in (60) corresponds to the capacity of a MIMO channel with an SNR degradation factor of $\left(1-\sigma^2_{e_{\text{pil}}}/\sigma_h^2\right) / \left(1+\rho\sigma^2_{e_{\text{pil}}}/\sigma_h^2\right)$ with $m = \min\{n_T, n_R\}$, $n = \max\{n_T, n_R\}$, and L_j^i are the associated Laguerre polynomials, see [30]. In the upper bound (61) the second term on the RHS accounts for the non-Gaussianity of the effective noise including the estimation error. Furthermore, in (61) we use that (51) is fulfilled with equality. The difference between the upper bound and the lower bound decreases with an increasing number of transmit antennas n_T.

D. Lower Bound on the Achievable Rate with Joint Processing of Pilot and Data Symbols

Now, we extend the lower bound on the achievable rate with joint processing of pilot and data symbols given in Section IV to the MIMO case. We mainly focus on the additional steps required in comparison to the SISO case. The first steps in

the derivation of the lower bound are exactly equivalent to the SISO case, cf. (28) to (29), yielding

$$\begin{aligned}&\mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{y}_P,\mathbf{x}_P)\\&\geq \mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{h})-h(\mathbf{e}_{\text{pil}}|\mathbf{x}_P)+h(\mathbf{e}_{\text{joint}}|\mathbf{x}_D,\mathbf{x}_P)\\&=\mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{h})-\log\det\left(\pi e\mathbf{R}_{\mathbf{e}_{\text{pil}}}\right)+\mathrm{E}_{\mathbf{x}_D}\left[\log\det\left(\pi e\mathbf{R}_{\mathbf{e}_{\text{joint}}}\right)\right]\end{aligned} \tag{62}$$

with the solely pilot based channel estimate and its estimation error correlation matrix given by

$$\hat{\mathbf{h}}_{\text{pil}}=\mathrm{E}\left[\mathbf{h}|\mathbf{y}_P,\mathbf{x}_P\right]=\mathbf{h}-\mathbf{e}_{\text{pil}} \tag{63}$$

$$\mathbf{R}_{\mathbf{e}_{\text{pil}}}=\mathrm{E}\left[\mathbf{e}_{\text{pil}}\mathbf{e}_{\text{pil}}^H|\mathbf{x}_P\right]. \tag{64}$$

Furthermore, $\mathbf{R}_{\mathbf{e}_{\text{joint}}}$ is the estimation error correlation matrix of the following channel estimate $\hat{\mathbf{h}}_{\text{joint}}$ based on all pilot and data symbols with

$$\hat{\mathbf{h}}_{\text{joint}}=\mathrm{E}\left[\mathbf{h}|\mathbf{y}_D,\mathbf{x}_D,\mathbf{y}_P,\mathbf{x}_P\right]=\mathbf{h}-\mathbf{e}_{\text{joint}} \tag{65}$$

$$\mathbf{R}_{\mathbf{e}_{\text{joint}}}=\mathrm{E}\left[\mathbf{e}_{\text{joint}}\mathbf{e}_{\text{joint}}^H|\mathbf{x}_P,\mathbf{x}_D\right]. \tag{66}$$

Note that differently to the SISO case, here the estimates $\hat{\mathbf{h}}_{\text{pil}},\hat{\mathbf{h}}_{\text{joint}}$ contain also the spatial dimension in addition to the temporal one. Thus, $\mathbf{R}_{\mathbf{e}_{\text{pil}}}$ and $\mathbf{R}_{\mathbf{e}_{\text{joint}}}$ are of size $Nn_Tn_R\times Nn_Tn_R$.

First, we will derive a lower bound on $h(\mathbf{e}_{\text{joint}}|\mathbf{x}_D,\mathbf{x}_P)$, i.e., the third term on the RHS of (62). The estimation error covariance matrix $\mathbf{R}_{\mathbf{e}_{\text{joint}}}$ of the LMMSE estimate $\hat{\mathbf{h}}_{\text{joint}}$ in (66) can be expressed as follows

$$\begin{aligned}\mathbf{R}_{\mathbf{e}_{\text{joint}}}&=\mathbf{R}_{\mathbf{h}}-\mathbf{R}_{\mathbf{h}}\mathbf{X}^H\left(\mathbf{X}\mathbf{R}_{\mathbf{h}}\mathbf{X}^H+\sigma_n^2\mathbf{I}_{n_RN}\right)^{-1}\mathbf{X}\mathbf{R}_{\mathbf{h}}\\&\overset{(a)}{=}\left(\mathbf{R}_{\mathbf{h}}^{-1}+\frac{1}{\sigma_n^2}\mathbf{X}^H\mathbf{X}\right)^{-1}\\&\overset{(b)}{=}\mathbf{I}_{n_R}\otimes\left(\mathbf{I}_{n_T}\otimes\mathbf{R}_{\bar{\mathbf{h}}}^{-1}+\frac{1}{\sigma_n^2}\left[\bar{\mathbf{X}}_1\ldots\bar{\mathbf{X}}_{n_T}\right]^H\left[\bar{\mathbf{X}}_1\ldots\bar{\mathbf{X}}_{n_T}\right]\right)^{-1}\end{aligned} \tag{67}$$

i.e., the estimation error processes of subchannels to different receive antennas are independent. Here, (a) is the matrix inversion lemma and for (b) we have used (47), (49), and the following identities for matrices $\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D}$ of appropriate sizes [31, Chapter 4.2]

$$(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D})=\mathbf{AC}\otimes\mathbf{BD} \tag{68}$$

$$(\mathbf{A}\otimes\mathbf{B})^H=\mathbf{A}^H\otimes\mathbf{B}^H \tag{69}$$

$$(\mathbf{A}\otimes\mathbf{B})^{-1}=\mathbf{A}^{-1}\otimes\mathbf{B}^{-1} \tag{70}$$

$$\mathbf{A}\otimes\mathbf{B}+\mathbf{A}\otimes\mathbf{C}=\mathbf{A}\otimes(\mathbf{B}+\mathbf{C}). \tag{71}$$

Thus, we get

$$\begin{aligned}&h(\mathbf{e}_{\text{joint}}|\mathbf{x}_D,\mathbf{x}_P)=\mathrm{E}_{\mathbf{x}_D,\mathbf{x}_P}\left[\log\det\left(\pi e\mathbf{R}_{\mathbf{e}_{\text{joint}}}\right)\right]\\&=-n_R\mathrm{E}_{\mathbf{x}}\left[\log\det\left((\pi e)^{-1}\left(\mathbf{I}_{n_T}\otimes\mathbf{R}_{\bar{\mathbf{h}}}^{-1}+\frac{1}{\sigma_n^2}\left[\bar{\mathbf{X}}_1\ldots\bar{\mathbf{X}}_{n_T}\right]^H\left[\bar{\mathbf{X}}_1\ldots\bar{\mathbf{X}}_{n_T}\right]\right)\right)\right]\\&\overset{(a)}{\geq}-n_R\mathrm{E}_{\mathbf{x}}\left[\sum_{j=1}^{n_T}\log\det\left((\pi e)^{-1}\left(\mathbf{R}_{\bar{\mathbf{h}}}^{-1}+\frac{1}{\sigma_n^2}\bar{\mathbf{X}}_j^H\bar{\mathbf{X}}_j\right)\right)\right]\\&\overset{(b)}{=}n_R\left[\sum_{j=1}^{n_T}\left\{\log\det(\pi e\mathbf{R}_{\bar{\mathbf{h}}})-\mathrm{E}_{\mathbf{x}}\left[\log\det\left(\mathbf{I}_N+\frac{\mathbf{R}_{\bar{\mathbf{h}}}\bar{\mathbf{Z}}_j}{\sigma_n^2}\right)\right]\right\}\right]\\&\overset{(c)}{\geq}n_R\left[\sum_{j=1}^{n_T}\left\{\log\det\left(\pi e\mathbf{R}_{\bar{\mathbf{h}}}\right)-\log\det\left(\mathbf{I}_N+\frac{\mathbf{R}_{\bar{\mathbf{h}}}\mathrm{E}\left[\bar{\mathbf{Z}}_j\right]}{\sigma_n^2}\right)\right\}\right].\end{aligned} \tag{72}$$

For simplification, without loss of generality, here we treat the pilot symbols as random. Inequality (a) is Fischer's inequality [32, Theorem 7.8.3], for (b) we use the substitution $\bar{\mathbf{Z}}_j=\bar{\mathbf{X}}_j^H\bar{\mathbf{X}}_j$, which is diagonal and contains the powers of the transmit symbols of antenna j at the individual time instants, and (c) holds due to Jensen's inequality and the concavity of the $\log\det$ expression in $\bar{\mathbf{Z}}_j$.

To get a lower bound on $\mathcal{I}'(\mathbf{x}_D;\mathbf{y}_D,\mathbf{y}_P,\mathbf{x}_P)$, cf. (62), we have to find the distribution $p(\mathbf{x}_D,\mathbf{x}_P)$ that lower-bounds the entropy rate $h'(\mathbf{e}_{\text{joint}}|\mathbf{x}_D,\mathbf{x}_P)$ over all input distributions fulfilling the power constraints in (51) and (54) and the constraint to orthogonal pilot sequences. With (72), it can be shown that $h'(\mathbf{e}_{\text{joint}}|\mathbf{x}_D,\mathbf{x}_P)$ is lower-bounded by

$$\begin{aligned}&h'(\mathbf{e}_{\text{joint}}|\mathbf{x}_D,\mathbf{x}_P)\\&\overset{(a)}{\geq}\lim_{N\to\infty}\frac{n_R}{N}\left[\sum_{j=1}^{n_T}\left\{\log\det(\pi e\mathbf{R}_{\bar{\mathbf{h}}})-\log\det\left(\mathbf{I}_N+\frac{P_j}{\sigma_n^2}\mathbf{R}_{\bar{\mathbf{h}}}\right)\right\}\right]\\&\overset{(b)}{\geq}\lim_{N\to\infty}\frac{n_Rn_T}{N}\left[\log\det(\pi e\mathbf{R}_{\bar{\mathbf{h}}})-\log\det\left(\mathbf{I}_N+\frac{P}{n_T\sigma_n^2}\mathbf{R}_{\bar{\mathbf{h}}}\right)\right]\\&\overset{(c)}{=}\lim_{N\to\infty}\frac{n_Rn_T}{N}\log\det\left(\pi e\left(\mathbf{R}_{\bar{\mathbf{h}}}-\mathbf{R}_{\bar{\mathbf{h}}}\left(\mathbf{R}_{\bar{\mathbf{h}}}+\frac{n_T\sigma_n^2}{P}\mathbf{I}_N\right)^{-1}\mathbf{R}_{\bar{\mathbf{h}}}\right)\right)\\&=\lim_{N\to\infty}\frac{n_Rn_T}{N}\log\det\left(\pi e\mathbf{R}_{\bar{\mathbf{e}}_{\text{joint,SISO,CM}}}\right)\end{aligned} \tag{73}$$

