Channel Estimation in Wireless OFDM Systems With Irregular Pilot Distribution

Peter Fertl, *Member, IEEE*, and Gerald Matz, *Senior Member, IEEE*

***Abstract*—This paper addresses pilot-assisted channel estimation for wireless orthogonal frequency division multiplexing (OFDM) systems with *irregular* pilot arrangements. Using nonuniform sampling techniques, we propose a novel channel estimator that is based on conjugate gradient iterations and features very low computational complexity. We investigate the impact of the pilot arrangement on the channel estimation mean square error and we provide a heuristic stopping criterion for early termination of the conjugate gradient iterations. One-dimensional and two-dimensional channel estimator implementations and applications of our method to multiple-antenna and multiuser OFDM systems are discussed. Extensive numerical simulations corroborate that the proposed method performs similarly to computationally much more expensive minimum mean square error estimators.**

***Index Terms*—Irregular sampling, multiple-input multiple-output (MIMO), multiuser access, orthogonal frequency division multiplexing (OFDM), pilot-assisted channel estimation.**

I. Introduction and Outline

A. Background and Motivation

ORTHOGONAL FREQUENCY division multiplexing (OFDM) [3] is an attractive solution for broadband wireless communications since it is resilient to multipath propagation and allows for simple and efficient implementation. For that reason, it is part of numerous applications and standards, e.g., IEEE 802.11a,g,n [4] for wireless local area networks (WLANs), IEEE 802.11p [5] for wireless access in vehicular environments (WAVE), IEEE 802.16a,e [6] for broadband wireless access systems (aka WiMAX), and 3GPP's long-term evolution [7], to name but a few.

Coherent detection in such systems requires accurate channel state information (CSI) at the receiver. Channel estimation based on pilot-symbol-assisted transmissions provides a reliable way to obtain CSI. Here, estimation of the unknown channel coefficients is enabled by embedding training data into the transmitted signal. In OFDM systems communicating over doubly dispersive channels, training symbols are usually scattered in time (over OFDM symbols) and in frequency (over subcarriers). To recover the unknown channel coefficients, two-dimensional (2-D) least squares (LS) interpolation methods [8]–[10] and minimum mean-square error (MMSE) filtering algorithms [11]–[13] have been proposed. The latter exploit the time and frequency correlation of the channel and thus require second-order channel statistics. Since 2-D MMSE channel estimators exhibit a high computational complexity, suboptimal but less complex approaches like low-rank 2-D MMSE estimation [11] and double one-dimensional (1-D) MMSE estimators [12] (i.e., separate 1-D filters in time and frequency) have been proposed. A comparison of these methods with 2-D MMSE estimation is provided in [12] and [14].

All of the previously mentioned methods have been designed for regular pilot lattices that satisfy a suitable 2-D Nyquist criterion [14]. If the latter is violated, techniques based on the fast Fourier transform (FFT) suffer from severe aliasing. If the pilot arrangement constitutes no regular lattice, 2-D MMSE interpolation can in principle still be performed but requires shift-variant filters [12] that are impractical due to their huge design and implementation complexity.

In this paper, we provide a practical framework for channel estimation based on irregular pilot arrangements. Our work is motivated by the fact that irregular pilots arise in several situations, such as the following, for example:

- adaptive orthogonal frequency division multiple access (OFDMA) exploits multiuser diversity via dynamic subcarrier allocation that results in irregular pilot distributions for each user (see [2], [15], and [16]);
- joint consideration of the preamble and the continual pilots in WLAN and WiMAX type systems amounts to nonuniform pilot arrangements that allow to track channels that vary within the data packet (cf. [1]);
- in iterative turbo-like systems performing joint channel estimation and data detection, it can be advantageous with regard to performance and complexity to feed the channel estimator only with reliably detected symbols [17], whose arrangement naturally is irregular;
- finally, nonuniform pilot arrangements were shown to be optimal in the mean-square error (MSE) sense for certain classes of time-varying channels [18].

B. Contributions

In this paper, we propose 2-D and 1-D pilot-assisted channel estimators for scenarios with irregular pilot distribution. Our de-

Manuscript received July 02, 2009; accepted January 28, 2010. Date of publication February 25, 2010; date of current version May 14, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Philippe Ciblat. This work was supported by the STREP project MASCOT (IST-026905) within the Sixth Framework Programme of the European Commission and by the WWTF project MOHAWI (MA 44). This work in part was previously presented at the Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, October–November 2006 [1] and the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Honolulu, Hawaii, April 2007 [2].

The authors are with the Institute of Communications and Radio-Frequency Engineering, Vienna University of Technology, A-1040 Vienna, Austria (e-mail: pfertl@nt.tuwien.ac.at; gmatz@nt.tuwien.ac.at).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2010.2044254

velopment is based on ideas from nonuniform sampling, in particular the approach in [19], [20], which can also be related to the theory of frames [21]. The problem of reconstructing 1-D *periodic* band-limited signals from nonuniform samples was recently also investigated in [22]. The main contributions of this paper can be summarized as follows.

- We apply nonuniform sampling techniques to channel estimation in OFDM systems with irregularly distributed pilots.
- We develop 1-D and 2-D low-complexity channel estimators that allow for flexible pilot arrangements under doubly selective channel conditions.
- Using algebraic arguments and an MSE analysis, we study the impact of the pilot arrangement on the channel estimation error.
- The computational complexity of our proposed method is assessed and a comparison with LS and MMSE channel estimation schemes is provided.
- Numerical results for different scenarios demonstrate the usefulness of our approach.

Our method has several practical advantages. It can deal with almost arbitrary pilot arrangements and features excellent performance (i.e., small channel estimation error). Furthermore, it has low computational complexity that is essentially determined by the degrees of freedom of the channel. Finally, the only prior knowledge required by our algorithm are coarse estimates of the maximum delay and the maximum Doppler (i.e., no second-order channel statistics). For irregular pilot sets, the last two properties are in strong contrast to MMSE channel estimation as discussed in [12].

C. Organization of the Paper

The paper is organized as follows. After presenting the underlying OFDM system model in Section II, we give a detailed description of the proposed channel estimation scheme based on 2-D irregular sampling in Section III. In Section IV, an MSE analysis is used to assess the influence of the pilot arrangement on the channel estimation error. Section V discusses various implementation aspects which simplify the estimation algorithm. The computational complexity of our method is discussed in Section VI and a 1-D version of the proposed channel estimation algorithm is presented in Section VII. In Section VIII, we consider an application to multiple-antenna OFDMA systems. In Section IX, we verify the performance of our estimator via extensive numerical simulations. Conclusions are provided in Section X.

II. System Model

We consider a packet-oriented OFDM system with K subcarriers and total transmit bandwidth B. A packet is formed by N consecutive OFDM symbols. At the transmitter, information bits are encoded with a channel encoder, randomly interleaved and mapped to complex-valued symbols $X(n,k)$ from an arbitrary modulation alphabet ($n \in \{0,\ldots,N-1\}$ denotes the symbol index and $k \in \{0,\ldots,K-1\}$ is the subcarrier index). To enable channel estimation at the receiver,

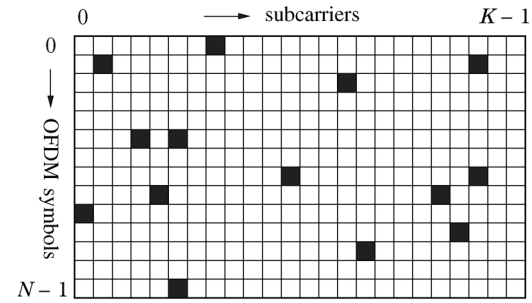

Fig. 1. Example of an irregular pilot distribution in an OFDM packet (pilot positions are indicated by black squares).

P pilot symbols are inserted into the stream of data symbols (leaving $KN - P$ positions for data symbols). Thus, certain subcarriers carry a pilot symbol rather than data. Different from most existing work, we do not impose any conditions on the pilot pattern, i.e., we explicitly allow for irregular pilot arrangements (see Fig. 1 for an example). We define the set of pilot positions as $\mathcal{P} = \{(n_p, k_p), p = 1,\ldots,P\}$. The OFDM modulator maps the data and pilot symbols to the transmit signal via a K-point inverse fast Fourier transform (IFFT) and prepends a cyclic prefix (CP) [23] of length L_{cp} (thus, the symbol duration equals $K + L_{\mathrm{cp}}$). As usual, the maximum delay of the channel is assumed to be shorter than the CP length L_{cp} in order to avoid intersymbol interference (ISI). We further assume that the channel remains approximately constant within one OFDM symbol so that intercarrier interference (ICI) is negligible. At the receiver, the OFDM demodulator removes the CP and calculates a received sequence $Y(n,k)$ via an FFT. Assuming perfect synchronization, the overall system model is then given by

$$Y(n,k) = H(n,k)X(n,k) + Z(n,k). \tag{1}$$

Here, $Z(n,k)$ is additive zero-mean, white, complex Gaussian noise with variance σ_Z^2. Furthermore, $H(n,k)$ denotes the complex channel coefficients, which are samples of the channel's time-varying transfer function [24]. The channel is assumed to be normalized such that $\mathcal{E}\{|H(n,k)|^2\} = 1$ ($\mathcal{E}\{\}$ is the expectation operator).

We consider doubly dispersive fading channels whose sampled transfer function $H(n,k)$ equals the 2-D discrete Fourier transform (DFT) of the spreading function $S(\tau,\nu)$ (τ and ν denote the discrete delay and Doppler index, respectively) [24]

$$H(n,k) = \frac{1}{\sqrt{KN}} \sum_{\tau=0}^{K-1} \sum_{\nu=-\frac{N}{2}}^{\frac{N}{2}-1} S(\tau,\nu) e^{-j2\pi\left(\frac{\tau k}{K} - \frac{\nu n}{N}\right)}. \tag{2}$$

Note that this DFT relationship implies that $H(n,k)$ is 2-D periodic. In Section V-C, we will discuss a simple approach to deal with the fact that in practice the channel is not periodic in time.

III. Proposed Channel Estimator

A. Basic Idea

In view of (1) it is straightforward to calculate (preliminary) LS channel estimates at the pilot positions according to

$$\hat{H}_{\rm pre}(n,k)=\frac{Y(n,k)}{X(n,k)}=H(n,k)+\frac{Z(n,k)}{X(n,k)},\quad (n,k)\in\mathcal{P}. \tag{3}$$

Since we are not assuming a regular pilot placement, the noisy pre-estimates in (3) may be distributed in a completely uneven, irregular fashion within each OFDM packet. Hence, (3) amounts to *noisy irregular 2-D sampling* of the transfer function $H(n,k)$.

For regular pilot lattices, conventional reconstruction methods like simple lowpass-filtering and FFT-based interpolation allow to obtain estimates of the channel coefficients $H(n,k)$, $(n,k)\notin\mathcal{P}$. However, these conventional schemes are not suited for nonuniform pilot arrangements since the irregular sampling in (3) may lead to severe aliasing.

Nevertheless, more advanced reconstruction methods for irregular sampling problems can be applied in such scenarios. Hence, we propose a channel estimation scheme that is based on irregular sampling and reconstruction techniques (cf. [25]). This amounts to viewing channel estimation as a (noisy) reconstruction problem for the 2-D lowpass function $H(n,k)$. More specifically, we adapt the so-called *ABC algorithm* [19], [20], originally developed for the noiseless reconstruction of band-limited images, to be applicable to our channel estimation problem. The acronym "ABC" signifies the main algorithm ingredients: **a**daptive weights, **b**lock Toeplitz matrices, and **c**onjugate gradient. The ABC method features excellent and stable reconstruction performance. Moreover, it allows for an efficient implementation based on FFTs and has a computational complexity that scales with the degrees of freedom of the underlying signal (image).

In the following, we show that these advantages carry over when applying these techniques to our channel estimation problem. The proposed channel estimation scheme furthermore does not require *a priori* information about the channel statistics.

B. Reconstruction Algorithm

We next provide a detailed description of the channel reconstruction problem and the proposed algorithm. To this end, we assume that the channel transfer function is a strictly band-limited 2-D lowpass function, i.e.,

$$H(n,k)=\frac{1}{\sqrt{KN}}\sum_{\tau=0}^{M_\tau-1}\sum_{\nu=-\frac{M_\upsilon}{2}}^{\frac{M_\upsilon}{2}}S(\tau,\nu)e^{-j2\pi\left(\frac{\tau k}{K}-\frac{\nu n}{N}\right)} \tag{4}$$

where $M_\tau \ll K$ and $M_\upsilon \ll N$ denote the channel's delay spread and Doppler spread,[1] respectively. The number of degrees of freedom of the channel (i.e., the number of nonzero values of $S(\tau,\nu)$) is given by $M = M_\tau(M_\upsilon+1)$. Relation (4) amounts to assuming that the spreading function $S(\tau,\nu)$ is supported within the rectangle $[0,M_\tau-1]\times[-M_\upsilon/2,M_\upsilon/2]$. Correspondingly, the channel transfer function belongs to the set of 2-D trigonometric polynomials of degree $(M_\tau, M_\upsilon+1)$ (the spreading function $S(\tau,\nu)$ corresponds to the coefficients of these polynomials). In the simulations section, we will relax this assumption and illustrate the performance of our method for channels where $H(n,k)$ is not strictly but only effectively band-limited, i.e., $S(\tau,\nu)$ is small outside the support region stated above.

