Achievable Rate with Receivers using Iterative Channel Estimation in Stationary Fading Channels

Meik Dörpinghaus
Institute for Theoretical Information Technology
RWTH Aachen University, 52056 Aachen, Germany
Doerpinghaus@ti.rwth-aachen.de

Adrian Ispas and Heinrich Meyr
Institute for Integrated Signal Processing Systems
RWTH Aachen University, 52056 Aachen, Germany
{Ispas, Meyr}@iss.rwth-aachen.de

***Abstract*—We study the achievable rate with receivers based on synchronized detection and iterative code-aided channel estimation for stationary Rayleigh flat-fading channels. The main idea behind this type of receivers is that—additionally to the pilot symbols which are used for the initial channel estimation and coherent detection/decoding—the channel estimation is enhanced by iteratively feeding back reliability information on the data symbols acquired by the channel decoder. For a specific type of such a receiver, we derive an upper bound on the achievable rate. Based on an approximation of the upper bound, which is not a closed-form expression, we are able to upper-bound the possible performance gain when using this specific receiver based on code-aided channel estimation in comparison to receivers using synchronized detection with a solely pilot based channel estimation. Furthermore, we compare this approximate upper bound with a lower bound on the achievable rate with joint processing of data and pilot symbols given in [1]. In addition, we show which part of the mutual information between the transmitter and the receiver cannot be exploited by the given receiver structure using synchronized detection in combination with iterative code-aided channel estimation.**

I. Introduction

In many mobile communication receivers, the channel is estimated based on pilot symbols to allow for a coherent detection and decoding (synchronized detection) in a separate step. However, in recent years receivers using iterative code-aided channel estimation got into the focus of research, see, e.g., [2], [3]. The main idea behind this type of receivers is that—additionally to the pilot symbols which are used for an initial channel estimation and detection/decoding—the channel estimation is enhanced by iteratively feeding back reliability information on the data symbols acquired by the channel decoder. Subsequently, this enhanced channel estimate is used in a further detection/decoding step, permitting enhanced decoding results, see Fig. 1. To evaluate the possible performance gain that can be achieved by receivers using iterative code-aided channel estimation and synchronized detection in comparison to receivers based on synchronized detection and a solely pilot based channel estimation, we are interested in the achievable data rate with such a type of receiver. Therefore, within this work, we study the achievable rate with iterative code-aided channel estimation based receivers. For a specific type of receiver based on iterative code-aided channel estimation, which is a slight modification of the typically studied code-aided channel estimation based receiver, we discuss which part of the mutual information between the transmitted data sequence and the observation sequence at the receiver side can be retrieved. Furthermore, we derive an upper bound on the achievable rate with this type of receiver for the case of a stationary Rayleigh flat-fading channel. The given upper bound is not a closed-form expression; however, for small channel dynamics we are able to give an approximation of this upper bound in closed form. This approximative upper bound enables us to evaluate the maximum possible gain when using this type of iterative code-aided channel estimation and synchronized detection based receiver in comparison to the conventional approach using synchronized detection in combination with a channel estimation which is solely based on pilot symbols. Furthermore, we compare this approximate upper bound with a lower bound on the achievable rate with joint processing of pilot and data symbols, i.e., with the optimum receiver processing, given in [1]. In addition, the derivation reveals why we apply the above mentioned slight modification of the receiver with respect to the typical iterative code-aided channel estimation based receiver and shows that this receiver cannot retrieve the complete mutual information between the transmitter and the receiver.

During the course of this work M. Dörpinghaus was with the Institute for Integrated Signal Processing Systems, RWTH Aachen University, Germany.
This work has been supported by the UMIC (Ultra High Speed Mobile Information and Communication) research centre.

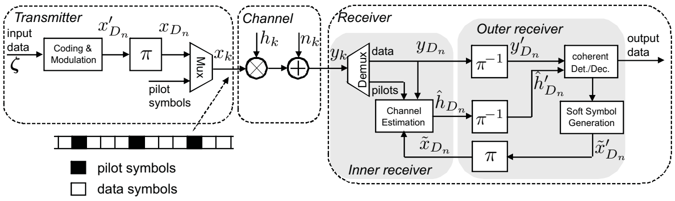

Fig. 1. Receiver based on iterative code-aided channel estimation

II. System Model

We consider a discrete-time zero-mean jointly proper Gaussian flat-fading channel with the I/O-relation

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

with $\mathbf{X} = \operatorname{diag}(\mathbf{x})$. Here, the $\operatorname{diag}(\cdot)$ operator generates a diagonal matrix whose diagonal elements are given by the argument vector. The vector $\mathbf{y}$ contains the channel output symbols in temporal order. Analogously, $\mathbf{x}$, $\mathbf{n}$, and $\mathbf{h}$ contain the transmit symbols, the additive noise samples, and the channel fading weights. All vectors are of length $2N+1$.

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 autocorrelation function $r_h(l) = \mathrm{E}[h_{k+l}h_k^*]$. Its variance is given by $r_h(0) = \sigma_h^2$ and its PSD is defined as

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

We assume that the PSD exists, which for a jointly proper Gaussian fading process implies ergodicity. We assume the PSD to be compactly supported within $[-f_d, f_d]$ with f_d being the maximum Doppler shift and $0 < f_d < 0.5$. I.e., $S_h(f) = 0$ 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 $f_d = 0$.