where for (a) we have used that the elements of the sum on the RHS of (72) (b) correspond to the differential entropy of the estimation error processes of the individual subchannels from transmit antenna j to an arbitrary receive antenna. For the SISO channel it has been shown in Appendix A that the differential entropy rate of the estimation error is minimized for a given average power constraint if all symbols have constant modulus, yielding (a) with P_j being the transmit power of transmit antenna j, cf. (138). Here it must hold that

$$\sum_{j=1}^{n_T}P_j\leq P. \tag{74}$$

Due to the concavity of the $\log\det$ in P_j it is easy to see that for all P_j fulfilling (74) the RHS of (a) is lower-bounded if $P_j=P/n_T$, resulting in inequality (b). Finally, (c) follows with the matrix inversion lemma. Thus, this lower bound corresponds to n_Rn_T times the channel estimation error entropy rate for a single-input single-output (SISO) channel with constant modulus (CM) input symbols of power P/n_T where the error correlation matrix is given by $\mathbf{R}_{\bar{\mathbf{e}}_{\text{joint,SISO,CM}}}$, cf. (139).

The RHS of (73) is a lower bound on $h'(\mathbf{e}_{\text{joint}}|\mathbf{x}_D,\mathbf{x}_P)$ corresponding to the case that all transmit symbols have con-

stant modulus. However, also for this case it is not necessarily achievable, as in (72) (a) we have used Fischer's inequality. This inequality corresponds to the ignorance of interference caused by the signals transmitted from the antennas different of the subchannel to be estimated. All other inequalities in the derivation of (73) hold with equality when using constant modulus input symbols with power P/n_T on all transmit antennas. Note that the lower bound in (73) holds also when using input sequences with orthogonal pilot sequences with a fixed pilot symbol power of P/n_T.

With (73) it holds that the mutual information rate $\mathcal{I}'(\mathbf{x}_D;\mathbf{y}_D,\mathbf{y}_P,\mathbf{x}_P)$ is lower-bounded by, cf. (62)

$$\mathcal{I}'(\mathbf{x}_D;\mathbf{y}_D,\mathbf{y}_P,\mathbf{x}_P) \geq \lim_{N\to\infty}\frac{1}{N}\big\{\mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{h}) - h(\mathbf{e}_{\text{pil}}|\mathbf{x}_P) + n_R n_T \log\det\left(\pi e \mathbf{R}_{\bar{\mathbf{e}}_{\text{joint,SISO,CM}}}\right)\big\}. \tag{75}$$

Now, we evaluate $h(\mathbf{e}_{\text{pil}}|\mathbf{x}_P)$ by exploiting the structure of $\mathbf{R}_{\mathbf{e}_{\text{pil}}}$ in (64). Therefor, note that the solely pilot based channel estimate $\hat{\mathbf{h}}_{\text{pil}}$ in (63) is given by

$$\hat{\mathbf{h}}_{\text{pil}} = \mathbf{R}_{\mathbf{h}\mathbf{h}_P}\mathbf{X}_P^H\left(\mathbf{X}_P\mathbf{R}_{\mathbf{h}_P}\mathbf{X}_P^H + \sigma_n^2\mathbf{I}_{N_P n_R}\right)^{-1}\mathbf{y}_P \tag{76}$$

where $\mathbf{X}_P$ corresponds to $\mathbf{X}$ in (47) except that only time instances used for the transmission of pilot symbols are contained and it holds that

$$\mathbf{X}_P = \mathbf{I}_{n_R}\otimes\left[\bar{\mathbf{X}}_{1,P},\ldots,\bar{\mathbf{X}}_{n_T,P}\right] \tag{77}$$

where the $\bar{\mathbf{X}}_{j,P}$, $1\leq j\leq n_T$ correspond to $\bar{\mathbf{X}}_j$ but also only contain pilot symbol time instances. Furthermore, N_P is the number of time instances used for pilot symbols and

$$\mathbf{R}_{\mathbf{h}\mathbf{h}_P} = \mathrm{E}\left[\mathbf{h}\mathbf{h}_P^H\right] \tag{78}$$

$$\mathbf{R}_{\mathbf{h}_P} = \mathrm{E}\left[\mathbf{h}_P\mathbf{h}_P^H\right]. \tag{79}$$

Using (76) the estimation error correlation matrix $\mathbf{R}_{\mathbf{e}_{\text{pil}}}$ in (64) can be expressed as

$$\begin{aligned}\mathbf{R}_{\mathbf{e}_{\text{pil}}} &= \mathrm{E}\left[\left(\mathbf{h}-\hat{\mathbf{h}}_{\text{pil}}\right)\left(\mathbf{h}-\hat{\mathbf{h}}_{\text{pil}}\right)^H\middle|\mathbf{x}_P\right]\\ &= \mathbf{R}_{\mathbf{h}} - \mathbf{R}_{\mathbf{h}\mathbf{h}_P}\mathbf{X}_P^H\left(\mathbf{X}_P\mathbf{R}_{\mathbf{h}_P}\mathbf{X}_P^H+\sigma_n^2\mathbf{I}_{N_P n_R}\right)^{-1}\mathbf{X}_P\mathbf{R}_{\mathbf{h}\mathbf{h}_P}^H.\end{aligned} \tag{80}$$

Using (49), $\mathbf{R}_{\mathbf{h}\mathbf{h}_P}$ and $\mathbf{R}_{\mathbf{h}_P}$ are given by

$$\mathbf{R}_{\mathbf{h}\mathbf{h}_P} = \mathbf{I}_{n_R}\otimes\mathbf{I}_{n_T}\otimes\mathbf{R}_{\bar{\mathbf{h}}\bar{\mathbf{h}}_P} \tag{81}$$

$$\mathbf{R}_{\mathbf{h}_P} = \mathbf{I}_{n_R}\otimes\mathbf{I}_{n_T}\otimes\mathbf{R}_{\bar{\mathbf{h}}_P} \tag{82}$$

with

$$\mathbf{R}_{\bar{\mathbf{h}}\bar{\mathbf{h}}_P} = \mathrm{E}\left[\bar{\mathbf{h}}_{ij}\bar{\mathbf{h}}_{ij,P}^H\right],\quad 1\leq i\leq n_R, 1\leq j\leq n_T \tag{83}$$

$$\mathbf{R}_{\bar{\mathbf{h}}_P} = \mathrm{E}\left[\bar{\mathbf{h}}_{ij,P}\bar{\mathbf{h}}_{ij,P}^H\right],\quad 1\leq i\leq n_R, 1\leq j\leq n_T \tag{84}$$

where $\bar{\mathbf{h}}_{ij,P}$ corresponds to $\bar{\mathbf{h}}_{ij}$ but contains only pilot symbol time instances.

With (49), (77), (81), and (82), we can rewrite (80) as follows:

$$\begin{aligned}\mathbf{R}_{\mathbf{e}_{\text{pil}}} = \mathbf{I}_{n_R}\otimes\Bigg[&\mathbf{I}_{n_T}\otimes\mathbf{R}_{\bar{\mathbf{h}}} - \left(\mathbf{I}_{n_T}\otimes\mathbf{R}_{\bar{\mathbf{h}}\bar{\mathbf{h}}_P}\right)\left[\bar{\mathbf{X}}_{1,P},\ldots,\bar{\mathbf{X}}_{n_T,P}\right]^H\\ &\times\Big\{\left[\bar{\mathbf{X}}_{1,P},\ldots,\bar{\mathbf{X}}_{n_T,P}\right]\left(\mathbf{I}_{n_T}\otimes\mathbf{R}_{\bar{\mathbf{h}}_P}\right)\left[\bar{\mathbf{X}}_{1,P},\ldots,\bar{\mathbf{X}}_{n_T,P}\right]^H\\ &+\sigma_n^2\mathbf{I}_{N_P}\Big\}^{-1}\left[\bar{\mathbf{X}}_{1,P},\ldots,\bar{\mathbf{X}}_{n_T,P}\right]\left(\mathbf{I}_{n_T}\otimes\mathbf{R}_{\bar{\mathbf{h}}\bar{\mathbf{h}}_P}^H\right)\Bigg]\end{aligned} \tag{85}$$

where we have again used (68)-(71). Obviously, the estimation error processes of subchannels to different receive antennas are independent and have the same error correlation. This is due to the fact that the fading on the individual subchannels as well as the additive noise at different receive antennas are independent and identically distributed.