For convenience, we rewrite (4) in matrix-vector form as

$$\mathbf{h}=\mathbf{V}\mathbf{s}, \tag{5}$$

with the length-NK vector[2] $\mathbf{h} = (\mathbf{h}_0^T \ldots \mathbf{h}_{N-1}^T)^T$ where $\mathbf{h}_n = (H(n,0) \ \ldots \ H(n,K-1))^T$, the length-$M$ vector $\mathbf{s} = (\mathbf{s}_0^T \ldots \mathbf{s}_{M_\tau-1}^T)^T$ where $\mathbf{s}_\tau = (S(\tau,-M_\upsilon/2) \ \ldots \ S(\tau,M_\upsilon/2))^T$, and the $NK\times M$ double Vandermonde matrix $\mathbf{V}$ with elements

$$[\mathbf{V}]_{l,m}=\frac{1}{\sqrt{KN}}e^{-j2\pi\left(\frac{\tau k}{K}-\frac{\nu n}{N}\right)} \tag{6}$$

where $l=k+Kn+1$ and $m=\tau+[(M_\upsilon/2)+\nu]M_\tau+1$. Hence, the span of the Vandermonde matrix $\mathbf{V}$, denoted $\mathcal{V}=\mathrm{span}\{\mathbf{V}\}$, corresponds to the set of all possible channels with delay spread M_τ and Doppler spread M_υ.

Let us define the $P\times NK$ pilot selection matrix $\mathbf{P}$ which has one nonzero entry in each row, i.e., it equals one in column $l_p = k_p + Kn_p + 1$ of row p, $p = 1,\ldots,P$, and is zero elsewhere. In the following, $\mathbf{h}_{\rm p} = \mathbf{P}\mathbf{h} = (H(n_1,k_1) \ \ldots \ H(n_P,k_P))^T$ denotes the length-P vector containing the channel coefficients at the pilot positions; similarly, $\hat{\mathbf{h}}_{\rm pre} = (\hat{H}_{\rm pre}(n_1,k_1) \ \ldots \ \hat{H}_{\rm pre}(n_P,k_P))^T$ and $\mathbf{V}_{\rm p}=\mathbf{P}\mathbf{V}$ such that $\mathbf{h}_{\rm p}=\mathbf{V}_{\rm p}\mathbf{s}$.

We now formulate the task of channel estimation/interpolation as a constrained weighted LS problem that looks for the trigonometric polynomial closest to the pre-estimate $\hat{H}_{\rm pre}(n,k)$ at the pilot positions, i.e.,

$$\hat{\mathbf{h}}=\arg\min_{\mathbf{h}\in\mathcal{V}}(\mathbf{P}\mathbf{h}-\hat{\mathbf{h}}_{\rm pre})^H\mathbf{W}(\mathbf{P}\mathbf{h}-\hat{\mathbf{h}}_{\rm pre}).$$

Here, $\mathbf{W} = \mathrm{diag}\{w_1,\ldots,w_P\}$ with $w_p > 0$ is a diagonal weight matrix whose elements can be adapted to the pilot set $\mathcal{P}$ in order to account for a strongly irregular pilot distribution (see Section III-C for a more detailed discussion). Without loss of generality, we will throughout assume the weight matrix to have normalized trace, i.e., $\mathrm{tr}\{\mathbf{W}\}=\sum_{p=1}^P w_p = 1$.

Standard LS theory implies that the channel estimate is given by $\hat{\mathbf{h}}=\mathbf{V}\hat{\mathbf{s}}$ (cf. (5)), where the spreading function estimate $\hat{\mathbf{s}}$ is obtained as the solution of

$$\mathbf{T}\hat{\mathbf{s}}=\tilde{\mathbf{s}}_{\rm pre}. \tag{7}$$

Here, the $M\times M$ matrix $\mathbf{T}$ and the length-M vector $\tilde{\mathbf{s}}_{\rm pre}$ are given by

$$\mathbf{T}=\mathbf{V}_{\rm p}^H\mathbf{W}\mathbf{V}_{\rm p},\qquad \tilde{\mathbf{s}}_{\rm pre}=\mathbf{V}_{\rm p}^H\mathbf{W}\hat{\mathbf{h}}_{\rm pre}. \tag{8}$$

The system matrix $\mathbf{T}$ can be shown to be block Toeplitz with Toeplitz blocks (BTTB), cf. [19].

[1]For simplicity, we assume that M_υ is even.

[2]Superscripts ^{T} and ^{H} denote transposition and Hermitian transposition, respectively.

For certain pilot sets $\mathcal{P}$, the condition number [26] of $\mathbf{T}$, denoted $\kappa(\mathbf{T})$, may become quite large (cf. Section IV). In this case, direct inversion of $\mathbf{T}$ to solve (7) is numerically unstable and leads to severe noise enhancement. This can be circumvented using regularization techniques; specifically, we propose to approximately solve (7) by using the iterative conjugate gradient (CG) algorithm (cf. [26]) with early termination in order to avoid noise enhancement. The CG algorithm permits to solve (7) efficiently (see the complexity discussion in Section VI) since it is also well suited for systems with Toeplitz structure.

In the following, we provide a short description of the CG algorithm when applied to our problem in (7). We first initialize the algorithm by setting $\mathbf{a}^{(0)} = \mathbf{b}^{(0)} = \tilde{\mathbf{s}}_{\text{pre}}$, $\mathbf{c}^{(0)} = \mathbf{T}\mathbf{b}^{(0)}$, $\alpha^{(0)} = \|\mathbf{a}^{(0)}\|^2$, and $\hat{\mathbf{s}}^{(0)} = \mathbf{0}$. The rth CG iteration ($r \geq 1$) then involves the following computational steps:

$$\beta^{(r)} = \frac{\alpha^{(r-1)}}{\left[\mathbf{b}^{(r-1)}\right]^H \mathbf{c}^{(r-1)}} \tag{9a}$$

$$\hat{\mathbf{s}}^{(r)} = \hat{\mathbf{s}}^{(r-1)} + \beta^{(r)}\mathbf{b}^{(r-1)} \tag{9b}$$

$$\mathbf{a}^{(r)} = \mathbf{a}^{(r-1)} - \beta^{(r)}\mathbf{c}^{(r-1)} \tag{9c}$$

$$\alpha^{(r)} = \|\mathbf{a}^{(r)}\|^2 \tag{9d}$$

$$\gamma^{(r)} = \frac{\alpha^{(r)}}{\alpha^{(r-1)}} \tag{9e}$$

$$\mathbf{b}^{(r)} = \mathbf{a}^{(r)} + \gamma^{(r)}\mathbf{b}^{(r-1)} \tag{9f}$$

$$\mathbf{c}^{(r)} = \mathbf{T}\mathbf{b}^{(r)}. \tag{9g}$$

These CG steps are repeated until a certain stopping criterion is satisfied (see Section V-A). Within each CG iteration, the vector $\hat{\mathbf{s}}^{(r)}$ constitutes an approximate solution of (7). The corresponding channel estimate can then be obtained as $\hat{\mathbf{h}}^{(r)} = \mathbf{V}\hat{\mathbf{s}}^{(r)}$.

The convergence speed of the CG iterations depends on the eigenvalue distribution of the matrix $\mathbf{T}$. If the eigenvalues are clustered, the CG iterations will converge fast. Methods to improve the clustering of the eigenvalues are discussed in the next section.

C. Adaptive Weights and Preconditioners

Adapting the weights w_p to the pilot set $\mathcal{P}$ is one way to deal with poorly conditioned $\mathbf{T}$ resulting from large variations of the pilot density (e.g., if the pilot pattern features clusters). The basic idea is to use larger weights for isolated pilots. One way to adapt the weights is by choosing w_p proportional to the area of the Voronoi region corresponding to the pilot position (n_p, k_p) (the Voronoi region associated to (n_p, k_p) consists of all points that are closer to (n_p, k_p) than to any other pilot position [27]). However, the calculation of the Voronoi region can be computationally expensive. A more efficient method (used in our simulations) to determine the adaptive weights is based on the inverse of the local pilot density and on FFTs [28]; see Appendix A for details.

If no weights are used (i.e., $\mathbf{W} = \mathbf{I}$), preconditioning is another way to deal with ill-conditioned $\mathbf{T} = \mathbf{V}_{\text{p}}^H\mathbf{V}_{\text{p}}$. Here, the CG iterations are applied to the system $\mathbf{C}^{-1}\mathbf{T}\hat{\mathbf{s}} = \mathbf{C}^{-1}\tilde{\mathbf{s}}_{\text{pre}}$ instead of (7). The preconditioner $\mathbf{C}$ is chosen such that the eigenvalues of $\mathbf{C}^{-1}\mathbf{T}$ are better clustered than those of $\mathbf{T}$. This also implies a smaller condition number, i.e., $\kappa(\mathbf{C}^{-1}\mathbf{T}) \ll \kappa(\mathbf{T})$. An efficient preconditioner for Toeplitz systems is presented in [29]. A comparison in [19] revealed that adaptive weights have the advantage of being computationally less expensive than preconditioning. Specifically, preconditioning usually results in fewer CG iterations but has a computational complexity per iteration that is significantly larger than with adaptive weights. However, if the condition number of $\mathbf{T}$ is very large (e.g., due to large gaps in the pilot arrangement), preconditioning may yield a performance improvement over adaptive weights.

IV. Pilot Arrangement and MSE Analysis

A. Nyquist Criterion

There is a fundamental difference between 1-D and 2-D (channel) reconstruction from irregular samples. For the noiseless 1-D case, the fundamental theorem of algebra implies that—independent of the pilot arrangement—the channel (i.e., a trigonometric polynomial of degree M) is uniquely determined if the number of samples/pilots satisfies $P \geq M$ (cf. also [9]). The system matrix in this case has always full rank and is invertible. However, this is no longer true for the 2-D case since the fundamental theorem of algebra does not apply to dimensions larger than one. For the 2-D case, we will say that the pilot set satisfies the Nyquist criterion if the region $[0, N-1] \times [0, K-1]$ is completely covered by rectangles of size $(N/(M_\upsilon + 1)) \times (K/M_\tau)$ centered at the pilot positions [19], i.e.,

$$[0, N-1] \times [0, K-1] \subseteq \bigcup_{p=1}^{P} [n_p - \Delta n, n_p + \Delta n] \times [k_p - \Delta k, k_p + \Delta k] \tag{10}$$

with $\Delta n = N/(2(M_\upsilon + 1))$, $\Delta k = K/(2M_\tau)$. Since the right-hand side of (10) has an area of at most NKP/M, it follows that a necessary (but not sufficient) condition for (10) is given by $P \geq M$. The same condition is necessary (but not sufficient) for the BTTB matrix $\mathbf{T}$ to be invertible; this means that regularity of $\mathbf{T}$ depends not only on the number of pilots but also on their arrangement. So far, there exists no general simple criterion for 2-D pilot/sampling sets that guarantees invertibility of $\mathbf{T}$.

B. MSE Analysis

We next study the impact of the pilot arrangement on the proposed channel estimation scheme via an MSE analysis. The iterative CG solution of (7) is difficult to analyze. Hence, in the following we assume that (7) is solved by direct inversion of the matrix $\mathbf{T}$ (in the noiseless case this is equivalent to performing M CG iterations). With appropriate early termination (see Section V-A), the CG algorithm will perform an implicit regularization, thereby achieving an even smaller MSE.

