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

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

***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 discuss the nature of 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. In addition, we derive a new lower bound on the achievable rate for joint processing of pilot and data symbols.**

I. Introduction

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

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], [2]. 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 receivers. To evaluate the payoff for the increased receiver complexity with joint processing, 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.

Therefore, in the present work we will evaluate the performance of a joint processing in comparison to synchronized detection with a solely pilot based channel estimation based on the achievable rate. Regarding the channel statistics we assume a stationary Rayleigh flat-fading channel as it is usually applied to model the fading in a mobile environment without a line of sight component. Furthermore, we assume that the power spectral density (PSD) of the channel fading process is compactly supported, and that the fading process is *non-regular* [3]. Moreover, 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 fading process.

For the case of synchronized detection with a solely pilot based channel estimation there exist already bounds on the achievable rate [4]. In contrast, for the case of joint processing there is not much knowledge on the achievable rate. Very recently, in [5] 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 give a lower bound on the achievable rate with joint processing of pilot and data symbols. Besides this lower bound on the achievable rate with a joint processing of pilot and data symbols, we identify the nature of the possible gain of a joint processing in comparison to a separate processing.

II. System Model

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

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

with the diagonal matrix $\mathbf{X} = \text{diag}(\mathbf{x})$. Here the 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, $\mathbf{x}$, $\mathbf{n}$, and $\mathbf{h}$ contain the channel input symbols, the additive noise samples, and the channel fading weights. All vectors are of length N.

The samples of the additive noise process are i.i.d. zero-mean jointly proper Gaussian with variance σ_n^2.

The channel fading process is zero-mean jointly proper Gaussian with the temporal correlation characterized by $r_h(l) = \text{E}[h_{k+l}h_k^*]$. Its variance is given by $r_h(0) = \sigma_h^2$, and, due to technical reasons, it is assumed to be absolutely summable, i.e., $\sum_{l=-\infty}^{\infty} |r_h(l)| < \infty$. The PSD of the channel fading process is defined as

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

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

The transmit symbol sequence consists of data symbols with an average power σ_x^2 and periodically inserted pilot symbols with a fixed power σ_x^2. Each L-th symbol is a pilot symbol. The pilot spacing is chosen such that the channel fading process is sampled at least with Nyquist rate, i.e.,

$$L < 1/(2f_d). \quad (3)$$

In the following we use the subvectors $\mathbf{x}_D$ containing all data symbols of $\mathbf{x}$ and $\mathbf{x}_P$ containing all pilot symbols of $\mathbf{x}$. Correspondingly, we define $\mathbf{h}_D$, $\mathbf{h}_P$, $\mathbf{y}_D$, $\mathbf{y}_P$, $\mathbf{n}_D$, and $\mathbf{n}_P$.

The processes $\{x_k\}$, $\{h_k\}$ and $\{n_k\}$ are assumed to be mutually independent. The mean SNR is given by $\rho = \sigma_x^2\sigma_h^2/\sigma_n^2$.

III. The Nature of the Gain by Joint Processing of Data and Pilot Symbols

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

$$\mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{y}_P,\mathbf{x}_P) \overset{(a)}{=} \mathcal{I}(\mathbf{x}_D;\mathbf{y}_D|\mathbf{y}_P,\mathbf{x}_P) + \mathcal{I}(\mathbf{x}_D;\mathbf{y}_P|\mathbf{x}_P) + \mathcal{I}(\mathbf{x}_D;\mathbf{x}_P) \overset{(b)}{=} \mathcal{I}(\mathbf{x}_D;\mathbf{y}_D|\mathbf{y}_P,\mathbf{x}_P) \quad (4)$$

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.,

$$\hat{\mathbf{x}}_D = \arg\max_{\mathbf{x}_D\in\mathcal{C}_D} p(\mathbf{y}|\mathbf{x}) = \arg\max_{\mathbf{x}_D\in\mathcal{C}_D} p(\mathbf{y}_D|\mathbf{x}_D,\mathbf{y}_P,\mathbf{x}_P) \quad (5)$$