With (85), $h(\mathbf{e}_{\text{pil}}|\mathbf{x}_P)$ can be upper-bounded as follows:

$$\begin{aligned}h(\mathbf{e}_{\text{pil}}|\mathbf{x}_P) &= \log\det\left(\pi e\mathbf{R}_{\mathbf{e}_{\text{pil}}}\right)\\ &\overset{(a)}{=} n_R\log\det\Bigg(\pi e\Big[\mathbf{I}_{n_T}\otimes\mathbf{R}_{\bar{\mathbf{h}}} - \left(\mathbf{I}_{n_T}\otimes\mathbf{R}_{\bar{\mathbf{h}}\bar{\mathbf{h}}_P}\right)\\ &\qquad\times\left[\bar{\mathbf{X}}_{1,P},\ldots,\bar{\mathbf{X}}_{n_T,P}\right]^H\left(P\mathbf{S}\odot\mathbf{R}_{\bar{\mathbf{h}}_P}+\sigma_n^2\mathbf{I}_{N_P}\right)^{-1}\\ &\qquad\times\left[\bar{\mathbf{X}}_{1,P},\ldots,\bar{\mathbf{X}}_{n_T,P}\right]\left(\mathbf{I}_{n_T}\otimes\mathbf{R}_{\bar{\mathbf{h}}\bar{\mathbf{h}}_P}^H\right)\Big]\Bigg)\\ &\overset{(b)}{\leq} n_R\sum_{j=1}^{n_T}\log\det\Big(\pi e\Big[\mathbf{R}_{\bar{\mathbf{h}}}\\ &\qquad-\mathbf{R}_{\bar{\mathbf{h}}\bar{\mathbf{h}}_P}\bar{\mathbf{X}}_{j,P}^H\left(P\mathbf{S}\odot\mathbf{R}_{\bar{\mathbf{h}}_P}+\sigma_n^2\mathbf{I}_{N_P}\right)^{-1}\bar{\mathbf{X}}_{j,P}\mathbf{R}_{\bar{\mathbf{h}}\bar{\mathbf{h}}_P}^H\Big]\Big)\\ &\overset{(c)}{=} n_R\sum_{j=1}^{n_T}\log\det\Bigg(\pi e\Bigg[\mathbf{R}_{\bar{\mathbf{h}}}\\ &\qquad-\mathbf{R}_{\bar{\mathbf{h}}\bar{\mathbf{h}}_P}\bar{\mathbf{X}}_{j,P}^H\Big(P(\mathbf{R}_{\bar{\mathbf{h}}_{P,1}}\otimes\mathbf{I}_{n_T})+\sigma_n^2\mathbf{I}_{N_P}\Big)^{-1}\bar{\mathbf{X}}_{j,P}\mathbf{R}_{\bar{\mathbf{h}}\bar{\mathbf{h}}_P}^H\Bigg]\Bigg)\\ &= n_R\sum_{j=1}^{n_T}\log\det\Bigg(\pi e\Bigg[\mathbf{R}_{\bar{\mathbf{h}}}-\mathbf{R}_{\bar{\mathbf{h}}\bar{\mathbf{h}}_P}\bar{\mathbf{X}}_{j,P}^H\\ &\qquad\times\left(\left(P\mathbf{R}_{\bar{\mathbf{h}}_{P,1}}+\sigma_n^2\mathbf{I}_{\frac{N}{L}}\right)^{-1}\otimes\mathbf{I}_{n_T}\right)\bar{\mathbf{X}}_{j,P}\mathbf{R}_{\bar{\mathbf{h}}\bar{\mathbf{h}}_P}^H\Bigg]\Bigg)\\ &\overset{(d)}{=} n_R\sum_{j=1}^{n_T}\log\det\left(\pi e\mathbf{R}_{\bar{\mathbf{e}}_{\text{pil},ij}}\right)\end{aligned} \tag{86}$$

where for (a) we have used the following substitution

$$\begin{aligned}\sum_{j=1}^{n_T}\bar{\mathbf{X}}_{j,P}\mathbf{R}_{\bar{\mathbf{h}}_P}\bar{\mathbf{X}}_{j,P}^H &= \left[\sum_{j=1}^{n_T}\mathrm{diag}(\bar{\mathbf{X}}_{j,P})[\mathrm{diag}(\bar{\mathbf{X}}_{j,P})]^H\right]\odot\mathbf{R}_{\bar{\mathbf{h}}_P}\\ &= P\mathbf{S}\odot\mathbf{R}_{\bar{\mathbf{h}}_P}.\end{aligned} \tag{87}$$

Here, $\odot$ denotes the Hadamard product and the $\mathrm{diag}(\cdot)$ operator generates a vector containing the elements of the diagonal of the argument matrix, which itself is diagonal. Due to the assumption of orthogonal pilot sequences fulfilling (53) and (54), $\mathbf{S}$ is a Hermitian Toeplitz matrix, where the main diagonal and every n_T-th off-diagonal is filled with ones and all other elements are zero. Furthermore, (b) is Fischer's inequality [32, Theorem 7.8.3] holding for positive definite matrices, which is fulfilled by the argument of the determinant on the LHS of (b). The inequality in (b) corresponds to discarding the correlation among the estimation error processes of the subchannels from different transmit antennas to the same receive antenna. These error processes are in general correlated as the estimation is based on the same channel output observations. In (c) the matrix $\mathbf{R}_{\bar{\mathbf{h}}_{P,1}}\in\mathbb{C}^{\frac{N}{L}\times\frac{N}{L}}$ is a submatrix of $\mathbf{R}_{\bar{\mathbf{h}}_P}$ containing every n_T-th row and column starting with the first row and column. I.e., it contains only one element of the channel autocorrelation function per pilot interval of length L. Finally,

(d) holds, as the argument of the $\log\det$-expression is nothing else than πe times the channel estimation error correlation matrix for the subchannel from transmit antenna j to an arbitrary receive antenna i, i.e.,

$$\begin{aligned}\mathbf{R}_{\bar{\mathbf{e}}_{\text{pil},ij}} &= \mathrm{E}\left[\bar{\mathbf{e}}_{\text{pil}_{ij}}\bar{\mathbf{e}}_{\text{pil}_{ij}}^{H}\middle|\mathbf{x}_P\right] \\ &= \mathrm{E}\left[(\bar{\mathbf{h}}_{ij}-\hat{\bar{\mathbf{h}}}_{\text{pil}_{ij}})(\bar{\mathbf{h}}_{ij}-\hat{\bar{\mathbf{h}}}_{\text{pil}_{ij}})^{H}\middle|\mathbf{x}_P\right], \\ &\qquad 1\le i\le n_R, 1\le j\le n_T. \end{aligned} \tag{88}$$

Due to the assumption that all subchannels are characterized by the same channel statistics, for a given transmit antenna j $\mathbf{R}_{\bar{\mathbf{e}}_{\text{pil},ij}}$ is equal for all receive antennas i. On the other hand, in general $\mathbf{R}_{\bar{\mathbf{e}}_{\text{pil},ij}}$ depends on the transmit antenna j as the pilot sequences transmitted from the different transmit antennas are different.

We want to derive a lower bound on the achievable rate for the average power constraint on the data symbols given in (51) and for pilot schemes taken out of the set $\mathcal{S}_P$ containing all orthogonal and periodic pilot sequences fulfilling (53) and (54). Obviously, such a lower bound is obtained by evaluating the RHS of (75) for any pilot scheme in the set $\mathcal{S}_P$. We next calculate the RHS of (75) by choosing a specific pilot scheme. The entropy of the estimation error when using a specific pilot scheme is always an upper bound for $h(\mathbf{e}_{\text{pil}}|\mathbf{x}_P)$ using the optimum orthogonal pilot scheme fulfilling (53) and (54). For this reason, we evaluate the upper-bound on $h(\mathbf{e}_{\text{pil}}|\mathbf{x}_P)$ on the RHS of (86) for a specific pilot scheme. The orthogonality of the pilot sequences can, e.g., be achieved by time-sharing. I.e., at each pilot symbol time instant only one antenna transmits a pilot symbol with constant power P while the other antennas do not transmit. For this pilot scheme it is also easy to verify that $\mathbf{R}_{\bar{\mathbf{e}}_{\text{pil},ij}}$ depends on the transmit antenna j as the pilot symbols from the individual transmit antennas are transmitted at different time instances. However, considering Nyquist sampling of the channel by the pilot symbols, see (52), for an infinite transmission length N the estimation error is stationary and, thus, all $\mathbf{R}_{\bar{\mathbf{e}}_{\text{pil},ij}}$ are asymptotically equal. Applying a time-sharing based pilot scheme the channel estimation of the individual subchannels corresponds exactly to the channel estimation in case of a SISO channel. Consequently, using a time-sharing based pilot scheme, e.g., introducing $\bar{\mathbf{X}}_{j,P}=\mathrm{diag}(\bar{\mathbf{x}}_{j,P})$ where the elements of $\bar{\mathbf{x}}_{j,P}$ are given by

$$[\bar{\mathbf{x}}_{j,P}]_l = \begin{cases} P & \text{for } l=(n-1)\cdot n_T+j, \\ & \text{with } n=\{1,2,\ldots,N_P/n_T\} \\ 0 & \text{otherwise} \end{cases} \tag{89}$$

into the RHS of (86) shows that the channel estimation error corresponds exactly to the channel estimation error in case of a SISO channel with the SNR ρ, for which the PSD of the error process is given by (35).

Hence, using (86) the entropy rate $h'(\mathbf{e}_{\text{pil}}|\mathbf{x}_P)|_{\text{TS}}$ for the time-sharing (TS) based pilot scheme in (89) is given by

$$\begin{aligned} h'(\mathbf{e}_{\text{pil}}|\mathbf{x}_P)\big|_{\text{TS}} &= \lim_{N\to\infty}\frac{1}{N}h(\mathbf{e}_{\text{pil}}|\mathbf{x}_P)\big|_{\text{TS}} \\ &\overset{(a)}{=} \lim_{N\to\infty}\frac{1}{N}n_R\sum_{j=1}^{n_T}\log\det\left(\pi e\mathbf{R}_{\bar{\mathbf{e}}_{\text{pil},ij}}\right)\Big|_{\text{TS}} \\ &= \lim_{N\to\infty}\frac{n_Rn_T}{N}\log\det\left(\pi e\mathbf{R}^{\text{TS}}_{\bar{\mathbf{e}}_{\text{pil}}}\right) \end{aligned} \tag{90}$$

where $\mathbf{R}^{\text{TS}}_{\bar{\mathbf{e}}_{\text{pil}}}$ corresponds to $\mathbf{R}_{\bar{\mathbf{e}}_{\text{pil},ij}}$ when using (89) and where we discard the index ij as all $\mathbf{R}_{\bar{\mathbf{e}}_{\text{pil},ij}}$ are asymptotically equal for an infinite transmission length. Note that for the time-sharing based pilot scheme where at each pilot symbol time instant only one antenna transmits, different observations are used for the estimation of the subchannels to the same receive antenna and, thus, also the estimation error processes of these subchannels are independent. Hence, for this specific pilot scheme $\mathbf{R}_{\mathbf{e}_{\text{pil}}}$ in (85) is a block-diagonal matrix, containing only the temporal error correlation matrices for the individual subchannels. Consequently, for this specific pilot scheme the inequality (b) in (86) holds with equality, yielding (a) in (90).