Combining $\hat{\mathbf{h}} = \mathbf{V}\hat{\mathbf{s}}$, $\hat{\mathbf{s}} = \mathbf{T}^{-1}\tilde{\mathbf{s}}_{\text{pre}}$ (cf. (7)), and $\tilde{\mathbf{s}}_{\text{pre}} = \mathbf{V}_{\text{p}}^H\mathbf{W}\hat{\mathbf{h}}_{\text{pre}}$ (cf. (8)) results in

$$\hat{\mathbf{h}} = \mathbf{V}\mathbf{T}^{-1}\mathbf{V}_{\text{p}}^H\mathbf{W}\hat{\mathbf{h}}_{\text{pre}}.$$

Writing the channel pre-estimate as $\hat{\mathbf{h}}_{\text{pre}} = \mathbf{h}_{\text{p}} + \tilde{\mathbf{z}}_{\text{p}}$ (cf. (3)) with

$$\tilde{\mathbf{z}}_{\text{p}} = (Z(n_1, k_1)/X(n_1, k_1) \ \ldots \ Z(n_P, k_P)/X(n_P, k_P))^T,$$

and recalling that $\mathbf{h}_\mathrm{p} = \mathbf{V}_\mathrm{p}\mathbf{s}$ and $\mathbf{T} = \mathbf{V}_\mathrm{p}^H\mathbf{W}\mathbf{V}_\mathrm{p}$, we further obtain

$$\begin{aligned}\hat{\mathbf{h}} &= \mathbf{V}\mathbf{T}^{-1}\mathbf{V}_\mathrm{p}^H\mathbf{W}(\mathbf{h}_\mathrm{p} + \tilde{\mathbf{z}}_\mathrm{p}) \\ &= \mathbf{V}\mathbf{T}^{-1}\mathbf{T}\mathbf{s} + \mathbf{V}\mathbf{T}^{-1}\mathbf{V}_\mathrm{p}^H\mathbf{W}\tilde{\mathbf{z}}_\mathrm{p} \\ &= \mathbf{h} + \mathbf{V}\mathbf{T}^{-1}\mathbf{V}_\mathrm{p}^H\mathbf{W}\tilde{\mathbf{z}}_\mathrm{p}.\end{aligned}$$

Based on the last expression it follows that the normalized MSE equals

$$\epsilon^2 \triangleq \frac{\mathcal{E}\left\{\|\mathbf{h} - \hat{\mathbf{h}}\|^2\right\}}{\mathcal{E}\left\{\|\mathbf{h}\|^2\right\}} = \frac{\sigma_Z^2}{NK}\mathrm{tr}\left\{\mathbf{T}^{-1}\mathbf{V}_\mathrm{p}^H\mathbf{W}\mathbf{\Sigma}_\mathrm{p}^{-1}\mathbf{W}\mathbf{V}_\mathrm{p}\mathbf{T}^{-1}\right\}.$$

Here, we used the fact that $\mathbf{V}^H\mathbf{V} = \mathbf{I}$ and $\mathcal{E}\{\tilde{\mathbf{z}}_\mathrm{p}\tilde{\mathbf{z}}_\mathrm{p}^H\} = \sigma_Z^2\mathbf{\Sigma}_\mathrm{p}^{-1}$ where the diagonal matrix $\mathbf{\Sigma}_\mathrm{p} \triangleq \mathrm{diag}\{\sigma_1^2, \ldots, \sigma_P^2\}$ contains the powers $\sigma_p^2 = |X(n_p, k_p)|^2$ of the pilot symbols. Assuming that the pilot symbol powers are proportional to the weights, i.e., $\sigma_p^2 = E_\mathrm{p}w_p$ and $\mathbf{\Sigma} = E_\mathrm{p}\mathbf{W}$ with E_p denoting the total pilot power, the MSE further simplifies to

$$\begin{aligned}\epsilon^2 &= \frac{\sigma_Z^2}{NKE_\mathrm{p}}\mathrm{tr}\left\{\mathbf{T}^{-1}\mathbf{V}_\mathrm{p}^H\mathbf{W}\mathbf{V}_\mathrm{p}\mathbf{T}^{-1}\right\} = \frac{\sigma_Z^2}{NKE_\mathrm{p}}\mathrm{tr}\{\mathbf{T}^{-1}\} \\ &= \frac{\sigma_Z^2}{NKE_\mathrm{p}}\sum_{m=1}^{M}\frac{1}{\lambda_m},\end{aligned} \tag{11}$$

where $\lambda_m, m = 1, \ldots, M$, denotes the eigenvalues of $\mathbf{T}$, which are positive and assumed sorted as $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_M$. To find pilot arrangements and pilot power allocations that minimize (11), we extend the proof presented in [9] to 2-D pilot arrangements. First, it can be shown using the definition of $\mathbf{V}$ (cf. (6)) and $\mathrm{tr}\{\mathbf{W}\} = 1$ that, independently of the chosen pilot arrangement, we have

$$\mathrm{tr}\{\mathbf{T}\} = \mathrm{tr}\left\{\mathbf{V}_\mathrm{p}^H\mathbf{W}\mathbf{V}_\mathrm{p}\right\} = \sum_{m=1}^{M}\lambda_m = \frac{M}{NK}. \tag{12}$$

Minimizing the MSE ϵ^2 thus amounts to solving the optimization problem

$$\min_{\lambda_p}\frac{\sigma_Z^2}{NKE_\mathrm{p}}\sum_{m=1}^{M}\frac{1}{\lambda_m} \quad \text{subject to} \quad \sum_{m=1}^{M}\lambda_m = \frac{M}{NK}.$$

The solution to this optimization problem is given by $\lambda_m = 1/(NK)$, $m = 1, \ldots, M$, i.e., $\mathbf{T} = 1/(NK)\mathbf{I}$ and the corresponding minimum MSE equals

$$\epsilon_\mathrm{min}^2 = \frac{\sigma_Z^2}{E_\mathrm{p}}M.$$

The condition $\mathbf{T} = \mathbf{I}/(NK)$ is satisfied if and only if i) $\mathbf{W} = \mathbf{I}/P$ and ii) the columns of $\mathbf{V}_\mathrm{p}$ are orthogonal. This in turn implies that the pilot set constitutes a regular lattice (e.g., rectangular or hexagonal) and that all pilots have the same power $\sigma_p^2 = E_\mathrm{p}/P$. The optimality of $\mathbf{T} = 1/(NK)\mathbf{I}$ suggests that the MSE is strongly influenced by the condition number $\kappa(\mathbf{T}) \triangleq \lambda_1/\lambda_M$ of $\mathbf{T}$. Indeed, by noting that

$$\sum_{m=1}^{M}\frac{1}{\lambda_m} \leq \frac{M}{\lambda_M} = \frac{M}{\lambda_1}\kappa(\mathbf{T})$$

and using the fact that $\lambda_1 \geq 1/(NK)$ (implied by (12)), it follows that

$$\epsilon_\mathrm{min}^2 \leq \epsilon^2 \leq \epsilon_\mathrm{min}^2\kappa(\mathbf{T}).$$

This confirms that a small condition number of the BTTB matrix $\mathbf{T}$ ensures that the MSE performance will be close to optimal. Note that small $\kappa(\mathbf{T})$ is also desirable for fast convergence of the CG iterations.

While few pilot patterns lead to singular $\mathbf{T}$, there are many pilot arrangements for which the condition number $\kappa(\mathbf{T})$ is large, specifically those which are far from uniform. Upper bounds on $\kappa(\mathbf{T})$ for pilot sets satisfying the Nyquist criterion (10) have been studied e.g., in [20]. These bounds confirm that the condition number tends to increase with increasing pilot spacing, maximum delay spread and maximum Doppler spread. While the results regarding $\kappa(\mathbf{T})$ do not apply if the Nyquist criterion is violated, we observed that the channel estimation performance in this case degrades only gradually. Note that a probabilistic approach to study the reconstruction quality of multidimensional fields from noisy irregular samples was recently presented in [30] and [31]. We finally remark that the discussion in this section concerns the overall channel estimation MSE. Specifically, even if ϵ^2 is large (e.g., due to "holes" in the pilot pattern), we observed that channel estimation still works locally in regions with sufficient pilots.

V. Implementation Details

A. Stopping Criterion

Early termination of the CG algorithm amounts to a regularized inversion of $\mathbf{T}$. Stopping the CG iterations before noise enhancement kicks in, allows to reduce the channel estimation error. In this section, we provide a simple heuristic stopping criterion that will be seen in Section IX to provide close-to-optimal performance.

After r CG iterations (see (9)), the approximate solution $\hat{\mathbf{s}}^{(r)}$ of (7) yields a channel estimate $\hat{\mathbf{h}}^{(r)} = \mathbf{V}\hat{\mathbf{s}}^{(r)}$ with corresponding normalized squared error

$$\epsilon^2(r) \triangleq \frac{\|\hat{\mathbf{h}}^{(r)} - \mathbf{h}\|^2}{\|\mathbf{h}\|^2}. \tag{13}$$

The squared error $\epsilon^2(r)$ decreases initially (as a function of r) and has a minimum at a certain optimal number of iterations r_opt, i.e., for $r > r_\mathrm{opt}$ noise enhancement causes performance degradation [32]. Unfortunately, r_opt cannot be determined in practical implementations since the true channel $\mathbf{h}$ in (13) is not available. Using $\hat{\mathbf{h}}_\mathrm{p}^{(r)} = \mathbf{V}_\mathrm{p}\hat{\mathbf{s}}^{(r)}$, we thus define the normalized squared error associated to the pilot positions as

$$\epsilon_\mathrm{p}^2(r) \triangleq \frac{\|\hat{\mathbf{h}}_\mathrm{p}^{(r)} - \hat{\mathbf{h}}_\mathrm{pre}\|^2}{\|\hat{\mathbf{h}}_\mathrm{p}^{(r)}\|^2}. \tag{14}$$

The main difference between (14) and (13) is the evaluation of the error only on the pilot positions and the replacement of the true channel with the pre-estimate $\hat{\mathbf{h}}_\mathrm{pre}$ in the numerator.

The normalization by $\|\hat{\mathbf{h}}_{\mathrm{p}}^{(r)}\|^2$ helps to avoid overfitting to the noisy pre-estimates $\hat{\mathbf{h}}_{\mathrm{pre}}$. We now propose to terminate the CG iterations if

$$\epsilon_{\mathrm{p}}^2(r) \geq \zeta \epsilon_{\mathrm{p}}^2(r-1) \tag{15}$$

where the design parameter ζ is typically chosen slightly smaller than 1. Thus, we stop the CG iterations if the squared error (14) no longer decreases fast enough. If (15) is not satisfied for any r, a maximum number of M iterations is performed to guarantee termination. This specific choice of the maximum number of CG iterations is motivated by the fact that in the noiseless case the CG algorithm (when applied to an $M \times M$ system) yields the exact solution after at most M steps. Moreover, this approach also guarantees a certain worst-case complexity. The iteration number at which the CG algorithm is terminated using the stopping criterion (15) is denoted r_{sc}. Depending on SNR, channel parameters, and pilot set, r_{sc} can be different from r_{opt}. Since the minimum of $\epsilon^2(r)$ often is not very sharp, deviations of r_{sc} from r_{opt} degrade the average performance only slightly. However, the impact on complexity can be significant. A more detailed discussion of the performance of our stopping rule is provided in Section IX.

B. Delay and Doppler Spread

The proposed reconstruction algorithm presupposes knowledge of the channel's maximum delay spread M_τ and maximum Doppler spread (characterized by M_υ) which are not known a priori. In practice, the unknown parameters M_τ and M_υ have to be replaced with appropriately chosen (estimated) values $\hat{M}_\tau$ and $\hat{M}_\upsilon$. If $\hat{M}_\tau$ and $\hat{M}_\upsilon$ are chosen slightly too large, the channel estimate doesn't degrade dramatically, partly due to the fact that actual channels are not strictly band-limited (cf. Section IX). However, larger $\hat{M}_\tau$ and $\hat{M}_\upsilon$ increase the condition number and the dimension of $\mathbf{T}$, thereby resulting in a larger number of more expensive CG iterations (see complexity discussion in Section VI).

A worst-case choice for the delay spread is the length (in samples) of the cyclix prefix, i.e., $\hat{M}_\tau = L_{\mathrm{cp}}$. A rule of thumb to determine the (effective) discrete Doppler spread is given by

$$\hat{M}_\upsilon = 2\left\lfloor \frac{f_{\upsilon,\max}}{B}(K+L_{\mathrm{cp}})N \right\rfloor. \tag{16}$$

Here, $\lfloor x \rfloor$ denotes the largest integer not larger than x and $f_{\upsilon,\max} = f_{\mathrm{c}} v_{\max}/c_0$ is the maximum Doppler frequency in Hertz, with $v_{\max}$, f_{c}, and c_0 respectively denoting the maximum terminal velocity, the carrier frequency, and the speed of light. Note that the expression $(K+L_{\mathrm{cp}})N/B$ in (16) equals the duration of an OFDM packet (in seconds). Since $f_{\upsilon,\max}$ is the maximum Doppler frequency in Hertz (cycles per second), $\hat{M}_\upsilon$ equals twice the number of entire "Doppler cycles" within an OFDM packet.