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

$$\mathrm{E}[\mathbf{y}_D|\mathbf{x}_D,\mathbf{y}_P,\mathbf{x}_P] = \mathbf{X}_D\mathrm{E}[\mathbf{h}_D|\mathbf{y}_P,\mathbf{x}_P] = \mathbf{X}_D\hat{\mathbf{h}}_{\text{pil},D} \quad (6)$$

$$\mathrm{cov}[\mathbf{y}_D|\mathbf{x}_D,\mathbf{y}_P,\mathbf{x}_P] = \mathbf{X}_D\mathbf{R}_{e_{\text{pil}},D}\mathbf{X}_D^H + \sigma_n^2\mathbf{I}_{N_D} \quad (7)$$

where $\mathbf{X}_D = \mathrm{diag}(\mathbf{x}_D)$ and $\mathbf{I}_{N_D}$ is an identity matrix of size $N_D \times N_D$ with N_D being the length of $\mathbf{n}_D$. The vector $\hat{\mathbf{h}}_{\text{pil},D}$ is an MMSE channel estimate at the data symbol time instances based on the pilot symbols, which is denoted by the index pil. Furthermore, the corresponding channel estimation error $\mathbf{e}_{\text{pil},D} = \mathbf{h}_D - \hat{\mathbf{h}}_D$ is zero-mean proper Gaussian and $\mathbf{R}_{e_{\text{pil}},D} = \mathrm{E}\left[\mathbf{e}_{\text{pil},D}\mathbf{e}_{\text{pil},D}^H|\mathbf{x}_P\right]$ is its correlation matrix, which is independent of $\mathbf{y}_P$ due to the principle of orthogonality.

Based on (6) and (7) 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.,

$$p(\mathbf{y}_D|\mathbf{x}_D,\mathbf{y}_P,\mathbf{x}_P) = p(\mathbf{y}_D|\mathbf{x}_D,\hat{\mathbf{h}}_{\text{pil},D},\mathbf{x}_P) \quad (8)$$

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, (5) can be written as

$$\hat{\mathbf{x}}_D = \arg\max_{\mathbf{x}_D\in\mathcal{C}_D} p(\mathbf{y}_D|\mathbf{x}_D,\hat{\mathbf{h}}_{\text{pil},D},\mathbf{x}_P) = \arg\max_{\mathbf{x}_D\in\mathcal{C}_D} p(\mathbf{y}|\mathbf{x}_D,\hat{\mathbf{h}}_{\text{pil}},\mathbf{x}_P). \quad (9)$$

For ease of notation in the following we will use the metric on the RHS of (9) 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., $\hat{\mathbf{h}}_{\text{pil}} = \mathrm{E}[\mathbf{h}|\mathbf{y}_P,\mathbf{x}_P]$. Based on $\hat{\mathbf{h}}_{\text{pil}}$, (1) can be expressed by

$$\mathbf{y} = \mathbf{X}(\hat{\mathbf{h}}_{\text{pil}} + \mathbf{e}_{\text{pil}}) + \mathbf{n} \quad (10)$$

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.,

$$\mathbf{R}_{e_{\text{pil}}} = \mathrm{E}\left[\mathbf{e}_{\text{pil}}\mathbf{e}_{\text{pil}}^H|\mathbf{x}_P\right] \quad (11)$$

is not diagonal, cf. (21). Thus, the PDF in (9) is given by

$$p(\mathbf{y}|\mathbf{x}_D,\hat{\mathbf{h}}_{\text{pil}},\mathbf{x}_P) = \mathcal{CN}\left(\mathbf{X}\hat{\mathbf{h}}_{\text{pil}}, \mathbf{X}\mathbf{R}_{e_{\text{pil}}}\mathbf{X}^H + \sigma_n^2\mathbf{I}_N\right) \quad (12)$$

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

Note that corresponding to (8), we can also rewrite (4) as

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

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

However, typical channel decoders like a Viterbi decoder are not able to exploit the temporal correlation of the channel estimation error. Therefore, the decoder performs mismatch 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

$$p(\mathbf{y}|\mathbf{x}_D,\hat{\mathbf{h}}_{\text{pil}},\mathbf{x}_P) \approx \mathcal{CN}\left(\mathbf{X}\hat{\mathbf{h}}_{\text{pil}}, \sigma_{e_{\text{pil}}}^2\mathbf{X}\mathbf{X}^H + \sigma_n^2\mathbf{I}_N\right). \quad (13)$$

As it is assumed that the channel is at least sampled with Nyquist frequency, see (3), 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 [4] and is given by

$$\sigma_{e_{\text{pil}}}^2 = \int_{f=-\frac{1}{2}}^{\frac{1}{2}} S_{e_{\text{pil}}}(f)df = \int_{f=-\frac{1}{2}}^{\frac{1}{2}} \frac{S_h(f)}{\frac{\rho}{L}\frac{S_h(f)}{\sigma_h^2}+1}df \quad (14)$$

where the PSD of the channel estimation error process $S_{e_{\text{pil}}}(f)$ is given in (21). Hence, the variance of the channel estimation process, i.e., of 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.