The transmit symbol sequence consists of i.i.d. zero-mean proper Gaussian data symbols with an average power σ_x^2 and each L-th symbol is a pilot symbol with power σ_x^2. The pilot spacing is chosen such that the channel fading process is sampled at least with Nyquist rate, i.e., $L < 1/(2f_d)$.

The processes $\{x_k\}$, $\{h_k\}$, and $\{n_k\}$ are assumed to be mutually independent. The mean SNR is given by

$$\rho = \sigma_x^2\sigma_h^2/\sigma_n^2. \tag{3}$$

A. Iterative Code-Aided Channel Estimation

Fig. 1 shows the structure of a receiver using iterative code-aided channel estimation with the iteratively coupled components, the channel estimator and the detection/decoding unit, which uses a coherent metric. In [3], it is shown that such a receiver structure can be systematically derived by expressing the joint ML detection and MAP parameter estimation[1]

$$\left\{\hat{\mathbf{x}}, \hat{\mathbf{h}}\right\} = \arg \max_{\mathbf{x}\in\mathcal{P}_{\mathbf{x}},\mathbf{h}} p\left(\mathbf{y}|\mathbf{x},\mathbf{h}\right) p\left(\mathbf{h}\right) \tag{5}$$

by a set of fixed point equations which can be solved iteratively. The set $\mathcal{P}_{\mathbf{x}}$ contains all possible transmit sequences. The set of fixed point equations consists of an equation describing the channel estimator and further fixed point equations for all code bits describing detection/decoding, which due to the processing of soft-information corresponds to a soft-demapper and MAP-decoder. Solving the fixed point equation set yields the following channel estimator (with iteration index (n))

$$\hat{\mathbf{h}}^{(n)} = \left\{\frac{1}{\sigma_n^2}\mathbf{\Gamma}_{\mathbf{s}}^{(n)} + \mathbf{R}_h^{-1}\right\}^{-1} \frac{1}{\sigma_n^2}\left(\bar{\mathbf{S}}^{(n)}\right)^* \mathbf{y}. \tag{6}$$

The channel estimator uses the reliability information on the symbols $[\mathbf{x}]_l$, i.e., the l-th element of $\mathbf{x}$. Here $\bar{\mathbf{S}}^{(n)} = \mathrm{diag}([\bar{s}_1^{(n)}, \ldots, \bar{s}_{2N+1}^{(n)}])$ contains the soft-symbols and $\mathbf{\Gamma}_{\mathbf{s}}^{(n)} = \mathrm{diag}([(\sigma_{s,1}^2)^{(n)}, \ldots, (\sigma_{s,2N+1}^2)^{(n)}])$ contains the expected powers of the symbols, which for the l-th symbol are given by

$$\bar{s}_l^{(n)} = \sum_{\mathbf{x}_i\in\mathcal{P}_x} p\left(\mathbf{x}_i|\mathbf{y}, \boldsymbol{\lambda}_I^{(n-1)}, \hat{\mathbf{h}}^{(n-1)}\right)[\mathbf{x}_i]_l^* \tag{7}$$

$$\left(\sigma_{s,l}^2\right)^{(n)} = \sum_{\mathbf{x}_i\in\mathcal{P}_x} p\left(\mathbf{x}_i|\mathbf{y}, \boldsymbol{\lambda}_I^{(n-1)}, \hat{\mathbf{h}}^{(n-1)}\right)\left|[\mathbf{x}_i]_l\right|^2. \tag{8}$$

[1]Note that (5) is in general not equal to ML sequence detection, i.e.,

$$\arg\max_{\mathbf{x}\in\mathcal{P}_{\mathbf{x}}} p(\mathbf{y}|\mathbf{x}) \neq \arg\max_{\mathbf{x}\in\mathcal{P}_{\mathbf{x}},\mathbf{h}} p(\mathbf{y}|\mathbf{x},\mathbf{h})p(\mathbf{h}). \tag{4}$$

However, the RHS is a high SNR approximation of the LHS. In the special case of constant modulus symbols the LHS and the RHS of (4) are equal.

Here the asterisk denotes complex conjugation. Moreover, $p(\mathbf{x}_i|\mathbf{y}, \boldsymbol{\lambda}_I^{(n-1)}, \hat{\mathbf{h}}^{(n-1)})$ gives the probability of the different sequences $\mathbf{x}_i$ based on the soft-information $\boldsymbol{\lambda}_I^{(n-1)}$ delivered by the decoder. In the derivation of the fixed point equations the independency of the individual code bits of all symbols $[\mathbf{x}]_l$ is assumed. This is achieved approximately by the ideal interleaver in Fig. 1. Therefore, the soft-symbol $\bar{s}_l$ depends only on the bits which define the l-th symbol. This greatly simplifies the computation of (7) and (8), i.e., of the reliability information in a practical receiver. The assumption of independent code bits has the consequence that the temporal correlation of the channel estimation error is neglected for detection.