Similar to the SISO case discussed in Section IV with (75) and (90), we get the following lower bound on $\mathcal{I}'(\mathbf{x}_D;\mathbf{y}_D,\mathbf{x}_P,\mathbf{y}_P)$, which is achievable with time-sharing based pilot symbols

$$\begin{aligned} &\mathcal{I}'(\mathbf{x}_D;\mathbf{y}_D,\mathbf{x}_P,\mathbf{y}_P) \\ &\ge \lim_{N\to\infty}\frac{1}{N}\Bigg\{\mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{h}) \\ &\qquad - n_Rn_T\left[\log\det\left(\mathbf{R}^{\text{TS}}_{\bar{\mathbf{e}}_{\text{pil}}}\right)-\log\det\left(\mathbf{R}_{\bar{\mathbf{e}}_{\text{joint,SISO,CM}}}\right)\right]\Bigg\} \\ &\overset{(a)}{=} \lim_{N\to\infty}\frac{1}{N}\mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{h}) \\ &\qquad - n_Rn_T\int_{-\frac{1}{2}}^{\frac{1}{2}}\log\left(\frac{S_{e_{\text{pil}}}(f)}{S_{e_{\text{joint,SISO,CM}}}(f)}\right)df \end{aligned} \tag{91}$$

where (a) is based on Szegö's theorem on the asymptotic eigenvalue distribution of Hermitian Toeplitz matrices [28]. $S_{e_{\text{pil}}}(f)$ and $S_{e_{\text{joint,SISO,CM}}}(f)$ are the PSDs corresponding to the error correlation matrices $\mathbf{R}^{\text{TS}}_{\bar{\mathbf{e}}_{\text{pil}}}$ and $\mathbf{R}_{\bar{\mathbf{e}}_{\text{joint,SISO,CM}}}$. As already stated, for the orthogonal pilot symbol sequences based on time-sharing that have been used for the derivation of (90), where at each pilot symbol time instant only one antenna transmits with power P, the channel estimation corresponds exactly to the channel estimation in case of a SISO channel with the SNR ρ given in (55), for which the PSD of the error process $S_{e_{\text{pil}}}(f)$ is given by (35). Furthermore, $S_{e_{\text{joint,SISO,CM}}}(f)$ is the PSD of the estimation error process for a SISO channel where all input symbols are known and of constant modulus with power P/n_T. Effectively, this corresponds to $S_{e_{\text{joint,CM}}}(f)$ in (36) but substituting ρ in (36) with $(P\sigma_h^2)/(n_T\sigma_n^2)$. Thus we get

$$S_{e_{\text{joint,SISO,CM}}}(f)=\frac{S_h(f)}{\frac{\rho}{n_T}\frac{S_h(f)}{\sigma_h^2}+1}. \tag{92}$$

As in the SISO case discussed in Section IV, the first term on the RHS of (91) is the mutual information rate in case of perfect channel knowledge, which for an average power constraint is maximized with i.i.d. zero-mean proper Gaussian data symbols in time and on the individual transmit antennas

[30]. Thus, we get the following lower bound on the achievable rate with joint processing of pilot and data symbols

$$\mathcal{R}_{L,\text{joint}} = \frac{L-n_T}{L} C_{\text{perf,MIMO}}(\rho) - n_R n_T \int_{-\frac{1}{2}}^{\frac{1}{2}} \log\left(\frac{\frac{\rho}{n_T}\frac{S_h(f)}{\sigma_h^2}+1}{\frac{\rho}{L}\frac{S_h(f)}{\sigma_h^2}+1}\right) df \quad (93)$$

where the factor $(L-n_T)/L$ arises, as n_T symbols per pilot interval of length L are pilot symbols. The coherent capacity $C_{\text{perf,MIMO}}(\rho)$ is given by [30]

$$C_{\text{perf,MIMO}}(\rho) = \int_0^\infty \log\left(1+\frac{\rho}{n_T}\lambda\right) \times \sum_{k=0}^{m-1} \frac{k![L_k^{n-m}(\lambda)]^2}{(k+n-m)!}\lambda^{n-m}e^{-\lambda}d\lambda \quad (94)$$

with $m=\min\{n_T,n_R\}$, $n=\max\{n_T,n_R\}$, and L_j^i are the associated Laguerre polynomials.

With (93) we have found a lower bound on the achievable rate with joint processing of pilot and data symbols, which holds for an arbitrary pilot spacing L fulfilling (52).

VI. Numerical Evaluation

A. SISO Channel

Fig. 1 shows a comparison of the bounds on the achievable rate for separate and joint processing of data and pilot symbols for the case of a SISO channel with a rectangular PSD as given in (39).

On the one hand, the lower bound on the achievable rate for joint processing in (40) 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., (40) in combination with (41), 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

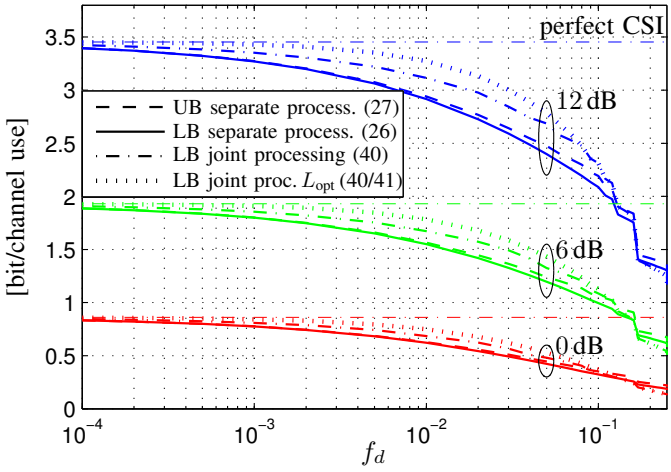

Fig. 1. Comparison of bounds on the achievable rate with separate processing of data and pilot symbols to lower bounds on the achievable rate with joint processing of data and pilot symbols; except of *LB joint proc.* L_{opt} the pilot spacing L is chosen such that the lower bound for separate processing (26) is maximized; the PSD $S_h(f)$ is assumed to be rectangular, see (39).

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.

Finally, for comparison also the capacity in case of perfect channel state information (CSI) at the receiver is shown. As this is also an upper bound on the achievable rate in the noncoherent case, this comparison shows that for small f_d the lower bound on joint processing is relatively tight.

From an engineering point of view our results give the indication that the prevalent choice of separate processing in combination with an optimized pilot spacing seems to be highly efficient in comparison to joint processing regarding the communication performance in terms of the achievable rate. This statement is additionally supported by the fact that typically observed channel dynamics f_d are smaller than 10^{-3}. Furthermore, the computational complexity of separate processing is feasible in contrast to the case of joint processing. Note that with iterative joint channel estimation and decoding schemes one tries to approximately solve the joint processing problem. These receivers are expected to come close to the achievable rate with joint processing while still being computationally feasible. Nevertheless, their computational complexity is higher than that of separate processing based receivers. However, regarding this interpretation keep in mind that we have only derived a lower bound on the achievable rate with joint processing.

Fig. 2 shows the lower bound on the achievable rate for joint processing of data and pilot symbols in (40) when choosing L as given in (41), which maximizes this lower bound. 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

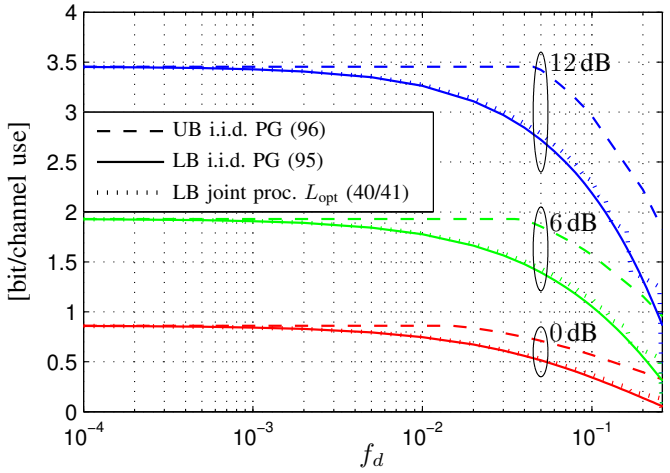

Fig. 2. Lower bound on the achievable rate with joint processing of data and pilot symbols and a pilot spacing L_{opt} that maximizes this bound, i.e., (40) in combination with (41); for comparison bounds on the achievable rate with i.i.d. zero-mean proper Gaussian (PG) input symbols are shown; rectangular PSD $S_h(f)$, see (39).

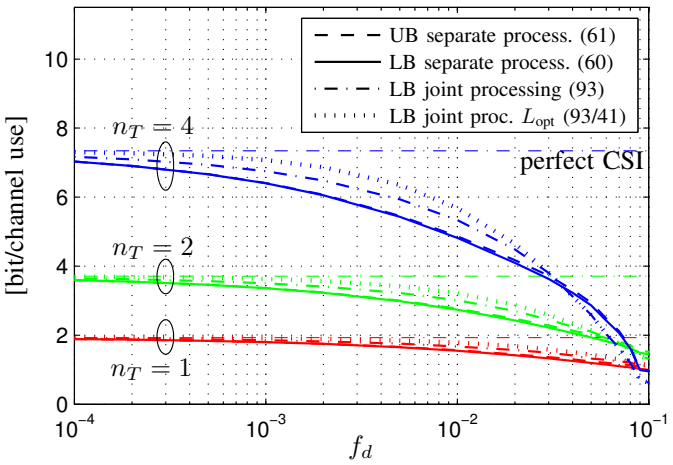

Fig. 3. Comparison of bounds on the achievable rate with separate processing to lower bounds on the achievable rate with joint processing of data and pilot symbols for the MIMO case, equal number of transmit and receive antennas $n_T = n_R$, SNR $\rho = 6$ dB; except of *LB joint proc.* L_{opt}, the pilot spacing L is chosen such that the lower bound for separate processing is maximized; rectangular $S_h(f)$, see (39).

process, see (39), which have been given in [12], [13] as

$$\mathcal{I}'_L(\mathbf{y};\mathbf{x})\big|_{\text{PG}} = \max\left\{C_{\text{perf}}(\rho) - 2f_d \log\left(1+\frac{\rho}{2f_d}\right), 0\right\} \tag{95}$$

$$\mathcal{I}'_U(\mathbf{y};\mathbf{x})\big|_{\text{PG}} = \min\Bigg\{\log(1+\rho) - 2f_d\int_0^\infty \log\left(1+\frac{\rho}{2f_d}z\right)e^{-z}dz, C_{\text{perf}}(\rho)\Bigg\}. \tag{96}$$

with $C_{\text{perf}}(\rho)$ being the coherent capacity of a Rayleigh flat-fading channel given in (38).