C. Zero Padding

The use of 2-D trigonometric polynomials in our channel model (4) implies that $H(n,k)$ is a bi-periodic function, i.e., $H(n,k) = H(n+lN, k+mK)$ for $l,m \in \mathbb{Z}$. Periodicity with respect to frequency k can be reasonably assumed due to the use of FFT processing in the OFDM modulator and demodulator. Here, the dedicated guard carriers protect against edge effects. However, periodicity in time is hardly justified, thus leading to increased channel estimation errors at the borders of the packet. To deal with this problem, we use zero padding at the receive side, i.e., before applying our proposed estimator we append all-zero OFDM symbols at the beginning and the end of the OFDM packet in order to create a larger processing block. The proposed channel estimator is then applied to this larger block. The idea is to shift the borders of this new processing block away from the useful OFDM symbols in order to prevent edge effects within the actual data packet.[3] Numerical simulations have shown that the number of OFDM symbols for zero padding, denoted N_{pad}, should be chosen proportional to the channel's coherence time, i.e.,

$$N_{\mathrm{pad}} = \left\lceil \frac{T_{\mathrm{c}} B}{K+L_{\mathrm{cp}}} \right\rceil \tag{17}$$

where the coherence time is given by $T_{\mathrm{c}} = 1/(2f_{\upsilon,\max})$ (here, $\lceil x \rceil$ denotes the smallest integer not smaller than x). Thus, channel estimation is performed over a block of $\bar{N} = N + N_{\mathrm{pad}}$ OFDM symbols by padding $\lceil N_{\mathrm{pad}}/2 \rceil$ zero OFDM symbols at the beginning and $\lfloor N_{\mathrm{pad}}/2 \rfloor$ zero OFDM symbols at the end of each OFDM packet. While the larger block length affects the convergence of the CG algorithm (due to an increase of the condition number of $\mathbf{T}$) and increases the computational complexity (see the next section), zero padding decreases the channel estimation error significantly. We further note that zero padding may also necessitate the use of a slightly larger Doppler spread parameter. Replacing N in (16) with $\bar{N}$ and using (17) shows that the discrete Doppler spread $\hat{M}_\upsilon$ increases by at most 2, thus having little influence on the overall computational complexity (see next section).

For very large coherence times (very small Doppler), (17) leads to processing block lengths $\bar{N}$ much larger than the original OFDM packet length N. In such situations, the channel changes only little over time and it may be more appropriate to avoid zero padding and rather use channel estimation schemes designed for time-invariant channels like the 1-D technique described in Section VII. From the perspective of a performance-complexity trade-off, 1-D channel estimation should be preferred over 2-D channel estimation with zero padding if the zero padding exceeds the length of the original OFDM packet, i.e., if $N_{\mathrm{pad}} \geq N$.

VI. Algorithm Summary and Complexity

Next, we summarize the individual steps of the overall channel estimation scheme and assess their computational complexity. We emphasize that two important structural properties result in very low computational complexity: i) the dimension of the system in (7) is independent of the number of pilots, and ii) the $M \times M$ matrix $\mathbf{T}$ has BTTB structure. Moreover, we argue that most computations can be efficiently implemented using 2-D FFTs.

[3]We note that this issue was not dealt with in our earlier publications [1], [2]. Moreover, we emphasize that the zero padding is performed only for channel estimation at the receiver and hence does not affect the transmitter or the spectral efficiency.

To simplify notation, we do not consider zero padding and assume that the effective discrete delay and Doppler spread are exactly known (i.e., $\bar{N} = N$, $\hat{M}_\tau = M_\tau$, and $\hat{M}_\upsilon = M_\upsilon$). The actual complexities are obtained by replacing N, M_τ, and M_υ with $\bar{N}$, $\hat{M}_\tau$, and $\hat{M}_\upsilon$, respectively.

Preprocessing: If the pilot arrangement is fixed and does not change from one packet to another, the adaptive weights w_p, $p = 1, \ldots, P$, and the matrix $\mathbf{T}$ can be precomputed and stored. Note, however, that w_p and $\mathbf{T}$ have to be computed anew each time the pilot arrangement changes. A simple FFT-based scheme [28] to calculate the adaptive weights can be found in Appendix A. Due to the BTTB structure of $\mathbf{T}$, only its first and M_τth column need to be computed and stored. The computational complexity for directly calculating the relevant entries scales as $\mathcal{O}(MP)$. Alternatively, these two matrix columns can be calculated using a 2-D FFT (see Appendix B), in total requiring $\mathcal{O}(NK\log(NK))$ operations. Which approach is more efficient depends on the percentage of pilots, denoted $\theta = P/NK$, and on the channel's degrees of freedom M. Specifically, for $\theta > \log(NK)/M$, the FFT approach tends to be more efficient than direct computation.

Channel interpolation: The following steps need to be performed for each OFDM packet.

1) *Calculation of pre-estimate*: Once $\hat{\mathbf{h}}_{\mathrm{pre}}$ is obtained via (3) using P operations, $\tilde{\mathbf{s}}_{\mathrm{pre}}$ in (7) can be computed with $\mathcal{O}(MP)$ operations, assuming the full $P \times M$ matrix $\mathbf{V}_{\mathrm{p}}$ has been pre-computed and stored. Alternatively, $\tilde{\mathbf{s}}_{\mathrm{pre}}$ can be obtained using $\mathcal{O}(NK\log(NK))$ operations via a 2-D FFT, similar to the computation of $\mathbf{T}$ (see Appendix B).
2) *CG iteration*: In each CG iteration, an intermediate estimate $\hat{\mathbf{s}}^{(r)}$ of the spreading function coefficients $\mathbf{s}$ is computed. The CG iteration steps (9a)–(9f) have a complexity of $\mathcal{O}(M)$. The dominating complexity is that of the matrix-vector multiplication (9g). Due to the BTTB structure of $\mathbf{T}$, this matrix-vector multiplication can be carried out efficiently using 2-D FFTs [29]. The idea is to embed $\mathbf{T}$ into a block-circulant matrix with circulant blocks which is diagonalized by a 2-D FFT (see Appendix C for details). The overall operation count of one CG iteration thus scales as $\mathcal{O}(M\log(M))$, i.e., it increases with the delay and Doppler spread but *not* with the number of pilots.
3) *Stopping criterion*: To check whether the CG algorithm should be terminated, $\hat{\mathbf{h}}_{\mathrm{p}}^{(r)} = \mathbf{V}_{\mathrm{p}}\hat{\mathbf{s}}^{(r)}$ is computed with a complexity that scales either with $\mathcal{O}(MP)$ (direct matrix-vector multiplication) or with $\mathcal{O}(NK\log(NK))$ (FFT-based implementation). The proposed stopping criterion (15) is then evaluated using $\mathcal{O}(P)$ operations. In case the reconstruction error is not small enough, another CG iteration (step 2) is performed. We note that the stopping criterion does not need to be checked after every iteration. It can even be completely omitted if the number of CG iterations is fixed. Of course, checking the stopping criterion less frequently or not at all may also have an effect on the estimation performance.
4) *Postprocessing*: The last step is the computation of the channel estimate $\hat{\mathbf{h}}^{(r_{\mathrm{sc}})}$ according to (5). This can be achieved using a 2-D FFT which requires $\mathcal{O}(NK\log(NK))$ operations.

It can be seen that almost all steps in the proposed algorithm can be computed either via 2-D FFTs or maybe even more efficient direct matrix-vector multiplications. Furthermore, the CG method solves (7) with only $\mathcal{O}(M\log(M))$ operations per iteration. For practical OFDM systems, our simulations showed that the required number of iterations is usually rather small (see Section IX).

Compared to our scheme, the computational complexity of the 2-D MMSE estimator in [12], which can also be applied in our scenario, is much higher. To allow for a fair complexity comparison we now replace N, M_τ, and M_υ with $\bar{N}$, $\hat{M}_\tau$, and $\hat{M}_\upsilon$, respectively; here, $\hat{M} = \hat{M}_\tau(\hat{M}_\upsilon + 1)$. The pre-processing complexity of the 2-D MMSE estimator scales as $\mathcal{O}(P^3/\theta)$, in comparison to $\mathcal{O}(\hat{M}P)$ or less operations for the pre-processing with our proposed method (note that $(P^3/\theta)/\hat{M}P = PNK/\hat{M} \geq NK \gg 1$). The MMSE channel interpolation requires $\mathcal{O}(PNK)$ operations for each OFDM packet, whereas the complexity order of our channel interpolation is $\mathcal{O}(\bar{N}K\log(\bar{N}K))$ or even less (recall that $\bar{N} \leq 2N$). We remind the reader that 2-D MMSE estimation additionally requires knowledge of the second-order statistics of the channel and the noise. The following section illustrates how a more structured (but still irregular) placement of the pilots can be exploited to further reduce the computational complexity of our scheme.

VII. 1-D Case

Our algorithm can be easily reformulated for 1-D channel estimation, relevant for the following:

- reconstructing a time-varying channel symbol-wise, i.e., over each OFDM symbol separately without taking into account temporal channel correlation;
- successively performing 1-D channel estimation with respect to frequency and with respect to time, i.e., double 1-D channel estimation as e.g., described in [12];
- estimating channels that are time-invariant or have very low Doppler from a single OFDM training symbol (if there are more OFDM training symbols, the 1-D estimates can be averaged or interpolated).

We note that an overview of irregular sampling in one dimension is given in [33]. The corresponding algorithm works similar to its 2-D equivalent but has significantly smaller computational complexity and is advantageous in terms of latency and memory requirements. However, it requires that a 1-D Nyquist criterion is satisfied, which for the first and second case above usually requires a larger overall number of pilots as compared to the 2-D approach.

The 1-D channel estimator over the frequency domain (k-domain) is obtained from the 2-D algorithm by formally setting $N = 1$ and $M_\upsilon = 0$.[4] For convenience, we thus omit the symbol time n and the discrete Doppler ν in the following. Then, (2) reduces to a 1-D DFT relating the channel coefficients $H(k)$ and the impulse response $S(\tau)$. We will assume that $S(\tau) = 0$ for $\tau \geq M_\tau$. Stacking the channel coefficients $H(k)$ into the length-K vector $\mathbf{h}$ and the impulse response $S(\tau)$

[4]The temporal (n-domain) version of the 1-D algorithm is obtained for $K = 1$ and $M_\tau = 0$.

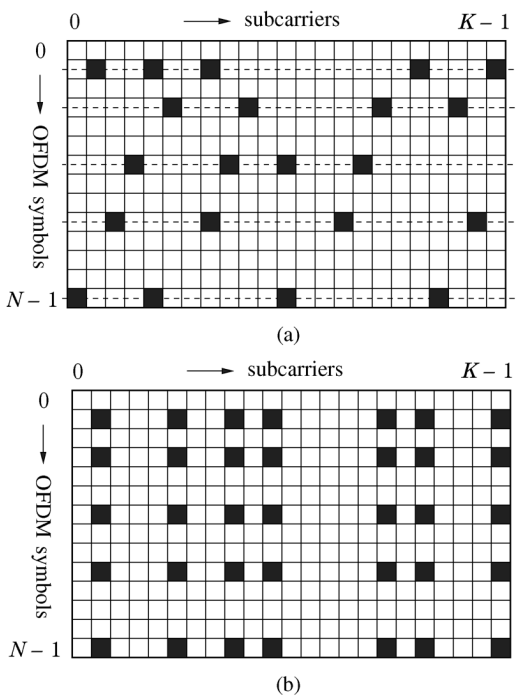

Fig. 2. Irregular pilot arrangement corresponding to (a) a line-type sampling set and (b) a product sampling set (pilot positions are indicated by black squares).

into the length-M_τ vector $\mathbf{s}$, we obtain a Vandermonde system $\mathbf{h} = \mathbf{V}\mathbf{s}$ analogous to (5) with $\mathbf{V}$ being a $K \times M_\tau$ Vandermonde matrix with elements $[\mathbf{V}]_{k+1,\tau+1} = (1/\sqrt{K})e^{-j2\pi k\tau/K}$, $k = 0, \ldots, K-1$, $\tau = 0, \ldots, M_\tau - 1$.

The pilot set consists of P pilot tones $k_p, p = 1, \ldots, P$, which are collected into the pilot set $\mathcal{P} = \{k_1, \ldots, k_P\}$. At the pilot positions, a channel pre-estimate can be calculated analogously to (3), and the weighted LS minimization problem leads to a system of equations similar to (7) with length-M_τ vectors $\hat{\mathbf{s}}$ and $\tilde{\mathbf{s}}_{\mathrm{pre}}$. Contrary to the BTTB structure of the 2-D case, the $M_\tau \times M_\tau$ matrix $\mathbf{T}$ here is a plain Toeplitz matrix. A major difference to the 2-D case is that here $\mathbf{T}$ is *guaranteed* to be invertible for $P \geq M_\tau$ independent of the pilot arrangement. However, the pilot distribution affects the condition number of $\mathbf{T}$.

The 1-D Toeplitz system is again solved with the CG method (cf. (9)). The computational complexity of the proposed 1-D channel estimator is based mainly on 1-D FFTs and involves $\mathcal{O}(M_\tau \log(M_\tau))$ operations per CG iteration. Altogether, with 1-D channel estimation the computational complexity and the memory requirements can be significantly reduced compared to 2-D channel estimation. This comes at the cost of an increased pilot overhead or reduced performance.