To understand the effect of the channel estimation error on detection/decoding, we discuss exemplarily initial detection using a solely pilot based channel estimation. The fact that coherent detection is not optimal can be easily understood when studying ML-detection, which for initial detection using the pilot based channel estimate $\hat{\mathbf{h}}$ is given by

$$\hat{\mathbf{x}}_{\mathrm{ML}} = \arg\max_{\mathbf{x}\in\mathcal{P}_{\mathbf{x}}} p(\mathbf{y}|\mathbf{x}, \hat{\mathbf{h}}) \qquad \text{with} \tag{9}$$

$$p(\mathbf{y}|\mathbf{x}, \hat{\mathbf{h}}) = \frac{\exp(-(\mathbf{y}-\mathbf{X}\hat{\mathbf{h}})^H(\mathbf{X}\mathbf{R}_{\mathbf{e}}\mathbf{X}^H + \sigma_n^2\mathbf{I}_{2N+1})^{-1}(\mathbf{y}-\mathbf{X}\hat{\mathbf{h}}))}{\pi^{2N+1}\det\left(\mathbf{X}\mathbf{R}_{\mathbf{e}}\mathbf{X}^H + \sigma_n^2\mathbf{I}_{2N+1}\right)} \tag{10}$$

where $\mathbf{I}_{2N+1}$ is a square identity matrix of size $(2N+1)$. The matrix $\mathbf{R}_{\mathbf{e}} = \mathrm{E}\left[\mathbf{e}\mathbf{e}^H\right]$ is the correlation matrix of the estimation error. As the channel estimation is an interpolation or, in further channel estimation iterations, a smoothing, the estimation error is correlated over time and, thus, $\mathbf{R}_{\mathbf{e}}$ is not diagonal. This makes the evaluation of (9) computationally complex. Therefore, in practical receivers coherent detection is performed, which means that the temporal correlation of the channel estimation error is ignored, i.e., $\mathbf{R}_{\mathbf{e}}$ is substituted by the diagonal matrix $\mathbf{I}_{2N+1} \odot \mathbf{R}_{\mathbf{e}}$, where $\odot$ is the Hadamard product. This corresponds to an approximation of $p(\mathbf{y}|\mathbf{x}, \hat{\mathbf{h}})$ in (10) by

$$p(\mathbf{y}|\mathbf{x}, \hat{\mathbf{h}}) \approx \prod_{k=1}^{2N+1} \frac{\exp\left(-\frac{|y_k - x_k\hat{h}_k|^2}{|x_k|^2[\mathbf{R}_{\mathbf{e}}]_{kk} + \sigma_n^2}\right)}{\pi(|x_k|^2[\mathbf{R}_{\mathbf{e}}]_{kk} + \sigma_n^2)} \tag{11}$$

enabling symbol-wise detection. For a more detailed discussion on this we refer to [1]. As the considered receiver uses synchronized detection, in the present work we always assume coherent detection based on the channel estimate $\hat{\mathbf{h}}^{(n)}$.

As for the calculation of the soft-symbols independent code-bits have been assumed, the set of fixed point equations derived in [3] also corresponds to a coherent, i.e., symbol-wise, metric for detection/demapping.

B. Modified Channel Estimation Unit

As stated before, we have not been able to derive an upper bound on the achievable rate with the receiver described by (6) and (11), but only for a modified version. The modified receiver has a slightly different channel estimator, which for the calculation of $\hat{h}_k$, i.e., the k-th entry of $\hat{\mathbf{h}}$, does not use the corresponding observation y_k. Thus (6) is substituted by

$$\hat{h}_k^{(n)} = \left[\left\{\frac{1}{\sigma_n^2}\mathbf{\Gamma}_{\mathbf{s}}^{(n)}\mathbf{U}_k + \mathbf{R}_h^{-1}\right\}^{-1}\frac{1}{\sigma_n^2}\left(\bar{\mathbf{S}}^{(n)}\right)^*\mathbf{U}_k\mathbf{y}\right]_k, \forall k. \tag{12}$$

with $[\mathbf{a}]_k$ the k-th element of the vector $\mathbf{a}$. $\mathbf{U}_k$ is the identity matrix except the k-th diagonal element is zero. Hence y_k and the soft-information on x_k is not used for $\hat{h}_k$. The hereby discarded amount of information is small for practical, i.e., small channel dynamics. Although not being the typically studied code-aided channel estimator, it is also known from literature [4].

III. Upper Bound on the Achievable Rate with Iterative Code-Aided Channel Estimation

We derive an upper bound on the achievable rate with the receiver based on iterative code-aided channel estimation as described by (12) and (11). For the derivation, until stated otherwise, we assume i.i.d. input symbols, i.e., we assume that there are no pilot symbols. The mutual information between the transmitter and receiver can be expressed by

$$\mathcal{I}(\mathbf{y};\mathbf{x}) = \mathcal{I}(y_0;x_0|\mathbf{y}_{\backslash 0},\mathbf{x}_{\backslash 0}) + \mathcal{I}(y_0;\mathbf{x}_{\backslash 0}|\mathbf{y}_{\backslash 0}) + \mathcal{I}(\mathbf{x}_{\backslash 0};\mathbf{y}_{\backslash 0}) \tag{13}$$

where we have used the chain rule for mutual information and the independency of the transmit symbols. Without loss of generality, we use the following mapping of the time instances $\mathbf{x} = [x_{-N}, \ldots, x_{-1}, x_0, x_1, \ldots, x_N]^T$. Here, $\mathbf{x}_{\backslash 0}$ corresponds to $\mathbf{x}$ without x_0, and analogously for $\mathbf{y}_{\backslash 0}$.