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

[1] Note that for the case of data transmission only (12) 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 $\mathbf{R}_{e_{\text{pil}}} = \mathbf{R}_h$.

the mutual information for each individual data symbol time instant. As, obviously, by this separate processing information is discarded, the following inequality holds

$$\lim_{N\to\infty}\frac{\mathcal{I}(\mathbf{x}_D;\mathbf{y}_D|\hat{\mathbf{h}}_{\text{pil}})}{N}=\mathcal{I}'(\mathbf{x}_D;\mathbf{y}_D|\hat{\mathbf{h}}_{\text{pil}})\geq\frac{L-1}{L}\mathcal{I}(x_{D_k};y_{D_k}|\hat{\mathbf{h}}_{\text{pil}}) \quad (15)$$

where $\mathcal{I}'$ denotes the mutual information rate and the index D_k denotes an arbitrarily chosen data symbol.

As the LHS of (15) is the mutual information of the channel and the RHS of (15) is the mutual information achievable with synchronized detection with a metric corresponding to (13) and a solely pilot based channel estimation, i.e., a separate processing, the difference of both terms upper-bounds the possible gain due to joint processing of data and pilot symbols. The additional information that can be gained by a joint processing in contrast to separate processing is contained in the temporal correlation of the channel estimation error process.

Regarding synchronized detection in combination with a solely pilot based channel estimation, i.e., separate processing, in [4] bounds on the achievable rate, i.e., on the RHS of (15), are given. In Fig. 1 these bounds are shown for i.i.d. zero-mean proper Gaussian data-symbols. These bounds show that the achievable rate with separate processing is decreased in comparison to perfect channel knowledge in two ways. First, time instances used for pilot symbols are lost for data symbols, and secondly, the average SNR is decreased due to the channel estimation error variance.

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 [5] for the case of a block-fading channel to the stationary Rayleigh flat-fading scenario discussed in the present work. Therefore, analogous to [5] we decompose and lower-bound the mutual information between the transmitter and the receiver as follows

$$\begin{aligned}\mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{y}_P,\mathbf{x}_P)&\overset{(a)}{=}\mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{y}_P,\mathbf{x}_P,\mathbf{h})-\mathcal{I}(\mathbf{x}_D;\mathbf{h}|\mathbf{y}_D,\mathbf{y}_P,\mathbf{x}_P)\\&=\mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{h})-h(\mathbf{h}|\mathbf{y}_D,\mathbf{y}_P,\mathbf{x}_P)+h(\mathbf{h}|\mathbf{x}_D,\mathbf{y}_D,\mathbf{y}_P,\mathbf{x}_P)\\&\overset{(b)}{\geq}\mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{h})-h(\mathbf{h}|\mathbf{y}_P,\mathbf{x}_P)+h(\mathbf{h}|\mathbf{x}_D,\mathbf{y}_D,\mathbf{y}_P,\mathbf{x}_P) \quad (16)\end{aligned}$$

where (a) follows from the chain rule for mutual information and (b) is due to the fact that conditioning reduces entropy. The first term on the RHS of (16) is the mutual information in case of perfect channel knowledge.

Now we deviate from [5] and rewrite the RHS of (16) as

$$\begin{aligned}(16)&\overset{(a)}{=}\mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{h})-h(\mathbf{h}|\hat{\mathbf{h}}_{\text{pil}},\mathbf{x}_P)+h(\mathbf{h}|\hat{\mathbf{h}}_{\text{joint}},\mathbf{x}_D,\mathbf{x}_P)\\&\overset{(b)}{=}\mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{h})-h(\mathbf{e}_{\text{pil}}|\mathbf{x}_P)+h(\mathbf{e}_{\text{joint}}|\mathbf{x}_D,\mathbf{x}_P)\\&\overset{(c)}{=}\mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{h})-\log\det\left(\pi e\mathbf{R}_{e_{\text{pil}}}\right)+\log\det\left(\pi e\mathbf{R}_{e_{\text{joint}}}\right) \quad (17)\end{aligned}$$

where for (a) we have substituted the conditioning 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.,

$$\hat{\mathbf{h}}_{\text{joint}}=\mathrm{E}\left[\mathbf{h}|\mathbf{y}_D,\mathbf{x}_D,\mathbf{y}_P,\mathbf{x}_P\right]. \quad (18)$$

