

# Iterative Pilot-Layer Aided Channel Estimation with Emphasis on Interleave-Division Multiple Access Systems

Hendrik Schoeneich and Peter Adam Hoehner

*Information and Coding Theory Lab, Faculty of Engineering, University of Kiel, Kaiserstrasse 2, 24143 Kiel, Germany*

Received 1 June 2005; Revised 22 May 2006; Accepted 4 June 2006

Channel estimation schemes suitable for interleave-division multiple access (IDMA) systems are presented. Training and data are superimposed. Training-based and semiblind linear channel estimators are derived and their performance is discussed and compared. Monte Carlo simulation results are presented showing that the derived channel estimators in conjunction with a superimposed pilot sequence and chip-by-chip processing are able to track fast-fading frequency-selective channels. As opposed to conventional channel estimation techniques, the BER performance even improves with increasing Doppler spread for typical system parameters. An error performance close to the case of perfect channel knowledge can be achieved with high power efficiency.

Copyright © 2006 H. Schoeneich and P. A. Hoehner. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1. INTRODUCTION

Spread-spectrum multiple access is a popular technique allowing several users to share the same bandwidth at the same time. Spread spectrum is often equated with direct-sequence code-division multiple access (DS-CDMA), where data detection is based on orthogonal or near-orthogonal spreading sequences. In [1–3], a spread-spectrum technique without the need for spreading sequences has been proposed. In this technique, data separation is based on chip-level interleavers. Therefore we refer to it as interleave-division multiple access (IDMA) [3, 4]. Processing is done on a chip-level basis. No orthogonal design is necessary. According to the results in [5, 6], the power and bandwidth efficiency of DS-CDMA can theoretically be maximized when devoting the entire bandwidth expansion (spreading) to FEC coding and removing the spreading sequences. IDMA fulfills this requirement and still allows for user separation. In conjunction with an optimized power allocation scheme, IDMA is able to reach the channel capacity—even when binary antipodal signaling is applied [7]. Like DS-CDMA, IDMA is well suited to make use of the diversity that is introduced by frequency-selective fading, as will be shown by the subsequent numerical results. IDMA is currently discussed as a candidate for upcoming 4G systems [8–11].

In this paper, channel estimation schemes for IDMA are proposed. Parts of this paper are published in [12]. Robust

channel estimation is especially important for spread-spectrum systems with iterative receiver structures, where channel estimation is performed before despreading, as in this case the signal-to-noise ratio is typically very low due to low-rate encoding. This is especially true for IDMA, where the despreading is completely done in the decoder and the detector works on a chip-level basis. Frequency-selective fading channels additionally pose a challenge to the channel estimator as the performance of the channel estimates typically degrades when the number of channel coefficients to be estimated increases and the correlation of neighboring channel coefficients decreases due to fading.

There exist two main training concepts: (a) time multiplexing (periodically or once per block) [13–15] and (b) superposition of training and data [4, 16, 17]. Combinations of (a) and (b) are possible and are used, for example, in UMTS [18]. The advantage of superimposed training is that the channel estimator is actually trained at the same time indices where the channel estimate is needed for detection. This method is therefore well suited for estimating fast-fading channels. In this paper, we apply superimposed training to IDMA. Superimposed training for IDMA is particularly simplified by the fact that—as opposed to DS-CDMA—the cross-correlations between the spreading sequences and the chip training sequence do not have to be taken into account as the data separation is based on different chip interleavers and not on (nearly) uncorrelated spreading sequences.

One training sequence, the so-called pilot layer, is superimposed per user. The scheme is referred to as pilot-layer aided channel estimation (PLACE). PLACE is well suited for semiblind channel estimation [19, 20], which allows for power and bandwidth efficient transmission, and is especially useful for multilayer IDMA, where the data of one user is transmitted using multiple data layers, as proposed in [8] for adaptive IDMA. PLACE is a generalization of the scheme in [4], where one layer is assigned to each user and channel estimation is performed by a simple correlation operation. In this paper, the number of layers per user is arbitrary and we concentrate on optimal and suboptimal joint channel estimators.

The rest of this paper is organized as follows. In Section 2, the system model is described. A short introduction to IDMA and the multilayer concept is provided. Section 3 introduces the iterative receiver structure. A detailed consideration of the channel estimation scheme under investigation in Section 4 is followed by a short description of the Gaussian multilayer detector in Section 5, which is used to obtain the numerical results in Section 6.

## 2. SYSTEM MODEL

Throughout this paper, the discrete-time complex baseband notation is used. The received sample at chip index  $k$ ,  $1 \leq k \leq K_c$ , can be written as

$$y[k] = \sum_{u=1}^U \sum_{l=0}^L h_{u,l}[k] \left( p_u[k-l] + \sum_{m=1}^{M_u} x_{u,m}[k-l] \right) + n[k], \quad (1)$$

where  $K_c$  is the block length in chips,  $U$  is the number of active users. The downlink case can be treated as  $U = 1$ . The Gaussian-distributed channel coefficients  $h_{u,l}[k] \sim \mathcal{N}_C(0, \sigma_{h_{u,l}}^2)$  describe the physical channel, pulse shaping, and sampling. The effective memory length is denoted by  $L$ . The average power of the channel of user  $u$  is denoted as  $\sigma_{h_u}^2$ . Channel coefficients with different delays and/or user indices are assumed to be statistically independent. Channel coefficients of different blocks are also assumed to be statistically independent. The  $M = \sum_{u=1}^U M_u$  sequences of interleaved chips  $x_{u,m}[k]$  are referred to as data layers. The  $M_u$  data layers and the associated chips of the pilot layer,  $p_u[k]$ , form the transmitted signal of user  $u$ . The average power of the pilot layer of user  $u$  is  $P_{p,u}$  and the total power of all pilot layers is  $P_p = \sum_{u=1}^U P_{p,u}$ . The noise samples  $n[k] \sim \mathcal{N}_C(0, \sigma_n^2)$  are statistically independent realizations of a zero-mean complex Gaussian process of variance  $\sigma_n^2$ .

The chips are assumed to be out of the set  $\{\pm a_{u,m} e^{j\varphi_{u,m}}\}$ , where  $a_{u,m}$  is the amplitude of the  $m$ th layer of user  $u$  and  $\varphi_{u,m}$  is a uniformly distributed phase, that is, in every layer BPSK modulation with a layer-specific phase offset is applied. This results in a fixed data rate per layer. Any user  $u$  can be assigned multiple layers  $M_u$ , so that the data rate of one particular user is proportional to the number of layers that is assigned to this user [8].

Though it may seem inefficient to use a binary modulation scheme at first glance, this is actually not true due to the layer-specific phase offsets. A system load near 4 bit/s/Hz is reported in [21] using this scheme. It is shown in [7] that IDMA with superimposed binary sequences (BPSK mapping) is actually capacity-approaching in combination with a suitable power allocation scheme—even for a moderate number of layers. The combination of BPSK and layer-specific phase offsets can itself be interpreted as a modulation scheme. For an even number of data layers, equivalence to QPSK is obtained. Therefore, binary modulated layers with uniformly distributed phase offsets do not lead to a performance loss nor to a complexity increase compared to QPSK. The main reason to use BPSK instead of QPSK is that the quantization of the system load is halved compared to QPSK (code rate  $R$  instead of  $2R$ ), and that therefore the granularity is minimized. This is an important aspect when adjusting the system load close to capacity and/or in a system with many users.

The amplitudes  $a_{u,m}$  include power control. For simplicity, all amplitudes are assumed to be the same throughout this paper. Further performance improvements can be obtained by an optimized power allocation as shown in [22].

Equation (1) can also be written in matrix form:

$$\mathbf{y} = (\mathbf{P} + \mathbf{X}) \cdot \mathbf{h} + \mathbf{n}, \quad (2)$$

where  $\mathbf{X}$  is the stacked data matrix of all  $M$  data layers and  $\mathbf{P}$  is the stacked data matrix of all pilot layers and  $\mathbf{h}$  is the stacked channel vector of length  $U \cdot K_c \cdot (L + 1)$ . All vectors in this paper are column vectors. Vectors and matrices are denoted as boldface small and capital letters, respectively.

IDMA can be interpreted as conventional DS-CDMA with interleaver and spreader in exchanged order, which is illustrated in Figure 1 for one data layer. The spreader becomes part of the encoder (ENC) and has no special meaning anymore. Note that no spreading sequences are applied. Nevertheless, the interleaved code symbols can be transmitted at a rate up to  $1/R$  times higher than the info bit rate, where  $R$  is the code rate of the encoder. Therefore the terms *code symbol* and *chip* are interchangeable with each other for IDMA. We will use the term chip throughout the rest of this paper. The bit load of user  $u$  is  $b_u = RM_u$  and the overall bit load (referred to as system load throughout this paper) is  $b = \sum_{u=1}^U b_u$ .

If not stated otherwise, a binary  $(1/R, 1)$  code with random code bits is used throughout this paper, that is, every info bit is mapped to a random binary sequence of length  $1/R$ . This code is equivalent to a repetition code with subsequent random scrambling. Therefore no coding gain can be achieved, but as shown in [21] a robust transmission with very high system loads near 4 bit/s/Hz can be achieved.

## 3. ITERATIVE RECEIVER STRUCTURE

In spread-spectrum systems, optimal detection is usually infeasible, because the computational complexity increases exponentially with the number of data layers. A suboptimal solution to this problem is an iterative approach performing

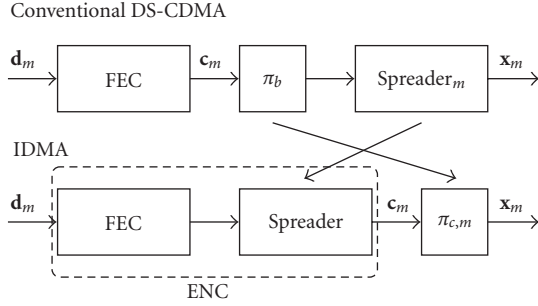


FIGURE 1: IDMA can be interpreted as conventional DS-CDMA with interleaver and spreader in exchanged order.

cross-layer multilayer chip detection (MLD)—thereby ignoring the code constraints—and layer-wise channel decoding (DEC)—thereby ignoring the channel interferences. Figure 2 depicts an iterative receiver structure for layer  $m$ ,  $1 \leq m \leq M_u$ , of user  $u$ ,  $1 \leq u \leq U$ . The received samples of (1) are fed into the MLD and PLACE unit. One iteration consists of an estimation of  $\mathbf{h}$  based on  $\mathbf{P}$  and the extrinsic information from the last decoding in the PLACE unit, a detection of all data layers in the MLD unit, and the MAP decoding in the DEC unit. A detailed description of the PLACE and the MLD units is given in Section 4 and Section 5, respectively.

For each layer, the decoder performs chip-by-chip maximum a posteriori (MAP) decoding, for example, by means of the well-known BCJR algorithm, to obtain extrinsic soft information about the chips. This soft information can be represented in different equivalent forms—as probabilities, log-likelihood ratios, or soft chips.<sup>1</sup> Since the subsequent processing is based on the reinterleaved soft chips, we concentrate on the latter, and denote the reinterleaved soft chip of user  $u$  and layer  $m$  at chip index  $k$  in iteration  $i$  as  $\tilde{x}_{u,m}^{(i)}[k]$ . The soft chips of iteration  $i$  can be stacked together to form the soft chip matrix  $\tilde{\mathbf{X}}^{(i)}$ . The iteration number is indicated by a superior number in brackets throughout this paper.