The difference of the LHS of (13) and the last term at the RHS of (13) is an upper bound on the mutual information rate $\mathcal{I}'(\mathbf{y};\mathbf{x})$. To prove this, we use the following inequality:

$$\begin{aligned}
\mathcal{I}(\mathbf{y};\mathbf{x}) - \mathcal{I}(\mathbf{y}_{\backslash 0};\mathbf{x}_{\backslash 0}) &\overset{(a)}{=} \mathcal{I}(y_0;\mathbf{x}|\mathbf{y}_{\backslash 0}) + \mathcal{I}(\mathbf{y}_{\backslash 0};\mathbf{x}) - \mathcal{I}(\mathbf{y}_{\backslash 0};\mathbf{x}_{\backslash 0}) \\
&\overset{(b)}{=} \mathcal{I}(y_0;\mathbf{x}|\mathbf{y}_{\backslash 0}) \overset{(c)}{\geq} \mathcal{I}(y_0;\mathbf{x}|\mathbf{y}_{-N}^{-1}) \overset{(d)}{=} \mathcal{I}(y_0;\mathbf{x}_{-N}^{0}|\mathbf{y}_{-N}^{-1}) \\
&= h(y_0|\mathbf{y}_{-N}^{-1}) - h(y_0|\mathbf{x}_{-N}^{0},\mathbf{y}_{-N}^{-1})
\end{aligned} \tag{14}$$

where (a) is the chain rule for mutual information and (b) follows from the independency of the transmit symbols. Inequality (c) holds as the knowledge on the observations $\mathbf{y}_1^N$ will increase the mutual information between y_0 and $\mathbf{x}$ as they contain information on h_0. Here, e.g., $\mathbf{y}_{-N}^{-1}$ is a subvector of $\mathbf{y}$ containing the symbols from time instant $-N$ to -1. (d) holds due to the independency of the transmit symbols.

The definition of the entropy rate for stationary ergodic processes, see [5, Chapter 4.2], and (13) and (14), yields[2]

$$\begin{aligned}
\mathcal{I}'(\mathbf{y};\mathbf{x}) &= \lim_{N\to\infty}\frac{\mathcal{I}(\mathbf{y};\mathbf{x})}{2N+1} = \lim_{N\to\infty}\left\{h(y_0|\mathbf{y}_{-N}^{-1}) - h(y_0|\mathbf{x}_{-N}^{0},\mathbf{y}_{-N}^{-1})\right\} \\
&\leq \lim_{N\to\infty}\left\{\mathcal{I}(y_0;x_0|\mathbf{y}_{\backslash 0},\mathbf{x}_{\backslash 0}) + \mathcal{I}(y_0;\mathbf{x}_{\backslash 0}|\mathbf{y}_{\backslash 0})\right\}.
\end{aligned} \tag{16}$$

[2]Inequality (16) results from the fact that the time instant 0 is in the middle of the transmit sequence. If it was the most recent symbol, the separation would hold with equality, as in this case (14c) holds with equality, i.e.,

$$\mathcal{I}'(\mathbf{y};\mathbf{x}) = \lim_{N\to\infty}\left\{\mathcal{I}(y_N;x_N|\mathbf{y}_{\backslash N},\mathbf{x}_{\backslash N}) + \mathcal{I}(y_N;\mathbf{x}_{\backslash N}|\mathbf{y}_{\backslash N})\right\}. \tag{15}$$

A. The Term $\mathcal{I}(y_0;x_0|\mathbf{y}_{\backslash 0},\mathbf{x}_{\backslash 0})$

We want to obtain an insightful interpretation of the term

$$\mathcal{I}(y_0;x_0|\mathbf{y}_{\backslash 0},\mathbf{x}_{\backslash 0}) = \mathrm{E}\left[\log\left(\frac{p(y_0|x_0,\mathbf{y}_{\backslash 0},\mathbf{x}_{\backslash 0})}{p(y_0|\mathbf{y}_{\backslash 0},\mathbf{x}_{\backslash 0})}\right)\right]. \tag{17}$$

The probability density function (PDF) $p(y_0|x_0,\mathbf{y}_{\backslash 0},\mathbf{x}_{\backslash 0})$ is proper Gaussian and, thus, completely described by its mean and its variance which are given by

$$\mathrm{E}\left[y_0|x_0,\mathbf{y}_{\backslash 0},\mathbf{x}_{\backslash 0}\right] = x_0\mathrm{E}\left[h_0|\mathbf{y}_{\backslash 0},\mathbf{x}_{\backslash 0}\right] = x_0\hat{h}_0 \tag{18}$$

$$\begin{aligned}
\mathrm{var}\left[y_0|x_0,\mathbf{y}_{\backslash 0},\mathbf{x}_{\backslash 0}\right] &= \mathrm{E}\left[|y_0 - x_0\hat{h}_0|^2|x_0,\mathbf{y}_{\backslash 0},\mathbf{x}_{\backslash 0}\right] \\
&= |x_0|^2\mathrm{E}\left[|h_0 - \hat{h}_0|^2|\mathbf{x}_{\backslash 0}\right] + \sigma_n^2 = |x_0|^2\sigma_{e_{\text{int}}}^2(\mathbf{x}_{\backslash 0}) + \sigma_n^2
\end{aligned} \tag{19}$$