The 1-D channel estimator can also be successively performed over time and frequency, provided that the pilot arrangement corresponds to a *line-type sampling set* [20]. Here, nonuniformly distributed OFDM training symbols contain irregularly distributed pilot symbols [see Fig. 2(a)]. In the first step, 1-D channel estimation is performed in the k-domain for those OFDM symbols which contain pilot tones. This requires the number of pilots in each OFDM symbol to be no less than M_τ. In the second step, 1-D channel estimation in the n-domain is performed for each subcarrier to obtain the channel coefficients of the remaining OFDM symbols. For this step, the number of OFDM symbols containing pilots has to be at least equal to $M_\upsilon + 1$. A special case of this separable 1-D time-domain and frequency-domain channel estimator is obtained when the OFDM training symbols all have identical (irregular) pilot arrangements. The resulting *product sampling sets* allow for further complexity reduction [34] [see Fig. 2(b)].

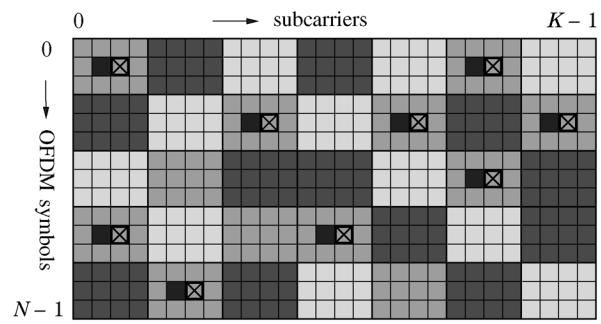

Fig. 3. Illustration of a 2×2 MIMO-OFDMA receive packet with $U = 3$ users (corresponding to three different gray shadings) and 4×3 tiles. The pilot symbols transmitted from the two antennas of one of the users are indicated by black squares and crosses.

VIII. Application to MIMO-OFDMA

In this section, we investigate how the proposed algorithm can be applied to multiuser OFDM systems with multiple antennas (MIMO-OFDMA). With OFDMA, multiple access is achieved via dynamic subcarrier assignments that avoid intra-cell interference and, in conjunction with frequency hopping and coding, allow to exploit diversity gains and to reduce inter-cell interference. Current OFDMA systems (e.g., WiMAX [6]) collect several adjacent subcarriers and time slots into subsets referred to as tiles. An exclusive collection of tiles is dynamically allocated to each user depending on his rate requirements. The distribution of a user's tiles may be highly irregular due to permutations and frequency hopping [6] (cf. Fig. 3). Channel estimation for the OFDMA uplink is challenging since the different user channels need to be simultaneously determined at the base station. In particular, 2-D pilot-assisted channel estimation is difficult since random tile allocations induce an overall pilot arrangement that is irregular (even if the pilot structure within a tile is fixed). This issue is currently dealt with by performing intra-tile channel estimation, i.e., the channel is estimated for each tile separately [35]. However, in highly selective channels intra-tile channel interpolation suffers from an error floor that severely degrades the performance (cf. Section IX-B). The error floor can be lowered only by increasing the number of pilots per tile, thereby reducing the users net data rate. As an example, for the uplink the IEEE 802.16a,e WiMAX standards [6] use 3×4 tiles with four pilots, i.e., a fraction of $\theta = 4/12 = 33\%$ pilots. Parametric inter-tile channel estimation schemes that can deal with irregular tile allocations have been presented in [15] and in [16]. However, both schemes require a fixed tile allocation over time and a specific fixed pilot structure *within* each tile; hence, they are not applicable to completely irregular pilot arrangements. In contrast, our proposed channel estimator does not impose any restrictions on the pilot structure, i.e., the pilot

number and arrangement within each tile and over time is completely arbitrary.[5]

In our MIMO-OFDMA uplink system model, U users have T transmit antennas each to transmit synchronously to a base station equipped with R receive antennas (the case where each user has a different number of transmit antennas is a straightforward generalization). Each user is exclusively allocated a set of symbol and subcarrier positions $\mathcal{U}^{(u)}$ for the transmission of its data and pilot symbols (see Fig. 3). The set of pilot positions on transmit antenna t of user u is denoted by $\mathcal{P}_t^{(u)}$ and $P_t^{(u)} = |\mathcal{P}_t^{(u)}|$ is the corresponding number of pilots. In addition to $\mathcal{P}_t^{(u)} \subseteq \mathcal{U}^{(u)}$, we assume that for different transmit antennas $t \neq t'$ the sets $P_t^{(u)}$ and $P_{t'}^{(u)}$ are disjoint. The latter means that the pilot sets for different transmit antennas of a user do not overlap, thereby allowing separate estimation of the various spatial channels at the base station.

Assuming perfect time and frequency synchronization, the demodulated base station receive signal vector $\mathbf{y}(n,k) = (Y_1(n,k)\dots Y_R(n,k))^T$ at symbol time n and subcarrier k is given by

$$\mathbf{y}(n,k) = \sum_{u=1}^{U} \mathbf{H}^{(u)}(n,k)\mathbf{x}^{(u)}(n,k) + \mathbf{z}(n,k) \tag{18}$$

where $\mathbf{x}^{(u)}(n,k) = \left(X_1^{(u)}(n,k)\dots X_T^{(u)}(n,k)\right)^T$ denotes the transmit vector of user u, $\mathbf{H}^{(u)}(n,k)$ is the $R \times T$ channel matrix whose elements $H_{q,t}^{(u)}(n,k)$ denote the fading coefficients between transmit antenna t of user u and receive antenna q; moreover, $\mathbf{z} = (Z_1(n,k)\dots Z_R(n,k))^T$ is a length-R zero-mean, white, circularly complex Gaussian noise vector. Since $\mathbf{x}^{(u)}(n,k) = \mathbf{0}$ for $(n,k) \notin \mathcal{U}^{(u)}$, we can rewrite (18) as

$$\mathbf{y}(n,k) = \mathbf{H}^{(u)}(n,k)\mathbf{x}^{(u)}(n,k) + \mathbf{z}(n,k), \quad (n,k) \in \mathcal{U}^{(u)}.$$

Taking into account that for each $(n,k) \in \mathcal{P}_t^{(u)}$ only the tth element of $\mathbf{x}^{(u)}$ corresponds to a pilot symbol and all other elements are zero, we further obtain

$$Y_q(n,k) = H_{q,t}^{(u)}(n,k)X_t^{(u)}(n,k) + Z_q(n,k), \quad (n,k) \in \mathcal{P}_t^{(u)}.$$

This equation holds for every $t = 1,\dots,T$, $q = 1,\dots,R$, and $u = 1,\dots,U$ and allows for computing the pre-estimates of the $T \times R \times U$ different scalar time-varying channels $H_{q,t}^{(u)}(n,k)$ at the positions $\mathcal{P}_t^{(u)}$. This provides the basis for the proposed reconstruction procedure described in Sections III and V. If only the channel coefficients for $(n,k) \in \mathcal{U}^{(u)}$ are desired, this can be achieved in the final reconstruction step by omitting those rows of the double Vandermonde matrix $\mathbf{V}$ (cf. (5) and (6)) that correspond to symbol and subcarrier positions not allocated to user u (i.e., $(n,k) \notin \mathcal{U}^{(u)}$). We emphasize once more that our "nonuniform" channel estimator is perfectly suited for this scenario since dynamic tile allocation in general results in highly irregular pilot arrangements. Furthermore, our approach can provide channel estimates for all (n,k) and not only for the user tiles that contain pilots. This enables adaptive resource allocation and even allows to transmit tiles without pilots (cf. Fig. 3).

[5]We recall the condition that the total number of pilots has to be larger than the degrees of freedom of the channel (cf. Section IV-A). Furthermore, using the same pilot pattern in each tile (as in Fig. 3) may improve the performance of our scheme.

IX. Simulation Results

We first apply our channel estimation scheme to a WLAN-like system [4]. Furthermore, we evaluate the performance of the 1-D version of our channel estimator in the context of a WiMAX-like MIMO-OFDMA system [6].

A. 2-D Channel Estimation for a WLAN-Like System

We consider an OFDM system with $K = 64$ subcarriers (of which 12 serve as guard and DC carriers), bandwidth $B = 400$ kHz, carrier frequency $f_c = 2.2$ GHz, cyclic prefix length $L_{cp} = 16$, and packet duration $N = 64$ OFDM symbols. We used a rate-1/2 (13,15) convolutional code, a random block interleaver, and a 16-QAM modulation. Note that this particular choice of the system parameters allows the assessment of the effect of severe Doppler dispersion. A Rayleigh fading channel with wide-sense stationary uncorrelated scattering (WSSUS) [24] was simulated based on [36] with modifications according to [37], using flat delay and Doppler profiles (note that within each OFDM packet the transfer function $H(n,k)$ of the simulated channel is not strictly but only approximately band-limited). Unless stated otherwise, the channel had a maximum delay spread of 17.5 μs (corresponding to $M_\tau = 8$) and a maximum Doppler frequency of $f_{\upsilon,\max} = 204$ Hz (corresponding to a terminal velocity of $\upsilon_{\max} = 100$ km/h or equivalently a relative Doppler of 3.2%). Note that according to the rule of thumb stated in (16), this results in a discrete Doppler spread of $M_\upsilon = 8$. At the receiver the output of our channel estimator was provided to a simple zero-forcing detector [38] followed by a hard-decision Viterbi decoder. All results shown were obtained by averaging over $2 \cdot 10^4$ OFDM packets.

An irregular arrangement of $P = 165$ or $P = 495$ pilots was chosen randomly according to a uniform distribution for each packet (this corresponds to respectively $\theta = 5\%$ and $\theta = 15\%$ of pilots). Despite such a small number of pilots, the resulting channel estimation MSE turned out be close to optimal (see below). We note that such a completely random arrangement corresponds to a worst-case scenario with regard to performance and complexity in the sense that—in contrast to many practical setups—the pilot pattern does not feature any structure whatsoever. An example for a more structured (but still irregular) pilot arrangement was considered in [1].

We show results obtained with two variants of our proposed method: the first (practical) variant (labeled "IP-SC") terminates the CG iterations according to the stopping criterion (15) with $\zeta = 0.999$; the other variant (labeled "IP-OT") serves as benchmark and uses optimal termination, i.e., it stops the CG iterations when the squared error $\epsilon^2(r)$ in (13) is minimal (of course this is not feasible in practice). Furthermore, we limited the number of CG iterations to maximally $M = M_\tau(M_\upsilon + 1) = 72$. Our proposed 2-D channel estimator applied adaptive weights as described in Appendix A and performed zero padding according to Section V-C.[6] Unless stated otherwise, the parameters $\hat{M}_\tau$ and $\hat{M}_\upsilon$ (required for our proposed algorithm) were chosen equal to the actual discrete channel delay M_τ and Doppler spread M_υ, respectively.

[6]Additional simulations (not shown here due to space limitations) revealed that in our transmission setup no further improvements in terms of performance and complexity can be achieved when using preconditioning instead of adaptive weights.

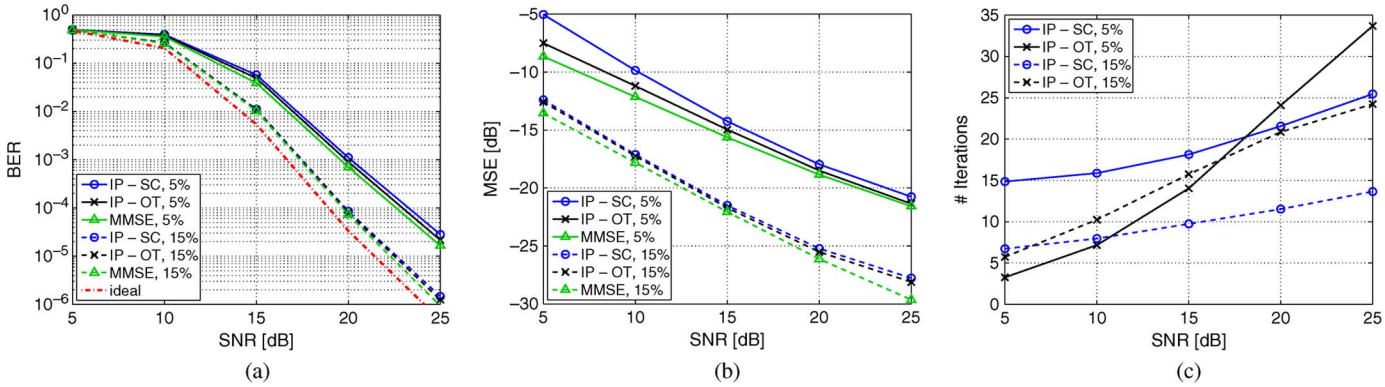

Fig. 4. Performance assessment of proposed channel estimator for WLAN-like system: (a) BER versus SNR; (b) MSE versus SNR; and (c) average number of iterations versus SNR.