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

B. MIMO Channel

Now, we compare the lower bound on the achievable rate with joint processing of pilot and data symbols in (93) with bounds on the achievable rate with a separate processing given in (60) and (61) for the case of a MIMO channel. Again, we assume that $S_h(f)$ is rectangular, i.e., given by (39).

For the case of a separate processing, we choose the pilot spacing L that maximizes the lower bound in (60), approximately yielding the maximum achievable rate with separate processing. This pilot spacing is smaller than or equal to the maximum possible pilot spacing fulfilling (52).

For the evaluation of the lower bound on the achievable rate with a joint processing of pilot and data symbols in (93) we choose on the one hand the same pilot spacing as for the bounds on separate processing. On the other hand, we evaluate the lower bound on the achievable rate with joint processing for the largest pilot spacing still fulfilling Nyquist sampling L_{opt} given in (41). For the parameters evaluated in Fig. 3 this choice maximizes the lower bound in (93) for a rectangular $S_h(f)$ and integer L fulfilling (52).

Fig. 3 shows the comparison of the achievable rate with separate and joint processing for an equal number of transmit and receive antennas. Obviously, when choosing the same pilot spacing in case of separate and joint processing for typically observed channel dynamics f_d there is a significant gain due to joint processing. This gain increases with the number of antennas and, up to some point, with the channel dynamic f_d. This gain is even larger when choosing the pilot spacing L_{opt}, except for very high channel dynamics. The comparison of the lower bound on the achievable rate with joint processing with the capacity in case of perfect channel state information shows that for small channel dynamics f_d the lower bound on the achievable rate of joint processing is relatively tight.

As the gain with joint processing of pilot and data symbols in comparison to separate processing strongly increases with the number of antennas even for small channel dynamics, differently to the SISO case for MIMO systems the use of joint processing seems to be more interesting from an engineering point of view.

VII. 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. This information 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, 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. Finally, we have extended the derivation of this lower bound to the case of MIMO channels.

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 a maximum average power of σ_x^2.

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

$$\begin{aligned}\hat{\mathbf{h}}_{\text{joint}} &= \arg\max_{\mathbf{h}} p(\mathbf{h}|\mathbf{y}, \mathbf{x}) \\ &= \arg\max_{\mathbf{h}} p(\mathbf{y}|\mathbf{h}, \mathbf{x}) p(\mathbf{h}) \\ &= \arg\max_{\mathbf{h}} \left\{\log(p(\mathbf{y}|\mathbf{h}, \mathbf{x})) + \log(p(\mathbf{h}))\right\} \end{aligned} \tag{97}$$

with

$$p(\mathbf{y}|\mathbf{h}, \mathbf{x}) = \frac{1}{\pi^N \sigma_n^{2N}} \exp\left(-\frac{|\mathbf{y} - \mathbf{X}\mathbf{h}|^2}{\sigma_n^2}\right) \tag{98}$$

$$p(\mathbf{h}) = \frac{1}{\pi^N \det(\mathbf{R}_h)} \exp\left(-\mathbf{h}^H \mathbf{R}_h^{-1} \mathbf{h}\right). \tag{99}$$

Thus, (97) becomes

$$\hat{\mathbf{h}}_{\text{joint}} = \arg\max_{\mathbf{h}} \left\{-\frac{1}{\sigma_n^2}|\mathbf{y} - \mathbf{X}\mathbf{h}|^2 - \mathbf{h}^H \mathbf{R}_h^{-1} \mathbf{h}\right\}. \tag{100}$$

Differentiating the argument of the maximum operation at the RHS of (100) with respect to $\mathbf{h}$ and setting the result equal to zero yields

$$-\frac{1}{\sigma_n^2}\left\{-\mathbf{X}^H\mathbf{y} + \mathbf{X}^H\mathbf{X}\mathbf{h}\right\} - \mathbf{R}_h^{-1}\mathbf{h} = \mathbf{0} \tag{101}$$

and, thus,[5]

$$\hat{\mathbf{h}}_{\text{joint}} = \mathbf{R}_h \left(\mathbf{R}_h + \sigma_n^2 \left(\mathbf{X}^H\mathbf{X}\right)^{-1}\right)^{-1} \mathbf{X}^{-1}\mathbf{y}. \tag{103}$$

With (103) the channel estimation error correlation matrix $\mathbf{R}_{e_{\text{joint}}}$ is given by

$$\begin{aligned}\mathbf{R}_{e_{\text{joint}}} &= \mathrm{E}\left[\left(\mathbf{h} - \hat{\mathbf{h}}_{\text{joint}}\right)\left(\mathbf{h} - \hat{\mathbf{h}}_{\text{joint}}\right)^H \middle| \mathbf{x}\right] \\ &= \mathbf{R}_h - \mathbf{R}_h \left(\mathbf{R}_h + \sigma_n^2 (\mathbf{X}^H\mathbf{X})^{-1}\right)^{-1} \mathbf{R}_h. \end{aligned} \tag{104}$$

[5]Note that the inverse of $\mathbf{X}$ in (103) 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

$$\hat{\mathbf{h}}_{\text{joint}} = \mathbf{R}_h \mathbf{X}^H \left(\mathbf{X}\mathbf{R}_h\mathbf{X}^H + \sigma_n^2 \mathbf{I}_N\right)^{-1} \mathbf{y} \tag{102}$$

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.

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

$$\begin{aligned}&h(\mathbf{e}_{\text{joint}}|\mathbf{x}_D, \mathbf{x}_P) = \mathrm{E}_{\mathbf{x}}\left[\log\det\left(\pi e \mathbf{R}_{e_{\text{joint}}}\right)\right] \\ &= \mathrm{E}_{\mathbf{x}}\left[\log\det\left(\pi e\left(\mathbf{R}_h - \mathbf{R}_h\left(\mathbf{R}_h + \sigma_n^2(\mathbf{X}^H\mathbf{X})^{-1}\right)^{-1}\mathbf{R}_h\right)\right)\right] \\ &\overset{(a)}{=} \mathrm{E}_{\mathbf{x}}\left[\log\det\left(\pi e\left(\mathbf{R}_h^{-1} + \frac{1}{\sigma_n^2}\mathbf{X}^H\mathbf{X}\right)^{-1}\right)\right] \\ &= \log\left((\pi e)^N \det(\mathbf{R}_h)\right) \\ &\quad - \mathrm{E}_{\mathbf{x}}\left[\log\det\left(\frac{1}{\sigma_n^2}\mathbf{R}_h\mathbf{X}^H\mathbf{X} + \mathbf{I}_N\right)\right]. \end{aligned} \tag{105}$$

where (a) follows from the matrix inversion lemma.

As the matrix $\mathbf{X} = \mathrm{diag}(\mathbf{x})$ 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

$$\mathbf{Z} = \mathbf{X}\mathbf{X}^H \tag{106}$$

and $\mathbf{z} = \mathrm{diag}(\mathbf{Z})$ 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 (105) 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.,

$$\mathrm{E}\left[\mathbf{x}^H\mathbf{x}\right] = \mathrm{E}\left[\sum_{k=1}^{N} z_k\right] \leq N\sigma_x^2 \tag{107}$$

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 (105), 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 (107). 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 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 distribution of $\mathbf{z}$ that minimizes (105) provided that the average power constraint is fulfilled. Therefor, we first show that the argument of the expectation operation on the RHS of (105), i.e.,

$$g(\mathbf{Z}) = \log\det\left(\frac{1}{\sigma_n^2}\mathbf{R}_h\mathbf{Z} + \mathbf{I}_N\right) \tag{108}$$

is concave in $\mathbf{Z}$. To verify the concavity of $g(\mathbf{Z})$, we follow along the lines of [33, 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

$$\begin{aligned}g(t) &= \log\det\left(\frac{1}{\sigma_n^2}\mathbf{R}_h\left(\bar{\mathbf{Z}} + t\boldsymbol{\Delta}\right) + \mathbf{I}_N\right) \\ &= \log\det\left(\frac{1}{\sigma_n^2}\mathbf{R}_h\right) + \log\det\left(\bar{\mathbf{Z}} + \sigma_n^2\mathbf{R}_h^{-1} + t\boldsymbol{\Delta}\right) \end{aligned}$$

$$\begin{aligned}
&\overset{(a)}{=} \log\det\left(\frac{1}{\sigma_n^2}\mathbf{R}_h\right) + \log\det\left(\mathbf{Q}+t\mathbf{\Delta}\right)\\
&= \log\det\left(\frac{\mathbf{R}_h}{\sigma_n^2}\right) + \log\det\Big(\mathbf{Q}^{\frac{H}{2}}\left(\mathbf{I}_N + t\mathbf{Q}^{-\frac{H}{2}}\mathbf{\Delta}\mathbf{Q}^{-\frac{1}{2}}\right)\mathbf{Q}^{\frac{1}{2}}\Big)\\
&= \log\det\Big(\frac{\mathbf{R}_h}{\sigma_n^2}\Big) + \log\det\left(\mathbf{Q}\right) + \log\det\Big(\mathbf{I}_N + t\mathbf{Q}^{-\frac{H}{2}}\mathbf{\Delta}\mathbf{Q}^{-\frac{1}{2}}\Big)\\
&= \log\det\left(\frac{1}{\sigma_n^2}\mathbf{R}_h\bar{\mathbf{Z}} + \mathbf{I}_N\right)\\
&\quad + \log\det\left(\mathbf{I}_N + t\left(\bar{\mathbf{Z}} + \sigma_n^2\mathbf{R}_h^{-1}\right)^{-\frac{H}{2}}\mathbf{\Delta}\left(\bar{\mathbf{Z}} + \sigma_n^2\mathbf{R}_h^{-1}\right)^{-\frac{1}{2}}\right)\\
&\overset{(b)}{=} \log\det\left(\frac{1}{\sigma_n^2}\mathbf{R}_h\bar{\mathbf{Z}} + \mathbf{I}_N\right) + \sum_{k=1}^{N}\log\left(1+t\lambda_k\right) \qquad (109)
\end{aligned}$$