where $\hat{h}_0$ is the MMSE estimate of h_0 which is linear, as the problem is jointly proper Gaussian. Thus, the estimation error is zero-mean proper Gaussian with variance $\sigma_{e_{\text{int}}}^2(\mathbf{x}_{\backslash 0})$. Here the index *int* denotes interpolation. $\hat{h}_0$ is also zero-mean proper Gaussian with variance $\sigma_{\hat{h}}^2(\mathbf{x}_{\backslash 0}) = \sigma_h^2 - \sigma_{e_{\text{int}}}^2(\mathbf{x}_{\backslash 0})$. Note that $\sigma_{e_{\text{int}}}^2(\mathbf{x}_{\backslash 0})$ is independent of $\mathbf{y}_{\backslash 0}$ due to the principle of orthogonality in MMSE estimation. The interpolation error variance depends on the past and future transmit symbols which is indicated by the notation $\sigma_{e_{\text{int}}}^2(\mathbf{x}_{\backslash 0})$. Using (18) and (19) we can use the following substitutions:

$$p(y_0|x_0,\mathbf{y}_{\backslash 0},\mathbf{x}_{\backslash 0}) = p(y_0|x_0,\hat{h}_0,\mathbf{x}_{\backslash 0}) \tag{20}$$

$$p(y_0|\mathbf{y}_{\backslash 0},\mathbf{x}_{\backslash 0}) = p(y_0|\hat{h}_0,\mathbf{x}_{\backslash 0}) \tag{21}$$

$$\text{and, hence} \quad \mathcal{I}(y_0;x_0|\mathbf{y}_{\backslash 0},\mathbf{x}_{\backslash 0}) = \mathcal{I}(y_0;x_0|\hat{h}_0,\mathbf{x}_{\backslash 0}). \tag{22}$$

Thus, the first term on the RHS of (16) is the mutual information between channel input and output at the arbitrarily chosen time instant 0 if an MMSE estimate $\hat{h}_0$ is available which is based on all past and future channel observations and the knowledge of all transmit symbols except the current one.

We want to explain why $\mathcal{I}(x_0;y_0|\hat{h}_0,\mathbf{x}_{\backslash 0})$ is an upper bound on the achievable rate with the iterative code-aided channel estimation based receiver given by (12) and (11). In this regard, consider that the only dependency between the individual time instances is established by the channel correlation. In case all past and all future transmit symbols are known, corresponding to $\mathcal{I}(x_0;y_0|\hat{h}_0,\mathbf{x}_{\backslash 0})$, all information on h_0 given by the past and future channel output observation $\mathbf{y}_{\backslash 0}$ and the knowledge on $\mathbf{x}_{\backslash 0}$ is carried by $\hat{h}_0$ and $\mathbf{x}_{\backslash 0}$. Note that for the calculation of $\hat{h}_0$, the observation y_0 is not used, which exactly corresponds to the channel estimator described by (12). Furthermore, observe that for the case of perfect knowledge of all past and all future transmit symbols the estimator in (12) exactly corresponds to the MMSE interpolator $\hat{h}_0$ in $\mathcal{I}(x_0;y_0|\hat{h}_0,\mathbf{x}_{\backslash 0})$ with the estimation error variance $\sigma_{e_{\text{int}}}^2(\mathbf{x}_{\backslash 0})$ given in (19). This means that $\mathcal{I}(x_0;y_0|\hat{h}_0,\mathbf{x}_{\backslash 0})$ corresponds to the mutual information at the arbitrarily chosen data symbol time instant 0 if all past and all future transmit symbols are perfectly known. Obviously, the assumption of perfect knowledge of all past and all future transmit symbols results in an upper bound to the actual achievable mutual information at the arbitrarily chosen data symbol time instant 0, as this yields the maximum amount of information on h_0 given by $\mathbf{y}_{\backslash 0}$.

Moreover, it is important to recognize that this argumentation only holds in case we assume a coherent, i.e., symbol-wise, detection as described in Section II-A. If detection would be performed over the whole sequence, evaluation of the mutual information for a single time instant, as it is done with $\mathcal{I}(x_0; y_0|\hat{h}_0, \mathbf{x}_{\backslash 0})$ would be meaningless. In addition, for optimal ML sequence detection, the type of channel estimator would be irrelevant for the achievable rate, as the detector has the whole observation sequence $\mathbf{y}$ at its input, see [1].

Now, we can interpret the upper bound on the mutual information rate in (16). It is the sum of two terms. The first term on the RHS of (16), which is the main contribution, is related to a *coherent Rayleigh flat-fading channel*, i.e., a channel whose optimum detection metric can be evaluated symbol-wise. In comparison to the genuine fading channel its fading variance is modified due to the estimation error and given by $\sigma_h^2 - \sigma_e^2$. Its noise variance is given by $|x_k|^2\sigma_e^2 + \sigma_n^2$. However, the fact that the *effective* noise variance depends on the transmit symbol x_k is a difference to a coherent fading channel. The second term on the RHS of (16), $\mathcal{I}(y_0; \mathbf{x}_{\backslash 0}|\mathbf{y}_{\backslash 0})$, can be viewed as a correction term accounting for the information contained in the temporal correlation of the channel estimation error, which cannot be exploited by the receiver structure described by (12) and (11). Note that this term partly arises due to the upper-bounding in (16), see also footnote 3.