#### 4. PILOT-LAYER AIDED CHANNEL ESTIMATION (PLACE)

The task of the PLACE unit in Figure 2 is to find an estimate  $\hat{\mathbf{h}}^{(i+1)}$  of the channel coefficients  $\mathbf{h}$  based on the received data  $\mathbf{y}$ , the perfectly known pilot data matrix  $\mathbf{P}$ , and the reinterleaved extrinsic information represented by the soft chip matrix  $\tilde{\mathbf{X}}^{(i)}$  that is obtained by the previous decoding step. By taking the soft chips properly into account, the channel estimates improve from iteration to iteration, which in turn improves chip detection.

There exist three major channel estimation concepts in this context: (1) training-based channel estimation (tb), (2) semiblind channel estimation (sb), and (3) blind channel estimation. For tb, channel estimation is only based on the knowledge of the pilot layer, which is illustrated in

Figure 3(a). As the updated soft chips are not used, this type of channel estimation can be taken out of the iterative process. The channel estimation is performed only once before the first detection and the resulting channel estimates are used without change in all detection steps. Therefore, the computational complexity of tb is the lowest of all channel estimation concepts listed above. Beside this advantage tb has two disadvantages. Firstly, without any knowledge about the data, the interference from data (due to the superposition) leads to a high noise level and consequently to unreliable channel estimates. A solution to this problem is to partially cancel the interference from data based on the soft chips from the decoder before tb (tb-IC) (cf. Figure 3). Note that in this case, channel estimation is still training-based, but the received data samples are modified before tb is performed:

$$\mathbf{y}^{(i)} = \mathbf{y} - \tilde{\mathbf{X}}^{(i)} \hat{\mathbf{h}}^{(i)}. \quad (3)$$

This modification depends on the decoder output of the  $i$ th iteration. Therefore the channel estimates obtained by tb-IC also depend on the iteration number, that is, tb-IC has to be performed once per iteration and cannot be taken out of the iterative process as tb.

The second disadvantage of tb is that the performance of the channel estimator is limited by the power of the pilot layer. Even if the data interference cancelation of (3) is perfect, the modified received data is still noisy. Note that the quality of the channel estimates depends on the training power and the noise power. Therefore, the quality of the channel estimates can be improved by making constructively use of the soft chips from the decoder for channel estimation. For sb, channel estimation is based on the knowledge of the pilot layer as for tb, but additionally based on the knowledge of the soft chips (cf. Figure 3). The data is not considered as interference, which has to be canceled as done for tb-IC—it is rather used as “virtual” training in combination with the pilot layer, which can improve the training power significantly. As tb-IC, sb is performed once every iteration between decoding and multilayer chip detection.

Blind channel estimation is treated as a special case of sb with  $\mathbf{P} = \mathbf{0}$  throughout this paper.

In the following, we focus on linear channel estimation schemes and present a detailed description of tb, tb-IC, and sb suitable for IDMA. Note that nonlinear channel estimation schemes can easily be used in the PLACE unit in a similar way as the PLACE structure is independent of the channel estimator type.

##### 4.1. Pilot layers

Throughout this paper, the “consecutive roots-of-unity phase difference” training sequences are used as pilot layers. In case  $U = 1$ , the pilot layer is (cf., e.g., [23])

$$p[k] = \sqrt{P_p} \cdot e^{j(2\pi/K_{CE})kr}, \quad 1 \leq k \leq K_c, \quad (4)$$

where  $r$  is relatively prime to the observation length  $K_{CE}$ . All subsequences of length  $K_{CE}$  exhibit perfect autocorrelation. In case  $U > 1$ , multiple training sequences with low

<sup>1</sup> Soft chips and soft code symbols are the same for IDMA.

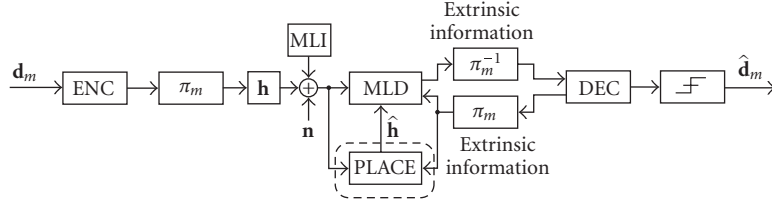


FIGURE 2: Iterative receiver structure for layer  $m$ . The user index is skipped. The layer-specific interleaver is denoted by  $\pi_m$ .

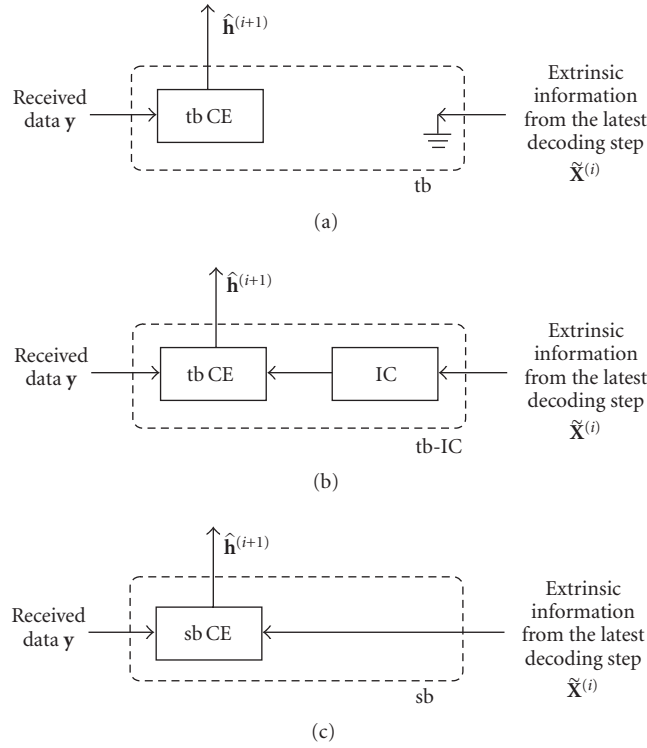


FIGURE 3: Illustration of training-based channel estimation (tb), training-based channel estimation with partial data interference cancellation (tb-IC), and semiblink channel estimation (sb) in the  $i$ th iteration. The PLACE unit in Figure 2 corresponds to one of these three.

cross-correlations are needed. We can construct a training sequence with observation length  $UK_{\text{CE}}$  based on (4) and sample this sequence with a sampling distance  $U$  and a user-specific sampling delay  $u-1$ . The resulting pilot layer for user  $u$  can be expressed as

$$\begin{aligned}
 p_u[k] &= \sqrt{P_{p,u}} \cdot e^{j(2\pi/UK_{\text{CE}})(Ukr+u-1)} \\
 &= \sqrt{P_{p,u}} \cdot e^{j(2\pi/K_{\text{CE}})kr} \cdot e^{(2\pi/UK_{\text{CE}})(u-1)} \\
 &= p[k] \cdot e^{(2\pi/UK_{\text{CE}})(u-1)}, \quad 1 \leq k \leq K_c, 1 \leq u \leq U,
 \end{aligned} \tag{5}$$

where  $r$  is relatively prime to  $U \cdot K_{\text{CE}}$ . The only difference to (4) is a user-specific phase offset. Note that (4) and (5) agree for  $U = 1$ . The latter result exhibits a perfect autocorrelation and a perfect cross-correlation property. It is used to obtain the numerical results with multiple users in Section 6.

#### 4.2. Joint least-squares channel estimation (JLSCE)

Least-squares channel estimation is a linear channel estimation technique that minimizes the average squared Euclidean distance between the received data and a replica of the received data based on channel estimates. Joint channel estimation is used to estimate multiple channels (in our case  $U$  channels) jointly. For JLSCE, the channels have to be assumed invariant over the observation length. To simplify the presentation, we firstly introduce different JLSCE schemes assuming block fading, that is, the channel coefficients are assumed to stay constant over the whole transmission block. In this case, the channel model (2) can be rewritten with a channel vector of length  $U \cdot (L + 1)$  as the channel coefficients are the same for all time indices. The resulting vectors and matrices are denoted with a subscript “ti.” Secondly, we will discuss how to approximate JLSCE in case of fast fading by means of sliding-window channel estimation. Finally, we present the minimum mean-squared error estimator taking time variations into account.

#### 4.2.1. Training-based JLSCE with and without partial data cancellation (tb-LS and tb-LS-IC)

The aim of tb-LS is to minimize  $E\{\|\mathbf{y} - \mathbf{P}_{\text{ti}} \cdot \hat{\mathbf{h}}_{\text{tb-LS}}\|_F^2\}$ , where  $F$  denotes the Frobenius norm. The channel estimates can be calculated as follows [24]:

$$\hat{\mathbf{h}}_{\text{tb-LS}} = \underbrace{(\mathbf{P}_{\text{ti}}^H \cdot \mathbf{P}_{\text{ti}})^{-1} \mathbf{P}_{\text{ti}}^H}_{\mathbf{P}_{\text{ti}}^\dagger} \cdot \mathbf{y}. \quad (6)$$

The mean-squared error (MSE) can be calculated as

$$\begin{aligned} v_{\text{tb-LS}} &= E\{\|\mathbf{h}_{\text{ti}} - \hat{\mathbf{h}}_{\text{tb-LS}}\|_F^2\} \\ &= \left( \sigma_n^2 + \sum_{u=1}^U M_u \cdot \sigma_{h_u}^2 \right) \cdot \text{trace}\{\mathbf{P}_{\text{ti}}^{\dagger H} \mathbf{P}_{\text{ti}}^\dagger\}. \end{aligned} \quad (7)$$

Note that for  $M = 0$ , this result collapses to the standard result for pure training. Note also that for  $M > 0$ , the MSE depends on the power profile of the estimated channel, which is not the case if the transmitted signal is perfectly known to the receiver.

In the case of partial data cancellation, the least-squares channel estimates can be calculated as

$$\hat{\mathbf{h}}_{\text{tb-LS-IC}}^{(i+1)} = \mathbf{P}_{\text{ti}}^\dagger \cdot \mathbf{y}^{(i)} \quad (8)$$

with MSE,

$$\begin{aligned} v_{\text{tb-LS-IC}}^{(i+1)} &= E\{\|\mathbf{h}_{\text{ti}} - \hat{\mathbf{h}}_{\text{tb-LS-IC}}^{(i+1)}\|_F^2\} \\ &= \left( \sigma_n^2 + \sum_{u=1}^U M_u \cdot \left( \sigma_{h_u}^2 \cdot \sigma_{x_u^{(i)}}^2 + v_{\text{tb-LS-IC}}^{(i)} \cdot (1 - \sigma_{x_u^{(i)}}^2) \right) \right) \\ &\quad \cdot \text{trace}\{\mathbf{P}_{\text{ti}}^{\dagger H} \mathbf{P}_{\text{ti}}^\dagger\}, \end{aligned} \quad (9)$$

where  $\sigma_{x_u^{(i)}}^2 \leq 1$  is the variance of the soft chips of user  $u$  in iteration  $i$ . Note that tb-LS and tb-LS-IC agree in the case that we have no information about the data, that is, all soft chips equal zero and  $\sigma_{x_u^{(i)}}^2 = 1$ . Different from (7), the MSE of tb-LS-IC depends on the variances of the soft chips.