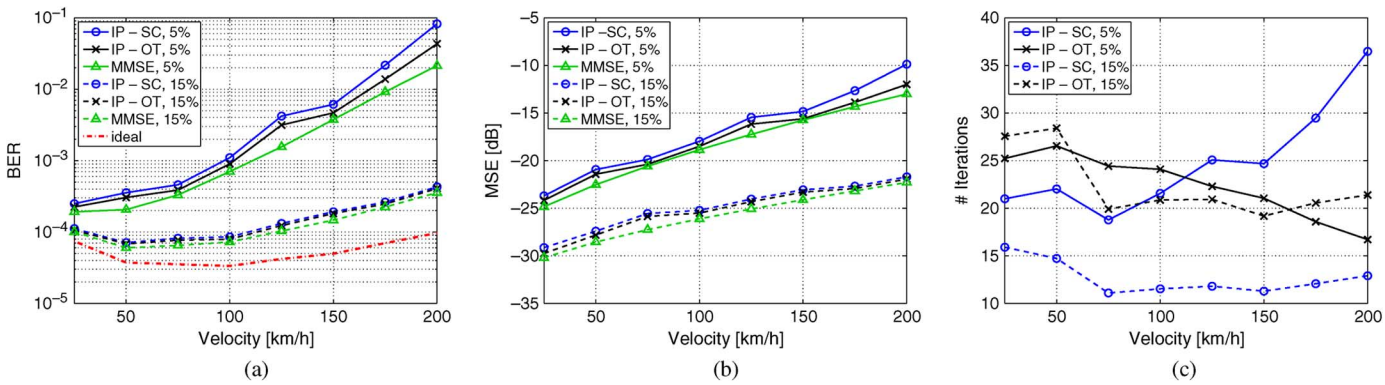

Fig. 5. Performance assessment of proposed channel estimator for WLAN-like system at an SNR of 20 dB: (a) BER versus velocity; (b) MSE versus velocity; and (c) average number of iterations versus velocity.

We further compare our algorithm with a 2-D MMSE estimator (labeled "MMSE"), which assumes knowledge of the channel covariance and the noise power (cf. [12]), and with an ideal receiver that has perfect channel knowledge available (labeled "ideal").

1) Performance Versus Signal-to-Noise Ratio (SNR): Fig. 4 shows bit error rate (BER), normalized MSE (obtained by averaging (13) over all OFDM packets), and average number of iterations versus the SNR $\rho = \mathcal{E}\{|X(n,k)|^2\}/\sigma_Z^2$. It is seen in Fig. 4(a) and (b) that the proposed scheme, both with optimal and heuristic stopping criterion, has virtually the same performance as the much more complex 2-D MMSE estimator. This also shows that our stopping criterion achieves essentially the same MSE as optimal CG termination. With 15% of pilots, our method comes to within a dB of the BER achieved by a receiver that uses perfect channel knowledge, whereas in the case of 5% pilots all schemes feature an SNR gap of about 5 dB to the ideal receiver.

Fig. 4(c) compares the average number of iterations of IP-SC and IP-OT. For both stopping rules, the number of iterations increases with increasing SNR. This can be explained by the fact that the residual error of the CG iterations initially decreases but deteriorates after a certain number of iterations due to noise enhancement. At high SNR the noise enhancement kicks in later (i.e., after more iterations) than at low SNR. Since both stopping criteria aim at terminating the CG algorithm when the residual error is (close to) minimal, more iterations are required at high SNR. We note that fewer iterations at high SNR could be enforced at the cost of larger residual errors by making the parameter ζ in (15) SNR dependent. For 15% of pilots and SNRs larger than 10 dB, our stopping criterion performs significantly fewer iterations than optimal termination without noticeable performance loss [cf. Fig. 4(b)]. For 5% pilots, IP-SC on average terminates earlier than IP-OT for SNRs larger than 18 dB. For lower SNRs, our stopping criterion leads to more iterations; however, the difference is significant only for quite low SNRs which are practically less relevant since they result in a rather high BER.

2) Impact of Receiver Velocity: We next investigate the effect of terminal velocity (equivalent to the channel's Doppler frequency) on the performance. Note that $\hat{M}_v$ is chosen, as a function of the terminal velocity, according to the rule of thumb (16). Fig. 5 shows the resulting BER, MSE, and average number of iterations at a fixed SNR of 20 dB. It is seen that the performance achieved with the proposed method again closely approaches that of the 2-D MMSE estimator. At certain velocities (e.g., $v_{\max} = 50$ km/h and $v_{\max} = 125$ km/h), we can see a slight performance degradation. This can be attributed to the fact that the proposed rule of thumb (16) here is slightly too pessimistic, i.e., $\hat{M}_v$ sometimes tends to be too large [compare with the observations for the mismatched case shown in Fig. 7(b)]. Larger velocities are seen to result in an increased MSE and eventually also a higher BER. This can be explained by the fact that ICI and the condition number of $\mathbf{T}$ both increase with in-

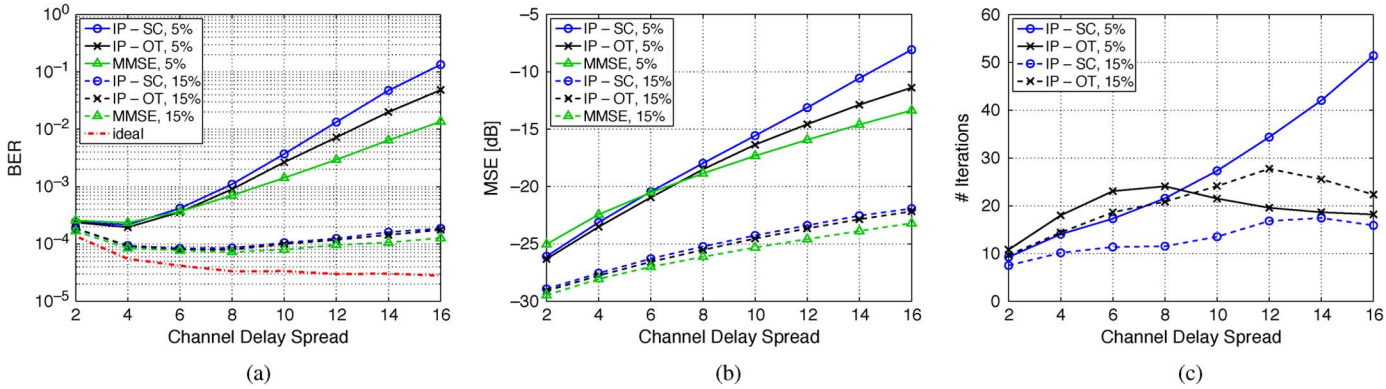

Fig. 6. Performance assessment of proposed channel estimator for WLAN-like system at an SNR of 20 dB: (a) BER versus delay spread M_τ; (b) MSE versus delay spread M_τ; and (c) average number of iterations versus delay spread M_τ.

creasing velocity. For 15% of pilots and velocities below 100 km/h, these effects are compensated by Doppler diversity gains [see Fig. 5(a)]. Fig. 5(c) confirms the statements about the proposed stopping criterion made above. With 15% of pilots, the proposed stopping rule terminates about twice as fast as optimal termination for all velocities. Significantly increased iteration counts with the heuristic stopping criterion are observed only for few pilots (5%) and velocities larger than 150 km/h.

3) Impact of Delay Spread: Fig. 6 shows that the effect of the channel delay spread on the performance is analogous to that of the Doppler spread discussed before (again we chose an SNR of 20 dB). We recall that $\hat{M}_\tau$ exactly equals the actual channel delay spread. Again, large delay spreads increase the MSE and thus degrade the BER performance (at low delay spreads, this degradation is compensated by the opposite trend of increasing delay diversity). We note that the slightly inferior performance of the 2-D MMSE estimator at low channel delay spreads $M_\tau \leq 6$ is due to edge effects in its design. Our stopping rule is again quickest to terminate for 15% pilots but iterates noticeably too long for 5% pilots and delay spreads beyond 10 [cf. Fig. 6(c)].

4) Mismatched Case: We finally study the impact of a mismatch between the actual delay and Doppler parameters M_τ and M_υ and their estimates $\hat{M}_\tau$ and $\hat{M}_\upsilon$. Fig. 7 shows the MSE versus $\hat{M}_\tau$ and $\hat{M}_\upsilon$ for a channel with $M_\tau = 8$ and $M_\upsilon = 8$ and an SNR of 20 dB. It can be seen that choosing the channel parameters too small (i.e., $\hat{M}_\tau < M_\tau$ and $\hat{M}_\upsilon < M_\upsilon$) has a highly detrimental effect on the MSE since the channel's spreading function is not being reconstructed in the region $\tau \geq \hat{M}_\tau$ and $\nu \geq \hat{M}_\upsilon$; this amounts to oversmoothing of the transfer function estimate. In contrast, for $\hat{M}_\tau > M_\tau$ and $\hat{M}_\upsilon > M_\upsilon$ the MSE degradation is more graceful, specifically for a large number of pilots and high SNRs; here, the parameter mismatch merely causes slightly poorer noise suppression, i.e., the channel estimator overfits the noise.

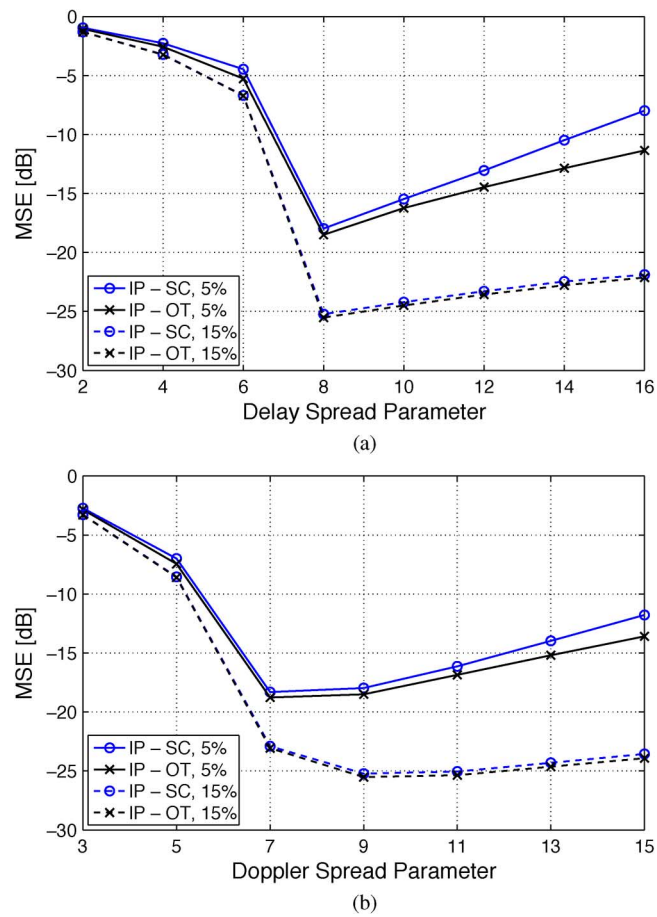

Fig. 7. Performance of proposed methods for channel parameter mismatch at an SNR of 20 dB and true delay and Doppler spreads of $M_\tau = 8$ and $M_\upsilon = 8$: (a) MSE versus $\hat{M}_\tau$ and (b) MSE versus $\hat{M}_\upsilon$.