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 the second term in equality (b) we have used (10), 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 $\hat{\mathbf{h}}_{\text{pil}}$. Analogously, for the third term we used the separation of $\mathbf{h}$ into the estimate $\hat{\mathbf{h}}_{\text{joint}}$ and the corresponding estimation error $\mathbf{e}_{\text{joint}}$ which depends on $\mathbf{x}_D$ and $\mathbf{x}_P$ and is independent of $\hat{\mathbf{h}}_{\text{joint}}$. Finally, (c) holds as the estimation error processes are zero-mean jointly proper Gaussian. The error correlation matrices are given by (11) and by

$$\mathbf{R}_{e_{\text{joint}}}=\mathrm{E}\left[\mathbf{e}_{\text{joint}}\mathbf{e}_{\text{joint}}^H|\mathbf{x}_D,\mathbf{x}_P\right]. \quad (19)$$

The estimation error $\mathbf{e}_{\text{joint}}$ depends on the distribution of the data symbols $\mathbf{x}_D$. It can be shown that the differential entropy rate $h'(\mathbf{e}_{\text{joint}}|\mathbf{x}_D,\mathbf{x}_P)=\lim_{N\to\infty}\frac{1}{N}h(\mathbf{e}_{\text{joint}}|\mathbf{x}_D,\mathbf{x}_P)$ is minimized for a given average transmit power σ_x^2 if the data symbols have constant modulus (CM). Due to lack of space the proof given in [6] is not shown here.

Thus, with (16) and (17) a lower bound for the achievable rate with joint processing of data and pilot symbols is given by

$$\begin{aligned}\mathcal{I}'(\mathbf{x}_D;\mathbf{y}_D,\mathbf{y}_P,\mathbf{x}_P)&=\lim_{N\to\infty}\frac{1}{N}\mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{y}_P,\mathbf{x}_P)\\&\geq\lim_{N\to\infty}\frac{1}{N}\left\{\mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{h})-\log\det\left(\mathbf{R}_{e_{\text{pil}}}\right)+\log\det\left(\mathbf{R}_{e_{\text{joint,CM}}}\right)\right\}\\&\overset{(a)}{=}\lim_{N\to\infty}\frac{1}{N}\mathcal{I}(\mathbf{x}_D;\mathbf{y}_D,\mathbf{h})-\int_{-\frac{1}{2}}^{\frac{1}{2}}\log\left(\frac{S_{e_{\text{pil}}}(f)}{S_{e_{\text{joint,CM}}}(f)}\right)df \quad (20)\end{aligned}$$

with $\mathbf{R}_{e_{\text{joint,CM}}}$ corresponding to (19), but under the assumption of CM data symbols with transmit power σ_x^2. Note that the CM assumption has only been used to lower bound the third term at the RHS of (17), and not the whole expression at the RHS of (17). For (a) in (20) we have used Szegö's theorem on the asymptotic eigenvalue distribution of Hermitian Toeplitz matrices [7]. $S_{e_{\text{pil}}}(f)$ 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 [6]

$$S_{e_{\text{pil}}}(f)=\frac{S_h(f)}{\frac{\rho}{L}\frac{S_h(f)}{\sigma_h^2}+1},\quad S_{e_{\text{joint,CM}}}(f)=\frac{S_h(f)}{\rho\frac{S_h(f)}{\sigma_h^2}+1}. \quad (21)$$

The first term on the RHS of (20) 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

$$\mathcal{R}_{L,\text{joint}}=\frac{L-1}{L}C_{\text{perf}}-\int_{-\frac{1}{2}}^{\frac{1}{2}}\log\left(\frac{\frac{\rho}{\sigma_h^2}S_h(f)+1}{\frac{\rho}{L\sigma_h^2}S_h(f)+1}\right)df \quad (22)$$

where C_{perf} corresponds to the coherent capacity with

$$C_{\text{perf}} = \mathrm{E}_h\left[\log\left(1+\rho\frac{|h|^2}{\sigma_h^2}\right)\right] = \int_{z=0}^{\infty}\log\left(1+\rho z\right)e^{-z}dz \quad (23)$$

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

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

Equation (22) is a lower bound on the achievable rate with joint processing of data and pilot symbols, for a given pilot spacing L and stationary Rayleigh flat-fading.