In both cases, the trace of  $\mathbf{P}_{\text{ti}}^{\dagger H} \mathbf{P}_{\text{ti}}^\dagger$  should be minimized to obtain optimal MSE. This can be achieved if the pseudoinverse  $\mathbf{P}_{\text{ti}}^\dagger$  is unitary up to a scalar factor ( $\mathbf{P}_{\text{ti}}^{\dagger H} \mathbf{P}_{\text{ti}}^\dagger \sim \mathbf{I}$ ). Then

the trace can be calculated (see [23, 25]) as

$$\text{trace}\{\mathbf{P}_{\text{ti}}^{\dagger H} \mathbf{P}_{\text{ti}}^\dagger\} = \frac{U \cdot (L+1)}{K_{\text{CE}} \cdot P_p}, \quad (10)$$

where  $K_{\text{CE}}$  is the training window length, which is  $K_c - L$  in the case of block fading.

With (7) and (10), we can get a lower bound for the MSE with tb-LS as

$$v_{\text{tb-LS}} \geq \left( \sigma_n^2 + \sum_{u=1}^U M_u \cdot \sigma_{h_u}^2 \right) \cdot \frac{U \cdot (L+1)}{K_{\text{CE}} \cdot P_p} \quad (11)$$

$$= v_{\text{LB,tb-LS}}$$

$$\approx \left( R\sigma_n^2 + \sum_{u=1}^U b_u \cdot \sigma_{h_u}^2 \right) \cdot \frac{U \cdot (L+1)}{K_b \cdot P_p}, \quad (12)$$

where  $K_b$  is the block length in info bits. The latter approximation holds if  $K_c \gg L$ , which is usually the case. The right-hand side of (11) is the Cramer-Rao lower bound (CRLB) for a training-based unbiased estimator [19].

Combining (9) and (10) leads us to the MSE lower bound for tb-LS-IC:

$$\begin{aligned} v_{\text{tb-LS-IC}}^{(i+1)} &\geq \left( \sigma_n^2 + \sum_{u=1}^U M_u \cdot \left( \sigma_{h_u}^2 \cdot \sigma_{x_u^{(i)}}^2 + v_{\text{tb-LS-IC}}^{(i)} \cdot (1 - \sigma_{x_u^{(i)}}^2) \right) \right) \\ &\quad \cdot \frac{U \cdot (L+1)}{K_{\text{CE}} \cdot P_p} = v_{\text{LB,tb-LS-IC}}^{(i+1)} \end{aligned} \quad (13)$$

$$\geq \left( \sigma_n^2 + \sum_{u=1}^U M_u \cdot \sigma_{h_u}^2 \cdot \sigma_{x_u^{(i)}}^2 \right) \cdot \frac{U \cdot (L+1)}{K_{\text{CE}} \cdot P_p} \quad (14)$$

$$\begin{aligned} &= v_{\text{LLB,tb-LS-IC}}^{(i+1)} \\ &\approx \left( R\sigma_n^2 + \sum_{u=1}^U b_u \cdot \sigma_{h_u}^2 \cdot \sigma_{x_u^{(i)}}^2 \right) \cdot \frac{U \cdot (L+1)}{K_b \cdot P_p}. \end{aligned} \quad (15)$$

The loose lower bound  $v_{\text{LLB,tb-LS-IC}}^{(i+1)}$  is the MSE in case that the previous channel estimates are perfect. The lower bound  $v_{\text{LB,tb-LS-IC}}^{(i+1)}$  takes the MSE of the previous channel estimates into account.

We compare the MSE of both training-based approaches by calculating the ratio

$$\begin{aligned} \frac{v_{\text{tb-LS}}}{v_{\text{tb-LS-IC}}^{(i+1)}} &= \frac{(\sigma_n^2 + \sum_{u=1}^U M_u \cdot \sigma_{h_u}^2) \cdot \text{trace}\{\mathbf{P}_{\text{ti}}^{\dagger H} \mathbf{P}_{\text{ti}}^\dagger\}}{(\sigma_n^2 + \sum_{u=1}^U M_u \cdot (\sigma_{h_u}^2 \cdot \sigma_{x_u^{(i)}}^2 + v_{\text{tb-LS-IC}}^{(i)} \cdot (1 - \sigma_{x_u^{(i)}}^2))) \cdot \text{trace}\{\mathbf{P}_{\text{ti}}^{\dagger H} \mathbf{P}_{\text{ti}}^\dagger\}} \\ &= \frac{\sigma_n^2 + \sum_{u=1}^U M_u \cdot \sigma_{h_u}^2 \cdot 1}{\sigma_n^2 + \sum_{u=1}^U M_u \cdot (\sigma_{h_u}^2 \cdot \sigma_{x_u^{(i)}}^2 + v_{\text{tb-LS-IC}}^{(i)} \cdot (1 - \sigma_{x_u^{(i)}}^2))} \geq 1, \end{aligned} \quad (16)$$

where the latter inequality holds because  $\sigma_{\tilde{x}_u}^2 \leq 1$  (which is the case for  $i \geq 1$ ) and  $v_{\text{tb-LS-IC}}^{(i)} \leq \sigma_{h_u}^2$  is assumed. In the very first iteration ( $i = 0$ ) tb-LS and tb-LS-IC agree:  $\sigma_{\tilde{x}_u}^2 = 1 \Rightarrow v_{\text{tb-LS}}/v_{\text{tb-LS-IC}}^{(1)} = 1$ . The MSE of tb-LS and tb-LS-IC is also the same in the case that  $v_{\text{tb-LS-IC}}^{(i)} = \sigma_{h_u}^2$ . We conclude from this comparison that tb-LS-IC outperforms tb-LS independent of the pilot data matrix  $\mathbf{P}_{\text{ti}}$ , that is,  $v_{\text{tb-LS-IC}}^{(i+1)} \leq v_{\text{tb-LS}}$ . As this conclusion is independent of the pilot data, it is especially true if the pilot data matrix is optimized to reach the MSE lower bound, that is, we can conclude that  $v_{\text{LLB,tb-LS-IC}}^{(i+1)} \leq v_{\text{LB,tb-LS-IC}}^{(i+1)} \leq v_{\text{LB,tb-LS}}$ .

#### 4.2.2. Semiblind JLSCE (sb-LS)

It is shown in [26] that for blind channel estimation, the least-squares channel estimates can be obtained by using soft data symbols instead of perfectly known pilot data. If we extend the result to joint estimation of multiple channels with a combined knowledge of pilot data and soft chips (which can be interpreted as “virtual” training), we obtain semiblind joint least-squares channel estimates as

$$\hat{\mathbf{h}}_{\text{sb-LS}}^{(i+1)} = \underbrace{((\tilde{\mathbf{X}}_{\text{ti}}^{(i)} + \mathbf{P}_{\text{ti}})^H (\tilde{\mathbf{X}}_{\text{ti}}^{(i)} + \mathbf{P}_{\text{ti}}))^{-1} (\tilde{\mathbf{X}}_{\text{ti}}^{(i)} + \mathbf{P}_{\text{ti}})^H}_{(\tilde{\mathbf{X}}_{\text{ti}}^{(i)} + \mathbf{P}_{\text{ti}})^\dagger} \cdot \mathbf{y}. \quad (17)$$

The MSE can be calculated as

$$\begin{aligned} v_{\text{sb-LS}}^{(i+1)} &= E\{\|\hat{\mathbf{h}}_{\text{sb-LS}}^{(i+1)} - \hat{\mathbf{h}}_{\text{sb-LS}}^{(i+1)}\|_F^2\} \\ &= \left( \sigma_n^2 + \sum_{u=1}^U M_u \cdot \sigma_{h_u}^2 \cdot \sigma_{\tilde{x}_u}^2 \right) \\ &\quad \cdot \text{trace}\left\{ (\tilde{\mathbf{X}}_{\text{ti}}^{(i)} + \mathbf{P}_{\text{ti}})^\dagger (\tilde{\mathbf{X}}_{\text{ti}}^{(i)} + \mathbf{P}_{\text{ti}}) \right\}. \end{aligned} \quad (18)$$

A lower bound of the MSE is obtained in the case where  $(\tilde{\mathbf{X}}_{\text{ti}}^{(i)} + \mathbf{P}_{\text{ti}})^\dagger$  is unitary up to a scaling factor, which leads to

$$\begin{aligned} v_{\text{sb-LS}}^{(i+1)} &\geq \left( \sigma_n^2 + \sum_{u=1}^U M_u \cdot \sigma_{h_u}^2 \cdot \sigma_{\tilde{x}_u}^2 \right) \\ &\quad \cdot \frac{U \cdot (L+1)}{K_{\text{CE}} \cdot (P_p + \sum_{u=1}^U M_u \cdot |\tilde{x}_u^{(i)}|^2)} \end{aligned} \quad (19)$$

$$= v_{\text{LB,sb-LS}}^{(i+1)} \quad (20)$$

$$\begin{aligned} &\approx \left( R\sigma_n^2 + \sum_{u=1}^U b_u \cdot \sigma_{h_u}^2 \cdot \sigma_{\tilde{x}_u}^2 \right) \\ &\quad \cdot \frac{U \cdot (L+1)}{K_b \cdot (P_p + \sum_{u=1}^U (b_u/R) \cdot |\tilde{x}_u^{(i)}|^2)}. \end{aligned} \quad (21)$$

Note that in the very first iteration,  $\tilde{\mathbf{X}}_{\text{ti}}^{(0)} = \mathbf{0}$  holds so that sb-LS reduces to tb-LS ((17) equals (6) and consequently (18) equals (7)) and the same conclusions for the choice of the pilot data matrix hold, especially the lower bound of (11) and its approximation (12).

A comparison of the lower bounds for tb-LS-IC and sb-LS,

$$\frac{v_{\text{LLB,tb-LS-IC}}^{(i+1)}}{v_{\text{LB,sb-LS}}^{(i+1)}} = \frac{P_p + \sum_{u=1}^U M_u \cdot |\tilde{x}_u^{(i)}|^2}{P_p} \geq 1, \quad (22)$$

reveals that sb-LS outperforms tb-LS-IC if  $v_{\text{LB,sb-LS}}^{(i+1)}$  is reached. For the training-based approaches, the MSE lower bounds can easily be reached by an optimal choice of the pilot sequence, for example, as proposed in [23]. In the case of semiblind channel estimation, such a design is impossible as the data is random. Even in the case of optimal pilot sequences, that is,  $\mathbf{P}_{\text{ti}}^\dagger$  is unitary up to a scalar factor, the lower bound cannot be reached due to the random data. Therefore it is interesting to investigate the MSE performance of sb-LS with random data and to compare it to the lower bound.