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

Based on (109) the derivatives of $g(t)$ with respect to t are given by

$$\frac{dg(t)}{dt} = \sum_{k=1}^{N}\frac{\lambda_k}{1+t\lambda_k} \qquad (110)$$

$$\frac{d^2g(t)}{dt^2} = -\sum_{k=1}^{N}\frac{\lambda_k^2}{\left(1+t\lambda_k\right)^2}. \qquad (111)$$

As the second derivative $\frac{d^2g(t)}{dt^2}$ 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 $h(\mathbf{e}_{\text{joint}}|\mathbf{x}_D,\mathbf{x}_P)$ in (105) by using Jensen's inequality as follows, cf. (108):

$$\begin{aligned}
&h(\mathbf{e}_{\text{joint}}|\mathbf{x}_D,\mathbf{x}_P) = \log\det\left((\pi e)^N\det(\mathbf{R}_h)\right) - \mathrm{E}_{\mathbf{z}}\left[g(\mathbf{Z})\right]\\
&\geq \log\det\left((\pi e)^N\det(\mathbf{R}_h)\right) - \log\det\left(\frac{1}{\sigma_n^2}\mathbf{R}_h\mathrm{E}\left[\mathbf{Z}\right] + \mathbf{I}_N\right). \qquad (112)
\end{aligned}$$

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

$$\begin{aligned}
h'(\mathbf{e}_{\text{joint}}|\mathbf{x}_D,\mathbf{x}_P) &= \lim_{N\to\infty}\frac{1}{N}h(\mathbf{e}_{\text{joint}}|\mathbf{x}_D,\mathbf{x}_P)\\
&\geq \lim_{N\to\infty}\frac{1}{N}\Bigg[\log\det\left((\pi e)^N\det(\mathbf{R}_h)\right)\\
&\qquad - \log\det\left(\frac{1}{\sigma_n^2}\mathbf{R}_h\mathrm{E}\left[\mathbf{Z}\right] + \mathbf{I}_N\right)\Bigg]. \qquad (113)
\end{aligned}$$

In the next step, we show for which kind of distribution of $\mathbf{z}$ fulfilling the maximum average power constraint in (107) the RHS of (113) is minimized. I.e., we have to find the distribution of $\mathbf{z}$ that maximizes

$$\lim_{N\to\infty}\frac{1}{N}\log\det\left(\frac{1}{\sigma_n^2}\mathbf{R}_h\mathrm{E}\left[\mathbf{Z}\right] + \mathbf{I}_N\right). \qquad (114)$$

For the evaluation of (114) 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 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

$$\begin{aligned}
&\lim_{N\to\infty}\frac{1}{N}\log\det\left(\frac{1}{\sigma_n^2}\mathbf{R}_h\mathrm{E}\left[\mathbf{Z}\right] + \mathbf{I}_N\right)\\
&\qquad = \lim_{N\to\infty}\frac{1}{N}\log\det\left(\frac{1}{\sigma_n^2}\mathbf{C}_h\mathrm{E}\left[\mathbf{Z}\right] + \mathbf{I}_N\right). \qquad (115)
\end{aligned}$$

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

$$\mathbf{R}_h = \mathbf{R}_h^{(N)} = \mathbf{U}^{(N)}\mathbf{\Lambda}_h^{(N)}\left(\mathbf{U}^{(N)}\right)^H \qquad (116)$$

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 [26, 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

$$c_k^{(N)} = \frac{1}{N}\sum_{l=0}^{N-1}\tilde{S}_h\left(\frac{l}{N}\right)e^{j2\pi\frac{lk}{N}}. \qquad (117)$$

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

$$\tilde{S}_h(f) = \sum_{k=-\infty}^{\infty}\delta(f-k)\star S_h(f) \qquad (118)$$

and $S_h(f)$ being zero outside the interval $|f|\leq 0.5$ for which it is defined in (4). The asterisk $\star$ in (118) 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

$$\begin{aligned}
\lim_{N\to\infty}c_k^{(N)} &= \lim_{N\to\infty}\frac{1}{N}\sum_{l=0}^{N-1}\tilde{S}_h\left(\frac{l}{N}\right)e^{j2\pi\frac{lk}{N}}\\
&= \int_0^1\tilde{S}_h(f)e^{j2\pi kf}df\\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}}S_h(f)e^{j2\pi kf}df = r_h(k) \qquad (119)
\end{aligned}$$

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

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

$$= \sum_{m=0}^{N-1} \tilde{S}_h\left(\frac{m}{N}\right) \left\{ \frac{1}{N} \sum_{l=0}^{N-1} e^{j2\pi \frac{l(m-(k-1))}{N}} \right\}$$
$$= \tilde{S}_h\left(\frac{k-1}{N}\right). \tag{120}$$

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

$$\mathbf{C}_h^{(N)} = \mathbf{F}^{(N)} \breve{\mathbf{\Lambda}}_h^{(N)} \left(\mathbf{F}^{(N)}\right)^H \tag{121}$$

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

$$\left[\mathbf{F}^{(N)}\right]_{k,l} = \frac{1}{\sqrt{N}} e^{j2\pi \frac{(k-1)(l-1)}{N}}. \tag{122}$$

Furthermore, the matrix $\breve{\mathbf{\Lambda}}_h^{(N)}$ is diagonal with the elements $\breve{\lambda}_k^{(N)}$ given in (120).

By this construction the circulant matrix $\mathbf{C}_h^{(N)}$ is asymptotically equivalent to the Toeplitz matrix $\mathbf{R}_h^{(N)}$, see [26, 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 [26, 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.,

$$\lim_{N\to\infty} \left| \mathbf{R}_h^{(N)} - \mathbf{C}_h^{(N)} \right| = 0 \tag{123}$$

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

$$|\mathbf{B}| = \left(\frac{1}{N} \mathrm{Tr}\left[\mathbf{B}^H \mathbf{B}\right] \right)^{\frac{1}{2}}. \tag{124}$$

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 (115), i.e.,

$$\mathbf{K}_1^{(N)} = \frac{1}{\sigma_n^2} \mathbf{R}_h^{(N)} \mathrm{E}\left[\mathbf{Z}\right] + \mathbf{I}_N \tag{125}$$

$$\mathbf{K}_2^{(N)} = \frac{1}{\sigma_n^2} \mathbf{C}_h^{(N)} \mathrm{E}\left[\mathbf{Z}\right] + \mathbf{I}_N \tag{126}$$

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 $N \to \infty$ [26, Section 2.3].

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

$$\left\| \mathbf{K}_1^{(N)} \right\| = \left\| \frac{1}{\sigma_n^2} \mathbf{R}_h^{(N)} \mathrm{E}\left[\mathbf{Z}\right] + \mathbf{I}_N \right\| < \infty \tag{127}$$

$$\left\| \mathbf{K}_2^{(N)} \right\| = \left\| \frac{1}{\sigma_n^2} \mathbf{C}_h^{(N)} \mathrm{E}\left[\mathbf{Z}\right] + \mathbf{I}_N \right\| < \infty \tag{128}$$

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

$$\|\mathbf{B}\|^2 = \max_k \gamma_k \tag{129}$$

where γ_k are the eigenvalues of the Hermitian nonnegative definite matrix $\mathbf{B}\mathbf{B}^H$. The diagonal matrix $\mathrm{E}\left[\mathbf{Z}\right]$ 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 [26, 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

$$\begin{aligned} \left| \mathbf{K}_1^{(N)} - \mathbf{K}_2^{(N)} \right| &= \left| \frac{1}{\sigma_n^2} \mathbf{R}_h^{(N)} \mathrm{E}\left[\mathbf{Z}\right] + \mathbf{I}_N - \frac{1}{\sigma_n^2} \mathbf{C}_h^{(N)} \mathrm{E}\left[\mathbf{Z}\right] - \mathbf{I}_N \right| \\ &= \left| \frac{1}{\sigma_n^2} \left(\mathbf{R}_h^{(N)} - \mathbf{C}_h^{(N)} \right) \mathrm{E}\left[\mathbf{Z}\right] \right| \\ &\overset{(a)}{\leq} \frac{1}{\sigma_n^2} \left| \mathbf{R}_h^{(N)} - \mathbf{C}_h^{(N)} \right| \left\| \mathrm{E}\left[\mathbf{Z}\right] \right\| \end{aligned} \tag{130}$$

where for (a) we have used [26, Lemma 2.3]. As $\|\mathrm{E}\left[\mathbf{Z}\right]\|$ is bounded, we get for $N \to \infty$

$$\begin{aligned} \lim_{N\to\infty} \left| \mathbf{K}_1^{(N)} - \mathbf{K}_2^{(N)} \right| &\leq \lim_{N\to\infty} \frac{1}{\sigma_n^2} \left| \mathbf{R}_h^{(N)} - \mathbf{C}_h^{(N)} \right| \left\| \mathrm{E}\left[\mathbf{Z}\right] \right\| \\ &= 0 \end{aligned} \tag{131}$$

due to (123). 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 [26, Theorem 2.4] the equality in (115) holds. For ease of notation, in the following we omit the use of the superscript (N) for all matrices and eigenvalues.

Based on (115) the term in (114) can be expressed by

$$\begin{aligned} &\lim_{N\to\infty} \frac{1}{N} \log\det\left(\frac{1}{\sigma_n^2} \mathbf{C}_h \mathrm{E}\left[\mathbf{Z}\right] + \mathbf{I}_N \right) \\ &\quad \overset{(a)}{=} \lim_{N\to\infty} \frac{1}{N} \log\det\left(\frac{1}{\sigma_n^2} \mathbf{F} \breve{\mathbf{\Lambda}}_h \mathbf{F}^H \mathrm{E}\left[\mathbf{Z}\right] + \mathbf{I}_N \right) \\ &\quad \overset{(b)}{=} \lim_{N\to\infty} \frac{1}{N} \log\det\left(\frac{1}{\sigma_n^2} \breve{\mathbf{\Lambda}}_h \mathbf{F}^H \mathrm{E}\left[\mathbf{Z}\right] \mathbf{F} + \mathbf{I}_N \right) \end{aligned} \tag{132}$$

where for (a) we have used (121) and (b) is based on the following relation

$$\det\left(\mathbf{A}\mathbf{B} + \mathbf{I}\right) = \det\left(\mathbf{B}\mathbf{A} + \mathbf{I}\right) \tag{133}$$

which holds as $\mathbf{A}\mathbf{B}$ has the same eigenvalues as $\mathbf{B}\mathbf{A}$ for $\mathbf{A}$ and $\mathbf{B}$ being square matrices [32, Theorem 1.3.20].