B. 1-D Channel Estimation for MIMO-OFDMA

We next consider the application of our scheme to 1-D channel estimation in a multiple antenna multiuser scenario. We simulated a coded MIMO-OFDMA uplink system with $U = 3$ users, where each user and the base station had $T = R = 2$ antennas. The system bandwidth and carrier frequency were chosen as $B = 2$ MHz and $f_c = 2.2$ GHz. Each user employed 16-QAM, a rate-1/2 (13,15) convolutional code, and a random block interleaver. The OFDMA system had $K = 256$ subcarriers (of which 40 served as guard and DC carriers), a cyclic prefix length of $L_{cp} = 32$, and a packet length of $N = 18$ OFDMA symbols. Each OFDMA packet was divided into six slots consisting of three consecutive OFDMA symbols each. Every slot was split into 54 tiles, each comprising four

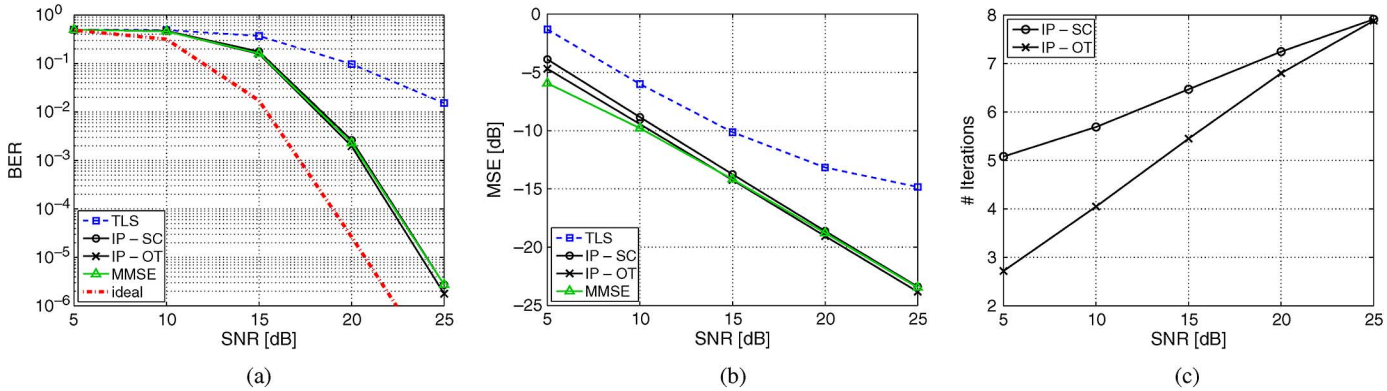

Fig. 8. Performance of 1-D channel estimation for a 2×2 MIMO-OFDMA system with $U = 3$ users and 16.6% pilots/user: (a) BER versus SNR; (b) MSE versus SNR; and (c) average number of iterations versus SNR.

subcarriers. The allocation of the 3×4 tiles to the users was performed randomly so that the tiles of each user were irregularly distributed within the packet. For simplicity, we assumed that each user was allocated the same number of tiles within a packet, resulting in 108 tiles per user and 18 tiles per user/slot. Within each tile, one pilot position was dedicated to each transmit antenna of a specific user, resulting in 18 pilots per user, per slot, per transmit antenna (i.e., a fraction of $\theta = 16.6\%$ pilots per user). Furthermore, the pilots of different antennas did not interfere. All pilot symbols were temporally aligned; thus, every third OFDMA symbol contained pilots which were irregularly distributed over all the subcarriers within an OFDMA symbol (cf. Fig. 3). This arrangement corresponds to a line-type sampling set as shown in Fig. 2(a).

All user channels were WSSUS Rayleigh fading (the channel model was the same as in Section IX-A). Here, the channel had a maximum delay spread of 4.5 μs (corresponding to $M_\tau = 10$) and a negligible Doppler frequency of 20.4 Hz (corresponding to terminal velocities of $v_{\max} = 10$ km/h). Thus, the channels were approximately constant within a slot of three OFDMA symbols.

We performed 1-D channel estimation in the frequency domain (as described in Section VII) by reconstructing all user channels for those OFDMA symbols which contain pilot symbols; then, these estimates were used for all OFDMA symbols within the corresponding slot. The received MIMO signal at each subcarrier was demodulated using a hard-output sphere decoder [39] and decoded using the Viterbi algorithm. All results were obtained by averaging over at least $3{\cdot}10^4$ OFDMA packets.

Our channel estimator used $\hat{M}_\tau = M_\tau = 10$ and both heuristic termination ("IP-SC") with $\zeta = 0.999$ and optimal ("IP-OT") termination of the CG iterations. Similarly to Section IX-A, we limited the number of CG iterations to maximally $\hat{M}_\tau = 10$ CG iterations and applied adaptive weights and zero padding. We compare our scheme with three reference methods[7]: an ideal receiver using perfect CSI (labeled "ideal"), a 1-D MMSE estimator (labeled "MMSE") designed according to [12], and an LS channel estimator applied on a per-tile basis (labeled "TLS"). The latter calculates the channel estimates at the pilot positions using (3) and then uses these estimates to detect all symbols within the corresponding tile. Note that in contrast to the per-tile LS estimator, our proposed estimator and the MMSE estimator allow to reconstruct the user channels for all the subcarriers in a packet.

Fig. 8(a) and (b) shows the BER of a user and the normalized MSE (computed over the allocated user tiles) versus the SNR. It can be seen that our scheme significantly outperforms per-tile LS channel estimation and achieves effectively the same performance as MMSE estimation. At high SNR, per-tile LS channel estimation suffers from an error floor [cf. Fig. 8(b)]. Moreover, at high SNR the performance of the 2-D MMSE estimator is again slightly degraded due to edge effects affecting its design. Note that there is almost no performance difference between the heuristic and the optimal stopping rule in our proposed scheme. There remains a 3 dB SNR gap between the BER of our scheme and that of the ideal receiver which is partly due to the fact that 1-D channel estimation ignores the time-variation of the channel within each slot. Finally, Fig. 8(c) compares the average number of iterations resulting from the two stopping rules. It is seen that even though the iteration count increases with the SNR, in both cases a small number of iterations is sufficient. In this scenario, our heuristic stopping criterion tends to perform one or two extra iterations in comparison to optimal termination.

X. Conclusion

We considered OFDM packet transmissions over doubly selective fading channels and introduced a novel two-dimensional pilot-assisted channel estimation scheme that is specifically tailored to nonuniform pilot arrangements. Our estimator requires coarse knowledge of the maximum delay and Doppler spread of the channel and is implemented in a computationally efficient manner using conjugate gradient iterations and fast Fourier transforms. The influence of the pilot distribution on the channel estimation error was assessed using a mean square error analysis. We further provided a detailed description of

[7]We note that the estimation techniques described in [15] and [16] are not applicable to our simulation scenario since they require at least two pilots per tile and transmit antenna.

the implementation details of the algorithm, presented a reformulation for the one-dimensional case, and discussed the application to MIMO-OFDMA systems. Extensive simulation results showed that our channel estimator performs as well as the minimum mean square error estimator, which is computationally much more expensive and requires second-order statistics of the channel and the noise.

Appendix A
Computation of Adaptive Weights

In order to make the following discussion compact, we define operators for the symplectic 2-D discrete Fourier transform (DFT) and its inverse:

$$\begin{aligned}\Psi(\tau,\nu) &= \mathrm{DFT}\{\psi(n,k)\}\\ &\triangleq \frac{1}{\sqrt{NK}}\sum_{n=0}^{N-1}\sum_{k=0}^{K-1}\psi(n,k)e^{-j2\pi\left(\frac{n\nu}{N}-\frac{k\tau}{K}\right)},\\ \psi(n,k) &= \mathrm{IDFT}\{\Psi(\tau,\nu)\}\\ &\triangleq \frac{1}{\sqrt{NK}}\sum_{\nu=0}^{N-1}\sum_{\tau=0}^{K-1}\Psi(\tau,\nu)e^{j2\pi\left(\frac{n\nu}{N}-\frac{k\tau}{K}\right)}.\end{aligned} \tag{19}$$

In practice these transformations are implemented using FFT algorithms. The 2-D cyclic convolution of the two functions $\psi_1(n,k)$ and $\psi_2(n,k)$ is then denoted as

$$(\psi_1 \circledast \psi_2)(n,k) \triangleq \mathrm{IDFT}\{\mathrm{DFT}\{\psi_1(n,k)\}\cdot\mathrm{DFT}\{\psi_2(n,k)\}\}. \tag{20}$$

In the following, we assume $\bar{N}=N$, $\hat{M}_\tau = M_\tau$, $\hat{M}_\upsilon = M_\upsilon$ (and consequently $\hat{M}=M$).

We next describe how to efficiently determine adaptive weights based on FFTs and the inverse of the local pilot density [28]. We first define the pilot indicator function $\Gamma_{\mathrm{p}}(n,k)=\sum_{p=1}^{P}\delta(n-n_p,k-k_p)$ (here, $\delta(.)$ denotes the Kronecker delta function) and a neighborhood function $\Upsilon(n,k)$ that describes the closeness of two points whose coordinates differ by (n,k), i.e., $\Upsilon(n,k)$ decays with increasing n and k and satisfies $\Upsilon(0,0)=0$.

The local pilot density can then be computed by the cyclic convolution (cf. (20)) of the neighborhood function $\Upsilon(n,k)$ with the pilot positions $\Gamma_{\mathrm{p}}(n,k)$; thus, the inverse of the local pilot density at the nonpilot positions is given by

$$\Omega(n,k)=\frac{1-\Gamma_{\mathrm{p}}(n,k)}{(\Gamma_{\mathrm{p}}\circledast\Upsilon)(n,k)}.$$

The adaptive weights are finally given by $w_p = w(n_p,k_p), p=1,\ldots,P$, where

$$w(n,k)=(\Omega\circledast\Upsilon)(n,k)+1.$$

We note that in our simulations, we used the neighborhood function

$$\Upsilon(n,k)=\begin{cases}3, & \text{for } (n,k)\in\{(0,1),(1,0),(N-1,0),\\ & \qquad (0,K-1)\}\\ 2, & \text{for } (n,k)\in\{(N-1,1),(1,K-1),\\ & \qquad (1,1),(N-1,K-1)\}\\ 0, & \text{else.}\end{cases}$$

Appendix B
Fast Computation of $\mathbf{T}$ and $\tilde{\mathbf{s}}_{\mathrm{pre}}$

To simplify notation, we assume $\bar{N}=N$, $\hat{M}_\tau=M_\tau$, $\hat{M}_\upsilon=M_\upsilon$ (and consequently $\hat{M}=M$) in the following. Due to its BTTB structure, $\mathbf{T}$ is completely described by its first and M_τth column. These columns can be efficiently calculated using 2-D FFTs (cf. [19]). Since $\mathbf{T}=\mathbf{V}_{\mathrm{p}}^H\mathbf{W}\mathbf{V}_{\mathrm{p}}$, the elements of the first column of $\mathbf{T}$ are given by (cf. (8))

$$[\mathbf{T}]_{m,1}=\frac{1}{NK}\sum_{p=1}^{P}w_p e^{-j2\pi\left(n_p\frac{\nu}{N}-k_p\frac{\tau}{K}\right)} \tag{21}$$

where $m=\tau+[(M_\upsilon/2)+\nu]M_\tau+1$ with $\tau=0,\ldots,M_\tau-1$, $\nu=-M_\upsilon/2,\ldots,M_\upsilon/2$ and $(n_p,k_p)\in\mathcal{P}$, $p=1,\ldots,P$, denotes the pilot positions. Comparing (21) to the 2-D symplectic DFT $\Psi(\tau,\nu)=\mathrm{DFT}\{\psi(n,k)\}$ defined in (19), it is seen that by choosing

$$\psi(n,k)=\begin{cases}w_p/\sqrt{NK}, & \text{for } (n,k)=(n_p,k_p)\\ 0, & \text{else}\end{cases}$$

it follows that $[\mathbf{T}]_{m,1}=\Psi(\tau,\nu+M_\upsilon/2)$. In a similar manner, the elements of the M_τth column can be obtained as $[\mathbf{T}]_{m,M_\tau}=\Psi(\tau-M_\tau+1,\nu+M_\upsilon/2)$.

Since the elements of $\tilde{\mathbf{s}}_{\mathrm{pre}}$ are given by (cf. (8))

$$[\tilde{\mathbf{s}}_{\mathrm{pre}}]_m=\sum_{p=1}^{P}w_p\hat{H}_{\mathrm{pre}}(n_p,k_p)e^{-j2\pi\left(n_p\frac{\nu}{N}-k_p\frac{\tau}{K}\right)},$$

it can be analogously calculated using a 2-D FFT, i.e., $[\tilde{\mathbf{s}}_{\mathrm{pre}}]_m=\Psi(\tau,\nu+M_\upsilon/2)$ with

$$\psi(n,k)=\begin{cases}\sqrt{NK}w_p\hat{H}_{\mathrm{pre}}(n_p,k_p), & \text{for } (n,k)=(n_p,k_p)\\ 0, & \text{else.}\end{cases}$$

Appendix C
Fast Computation of Matrix-Vector Multiplication

Exploiting the BTTB structure of the $M\times M$ matrix $\mathbf{T}$, we next discuss an efficient implementation of the matrix-vector multiplication $\mathbf{c}=\mathbf{T}\mathbf{b}$ (cf. (9g)). This implementation is based on FFTs and requires only $\mathcal{O}(M\log(M))$ operations instead of $\mathcal{O}(M^2)$ (recall that $M=M_\tau(M_\upsilon+1)$). The main idea is to embed the $M\times M$ BTTB matrix $\mathbf{T}$ into a $4M\times 4M$ block-circulant matrix with circulant blocks that can be diagonalized

by a 2-D FFT [29]. We start by writing $\mathbf{T} = \mathbf{V}_{\mathrm{p}}^H \mathbf{W} \mathbf{V}_{\mathrm{p}}$ (as defined in (8)) as a $(M_\upsilon+1)\times(M_\upsilon+1)$ block Toeplitz matrix of the form

$$\mathbf{T} = \begin{pmatrix} \mathbf{T}^{(0)} & \mathbf{T}^{(-1)} & \dots & \mathbf{T}^{(-M_\upsilon)} \\ \mathbf{T}^{(1)} & \mathbf{T}^{(0)} & \dots & \mathbf{T}^{(-M_\upsilon+1)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{T}^{(M_\upsilon)} & \mathbf{T}^{(M_\upsilon-1)} & \dots & \mathbf{T}^{(0)} \end{pmatrix}$$

with each block $\mathbf{T}^{(\nu)}$, $\nu = -M_\upsilon, \dots, M_\upsilon$, being a $M_\tau \times M_\tau$ Toeplitz matrix. It can be shown that $\mathbf{T}$ has Hermitian block structure, i.e., $[\mathbf{T}^{(\nu)}]^H = \mathbf{T}^{(-\nu)}$. However, except for $\nu \neq 0$, the blocks $\mathbf{T}^{(\nu)}$ themselves are not Hermitian [19]. We rearrange the distinct elements of each Toeplitz block $\mathbf{T}^{(\nu)}$ into length-$2M_\tau$ vectors $\tilde{\mathbf{t}}^{(\nu)}$ according to

$$\tilde{\mathbf{t}}^{(\nu)} = \left(T_{1,1}^{(\nu)}\ T_{2,1}^{(\nu)}\ \dots\ T_{M_\tau,1}^{(\nu)}\ 0\ T_{1,M_\tau}^{(\nu)}\ T_{1,M_\tau-1}^{(\nu)}\ \dots\ T_{1,2}^{(\nu)}\right)^T.$$