#### 4.2.3. Comparison of MSE performances

As a conclusion to the discussion above, we can state that  $v_{\text{LB,sb-LS}}^{(i+1)} \leq v_{\text{LB,tb-LS-IC}}^{(i+1)} \leq v_{\text{LB,tb-LS}}$  and that  $v_{\text{tb-LS-IC}}^{(i+1)} \leq v_{\text{tb-LS}}$ . In this subsection, we illustrate the results obtained so far. To concentrate on the main aspects, we choose  $U = 1$  and skip the user index. We assume a frequency-flat channel ( $L = 0$ ) so that the overall number of channel coefficients is  $U \cdot (L+1) = 1$ . The data is modeled as Gaussian-distributed noise with zero mean and variance 10, that is,  $M = 10$ . The average channel power  $\sigma_h^2$ , the system load  $b$ , the noise variance  $\sigma_n^2$ , and the pilot layer power  $P_p$  are chosen to be 1, that is, the code rate is  $R = b/M = 1/10$ . Simulated MSE results for the different channel estimators and the corresponding lower bounds are depicted in Figure 4 for an observation length of  $K_{\text{CE}} = 10$  (or equivalently  $K_b = 1$ ). Optimal pilot sequences are used. We can see that all curves match if the channel estimator does not have any information about the data ( $M \cdot |\tilde{x}|^2 = 0$ ). The tb-LS cannot make use of the information about the data, its MSE is constant. The tb-LS-IC outperforms tb-LS and sb-LS outperforms tb-LS-IC in all cases, which coincides with the discussion above. The MSE of tb-LS-IC depends on the MSE of the previous channel estimates, which is also shown in Figure 4. But even in the best case with  $v_{\text{tb-LS-IC}}^{(i)} = 0$ , sb-LS significantly outperforms tb-LS-IC. Due to the choice of the pilot sequence, the training-based schemes both reach the lower bound. This is not the case for sb-LS, because the random data does not lead to an optimal matrix  $\tilde{\mathbf{X}}_{\text{ti}}^{(i)} + \mathbf{P}_{\text{ti}}$ .

In Figure 5, we depict a comparison between the lower bound and the simulated MSE for sb-LS with different training lengths. All other parameters are as described before. We can see that sb-LS reaches its lower bound even for random data if the observation length is long enough, that is, at least 20 chips. In other words, for an observation length above 20 chips, Gaussian-distributed data is optimal in the sense of minimizing the MSE of the bias-free channel estimates. This result is especially interesting in the context of IDMA, where the superimposed data layers can be well approximated as a Gaussian random variable due to the central limit theorem.

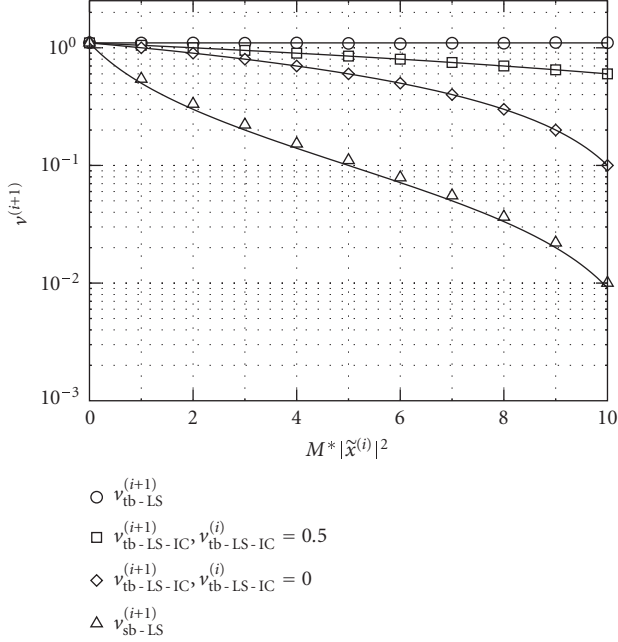


FIGURE 4: MSE versus soft chip power for training-based and semiblind LS channel estimators with optimal pilot data matrix. Results for  $U = 1, R = 1/10, b = 1, \sigma_n^2 = 1, \sigma_h^2 = 1, P_p = 1, K_{CE} = 10$ . Symbols show the simulated MSE values and lines show the corresponding lower bounds using (11), (13), (14), and (19), respectively.

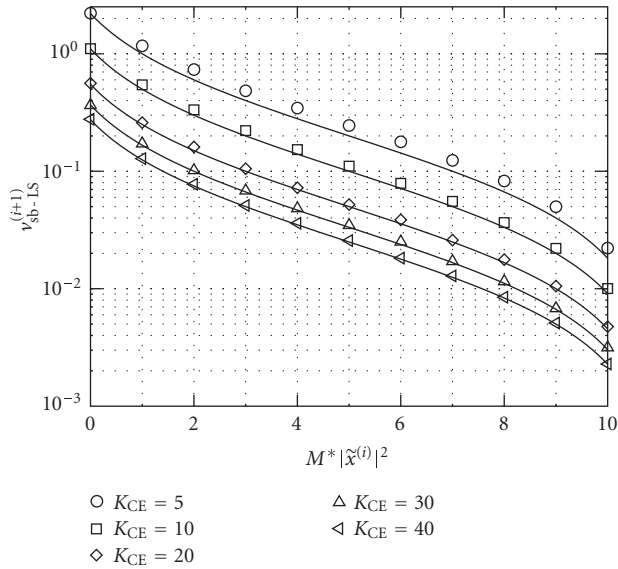


FIGURE 5: MSE versus soft chip power for sb-LS with optimal pilot data matrix. Results for  $U = 1, R = 1/10, b = 1, \sigma_n^2 = 1, \sigma_h^2 = 1, P_p = 1$ . Symbols show the simulated MSE values and lines show the corresponding lower bounds using (19).

4.2.4. Sliding-window JLSCE (sw-LS)

As mentioned before, JLSCE is only suitable for time-invariant channels. However, our goal is to estimate fast-fading channels. If we still want to apply JLSCE, we have to

make sure that the channel is approximately invariant over the observation length  $K_{CE}$ . This is actually possible if we can assume the fading rate of the channel to be upper-limited. Let us assume that the length of each chip  $x_{u,m}[k]$  is  $T_c$ . Let  $f_c$  denote the carrier frequency. Let furthermore  $v$  be the velocity of the mobile user, and let  $c_0$  be the speed of light in vacuum. Then the maximum possible frequency shift due to the Doppler effect normalized by the chip rate is

$$f_{D,max} \cdot T_c = f_c \cdot \frac{v}{c_0} \cdot T_c. \tag{23}$$

In the case that  $K_{CE} \ll (f_{D,max} \cdot T_c)^{-1}$ , the channel can approximately be assumed to be invariant over the observation length  $K_{CE}$ . Therefore, the derived LS channel estimators can be applied to a window of the received sequence. The estimated channel coefficients of a window are of course only valid for this particular window. Therefore, we have to shift the window and perform JLSCE for every shifted window to obtain channel estimates for the complete received sequences. We refer to this as sliding-window JLSCE (sw-LS). This approach can be used for tb-LS(-IC) and sb-LS and we will refer to it as sw-tb-LS(-IC) and sw-sb-LS, respectively. Note that the results obtained in the discussion above are also valid for sw-LS.

Another alternative is to take the fading characteristics of the channel properly into account, which is optimally done in Section 4.3. The drawback of doing this is the high computational complexity compared to sw-LS, which keeps the sliding-window method attractive from a practical point of view.

4.3. Semiblind joint minimum mean-squared error channel estimation (sb-MMSE)

In the following, the optimal semiblind linear joint channel estimator is derived in the sense of minimizing the mean-squared error of the channel estimates. This optimization criterion is different from the LS approach in Section 4.2.2 and allows us to take the statistical fading characteristics properly into account. Due to its linearity, the derived channel estimator is optimal in the MMSE sense if and only if the channel coefficients to be estimated are Gaussian distributed. As we concentrate on Rayleigh fading channels, this assumption is fulfilled throughout this paper. For other distributions, a nonlinear approach might be necessary to find the MMSE solution. This issue is out of the scope of this paper. However, as already mentioned in the beginning of this section, the channel estimator type does not influence the general PLACE structure proposed in this paper.

The iteration number is skipped throughout this subsection to enhance the readability. Let  $\mathbf{h}[k]$  consist of the  $U \cdot (L + 1)$  elements of  $\mathbf{h}$  with chip index  $k$  and let  $\hat{\mathbf{h}}[k]$  denote its MMSE estimate, which is the solution to the well-known Wiener-Hopf equation:

$$\hat{\mathbf{h}}_{sb-MMSE}[k] = \underbrace{\mathbf{R}_{hy}[k] \cdot \mathbf{R}_{yy}^{-1}}_{\mathbf{W}[k]} \cdot \mathbf{y}. \tag{24}$$

The matrices in (24) are calculated as follows:

$$\begin{aligned} \mathbf{R}_{hy}[k] &= E_{h,y}\{\mathbf{h}[k] \cdot \mathbf{y}^H\} \\ &= E_{h,X}\{\mathbf{h}[k] \cdot \mathbf{h}^H \cdot (\mathbf{X} + \mathbf{P})^H\} \\ &= E_h\{\mathbf{h}[k] \cdot \mathbf{h}^H\} \cdot (E_X\{\mathbf{X}^H\} + \mathbf{P}^H) \\ &= \mathbf{R}_{hh}[k] \cdot (\tilde{\mathbf{X}}^H + \mathbf{P}^H), \end{aligned} \quad (25)$$

$$\begin{aligned} \mathbf{R}_{yy} &= E_y\{\mathbf{y} \cdot \mathbf{y}^H\} \\ &= E_{X,h}\{(\mathbf{X} + \mathbf{P}) \cdot \mathbf{h} \cdot \mathbf{h}^H \cdot (\mathbf{X} + \mathbf{P})^H\} + \sigma_n^2 \mathbf{I} \\ &= E_X\{(\mathbf{X} + \mathbf{P}) \cdot \mathbf{R}_{hh} \cdot (\mathbf{X} + \mathbf{P})^H\} + \sigma_n^2 \mathbf{I} \\ &= E_X\{\mathbf{X} \cdot \mathbf{R}_{hh} \cdot \mathbf{X}^H\} + \tilde{\mathbf{X}} \cdot \mathbf{R}_{hh} \cdot \mathbf{P}^H \\ &\quad + \mathbf{P} \cdot \mathbf{R}_{hh} \cdot \tilde{\mathbf{X}}^H + \mathbf{P} \cdot \mathbf{R}_{hh} \cdot \mathbf{P}^H + \sigma_n^2 \mathbf{I}, \end{aligned} \quad (26)$$

where  $\mathbf{X} = \tilde{\mathbf{X}} + \Delta\mathbf{X}$  is the sum of the fixed soft chip matrix  $\tilde{\mathbf{X}}$  based on the decoder output values and  $\Delta\mathbf{X}$ , which is a random variable. The remaining term in (26) is

$$\begin{aligned} &E_X\{\mathbf{X} \cdot \mathbf{R}_{hh} \cdot \mathbf{X}^H\} \\ &= E_{\Delta X}\{\tilde{\mathbf{X}} \cdot \mathbf{R}_{hh} \cdot \tilde{\mathbf{X}}^H + \tilde{\mathbf{X}} \cdot \mathbf{R}_{hh} \cdot \Delta\mathbf{X}^H \\ &\quad + \Delta\mathbf{X} \cdot \mathbf{R}_{hh} \cdot \tilde{\mathbf{X}}^H + \Delta\mathbf{X} \cdot \mathbf{R}_{hh} \cdot \Delta\mathbf{X}^H\} \\ &= \tilde{\mathbf{X}} \cdot \mathbf{R}_{hh} \cdot \tilde{\mathbf{X}}^H + E_{\Delta X}\{\Delta\mathbf{X} \cdot \mathbf{R}_{hh} \cdot \Delta\mathbf{X}^H\} \\ &= \tilde{\mathbf{X}} \cdot \mathbf{R}_{hh} \cdot \tilde{\mathbf{X}}^H + \mathbf{\Gamma}', \end{aligned} \quad (27)$$

where  $\mathbf{\Gamma}'$  is a diagonal matrix with entries

$$\sum_{u=1}^U \sum_{m=1}^{M_u} \sum_{l=0}^L \sigma_{h_{u,l}}^2 \cdot \sigma_{\Delta x_{u,m}}^2 [k-l], \quad L \leq k \leq K_c. \quad (28)$$