As the matrix $\frac{1}{\sigma_n^2} \breve{\mathbf{\Lambda}}_h \mathbf{F}^H \mathrm{E}\left[\mathbf{Z}\right] \mathbf{F} + \mathbf{I}_N$ in the argument of the logarithm on the RHS of (132) is positive definite, using Hadamard's inequality we can upper-bound the argument of the limit on the RHS of (132) as follows

$$\begin{aligned} &\frac{1}{N} \log\det\left(\frac{1}{\sigma_n^2} \breve{\mathbf{\Lambda}}_h \mathbf{F}^H \mathrm{E}\left[\mathbf{Z}\right] \mathbf{F} + \mathbf{I}_N \right) \\ &\quad \leq \frac{1}{N} \sum_{k=1}^{N} \log\left(\frac{1}{\sigma_n^2} \breve{\lambda}_k \left[\mathbf{F}^H \mathrm{E}\left[\mathbf{Z}\right] \mathbf{F}\right]_{k,k} + 1 \right) \end{aligned} \tag{134}$$

where $\left[\mathbf{F}^H \mathrm{E}\left[\mathbf{Z}\right] \mathbf{F}\right]_{k,k}$ are the diagonal entries of the matrix $\mathbf{F}^H \mathrm{E}\left[\mathbf{Z}\right] \mathbf{F}$. Note, this means that for distributions of the input sequences $\mathbf{z}$ which lead to the case that the matrix $\mathbf{F}^H \mathrm{E}\left[\mathbf{Z}\right] \mathbf{F}$ is diagonal (134) is fulfilled with equality. Using (134), the

RHS of (132) is upper-bounded by

$$\begin{aligned}
&\lim_{N\to\infty}\frac{1}{N}\log\det\left(\frac{1}{\sigma_n^2}\breve{\mathbf{\Lambda}}_h\mathbf{F}^H\mathrm{E}\left[\mathbf{Z}\right]\mathbf{F}+\mathbf{I}_N\right)\\
&\leq\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}\log\left(\frac{1}{\sigma_n^2}\breve{\lambda}_k\left(\frac{1}{N}\sum_{l=1}^{N}\mathrm{E}\left[z_l\right]\right)+1\right)\\
&=\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}\log\left(\frac{1}{\sigma_n^2}\breve{\lambda}_k\left(\mathrm{E}\left[\frac{1}{N}\sum_{l=1}^{N}z_l\right]\right)+1\right). \quad (135)
\end{aligned}$$

As the logarithm is a monotonically increasing function, with the maximum average power constraint in (107) the RHS of (135) is upper-bounded by

$$\begin{aligned}
&\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}\log\left(\frac{1}{\sigma_n^2}\breve{\lambda}_k\left(\mathrm{E}\left[\frac{1}{N}\sum_{l=1}^{N}z_l\right]\right)+1\right)\\
&\quad\leq\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N}\log\left(\frac{\sigma_x^2}{\sigma_n^2}\breve{\lambda}_k+1\right)\\
&\quad\overset{(a)}{=}\lim_{N\to\infty}\frac{1}{N}\log\det\left(\frac{\sigma_x^2}{\sigma_n^2}\mathbf{C}_h+\mathbf{I}_N\right)\\
&\quad\overset{(b)}{=}\lim_{N\to\infty}\frac{1}{N}\log\det\left(\frac{\sigma_x^2}{\sigma_n^2}\mathbf{R}_h+\mathbf{I}_N\right) \quad (136)
\end{aligned}$$

where (a) is based on (121) 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 (115), (132), (135), and (136) the term in (114) is upper-bounded by

$$\begin{aligned}
&\lim_{N\to\infty}\frac{1}{N}\log\det\left(\frac{1}{\sigma_n^2}\mathbf{R}_h\mathrm{E}\left[\mathbf{Z}\right]+\mathbf{I}_N\right)\\
&\quad\leq\lim_{N\to\infty}\frac{1}{N}\log\det\left(\frac{\sigma_x^2}{\sigma_n^2}\mathbf{R}_h+\mathbf{I}_N\right). \quad (137)
\end{aligned}$$

However, this means that the entropy rate $h'(\mathbf{e}_{\text{joint}}|\mathbf{x}_D,\mathbf{x}_P)$ in (113) is lower-bounded by

$$\begin{aligned}
&h'(\mathbf{e}_{\text{joint}}|\mathbf{x}_D,\mathbf{x}_P)\\
&\quad\geq\lim_{N\to\infty}\frac{1}{N}\Bigg[\log\det\left((\pi e)^N\det(\mathbf{R}_h)\right)\\
&\qquad\qquad-\log\det\left(\frac{1}{\sigma_n^2}\mathbf{R}_h\mathrm{E}\left[\mathbf{Z}\right]+\mathbf{I}_N\right)\Bigg]\\
&\quad\geq\lim_{N\to\infty}\frac{1}{N}\Bigg[\log\det\left((\pi e)^N\det(\mathbf{R}_h)\right)\\
&\qquad\qquad-\log\det\left(\frac{\sigma_x^2}{\sigma_n^2}\mathbf{R}_h+\mathbf{I}_N\right)\Bigg]\\
&\quad\overset{(a)}{=}\lim_{N\to\infty}\frac{1}{N}\log\det\left(\pi e\mathbf{R}_{e_{\text{joint,CM}}}\right) \quad (138)
\end{aligned}$$

where for (a) we have used similar steps as in (105) and where $\mathbf{R}_{e_{\text{joint,CM}}}$ is the estimation error correlation matrix in case all input symbols have a constant modulus with power σ_x^2, cf. (104)

$$\mathbf{R}_{e_{\text{joint,CM}}}=\mathbf{R}_h-\mathbf{R}_h\left(\mathbf{R}_h+\frac{\sigma_n^2}{\sigma_x^2}\mathbf{I}_N\right)^{-1}\mathbf{R}_h. \quad (139)$$

This means, that the entropy rate $h'(\mathbf{e}_{\text{joint}}|\mathbf{x}_D,\mathbf{x}_P)$ 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. Furthermore, note that for constant modulus input distributions with power σ_x^2 the inequalities in (113), (135), and (136) hold with equality.

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

$$\begin{aligned}
h'(\mathbf{e}_{\text{joint}}|\mathbf{x}_D,\mathbf{x}_P)&\geq h'(\mathbf{e}_{\text{joint}}|\mathbf{x}_D,\mathbf{x}_P)\big|_{\text{CM}}\\
&=\lim_{N\to\infty}\frac{1}{N}\log\det\left(\pi e\mathbf{R}_{e_{\text{joint,CM}}}\right). \quad (140)
\end{aligned}$$

Appendix B
Application of Szegö's Theorem on the Asymptotic Eigenvalue Distribution in (34)

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

$$\begin{aligned}
&\lim_{N\to\infty}\frac{1}{N}\left\{\log\det\left(\mathbf{R}_{e_{\text{pil}}}\right)-\log\det\left(\mathbf{R}_{e_{\text{joint,CM}}}\right)\right\}\\
&\overset{(a)}{=}\lim_{N\to\infty}\frac{1}{N}\left\{\log\det\left(\mathbf{R}_h^{-1}\mathbf{R}_{e_{\text{pil}}}\right)-\log\det\left(\mathbf{R}_h^{-1}\mathbf{R}_{e_{\text{joint,CM}}}\right)\right\}\\
&\overset{(b)}{=}\lim_{N\to\infty}\frac{1}{N}\left\{-\log\det\left(\frac{\sigma_x^2}{L\sigma_n^2}\mathbf{R}_h+\mathbf{I}_N\right)\right\}\\
&\qquad\qquad+\lim_{N\to\infty}\frac{1}{N}\log\det\left(\frac{\sigma_x^2}{\sigma_n^2}\mathbf{R}_h+\mathbf{I}_N\right)\\
&\overset{(c)}{=}-\int_{-\frac{1}{2}}^{\frac{1}{2}}\log\left(\frac{\sigma_x^2}{L\sigma_n^2}S_h(f)+1\right)df+\int_{-\frac{1}{2}}^{\frac{1}{2}}\log\left(\frac{\sigma_x^2}{\sigma_n^2}S_h(f)+1\right)df\\
&=\int_{-\frac{1}{2}}^{\frac{1}{2}}\log\left(\frac{\rho\frac{S_h(f)}{\sigma_n^2}+1}{\frac{\rho}{L}\frac{S_h(f)}{\sigma_n^2}+1}\right)df\\
&\overset{(d)}{=}\int_{-\frac{1}{2}}^{\frac{1}{2}}\log\left(\frac{S_{e_{\text{pil}}}(f)}{S_{e_{\text{joint,CM}}}(f)}\right)df \quad (141)
\end{aligned}$$

where the premultiplication of both terms by $\mathbf{R}_h^{-1}$ in (a) assures that the limits of the individual logarithms exist. Furthermore, the second term in (b) follows with (139) and the first term in (b) can be shown analogously by using some additional mathematical steps, which we have omitted here, to take the lower sampling rate of the channel due to the pilot symbols into account. Moreover, (c) follows from Szegö's theorem on the asymptotic eigenvalue distribution of Hermitian Toeplitz matrices [28], [26, Theorem 4.2]. Finally, in (d) we express the term based on the spectra of the channel estimation error processes $S_{e_{\text{pil}}}(f)$ and $S_{e_{\text{joint,CM}}}(f)$ given in (35) and (36), proving (34) (a). We have thus also shown that the limit in the first line of (141) exists. Therefore, the use of the differential entropy rates $h'(\mathbf{e}_{\text{joint}}|\mathbf{x}_D,\mathbf{x}_P)$ and $h'(\mathbf{e}_{\text{pil}}|\mathbf{x}_P)$ in the derivation of the lower bound on the achievable rate is justified.