B. The Term $\mathcal{I}(y_0; \mathbf{x}_{\backslash 0}|\mathbf{y}_{\backslash 0})$

The only relation between the individual symbol time instances is given by the fading correlation. Thus, using the chain rule for mutual information, we rewrite $\mathcal{I}(y_0; \mathbf{x}_{\backslash 0}|\mathbf{y}_{\backslash 0})$ such that its relation to the fading weight h_0 becomes explicit

$$\begin{aligned}\mathcal{I}(y_0; \mathbf{x}_{\backslash 0}|\mathbf{y}_{\backslash 0}) &= \mathcal{I}(h_0, y_0; \mathbf{x}_{\backslash 0}|\mathbf{y}_{\backslash 0}) - \mathcal{I}(h_0; \mathbf{x}_{\backslash 0}|\mathbf{y})\\ &= \mathcal{I}(y_0; \mathbf{x}_{\backslash 0}|h_0, \mathbf{y}_{\backslash 0}) + \mathcal{I}(h_0; \mathbf{x}_{\backslash 0}|\mathbf{y}_{\backslash 0}) - \mathcal{I}(h_0; \mathbf{x}_{\backslash 0}|\mathbf{y})\\ &\overset{(a)}{=} \mathcal{I}(h_0; \mathbf{x}_{\backslash 0}|\mathbf{y}_{\backslash 0}) - \mathcal{I}(h_0; \mathbf{x}_{\backslash 0}|\mathbf{y}) \qquad (23)\end{aligned}$$

where (a) follows from the fact that y_0 is independent of $\mathbf{y}_{\backslash 0}$ and of $\mathbf{x}_{\backslash 0}$ while conditioning on h_0. The RHS of (23) indicates that $\mathcal{I}(y_0; \mathbf{x}_{\backslash 0}|\mathbf{y}_{\backslash 0})$ is related to the additional information on h_0 contained in y_0 while *knowing* $\mathbf{y}_{\backslash 0}$. The fact that this additional information on h_0 contained in y_0 cannot be exploited by the given receiver structure, i.e., using the channel estimator in (12) in combination with the symbol-wise detection metric in (11), is supported by the following observation. First, consider that the observation y_0 is also used at the input of the detection unit. Nevertheless, with the given structure, where y_0 is not used for channel estimation, the information corresponding to $\mathcal{I}(y_0; \mathbf{x}_{\backslash 0}|\mathbf{y}_{\backslash 0})$ cannot be exploited. Therefor, consider that detection works symbol-wise, and that y_0 contains additional information on h_0, which can only be exploited when using it in combination with $\mathbf{y}_{\backslash 0}$ and $\mathbf{x}_{\backslash 0}$, indicated by the conditioning on $\mathbf{y}_{\backslash 0}$ in $\mathcal{I}(y_0; \mathbf{x}_{\backslash 0}|\mathbf{y}_{\backslash 0})$. However, this is not possible for the detector due to its symbol-wise metric. This supports the statement at the end of Section III-A that $\mathcal{I}(y_0; \mathbf{x}_{\backslash 0}|\mathbf{y}_{\backslash 0})$ is a correction term to $\mathcal{I}(y_0; x_0|\mathbf{y}_{\backslash 0}, \mathbf{x}_{\backslash 0})$ on the RHS of (16) accounting for the fact that the detector cannot exploit the temporal correlation of the channel estimation error.

The observation that $\mathcal{I}(y_0; \mathbf{x}_{\backslash 0}|\mathbf{y}_{\backslash 0})$ accounts for the temporal correlation of the channel estimation error is also supported by the fact that in case of an uncorrelated channel, i.e., $\mathrm{E}[h_k h_l^*] = 0, \forall k \neq l$, the term $\mathcal{I}(y_0; \mathbf{x}_{\backslash 0}|\mathbf{y}_{\backslash 0})$ becomes zero.

C. An Upper Bound on $\mathcal{I}(x_0; y_0|\hat{h}_0, \mathbf{x}_{\backslash 0})$

To upper-bound $\mathcal{I}(x_0; y_0|\hat{h}_0, \mathbf{x}_{\backslash 0})$ and, thus, the achievable rate with the iterative code-aided channel estimation based receiver described by (12) and (11), we use the expression

$$\mathcal{I}(y_0; x_0|\hat{h}_0, \mathbf{x}_{\backslash 0}) = h(y_0|\hat{h}_0, \mathbf{x}_{\backslash 0}) - h(y_0|\hat{h}_0, \mathbf{x}). \qquad (24)$$

1) Calculation of $h(y_0|\hat{h}_0, \mathbf{x})$: Based on $\hat{h}_0$ the channel output y_0 can be written as

$$y_0 = x_0 h_0 + n_0 = x_0\left(\hat{h}_0 + e_0\right) + n_0 \qquad (25)$$

where e_0 is the estimation error, which is zero-mean proper Gaussian. Thus, y_0 conditioned on $\hat{h}_0$ and $\mathbf{x}$ is proper Gaussian. Hence, $h(y_0|\hat{h}_0, \mathbf{x})$ is completely described by the conditional variance of y_0, which is given in (19), and we get

$$h(y_0|\hat{h}_0, \mathbf{x}) = \mathrm{E}_{\mathbf{x}}\left[\log\left(\pi e\left(\sigma_n^2 + \sigma_{e_{\text{int}}}^2(\mathbf{x}_{\backslash 0})|x_0|^2\right)\right)\right]. \qquad (26)$$