Let  $\mathbf{\Gamma} \doteq \mathbf{\Gamma}' + \sigma_n^2 \mathbf{I}$  be the diagonal noise matrix. Then (26) and (27) can be combined to obtain

$$\begin{aligned} \mathbf{R}_{yy} - \mathbf{\Gamma} &= \tilde{\mathbf{X}} \cdot \mathbf{R}_{hh} \cdot \tilde{\mathbf{X}}^H + \tilde{\mathbf{X}} \cdot \mathbf{R}_{hh} \cdot \mathbf{P}^H \\ &\quad + \mathbf{P} \cdot \mathbf{R}_{hh} \cdot \tilde{\mathbf{X}}^H + \mathbf{P} \cdot \mathbf{R}_{hh} \cdot \mathbf{P}^H \\ &= (\tilde{\mathbf{X}} + \mathbf{P}) \cdot \mathbf{R}_{hh} \cdot (\tilde{\mathbf{X}} + \mathbf{P})^H. \end{aligned} \quad (29)$$

Combining the intermediate results from (24) to (29), the MMSE channel estimates are obtained as

$$\begin{aligned} \hat{\mathbf{h}}_{\text{sb-MMSE}}[k] &= \mathbf{W}[k] \cdot \mathbf{y} = \mathbf{R}_{hh}[k] (\tilde{\mathbf{X}}^H + \mathbf{P}^H) \\ &\quad \cdot \left( (\tilde{\mathbf{X}} + \mathbf{P}) \mathbf{R}_{hh} (\tilde{\mathbf{X}} + \mathbf{P})^H + \mathbf{\Gamma} \right)^{-1} \cdot \mathbf{y}. \end{aligned} \quad (30)$$

Equation (30) corresponds to the optimal semiblind channel estimator. The computational complexity of this estimator is dominated by the inversion of a matrix with a row/column length growing linearly with the number of chips per layer. Note that the computational complexity of sw-LS (cf. Section 4.2.4) is also dominated by a matrix inversion, but with a row/column length only growing linearly with the channel memory. Therefore, the computational complexity is typically much lower for sw-LS.

Note that in the case that no information about the data is used, the result degenerates to purely training-based joint MMSE channel estimation. The MSE performance of

training-based joint MMSE channel estimation can be improved by partially canceling the data interference before channel estimation—just like for tb-LS-IC. We refer to this as training-based joint MMSE channel estimation with partial data interference cancellation (tb-MMSE-IC).

Let  $\mathbf{v}_h[k] = E\{\mathbf{h}[k] \odot \mathbf{h}^*[k]\}$ —where  $\odot$  denotes the scalar product—be a vector containing the channel variances at chip index  $k$ . Then the MSE of the channel coefficients obtained by MMSE channel estimation can easily be shown to be

$$\mathbf{v}_{\text{sb-MMSE}}[k] = \mathbf{v}_h[k] - \text{diag}\{\mathbf{R}_{hy}[k] \cdot \mathbf{R}_{yy}^{-1} \cdot \mathbf{R}_{hy}^H[k]\} \quad (31)$$

and the overall MSE of the channel estimates at chip index  $k$  is

$$\begin{aligned} v_{\text{sb-MMSE}}[k] &= E\{ \|\mathbf{h}[k] - \hat{\mathbf{h}}[k]\|_F^2 \} \\ &= \underbrace{\sum_{u=1}^U \sum_{l=0}^L \sigma_{h_{u,l}}^2}_{\sigma_h^2} - \text{trace}\{\mathbf{R}_{hy}[k] \cdot \mathbf{R}_{yy}^{-1} \cdot \mathbf{R}_{hy}^H[k]\}. \end{aligned} \quad (32)$$

In case of block fading, the channel coefficients agree for all time indices and (30) can be rewritten as

$$\begin{aligned} \hat{\mathbf{h}}_{\text{sb-MMSE}} &= \mathbf{I}(\tilde{\mathbf{X}} + \mathbf{P})^H \\ &\quad \cdot ((\tilde{\mathbf{X}}_{\text{ti}} + \mathbf{P}_{\text{ti}}) \mathbf{I}(\tilde{\mathbf{X}}_{\text{ti}} + \mathbf{P}_{\text{ti}})^H + \mathbf{\Gamma}_{\text{ti}})^{-1} \cdot \mathbf{y} \\ &= ((\tilde{\mathbf{X}}_{\text{ti}} + \mathbf{P}_{\text{ti}})^H \mathbf{\Gamma}_{\text{ti}}^{-1} (\tilde{\mathbf{X}}_{\text{ti}} + \mathbf{P}_{\text{ti}}) + \mathbf{I})^{-1} \\ &\quad \times (\tilde{\mathbf{X}}_{\text{ti}} + \mathbf{P}_{\text{ti}})^H \mathbf{\Gamma}_{\text{ti}}^{-1} \cdot \mathbf{y}, \end{aligned} \quad (33)$$

where we applied the matrix inversion lemma to obtain the last equation. The latter expression has significantly lower computational complexity than the former one as the size of the inverse matrix is significantly lower but it can only be applied to estimate time-invariant channels. For the expression with time-varying channel coefficients (30), the application of the matrix inversion lemma does not lead to decreased computational complexity, which makes sb-MMSE rarely attractive from a complexity point of view if the block lengths are not short. Its significance rather lies in its optimality and we will use sb-MMSE to verify the performance of the suboptimal but low-complexity sw-LS in Section 6.

## 5. MULTILAYER DETECTION (MLD): INTERFERENCE CANCELATION AND DETECTION

After channel estimation, multilayer detection (MLD) is performed. A common low-complexity approach for MLD is to cancel out interfering layers before detection and to perform the detection only on the layer of interest. The same concept is used for all numerical results in Section 6. We therefore give a short description of this type of MLD for convenience.

The interference cancelation is done in a parallel fashion and is based on soft chip values from the decoder. All layers from all active users are simultaneously taken into account. In case of perfect channel knowledge and soft chips matching the transmitted chips, the transmission is interference-free for all layers. In this ideal case, the performance is the



same as if only one single layer would access the channel. The single-layer bit error probability (single-layer performance, SLP) therefore provides a lower bound. As the interference cancelation is not perfect, some remaining interference still disturbs the detection. This remaining interference may be modeled as Gaussian-distributed noise, which is the so-called Gaussian assumption. The computational complexity of this suboptimal MLD grows only linearly with the number of layers  $M$  and the number of channel coefficients  $L + 1$ . Note that the computational complexity of the optimal MLD in the MAP sense grows exponentially with both parameters which is infeasible and makes a suboptimal MLD inevitable.

### 5.1. Interference cancelation and Gaussian assumption

The estimated received value for chip index  $k$  in iteration  $i + 1$  is

$$\begin{aligned} \hat{y}^{(i+1)}[k] &= \sum_{u=1}^U \sum_{l=0}^L \hat{h}_{u,l}^{(i+1)}[k] \cdot \sum_{m=1}^{M_u} \tilde{x}_{u,m}^{(i)}[k-l] \\ &+ \sum_{u=1}^U \sum_{l=0}^L \hat{h}_{u,l}^{(i+1)}[k] \cdot p_u[k-l]. \end{aligned} \quad (34)$$

The task of the interference canceler (IC) is to subtract interference from the received signal. Which part of the received signal is to be interpreted as interference depends on the detector. Throughout this paper, we concentrate on the low-complexity soft rake detector [4]. The derivations of ICs for other detector types are similar.

If the soft rake detector is used, the detector input for layer  $\mu$  of user  $v$  with delay  $\lambda$  at chip index  $k$  in iteration  $i + 1$  is

$$\begin{aligned} \check{y}_{v,\mu,\lambda}^{(i+1)}[k] &= y[k] - (\hat{y}^{(i+1)}[k] - \hat{h}_{v,\lambda}^{(i+1)}[k] \cdot \tilde{x}_{v,\mu}^{(i)}(k-\lambda)) \\ &= h_{v,\lambda}[k] \cdot x_{v,\mu}[k-\lambda] + \eta_{v,\mu,\lambda}^{(i+1)}[k], \end{aligned} \quad (35)$$

where  $\eta_{v,\mu,\lambda}^{(i+1)}[k]$  is the noise at the detector input. Let furthermore  $\sigma_{\tilde{x}_{u,m}^{(i)}}^2[k]$  denote the variance of the soft chip  $\tilde{x}_{u,m}^{(i)}[k]$ , which in our case can be calculated as  $1 - |\tilde{x}_{u,m}^{(i)}[k]|^2$ , and let  $\hat{P}_{h_{u,l}}^{(i+1)}[k] \doteq |\hat{h}_{u,l}^{(i+1)}[k]|^2$  be the power estimate of the channel coefficient  $h_{u,l}[k]$  in iteration  $i + 1$ . If we assume the channel estimates to be perfect, the expectation and variance of the noise at the detector input can be calculated as

$$\begin{aligned} E\{\eta_{v,\mu,\lambda}^{(i+1)}[k]\} &= 0, \\ E\{|\eta_{v,\mu,\lambda}^{(i+1)}[k]|^2\} &= \sum_{u=1}^U \sum_{l=0}^L \hat{P}_{h_{u,l}}^{(i+1)}[k] \cdot \sum_{m=1}^{M_u} \sigma_{\tilde{x}_{u,m}^{(i)}}^2[k-l] \\ &\quad - \hat{P}_{h_{v,\lambda}}^{(i+1)}[k] \cdot \sigma_{\tilde{x}_{v,\mu}^{(i)}}^2[k-\lambda] + \sigma_n^2. \end{aligned} \quad (36)$$