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

First, we calculate the PSD $S_{e_{\text{pil}}}(f)$ 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

$$E_N(e^{j2\pi f}) = \sum_{k=1}^{N} e_{\text{pil},k} \cdot e^{-j2\pi fk} \tag{142}$$

where $e_{\text{pil},k}$ are the elements of the vector $\mathbf{e}_{\text{pil}}$. In the following we are interested in the case $N \to \infty$. As in this case the sum in (142) 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

$$\begin{aligned}
\lim_{N\to\infty} \frac{1}{N} E_N(e^{j2\pi f}) &\overset{(a)}{=} \lim_{N\to\infty} \frac{1}{N} \sum_{l=1}^{L} E_{N,l}(e^{j2\pi Lf}) e^{-j2\pi lf} \\
&\overset{(b)}{=} \lim_{N\to\infty} \frac{1}{N} \sum_{l=1}^{L} \Bigg[H_{N,l}(e^{j2\pi Lf}) \\
&\qquad - W_l(e^{j2\pi Lf}) \frac{Y_{N,P}(e^{j2\pi Lf})}{\sigma_x} \Bigg] e^{-j2\pi lf} \\
&\overset{(c)}{=} \lim_{N\to\infty} \frac{1}{N} \Bigg[H_N(e^{j2\pi f}) \\
&\qquad - \sum_{l=1}^{L} W(e^{j2\pi Lf}) e^{j2\pi lf} \frac{Y_{N,P}(e^{j2\pi Lf})}{\sigma_x} e^{-j2\pi lf} \Bigg] \\
&= \lim_{N\to\infty} \frac{1}{N} \left[H_N(e^{j2\pi f}) - L \cdot W(e^{j2\pi Lf}) \frac{Y_{N,P}(e^{j2\pi Lf})}{\sigma_x} \right] \\
&\overset{(d)}{=} \lim_{N\to\infty} \frac{1}{N} \Bigg[H_N(e^{j2\pi f}) \\
&\qquad - L \cdot W(e^{j2\pi Lf}) \left[H_{N,P}(e^{j2\pi Lf}) + \frac{N_{N,P}(e^{j2\pi Lf})}{\sigma_x} \right] \Bigg].
\end{aligned} \tag{143}$$

For (a) we have used that the estimation error in frequency domain is the sum of the interpolation errors at the individual symbol time instances between and at 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 symbol, i.e.,

$$E_{N,l}(e^{j2\pi Lf}) = \sum_{k=1}^{\frac{N}{L}} e_{\text{pil},(k-1)L+1+l} \cdot e^{-j2\pi fkL}, \qquad \text{for } l = 0, \ldots, L-1 \tag{144}$$

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 symbol time instances 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

$$W_l(e^{j2\pi Lf}) = W(e^{j2\pi Lf}) e^{j2\pi lf} \tag{145}$$

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

$$\begin{aligned}
S_{e_{\text{pil}}}(f) &= \lim_{N\to\infty} \frac{1}{N} \mathrm{E}\left[|E_N(e^{j2\pi f})|^2 \right] \\
&= \lim_{N\to\infty} \frac{1}{N} \mathrm{E}\Bigg[|H_N(e^{j2\pi f})|^2 \\
&\quad - L \cdot H_N(e^{j2\pi f}) W^*(e^{j2\pi Lf}) H^*_{N,P}(e^{j2\pi Lf}) \\
&\quad - L \cdot H^*_N(e^{j2\pi f}) W(e^{j2\pi Lf}) H_{N,P}(e^{j2\pi Lf}) \\
&\quad + L^2 |W(e^{j2\pi Lf})|^2 \left[|H_{N,P}(e^{j2\pi Lf})|^2 + \left| \frac{N_{N,P}(e^{j2\pi Lf})}{\sigma_x} \right|^2 \right] \Bigg] \\
&= S_h(e^{j2\pi f}) - \lim_{N\to\infty} \frac{1}{N} \mathrm{E}\Bigg[L \cdot W^*(e^{j2\pi Lf}) \\
&\quad \times \sum_{l=1}^{L} H_{N,l}(e^{j2\pi Lf}) e^{-j2\pi lf} H^*_{N,P}(e^{j2\pi Lf}) \\
&\quad + L \cdot W(e^{j2\pi Lf}) \sum_{l=1}^{L} H^*_{N,l}(e^{j2\pi Lf}) e^{j2\pi lf} H_{N,P}(e^{j2\pi Lf}) \Bigg] \\
&\quad + L^2 |W(e^{j2\pi Lf})|^2 \left[\frac{1}{L} S_h(e^{j2\pi Lf}) + \frac{1}{L} \frac{\sigma_n^2}{\sigma_x^2} \right] \\
&= S_h(e^{j2\pi f}) - L \cdot W^*(e^{j2\pi Lf}) \sum_{l=1}^{L} \frac{1}{L} \cdot S_h(e^{j2\pi Lf}) \\
&\quad - L \cdot W(e^{j2\pi Lf}) \sum_{l=1}^{L} \frac{1}{L} \cdot S^*_h(e^{j2\pi Lf}) \\
&\quad + L^2 |W(e^{j2\pi Lf})|^2 \left[\frac{1}{L} S_h(e^{j2\pi Lf}) + \frac{1}{L} \frac{\sigma_n^2}{\sigma_x^2} \right] \\
&\overset{(a)}{=} S_h(e^{j2\pi f}) - 2L \cdot W(e^{j2\pi Lf}) S_h(e^{j2\pi Lf}) \\
&\quad + L |W(e^{j2\pi Lf})|^2 \left[S_h(e^{j2\pi Lf}) + \frac{\sigma_n^2}{\sigma_x^2} \right]
\end{aligned} \tag{146}$$

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

$$W(e^{j2\pi Lf}) = \frac{S_h(e^{j2\pi Lf})}{S_h(e^{j2\pi Lf}) + \frac{\sigma_n^2}{\sigma_x^2}} = \frac{\frac{1}{L} S_h(e^{j2\pi f})}{\frac{1}{L} S_h(e^{j2\pi f}) + \frac{\sigma_n^2}{\sigma_x^2}} \tag{147}$$

where we have used that

$$S_h(e^{j2\pi Lf}) = \frac{1}{L} S_h(e^{j2\pi f}). \tag{148}$$

Inserting (147) into (146) yields

$$\begin{aligned}
S_{e_{\text{pil}}}(f) &= S_h(e^{j2\pi f}) - 2L \frac{S_h(e^{j2\pi f})}{S_h(e^{j2\pi f}) + L\frac{\sigma_n^2}{\sigma_x^2}} S_h(e^{j2\pi Lf}) \\
&\quad + L \left| \frac{S_h(e^{j2\pi f})}{S_h(e^{j2\pi f}) + L\frac{\sigma_n^2}{\sigma_x^2}} \right|^2 \left[S_h(e^{j2\pi Lf}) + \frac{\sigma_n^2}{\sigma_x^2} \right] \\
&\overset{(a)}{=} S_h(e^{j2\pi f}) - 2 \cdot \frac{S_h(e^{j2\pi f})}{S_h(e^{j2\pi f}) + L\frac{\sigma_n^2}{\sigma_x^2}} S_h(e^{j2\pi f}) \\
&\quad + \frac{\left| S_h(e^{j2\pi f}) \right|^2}{S_h(e^{j2\pi f}) + L\frac{\sigma_n^2}{\sigma_x^2}} \\
&= \frac{S_h(e^{j2\pi f}) L\frac{\sigma_n^2}{\sigma_x^2}}{S_h(e^{j2\pi f}) + L\frac{\sigma_n^2}{\sigma_x^2}} \\
&= \frac{S_h(e^{j2\pi f})}{\frac{\rho}{L}\frac{S_h(e^{j2\pi f})}{\sigma_h^2} + 1} \\
&\overset{(b)}{=} \frac{S_h(f)}{\frac{\rho}{L}\frac{S_h(f)}{\sigma_h^2} + 1}
\end{aligned} \tag{149}$$

where (a) results from (148) 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 (149), i.e.,

$$S_{e_{\text{joint,CM}}}(f) = \frac{S_h(f)}{\rho\frac{S_h(f)}{\sigma_h^2} + 1} \tag{150}$$

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

ACKNOWLEDGMENT

The authors would like to thank Giuseppe Durisi for fruitful discussions.

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

[7] 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.

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

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

[10] J. Chen and V. V. Veeravalli, "Capacity results for block-stationary Gaussian fading channels with a peak power constraint," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4498–4520, Dec. 2007.

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

[12] X. Deng and A. M. Haimovich, "Achievable rates over time-varying Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 55, no. 7, pp. 1397–1406, Jul. 2007.

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

[14] M. Dörpinghaus, H. Meyr, and G. Ascheid, "The achievable rate of stationary Rayleigh flat-fading channels with IID input symbols," in *Proc. Int. Symp. Inf. Theory and its Applications (ISITA)*, Taichung, Taiwan, Oct. 2010, pp. 812–817.

[15] 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.

[16] S. Furrer and D. Dahlhaus, "Multiple-antenna signaling over fading channels with estimated channel state information: Capacity analysis," *IEEE Trans. Inf. Theory*, vol. 53, no. 6, pp. 2028–2043, Jun. 2007.

[17] 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.

[18] 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.

[19] 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.

[20] ——, "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.

[21] N. Jindal and A. Lozano, "A unified treatment of optimum pilot overhead in multipath fading channels," *IEEE Trans. Commun.*, vol. 58, no. 10, pp. 2939–2948, Oct. 2010.

[22] A. Lozano, "Interplay of spectral efficiency, power and Doppler spectrum for reference-signal-assisted wireless communication," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 5020–5029, Dec. 2008.

[23] M. Dong, L. Tong, and B. Sadler, "Optimal insertion of pilot symbols for transmissions over time-varying flat fading channels," *IEEE Trans. on Signal Process.*, vol. 52, no. 5, pp. 1403–1418, May 2004.

[24] G. Taricco and E. Biglieri, "Space-time decoding with imperfect channel estimation," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1874–1888, Jul. 2005.

[25] 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.

[26] 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.

[27] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2001.

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

[29] 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.

[30] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. on Telecommunications*, vol. 10, no. 6, pp. 585–595, Nov./Dec. 1999.

[31] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1991.

[32] ——, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

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

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

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

Adrian Ispas (S'11) obtained the Dipl.-Ing. degree in electrical engineering and information technology (with distinction) from RWTH Aachen University, Germany, in 2007. He is currently working as a research assistant and pursuing a doctorate degree at the Chair for Integrated Signal Processing Systems, RWTH Aachen University. His main research interests lie in the areas of channel modeling and characterization and statistical signal processing.

Adrian Ispas received the Henry Ford II Award 2005 for academic achievements during his studies and the SEW-EURODRIVE-Stiftung Award 2008 in recognition of his diploma thesis.

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

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

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