These vectors are then grouped into the following $2M_\tau \times 2(M_\upsilon+1)$ matrix

$$\tilde{\mathbf{T}} = \left(\tilde{\mathbf{t}}^{(0)}\ \dots\ \tilde{\mathbf{t}}^{(M_\upsilon)}\ \mathbf{0}\ \tilde{\mathbf{t}}^{(-M_\upsilon)}\ \dots\ \tilde{\mathbf{t}}^{(-1)}\right).$$

We next define the functions

$$\tilde{T}(\tau,\nu) = \begin{cases} [\tilde{\mathbf{T}}]_{\tau+1,\nu+1}, & \text{for } \tau = 0,\dots,2M_\tau-1 \\ & \quad \nu = 0,\dots,2M_\upsilon+1 \\ 0, & \text{else} \end{cases}$$

and

$$b(\tau,\nu) = \begin{cases} [\mathbf{b}]_{\tau+\nu M_\tau+1}, & \text{for } \tau = 0,\dots,M_\tau-1 \\ & \quad \nu = 0,\dots,M_\upsilon \\ 0, & \text{else.} \end{cases}$$

The matrix-vector multiplication now amounts to a cyclic convolution over the domain $[0, 2M_\tau-1]\times[0, 2M_\upsilon+1]$, computed via 2-D FFTs of dimensions $2M_\tau \times 2(M_\upsilon+1) = 4M$ according to (20), i.e.,

$$c(\tau,\nu) = (\tilde{T} \circledast b)(\tau,\nu).$$

The vector $\mathbf{c}$ is finally obtained as $[\mathbf{c}]_m = c(\tau, \nu + M_\upsilon/2)$, where $m = \tau + [(M_\upsilon/2)+\nu]M_\tau + 1$ with $\tau = 0,\dots,M_\tau-1$, $\nu = -M_\upsilon/2,\dots,M_\upsilon/2$.

ACKNOWLEDGMENT

The authors thank T. Strohmer and H. G. Feichtinger for useful discussions and for providing Matlab code.

REFERENCES

[1] P. Fertl and G. Matz, "Efficient OFDM channel estimation in mobile environments based on irregular sampling," in *Proc. Asilomar Conf. Signals, Systems, Computers*, Pacific Grove, CA, Oct.–Nov. 2006, pp. 1777–1781.

[2] P. Fertl and G. Matz, "Multi-user channel estimation in OFDMA uplink systems based on irregular sampling and reduced pilot overhead," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Honolulu, HI, Apr. 2007, vol. 3, pp. 297–300.

[3] J. A. C. Bingham, "Multicarrier modulation for data transmission: An idea whose time has come," *IEEE Commun. Mag.*, vol. 28, no. 5, pp. 5–14, May 1990.

[4] IEEE P802 LAN/MAN Committee, The Working Group for Wireless Local Area Networks (WLANs) [Online]. Available: http://grouper.ieee.org/groups/802/11/

[5] *Wireless Access in Vehicular Environments (WAVE)*, IEEE P802.11, IEEE LAN/MAN Standards Committee, Jan. 2006, Draft Amendment.

[6] IEEE P802 LAN/MAN Committee, The Working Group on Broadband Wireless Access Standards [Online]. Available: http://grouper.ieee.org/groups/802/16/

[7] 3GPP TR 25.913, "Requirements for Evolved UTRA (E-UTRA) and Evolved UTRAN (E-UTRAN)," V7.3.0 Mar. 2006.

[8] J.-J. van de Beek, O. Edfors, M. Sandell, S. K. Wilson, and P. O. Börjesson, "On channel estimation in OFDM systems," in *Proc. IEEE Vehicular Technology Conf. (VTC)*, Chicago, IL, Jul. 1995, pp. 815–819.

[9] R. Negi and J. Cioffi, "Pilot tone selection for channel estimation in a mobile OFDM system," *IEEE Trans. Consum. Electron.*, vol. 44, no. 3, pp. 1122–1128, Aug. 1998.

[10] Y. Li, "Pilot-symbol-aided channel estimation for OFDM in wireless systems," *IEEE Trans. Veh. Technol.*, vol. 49, no. 4, pp. 1207–1215, Jul. 2000.

[11] O. Edfors, M. Sandell, J.-J. van de Beek, S. K. Wilson, and P. O. Börjesson, "OFDM channel estimation by singular value decomposition," *IEEE Trans. Commun.*, vol. 46, no. 7, pp. 931–939, Jul. 1998.

[12] P. Hoeher, S. Kaiser, and P. Robertson, "Pilot-symbol-aided channel estimation in time and frequency," in *Proc. IEEE GLOBECOM*, Phoenix, AZ, Nov. 1997, pp. 90–96.

[13] Y. Li, L. Cimini, and N. Sollenberger, "Robust channel estimation for OFDM systems with rapid dispersive fading channels," *IEEE Trans. Commun.*, vol. 46, no. 7, pp. 902–915, Jul. 1998.

[14] M. Sandell and O. Edfors, "A comparative study of pilot-based channel estimators for wireless OFDM," Luleå Univ. Technol., Luleå, Sweden, Div. of Signal Processing, Res. Rep. TULEA 1996:19, Sep. 1996.

[15] M. R. Raghavendra, E. Lior, S. Bhashyam, and K. Giridhar, "Parametric channel estimation for pseudo-random tile-allocation in uplink OFDMA," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5370–5381, Nov. 2007.

[16] X. Zhao, T. Peng, and W. Wang, "Parametric channel estimation by exploring hopping pilots in uplink OFDMA," in *Proc. IEEE PIMRC*, Cannes, France, Sep. 2008, pp. 1–5.

[17] O. Ureten and N. Serinken, "Decision directed iterative equalization of OFDM symbols using non-uniform interpolation," in *Proc. IEEE Vehicular Technology Conf.*, Montreal, QC, Canada, Sep. 2006, pp. 1–5.

[18] Z. Tang and G. Leus, "Time-multiplexed training for time-selective channels," *IEEE Signal Process. Lett.*, vol. 14, no. 9, pp. 585–588, Sep. 2007.

[19] T. Strohmer, "Computationally attractive reconstruction of bandlimited images from irregular samples," *IEEE Trans. Image Process.*, vol. 6, no. 4, pp. 540–548, Apr. 1997.

[20] K. Gröchenig and T. Strohmer, "Numerical and theoretical aspects of non-uniform sampling of band-limited images," in *Nonuniform Sampling: Theory*, F. Marvasti, Ed. New York: Kluwer, 2001, ch. 6, pp. 283–324.

[21] J. Kovačević and A. Chebira, "Life beyond bases: The advent of frames (Part I)," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 86–104, Jul. 2007.

[22] E. Margolis and Y. C. Eldar, "Nonuniform sampling of periodic bandlimited signals," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 2728–2745, Jul. 2008.

[23] A. Peled and A. Ruiz, "Frequency domain data transmission using reduced computational complexity algorithms," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Denver, CO, Apr. 1980, pp. 964–967.

[24] P. A. Bello, "Characterization of randomly time-variant linear channels," *IEEE Trans. Commun. Syst.*, vol. 11, no. 4, pp. 360–393, Dec. 1963.
[25] *Nonuniform Sampling: Theory and Practice*, F. Marvasti, Ed. New York: Kluwer Academic/Plenum, 2001.
[26] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: The Johns Hopkins Univ. Press, 1996.
[27] A. Okabe, B. Boots, K. Sugihara, and S. N. Chiu, *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*, 2nd ed. Chichester, U.K.: Wiley, 2000.
[28] H. Feichtinger and T. Strohmer, "Fast iterative reconstruction of band-limited images from non-uniform sampling values," in *Conf. Computer Analysis Images Patterns*, Budapest, Hungary, 1993, pp. 82–91.
[29] R. H. Chan, J. G. Nagy, and R. J. Plemmons, "FFT-based preconditioners for Toeplitz-block least squares problems," *SIAM J. Numer. Anal.*, vol. 30, no. 6, pp. 1740–1768, Dec. 1993.
[30] A. Nordio, C.-F. Chiasserini, and E. Viterbo, "Reconstruction of multidimensional signals from irregular noisy samples," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4274–4285, Sep. 2008.
[31] A. Nordio, C.-F. Chiasserini, and E. Viterbo, "Performance of linear field reconstruction techniques with noise and uncertain sensor locations," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3535–3547, Aug. 2008.
[32] M. Hanke, *Conjugate Gradient Type Methods for Ill-Posed Problems*. Harlow, U.K.: Longman Scientific & Technical, 1995.
[33] H. Feichtinger, K. Gröchenig, and T. Strohmer, "Efficient numerical methods in non-uniform sampling theory," *Numer. Math.*, vol. 69, no. 4, pp. 423–440, Feb. 1995.
[34] H. G. Feichtinger and T. Strohmer, "Recovery of missing segments and lines in images," *Opt. Eng. (Special Issue on Digital Image Recovery and Synthesis)*, vol. 33, no. 10, pp. 3283–3289, Oct. 1994.
[35] J. Bonnet and G. Auer, "Chunk-based channel estimation for uplink OFDM," in *Proc. IEEE Vehicular Technology Conf. (Spring)*, Melbourne, Australia, May 2006, vol. 4, pp. 1555–1559.
[36] Y. R. Zheng and C. Xiao, "Simulation models with correct statistical properties for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 51, no. 6, pp. 920–928, Jun. 2003.
[37] T. Zemen and C. F. Mecklenbräuker, "Time-variant channel estimation using discrete prolate spheroidal sequences," *IEEE Trans. Signal Process.*, vol. 53, no. 9, pp. 3597–3607, Sep. 2005.
[38] A. Paulraj, R. U. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
[39] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, "Closest point search in lattices," *IEEE Trans. Inf. Theory*, vol. 48, no. 8, pp. 2201–2214, Aug. 2002.

Peter Fertl (M'09) received the Dipl.-Ing. and Dr.techn. degrees in electrical/communication engineering from the Vienna University of Technology, Vienna, Austria, in 2005 and 2009, respectively.

From 2006 to 2009, he was a Research and Teaching Assistant at the Institute of Communications and Radio-Frequency Engineering, Vienna University of Technology. He also worked at Nokia Research Center, Helsinki, as a Visiting Researcher from July to September 2007. His research interests are in wireless communications with emphasis on MIMO multiuser signal processing and wireless relay networks.

Gerald Matz (S'95–M'01–SM'07) received the Dipl.-Ing. and Dr. techn. degrees in electrical engineering in 1994 and 2000, respectively, and the Habilitation degree for communication systems in 2004, all from Vienna University of Technology, Vienna, Austria.

Since 1995, he has been with the Institute of Communications and Radio-Frequency Engineering, Vienna University of Technology, where he currently holds a tenured position as Associate Professor. From March 2004 to February 2005, he was on leave as an Erwin Schrödinger Fellow with the Laboratoire des Signaux et Systèmes, Ecole Supérieure d'Electricité, France. During summer 2007, he was a Guest Researcher with the Communication Theory Lab at ETH Zurich, Switzerland. He has directed or actively participated in several research projects funded by the Austrian Science Fund (FWF) and by the European Union. He has published some 120 papers in international journals, conference proceedings, and edited books. His research interests include wireless communications, statistical signal processing, and information theory.

Prof. Matz serves as a member of the IEEE Signal Processing Society (SPS) Technical Committee on Signal Processing for Communications and Networking and of the IEEE SPS Technical Committee on Signal Processing Theory and Methods. He is an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and for the EURASIP journal SIGNAL PROCESSING. From 2004 to 2008, he was an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS. He was Technical Program Co-Chair of EUSIPCO 2004 and has been on the Technical Program Committee of numerous international conferences. In 2006, he received the Kardinal Innitzer Most Promising Young Investigator Award.