### 5.2. Soft rake detection

Based on the remaining signal after IC, the soft detector calculates the log-likelihood ratios (LLRs) of the chips given the Gaussian assumption (i.e., the remaining interference is modeled as Gaussian noise) and the channel knowledge/estimates. For soft rake detection,  $L + 1$  log-likelihood ratios per chip are calculated (one for each received sample influenced by this chip) and summed up to obtain the LLR of the chip. The LLR of the chip in layer  $m$  of user  $u$  at chip index  $k$  in iteration  $i + 1$  is

$$\begin{aligned} L_{u,m}^{(i+1)}[k] &\doteq \sum_{l=0}^L L_{l,u,m}^{(i+1)}[k] \\ &\doteq \sum_{l=0}^L L_l^{(i+1)}(X_{u,m}[k] | \check{y}_{u,m,l}^{(i+1)}[k+l], \hat{h}_{u,l}[k+l]) \\ &= \sum_{l=0}^L 4 \cdot \frac{\text{Re}\{\hat{h}_{u,l}^{*(i+1)}[k+l] \cdot \check{y}_{u,m,l}^{(i+1)}[k+l]\}}{E\{|\eta_{u,m,l}^{(i+1)}[k+l]|^2\}}. \end{aligned} \quad (37)$$

## 6. NUMERICAL RESULTS

In this section, the performance of the iterative MLD introduced in Section 5 with the channel estimators derived in Section 4 is investigated by means of Monte Carlo bit error rate simulations. Results for perfect channel knowledge serve as a reference. The channel codewords and the layer-specific interleavers are chosen randomly as described in Section 2. All results in this section are obtained by performing 10 iterations. If not stated explicitly, the ratio of power per info bit and noise power is fixed to  $E_b/N_0 = 10$  dB and a block length of  $K_b = 20$  is used. We concentrate on a fully loaded system ( $b = 1$ ) with a code rate of  $R = 1/10$ . This results in  $K_c = 200$  chips per layer and  $M \cdot K_c \cdot R = 200$  info bits are transmitted per block. This very short block length is particularly interesting in systems asking for low latency, for example, link adaptation [8]. Note that iterative detection, decoding, and channel estimation for such short block lengths are only possible if the interleaver length is long enough to break the correlations between the soft information that is shuffled between the receiver stages. A unique feature of IDMA is that the interleaver length is maximized, that is, the interleaver length is equal to  $K_c$ . Note that a comparable DS-CDMA system with the same system load uses an interleaver length of  $K_c \cdot R$ , which would be only 20 in our example. Such a small interleaver length leads to high correlations in the iterative receiver and is therefore not suitable, which motivates IDMA for low-latency transmissions.

The pilot layers are designed as described in Section 4.1. For the Rayleigh fading channels, Jakes spectrum is assumed. For frequency-selective channels,  $L = 4$  with a constant power profile is used. Channel coefficients with different delays and/or different user indices are assumed to be statistically independent. The number of receive antennas is fixed to be one throughout this paper. For the sliding-window

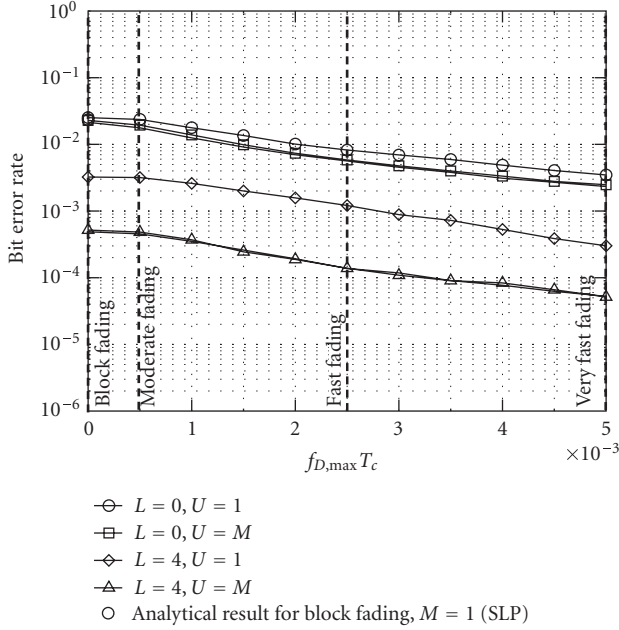


FIGURE 6: Bit error rates with perfect channel knowledge versus fading rate at  $E_b/N_0 = 10$  dB with 10 receiver iterations. The block length is  $K_b = 20$ , the code rate is  $R = 1/10$ , and the system load is  $b = 1$ . Time, frequency, and multiuser diversity effects improve the bit error performance. The maximum Doppler frequency is normalized to the chip rate. The thick lines show the SLP (upper line for frequency-flat ( $L = 0$ ), and lower line for frequency-selective ( $L = 4$ ) fading).

method, we choose a window length of  $1/10 \cdot f_{D,max} \cdot T_c$ . In case of block fading, the window length equals the block length in chips  $K_c$ .

Firstly, we investigate different diversity effects with perfect channel knowledge. Afterwards, we turn to results with the high-complexity MMSE channel estimator derived in Section 4.3. Finally, we investigate the performance of the low-complexity suboptimal sliding-window channel estimator from Section 4.2.4.

### 6.1. Perfect channel knowledge

Let us first consider perfect channel knowledge. The following results lead us to some interesting conclusions regarding the impact of different diversity effects on the bit error performance. In Figure 6, the bit error rates for different fading rates of different fading channels are depicted. The maximum Doppler frequency is normalized with respect to the chip rate. The analytical result for the bit error probability of BPSK transmission over a Rayleigh block fading channel is also depicted for comparison.

It can clearly be seen that the performance improves with the fading rate, which can be explained by the time diversity effect. Due to the chip-by-chip processing, reliable chip decisions help to improve weak chip decisions in subsequent iterations. This effect is even stronger when transmitting over a

frequency-selective channel. In this case, the iterative receiver can make use of diversity in time and in frequency. Figure 6 also shows the result for the case of independent fading channels ( $U = M$ ,  $M_u = 1$  for all  $u$ ) with different memory lengths. The independency of the channel coefficients of the single users (multiuser diversity) can be interpreted as space diversity, which improves the error performance compared to the case of a common channel.

Note that single-layer performance (SLP) is obtained in all depicted cases with multiple users, that is, there is virtually no loss in power efficiency compared to the case without MAI. We therefore obtain a quasiorthogonal multiple access without the need for orthogonal design—even for frequency-selective fading channels.

For convenience, we refer to some fading rates with the terms given in Table 1. The velocities are calculated assuming a chip duration of  $T_c \approx 260$  nanoseconds like that used in UMTS [18] and a carrier frequency of  $f_c = 2$  GHz using (23). These velocities are interpretations of the normalized maximum Doppler frequency for typical 3G system parameters in use today. We use these high values to demonstrate that the proposed semiblind scheme is not only able to track fast-fading channels and make use of the inherent diversity, but also to show the limits of the different channel estimators under consideration. An alternative interpretation are transmissions with a significantly higher carrier frequency and/or shorter chip duration. If we increase the carrier frequency to  $f_c = 50$  GHz and decrease the chip duration by a factor of 4, the resulting velocities are 100 times lower than in the example above. This would allow for mobile radio with mm-waves. Another example is acoustical underwater communication, where the speed of light ( $\approx 3 \cdot 10^8$  m/s) has to be exchanged by the speed of sound, which is typically  $\approx 1500$  m/s and therefore much less. This also leads to significantly reduced velocities in combination with typical values for the carrier frequency and the chip duration.

### 6.2. MMSE channel estimation

Let us now turn to MMSE channel estimation. Numerical bit error results for frequency-flat and frequency-selective Rayleigh fading channels are depicted in Figure 7 for tb-MMSE-IC and in Figure 8 for sb-MMSE and sw-sb-LS, respectively. In both plots, the bit error rates for perfect channel knowledge are depicted as well, which serve as a lower bound of the bit error rates with channel estimation. To allow for a fair comparison, the power loss due to the pilot layer is considered in these and all the following results for perfect channel knowledge.

As observed before, the bit error performance again improves for higher fading rates. Note that the bit error performance degrades for higher pilot-layer power. This is due to the constant  $E_b/N_0$ . When assuming a constant noise level, the power per transmitted info bit is kept constant. This also includes the power of the pilot layer. So the power of the data layers is reduced by the power that is spent for the pilot layer which results in a higher bit error rate. The improvement of the channel estimates and the power loss due to the pilot layer

TABLE 1: Normalized maximum Doppler frequencies and corresponding velocities. A chip duration of  $T_c \approx 260$  nanoseconds and a carrier frequency of  $f_c = 2$  GHz is assumed.

Term	$f_{D,\max} \cdot T_c$	$v$ in km/h at 2 GHz
Block fading	0	0
Moderate fading	0.0005	1038
Fast fading	0.0025	5189
Very fast fading	0.005	10377

lead to a local optimum of the bit error rate in all results with channel estimation.

Both MMSE estimators reach the performance with perfect channel knowledge, but sb-MMSE is significantly more power-efficient. The optimum is reached at 0.4 dB loss.

The results for the case of frequency-selective Rayleigh fading show that tb-MMSE-IC only gets close to the performance with perfect channel knowledge for block fading and a power loss of more than 2 dB. On the other hand, sb-MMSE is able to track even the fast channel virtually perfectly with a power loss of less than 0.4 dB.

### 6.3. LS channel estimation

As mentioned in Section 4.3, the computational complexity of sb-MMSE limits its practical usability—especially for larger block lengths. Therefore we apply sw-sb-LS and compare the resulting bit error performance on a Rayleigh fading channel in Figure 8.

For the frequency-flat block fading channel and also for the fast-fading channel, the bit error performance with sb-MMSE and sw-sb-LS is very similar. For very fast fading, sb-MMSE outperforms sw-sb-LS. The sliding-window channel estimator suffers from a short observation/window length of only  $K_{CE} = 20$ , which leads to suboptimal pilot data matrices (cf. Section 4) and additionally to a reduced training power. However, the difference of the power loss compared to sb-MMSE is only about 0.3 dB.

For a frequency-selective channel, sw-sb-LS is competitive to sb-MMSE in case of block fading. For fast fading, a loss of less than 1 dB occurs. Very fast training cannot be tracked.

We conclude from this comparison that sb-MMSE outperforms sw-sb-LS in terms of bit error probability and that sw-sb-LS is a competitive channel estimator for reduced requirements regarding fading rate and/or channel memory. We observe a tradeoff between performance and complexity.

### 6.4. Block length and number of users

By increasing the block length to  $K_b = 200$  or equivalently  $K_c = 2000$  and assuming the same fading rates as before, one obtains the results depicted in Figure 9. Results for sb-MMSE are not depicted due to its computational complexity.

Even very fast frequency-flat fading channels can be tracked with marginal loss by sw-sb-LS. As observed for  $K_b = 20$ , in case of frequency-selective fading a loss of less

than 1 dB occurs for fast fading, and very fast fading can not be tracked. For moderate fading, sw-sb-LS leads to virtually the same performance as perfect channel knowledge. Note that “moderate” still means more than 1000 km/h and that for fast fading, the bit error rates are significantly decreased compared to the results with shorter block length due to the improved time diversity.