For the special case of a rectangular PSD[2] $S_h(f)$, i.e.,

$$S_h(f) = \begin{cases} \frac{\sigma_h^2}{2f_d} & \text{for } |f| \le f_d \\ 0 & \text{otherwise} \end{cases} \quad (24)$$

the lower bound in (22) becomes

$$\mathcal{R}_{L,\text{joint}}\Big|_{\text{rect.}S_h(f)} = \frac{L-1}{L}\int_{z=0}^{\infty}\log(1+\rho z)e^{-z}dz - 2f_d\log\left(\frac{\rho+2f_d}{\frac{\rho}{L}+2f_d}\right). \quad (25)$$

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

The lower bound in (25) depends on the pilot spacing L and can be enhanced by calculating the supremum of (25) 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 (25) is based on the assumption that the pilot spacing is chosen such that the channel fading process is at least sampled with Nyquist rate, see (3). For larger L the estimation error process is no longer stationary, which is required for our derivation.[3]

For these conditions, the lower bound (25) is maximized for

$$L_{\text{opt}} = \lfloor 1/(2f_d)\rfloor \quad (26)$$

which can be observed based on differentiation of (25) w.r.t. L and numerical evaluation. Note that L_{opt} is not necessarily the L which maximizes the achievable rate.

V. Numerical Evaluation

Fig. 1 shows a comparison of the bounds on the achievable rate for separate and joint processing of data and pilot symbols. On the one hand, the lower bound on the achievable rate for joint processing in (25) is compared to bounds on the achievable rate with separate processing of data and pilot symbols for a fixed pilot spacing, i.e., [4,(22)] and [4,(23)] for zero-mean proper Gaussian data symbols. 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 [4,(22)] is maximized. Except of 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. The observation that the lower bound for joint processing for very large f_d is smaller than the achievable rate with separate processing indicates that the lower bound is not tight for these parameters.

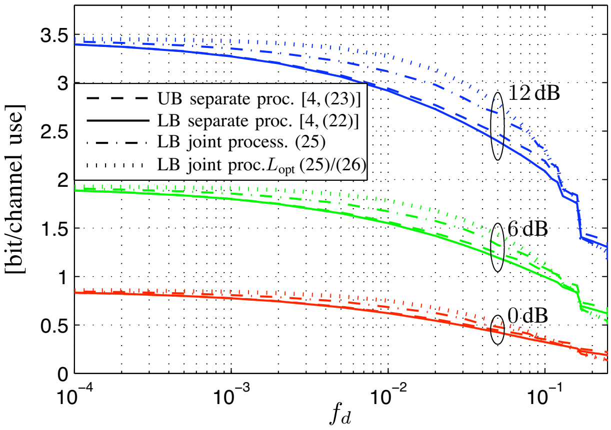

Fig. 1. Comparison of bounds on the achievable rate with separate processing to lower bounds on the achievable rate with joint processing of data and pilot symbols; except of *LB joint proc.* L_{opt}, the pilot spacing L is chosen such that the lower bound for separate processing is maximized; rectangular $S_h(f)$ (24)

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., (25) with (26), is shown. Obviously, this lower bound is larger than or equal to the lower bound for joint processing while choosing the pilot spacing as it is optimal for separate processing of data and pilot symbols. This behavior arises from the effect that for separate processing in case of small f_d a pilot rate is chosen that is higher than the Nyquist rate of the channel fading process to enhance the channel estimation quality. In case of a joint processing all symbols are used for channel estimation anyway. Therefore, a pilot rate higher than Nyquist rate always leads to an increased loss in the achievable rate as less symbols can be used for data transmission.

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

[3]Periodically inserted pilot symbols do not maximize the achievable rate. However, we restrict to periodical pilot symbols with a spacing fulfilling (3), as this enables detection with manageable complexity.

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] L. Schmitt, H. Meyr, and D. Zhang, "Systematic design of iterative ML receivers for flat fading channels," *IEEE Trans. Commun.*, submitted.

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

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

[5] 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.

[6] M. Dörpinghaus, A. Ispas, G. Ascheid, and H. Meyr, "On the gain of joint processing of pilot and data symbols in stationary Rayleigh fading channels," in preparation.

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