2) Upper Bound on $h(y_0|\hat{h}_0, \mathbf{x}_{\backslash 0})$: In contrast to the prior case, y_0 conditioned on $\hat{h}_0$ and $\mathbf{x}_{\backslash 0}$ is not proper Gaussian. Nevertheless, its entropy is upper-bounded by the entropy of a proper Gaussian random variable with the same variance, which is given by $\sigma_x^2(|\hat{h}_0|^2 + \sigma_{e_{\text{int}}}^2(\mathbf{x}_{\backslash 0})) + \sigma_n^2$. Thus, we get

$$\begin{aligned}h(y_0|\hat{h}_0, \mathbf{x}_{\backslash 0}) &\leq \mathrm{E}_{\mathbf{x}_{\backslash 0}}\left[\mathrm{E}_{\hat{h}_0}\left[\log\left(\pi e\left(\sigma_n^2 + \sigma_x^2\sigma_{e_{\text{int}}}^2(\mathbf{x}_{\backslash 0}) + \sigma_x^2|\hat{h}_0|^2\right)\right)\right]\right]\\ &= \mathrm{E}_{\mathbf{x}_{\backslash 0}}\left[\int_0^\infty \log\left(\pi e\left(\sigma_n^2 + \sigma_x^2\sigma_{e_{\text{int}}}^2(\mathbf{x}_{\backslash 0}) + \sigma_x^2\left(\sigma_h^2 - \sigma_{e_{\text{int}}}^2(\mathbf{x}_{\backslash 0})\right)u\right)\right)e^{-u}du\right]\end{aligned} \qquad (27)$$

using that $\hat{h}_0$ is zero-mean proper Gaussian with variance $\sigma_h^2 - \sigma_{e_{\text{int}}}^2(\mathbf{x}_{\backslash 0})$.

3) Derivation of an Upper Bound on $\mathcal{I}(x_0; y_0|\hat{h}_0, \mathbf{x}_{\backslash 0})$: With (24), (26), and (27), we get for a zero-mean proper Gaussian data symbol x_0 with variance σ_x^2 the upper bound

$$\begin{aligned}\mathcal{I}(x_0; y_0|\hat{h}_0, \mathbf{x}_{\backslash 0}) \leq \mathrm{E}_{\mathbf{x}_{\backslash 0}}\Bigg[\int_0^\infty \Bigg[&-\log\left(1 + \rho\frac{\sigma_{e_{\text{int}}}^2(\mathbf{x}_{\backslash 0})}{\sigma_h^2}u\right)\\ &+\log\left(1 + \rho\frac{\sigma_{e_{\text{int}}}^2(\mathbf{x}_{\backslash 0})}{\sigma_h^2} + \rho\left(1 - \frac{\sigma_{e_{\text{int}}}^2(\mathbf{x}_{\backslash 0})}{\sigma_h^2}\right)u\right)\Bigg]e^{-u}du\Bigg]\end{aligned} \qquad (28)$$

with the mean SNR ρ in (3). This upper bound depends on the interpolation error variance $\sigma_{e_{\text{int}}}^2(\mathbf{x}_{\backslash 0})$, which itself is a random variable. Its distribution depends on the distribution of the past and future transmit symbols. It can be expressed by

$$\sigma_{e_{\text{int}}}^2(\mathbf{x}_{\backslash 0}) = \sigma_h^2 - \mathbf{r}_{h,\text{int}}^H\left(\mathbf{R}_{h\backslash 0} + \sigma_n^2\mathbf{Z}_{\backslash 0}^{-1}\right)^{-1}\mathbf{r}_{h,\text{int}} = \sigma_{e_{\text{int}}}^2(\mathbf{z}_{\backslash 0})$$

with $\mathbf{R}_{h\backslash 0} = \mathrm{E}[\mathbf{h}_{\backslash 0}\mathbf{h}_{\backslash 0}^H]$, $\mathbf{r}_{h,\text{int}} = [r_h(-N), \ldots, r_h(-1), r_h(1), \ldots, r_h(N)]^T$, and $\mathbf{Z}_{\backslash 0} = \mathbf{X}_{\backslash 0}^H\mathbf{X}_{\backslash 0}$ is a diagonal matrix containing the powers of the past and future transmit symbols. The vector $\mathbf{z}_{\backslash 0}$ contains the diagonal elements of $\mathbf{Z}_{\backslash 0}$.

If it would be possible to show that the argument of the expectation operation on the RHS of (28) is concave with respect to each individual element of the diagonal of $\mathbf{Z}_{\backslash 0}$, considering i.i.d. data symbols, and using Jensen's inequality, this would mean that the RHS of (28) is maximized in case $\sigma^2_{e_{\text{int}}}(\mathbf{z}_{\backslash 0})$ is calculated under the assumption that all past and future transmit symbols are constant modulus symbols with power σ_x^2. Unfortunately, we have not been able to prove this concavity. Nevertheless, for small channel dynamics it is reasonable to approximate the channel interpolation error variance $\sigma^2_{e_{\text{int}}}(\mathbf{z}_{\backslash 0})$ by the channel interpolation error variance calculated under the assumption that all past and all future transmit symbols are constant modulus symbols. In this regard, consider that in case of small channel dynamics, the calculation of the channel estimate corresponds to a weighted averaging of many channel output observations, where in the limit of an asymptotically small channel dynamic all observations are equally weighted. Furthermore, it can be shown that in case of a constant channel the distribution of the past and future input symbols is irrelevant, only their average power has an influence on $\sigma^2_{e_{\text{int}}}(\mathbf{z}_{\backslash 0})$. Thus, we are able to approximate $\sigma^2_{e_{\text{int}}}(\mathbf{z}_{\backslash 0})$ by assuming that all past and all future transmit symbols are constant modulus symbols (CM) with power σ_x^2. The advantage of this specific assumption on the distribution of the past and future transmit symbols, i.e., CM symbols with power σ_x^2, is that, when additionally considering an infinite long observation horizon in the past and in the future, we are able to give a closed form expression for the interpolation error variance, which is given by

$$\sigma^2_{e_{\text{int,CM},\infty}} = \frac{\sigma_h^2}{\rho}\left[\left\{\int_{-\frac{1}{2}}^{\frac{1}{2}}\left[\frac{\rho}{\sigma_h^2}S_h(f)+1\right]^{-1}df\right\}^{-1}-1\right]. \quad (29)$$