Figure 10 shows the bit error rates for independent frequency-flat block fading channels and different block lengths. We consider the case  $U = M$ , that is, every user is assigned one data layer. This is the worst case from the viewpoint of channel estimation as the number of channels is maximized. The bit error performance improves for larger block lengths. For block fading, the window length agrees with the block length  $K_c$ . Therefore, the MSE of the initial estimate decreases reciprocally (cf. (18)) with the block length when fixing the power loss.

### 6.5. Channel coding

The numerical results presented so far are for a (10,1) random channel code without coding gain. In this section, additional results for a rate 1/2 convolutional code with generators  $\{5, 7\}$  followed by a rate 1/5 repetition code are presented, referred to as  $\{5, 7\} \times \{5, 5\}$  code. The generators are given in octal notation. The overall rate of both codes is 1/10, that is, the system load is still  $b = 1$ . As done in [27], the chip sequences are scrambled before transmission to ensure zero-mean receive samples.

Figure 11 shows the improvements due to coding compared to the  $\{5, 7\} \times \{5, 5\}$  code, which provides no coding gain. It can be observed that the loss compared to the case of perfect channel knowledge is negligible in both depicted cases provided that  $E_b/N_0$  is high enough.

## 7. CONCLUSIONS

IDMA is a power- and bandwidth-efficient multiple-access scheme. The diversity of frequency-selective (fast-) fading channels can be constructively used by a low-complexity iterative receiver to improve the bit error performance, provided that the channel is perfectly known at the receiver. In this paper, it is shown that this is also possible in the case of channel estimation. The proposed channel estimation scheme makes use of a superimposed pilot layer (PLACE). Linear LS and MMSE as well as sliding-window channel estimators are derived and compared. It is analytically shown that semiblind channel estimation (the channel estimation is based on the training and the soft information about the estimated data from the previous iteration) always outperforms training-based channel estimation (the channel estimation is based only on the training) and the MMSE channel estimators always outperform the LS channel estimators with respect to the mean-squared error of the channel estimates, power efficiency, and bit error performance. However, sliding-window LS channel estimators are an interesting alternative due to the significant decrease of computational load compared to MMSE channel estimators—especially for larger block lengths.

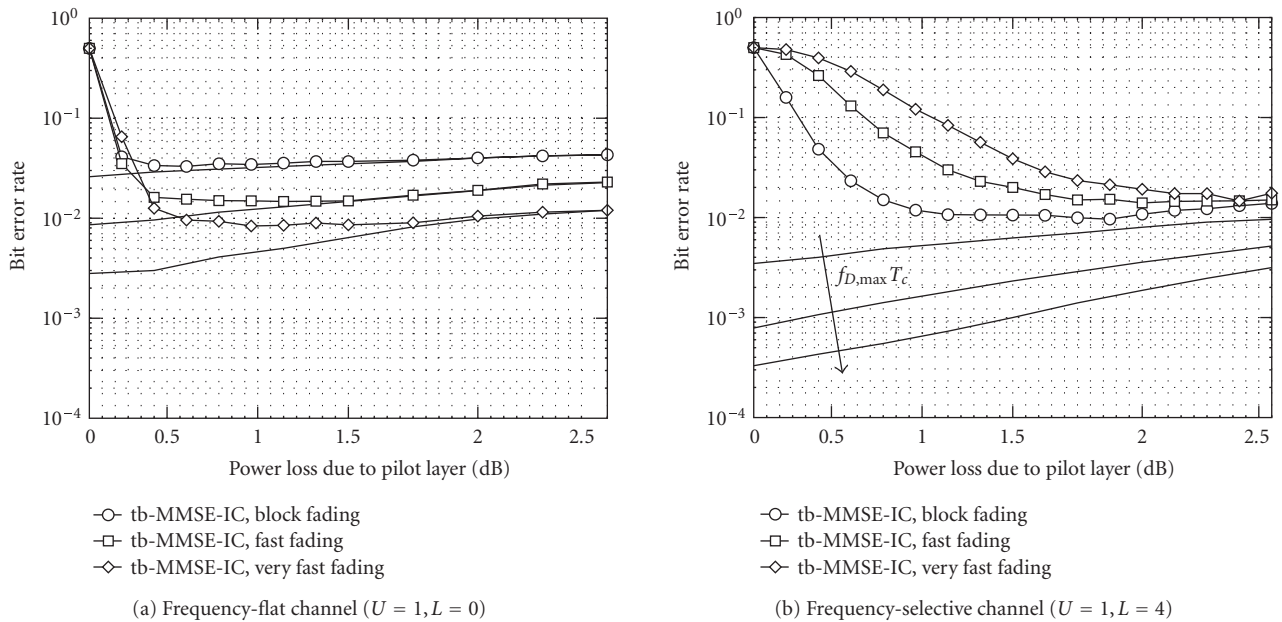


FIGURE 7: tb-MMSE-IC for common Rayleigh fading channel with block length  $K_b = 20$ . The degraded performance to the right is due to the data layer power loss that results from the increase of pilot layer power for constant  $E_b/N_0 = 10$  dB. Thick lines are for perfect channel knowledge. The bit error rate for perfect channel knowledge decreases with the fading rate.

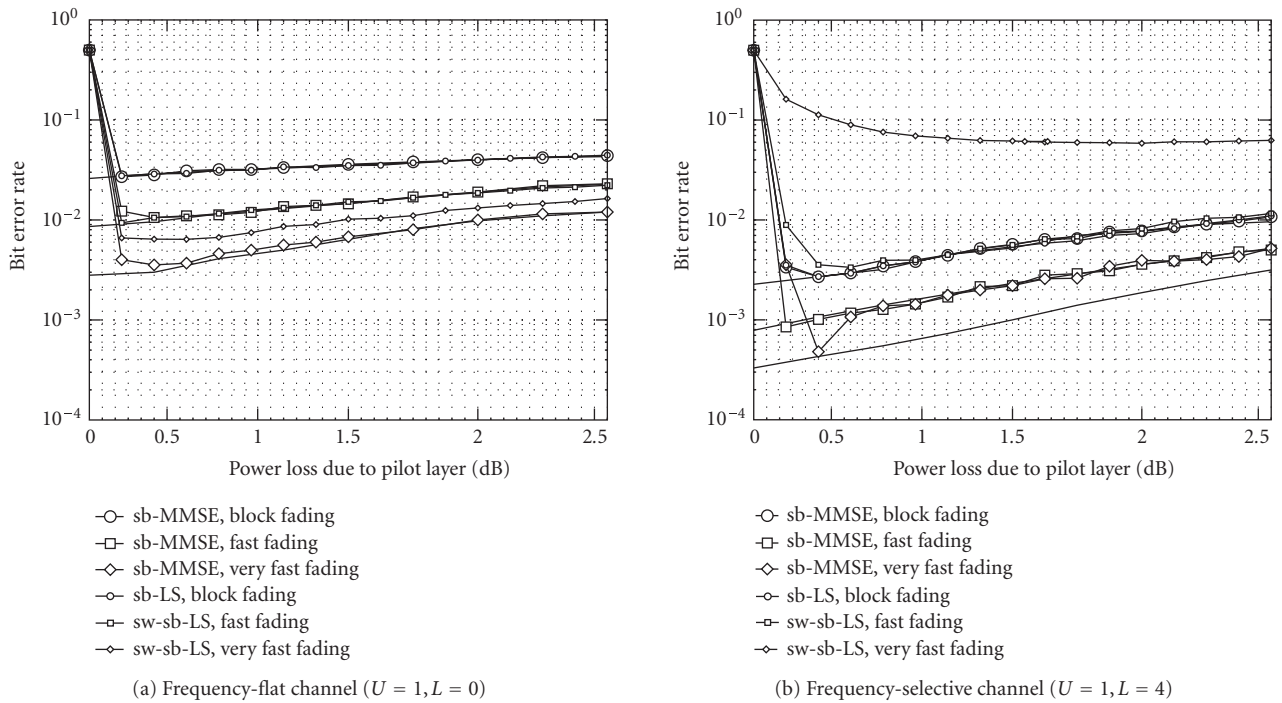
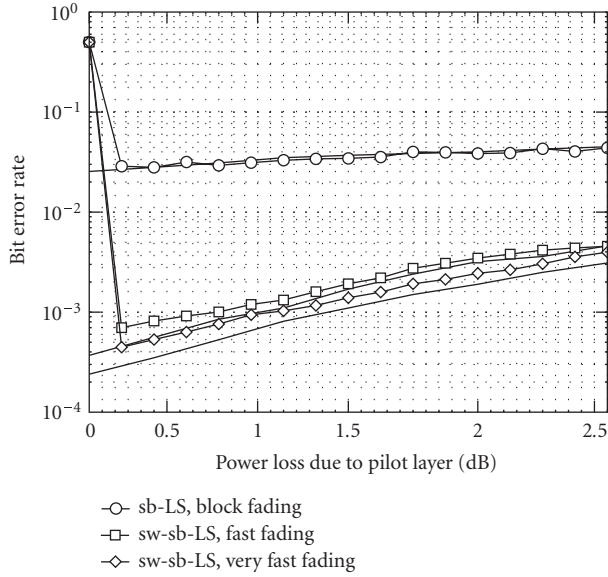
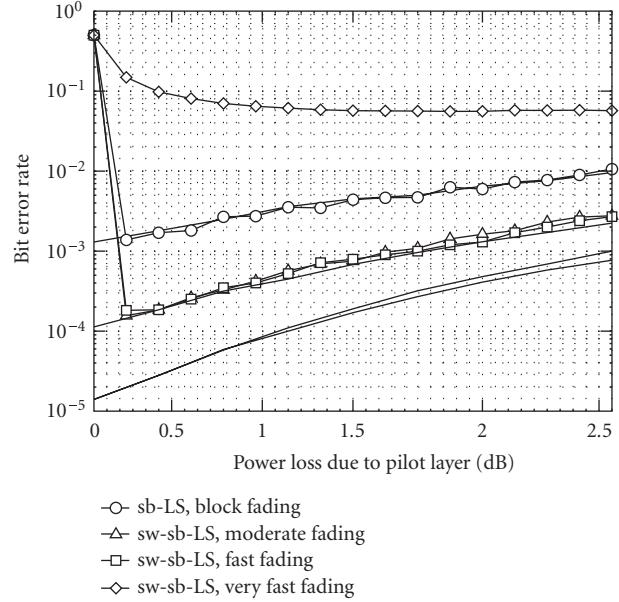


FIGURE 8: sb-MMSE and sw-sb-LS for common Rayleigh fading channel with block length  $K_b = 20$ . Thick lines are for perfect channel knowledge.



(a) Frequency-flat channel ( $U = 1, L = 0$ )



(b) Frequency-selective channel ( $U = 1, L = 4$ )

FIGURE 9: sw-sb-LS for common Rayleigh fading channel with block length  $K_b = 200$ . Thick lines are for perfect channel knowledge. The bit error rate for perfect channel knowledge decreases with the fading rate.

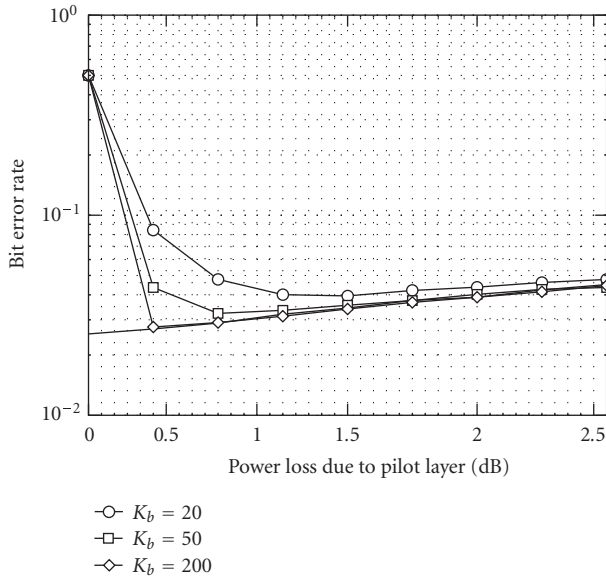


FIGURE 10: sw-sb-LS for independent frequency-flat block fading channels ( $U = M = 10, L = 0$ ). The thick line is for perfect channel knowledge.

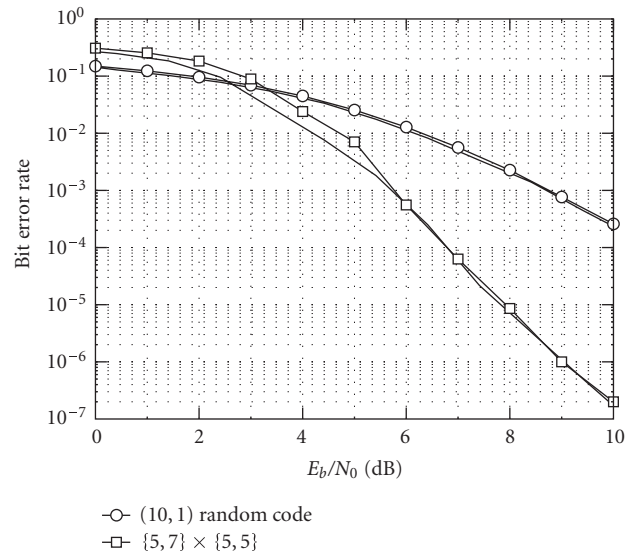


FIGURE 11: sw-sb-LS for common fast frequency-flat Rayleigh channel with block length  $K_b = 1000$ . Thick lines are for perfect channel knowledge. The power loss due to pilot layer is  $M + P_p/M \approx 0.414$  dB.

Numerical results motivate the use of IDMA/PLACE for transmission systems with short latency if semiblind channel estimation is performed. The capability of IDMA/PLACE to track channels with high Doppler spread offers possible ap-

plications for mobile radio with higher carrier frequencies and/or shorter chip durations than used in 3G systems. Note that a carrier frequency of  $f_C = 5$  GHz is anticipated for 4G systems, which is much higher than the carrier frequencies in

today's 3G systems. A further possible application is acoustic underwater communications.

## ACKNOWLEDGMENT

This work has been supported by the German Research Foundation (DFG) under Contract no. Ho2226/2.

## REFERENCES

- [1] P. Frenger, P. Orten, and T. Ottosson, "Code-spread CDMA using maximum free distance low-rate convolutional codes," *IEEE Transactions on Communications*, vol. 48, no. 1, pp. 135–144, 2000.
- [2] R. H. Mahadevappa and J. G. Proakis, "Mitigating multiple access interference and intersymbol interference in uncoded CDMA systems with chip-level interleaving," *IEEE Transactions on Wireless Communications*, vol. 1, no. 4, pp. 781–792, 2002.
- [3] P. Li, L. Liu, K. Wu, and W. Leung, "A unified approach to multiuser detection and space-time coding with low complexity and nearly optimal performance," in *Proceedings of the 40th Annual Allerton Conference on Communication, Control and Computing*, pp. 170–179, Monticelli, Ill, USA, October 2002.
- [4] P. Li, L. Liu, K. Wu, and W. Leung, "Interleave-division multiple-access (IDMA) communication systems," in *Proceedings of the 3rd International Symposium on Turbo Codes & Related Topics*, pp. 173–180, Brest, France, September 2003.
- [5] S. Verdú and S. Shamai, "Spectral efficiency of CDMA with random spreading," *IEEE Transactions on Information Theory*, vol. 45, no. 2, pp. 622–640, 1999.
- [6] A. J. Viterbi, "Very low rate convolutional codes for maximum theoretical performance of spread-spectrum multiple-access channels," *IEEE Journal on Selected Areas in Communications*, vol. 8, no. 4, pp. 641–649, 1990.
- [7] P. A. Hoeher and H. Schoeneich, "Interleave-division multiple access from a multiuser theory point of view," in *Proceedings of 4th International Symposium on Turbo Codes & Related Topics in Connection with the 6th International ITG-Conference on Source and Channel Coding*, Munich, Germany, April 2006.
- [8] H. Schoeneich and P. A. Hoeher, "Adaptive interleave-division multiple access—a potential air interface for 4G bearer services and wireless LANs," in *Proceedings of the 1st IFIP International Conference on Wireless and Optical Communications Networks (WOCN '04)*, pp. 179–182, Muscat, Oman, June 2004.
- [9] S. Zhou, Y. Li, M. Zhao, X. Xu, J. Wang, and Y. Yao, "Novel techniques to improve downlink multiple access capacity for Beyond 3G," *IEEE Communications Magazine*, vol. 43, no. 1, pp. 61–69, 2005.
- [10] J. C. Fricke, H. Schoeneich, and P. A. Hoeher, "An interleave-division multiple access based system proposal for the 4G up-link," in *Proceedings of 14th IST Mobile & Wireless Communications Summit*, Dresden, Germany, June 2005.
- [11] H. Schoeneich, J. C. Fricke, and P. A. Hoeher, "Adaptive 4G up-link proposal based on interleave-division multiple access," in *Proceedings of General Assembly of International Union of Radio Science (URSI '05)*, New Delhi, India, October 2005.
- [12] H. Schoeneich and P. A. Hoeher, "Semi-blind pilot-layer aided channel estimation with emphasis on interleave-division multiple access systems," in *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM '05)*, vol. 6, pp. 3513–3517, St. Louis, Mo, USA, November-December 2005, WC34-22.
- [13] M. Tüchler, R. Otnes, and A. Schmidbauer, "Performance of soft iterative channel estimation in turbo equalization," in *IEEE International Conference on Communications (ICC '02)*, vol. 3, pp. 1858–1862, New York, NY, USA, April-May 2002.
- [14] A. Kocian, B. Hu, C. Rom, P. Sørensen, B. H. Fleury, and E. K. Poulson, "Iterative joint data detection and channel estimation of DS/CDMA signals in multipath fading using the SAGE algorithm," in *Conference Record of the Asilomar Conference on Signals, Systems and Computers*, vol. 1, pp. 443–447, Pacific Grove, Calif, USA, November 2003.
- [15] J. Wehinger, C. F. Mecklenhräuker, R. R. Müller, T. Zemen, and M. Lončar, "On channel estimators for iterative CDMA multiuser receivers in flat Rayleigh fading," in *Proceedings of IEEE International Conference on Communications (ICC '04)*, vol. 5, pp. 2497–2501, Paris, France, June 2004.
- [16] P. A. Hoeher and F. Tufvesson, "Channel estimation with superimposed pilot sequence," in *Proceedings of IEEE Global Telecommunication Conference (GLOBECOM '99)*, vol. 4, pp. 2162–2166, Rio de Janeiro, Brazil, December 1999.
- [17] A. J. Weiss and B. Friedlander, "Channel estimation for DS-SS downlink with aperiodic spreading codes," *IEEE Transactions on Communications*, vol. 47, no. 10, pp. 1561–1569, 1999.
- [18] H. Holma and A. Toskala, Eds., *WCDMA for UMTS*, John Wiley & Sons, New York, NY, USA, 2000.
- [19] E. De Carvalho and D. T. M. Slock, "Cramer-Rao bounds for semi-blind, blind and training sequence based channel estimation," in *1st IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications (SPAWC '97)*, pp. 129–132, Paris, France, April 1997.
- [20] E. De Carvalho and D. T. M. Slock, "Blind and semi-blind FIR multichannel estimation: (global) identifiability conditions," *IEEE Transactions on Signal Processing*, vol. 52, no. 4, pp. 1053–1064, 2004.
- [21] H. Schoeneich and P. A. Hoeher, "A hybrid multiple access scheme delivering reliability information," in *Proceedings of 5th International ITG Conference on Source and Channel Coding*, pp. 437–442, Erlangen, Germany, January 2004.
- [22] L. Ping and L. Liu, "Analysis and design of IDMA systems based on SNR evolution and power allocation," in *IEEE Vehicular Technology Conference (VTC '04)*, vol. 2, pp. 1068–1072, Los Angeles, Calif, USA, September 2004.
- [23] J. C. L. Ng, K. B. Letaief, and R. D. Murch, "Complex optimal sequences with constant magnitude for fast channel estimation initialization," *IEEE Transactions on Communications*, vol. 46, no. 3, pp. 305–308, 1998.
- [24] P. A. Ranta, A. Hottinen, and Z.-C. Honkasalo, "Co-channel interference cancelling receiver for TDMA mobile systems," in *Proceedings of the IEEE International Conference on Communications (ICC '95)*, vol. 1, pp. 17–21, Seattle, Wash, USA, June 1995.
- [25] G. Caire and U. Mitra, "Training sequence design for adaptive equalization of multi-user systems," in *Conference Record of the Asilomar Conference on Signals, Systems and Computers*, vol. 2, pp. 1479–1483, Pacific Grove, Calif, USA, November 1998.
- [26] S. Badri-Hoher, *Digitale Empfängeralgorithmen für TDMA-Mobilfunksysteme mit besonderer Berücksichtigung des EDGE-Systems*, Ph.D. thesis, University Erlangen-Nuremberg, Erlangen, Germany, 2001.

- [27] P. K. Frenger, P. Orten, and T. Ottosson, "Code-spread CDMA with interference cancellation," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 12, pp. 2090–2095, 1999.

**Hendrik Schoeneich** received his Dipl.-Ing. degree in electrical and information engineering from the University of Kiel in 2001 for a Diploma thesis on cochannel interference cancellation in the GSM system. In 2000, he joined the Communications Research Centre (CRC) Canada during an internship on multiuser detection. Since 2001, he is with the Information and Coding Theory Lab (ICT), where he is currently working as a Research Assistant. His research interests include multiple access techniques, iterative multiuser detection, turbo equalization, semiblind channel estimation, and adaptive transmission.



**Peter Adam Hoehner** received Dipl.-Ing. and Dr.-Ing. (Ph.D.) degrees in electrical engineering from the Technical University of Aachen, Germany, and the University of Kaiserslautern, Germany, in 1986 and 1990. From 1986 to 1998, he was with the German Aerospace Research Establishment (DLR), Oberpfaffenhofen. In 1992, he was on leave at AT&T Bell Laboratories, Murray Hill, NJ. Since 1998 he is a Professor at the University of Kiel, Germany.