The assumption on constant modulus transmit symbols is in contrast to the assumption on i.i.d. zero-mean proper Gaussian input symbols. However, we use the assumption on constant modulus input symbols only in the context of the channel interpolation error variance $\sigma^2_{e_{\text{int}}}(\mathbf{z}_{\backslash 0})$ for bounding.

Finally, with (28) and (29) we have found an approximate upper bound on the achievable rate with the iterative code-aided channel estimation based receiver described by (12) and (11), and i.i.d. zero-mean proper Gaussian data-symbols. Notice that this bound holds without any assumption on the use of pilot symbols. If we additionally account for the rate loss due to the deterministic pilot symbols for the pilot spacing L, we get the following approximate upper bound[3]

$$\mathcal{I}'(x_0;y_0|\hat{h}_0,\mathbf{x}_{\backslash 0}) \lesssim \frac{L-1}{L}\text{RHS}(28)\Big|_{\sigma^2_{e_{\text{int}}}(\mathbf{x}_{\backslash 0})=\sigma^2_{e_{\text{int,CM},\infty}}}. \quad (30)$$

D. Numerical Evaluation

In Fig. 2 the approximate upper bound on the achievable rate with the receiver based on iterative code-aided channel estimation (using (12)) in (30) is compared to a lower bound on the achievable rate with joint processing of pilot and data symbols given in [1, (25)]. For both bounds the pilot spacing is chosen such that the fading process is sampled with Nyquist rate, i.e., $L = \lfloor 1/(2f_d)\rfloor$. We assume a rectangular PSD of the fading process $S_h(f)=\sigma_h^2/(2f_d)$ for $|f|\leq f_d$ and 0 otherwise.

[3]The upper bound on the achievable rate can also be applied when using pilot symbols, as it is based on $\mathcal{I}(x_0;y_0|\hat{h}_0,\mathbf{x}_{\backslash 0})$, which implicitly means that all past and future transmit symbols are known.

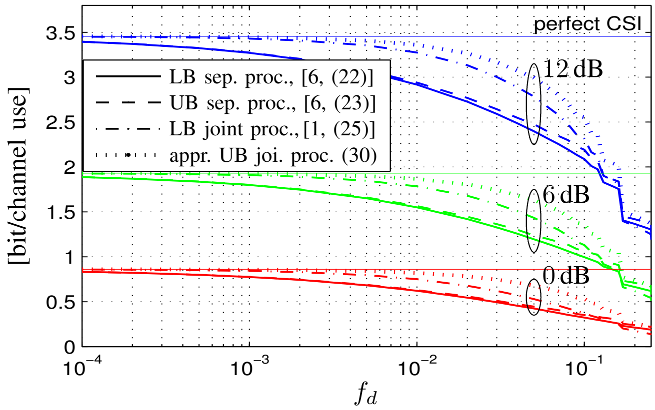

Fig. 2. Comparison of the approximate upper bound (30) on the achievable rate with the iterative code-aided channel estimation based receiver described by (12) and (11) with the lower bound on the achievable data rate with joint processing given in [1, (25)] ($L = \lfloor 1/(2f_d)\rfloor$ for both) and with bounds on the achievable rate with synchronized detection and solely pilot based channel estimation (separate processing) given in [6] for i.i.d. zero-mean proper Gaussian data symbols; additionally, the capacity in case of perfect channel state information (CSI) is shown; rectangular PSD $S_h(f)$

The lower bound given in [1, (25)] is an actual lower bound on the achievable rate with joint processing, i.e., with the optimal receiver. In contrast, we have shown that the iterative code-aided channel estimation based receiver in the present paper is not able to exploit the complete mutual information between the transmitter and the receiver. Nevertheless the upper bound on the achievable rate for the receiver based on iterative code-aided channel estimation described by (12) and (11) is larger than the lower bound for joint processing.

Additionally, these bounds are compared to bounds on the achievable rate with synchronized detection and a solely pilot based channel estimation, named separate processing in Fig. 2, given in [6]. As the upper and lower bound on the achievable rate with separate processing are relatively tight, we choose the pilot spacing in this case such that the lower bound for separate processing in [6, (22)] is maximized. The gap between the bounds on the achievable rate with separate processing and the bounds for joint processing/iterative code-aided channel estimation gives an indication on the possible gain by using code-aided channel estimation in comparison to a solely pilot based channel estimation.

References

[1] 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 *Proc. of the 2010 International Zurich Seminar on Communications*, Zurich, Switzerland, Mar. 2010, pp. 74–77.

[2] 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.

[3] 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.

[4] F. Sanzi, S. Jelting, and J. Speidel, "A comparative study of iterative channel estimators for mobile OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 2, no. 5, pp. 849–859, Sep. 2003.

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

[6] 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.