Pilot-Assisted Time-Varying Channel Estimation for OFDM Systems

Zijian Tang, *Student Member, IEEE*, Rocco Claudio Cannizzaro, Geert Leus, *Senior Member, IEEE*, and Paolo Banelli, *Member, IEEE*

***Abstract*—In this paper, we deal with channel estimation for orthogonal frequency-division multiplexing (OFDM) systems. The channels are assumed to be time-varying (TV) and approximated by a basis expansion model (BEM). Due to the time-variation, the resulting channel matrix in the frequency domain is no longer diagonal, but approximately banded. Based on this observation, we propose novel channel estimators to combat both the noise and the out-of-band interference. In addition, the effect of a receiver window on channel estimation is also studied. Our claims are supported by simulation results, which are obtained considering Jakes' channels with fairly high Doppler spreads.**

***Index Terms*—Basis expansion model (BEM), orthogonal frequency-division multiplexing (OFDM), pilot-assisted modulation, time-varying (TV) channels.**

I. Introduction

In mobile communications, high speeds of terminals and/or scatterers cause Doppler effects that could seriously affect the performance. To understand this problem, we need accurate models of the Doppler-affected or time-varying (TV) channels, whose channel taps vary with time. A common approach is to describe the channel taps statistically by their Doppler spectrum, which typically is bathtub-shaped [1], bell-shaped [2], or a combination thereof. The statistics assumed in these channel models are usually based on physical propagation parameters such as the path delays, path phases, path frequencies, path angles of arrival, etc. [1]. Despite their accuracy, these statistical models are generally bulky and difficult to handle. Therefore, many existing works resort to a parsimonious channel model such as the basis expansion model (BEM).

The BEM that is optimal in terms of the mean square error (MSE) is the so-called discrete Karhuen–Loève BEM (DKL-BEM) [3]–[5], which is, in essence, a reduced-rank decomposition of a certain type of Doppler spectrum, e.g., the bathtub-shaped or bell-shaped spectrum. The problem though is that if the assumed channel statistics deviate from the true scenario, which is very likely in practice, the DKL-BEM will perform suboptimally. As a compromise, one can derive a BEM that is based on a general approximation for all kinds of channel statistics. For instance, the discrete prolate spheroidal BEM (DPS-BEM) corresponds to the DKL-BEM with a rectangular spectrum [6]. It is featured by a set of orthogonal spheroidal functions that are perfectly band-limited but have maximal time concentration within the considered interval. Note that it is also possible to construct BEMs that are not per se dependent on the channel statistics like the so-called complex-exponential BEM (CE-BEM) [7]. Its basis functions are complex exponentials that have a period equal to the length of the considered interval. The CE-BEM gained a great deal of attention thanks to its algebraic ease [8]–[13], but induces a larger modeling error. As we understand, the CE-BEM can actually be viewed as a special DKL-BEM but based on a white spectrum. An improved modeling performance is obtained by the so-called generalized CE-BEM (GCE-BEM) [14], which employs a set of complex exponentials that are more closely spaced in the frequency domain than those related to the traditional CE-BEM. Finally, a great deal of attention goes to the polynomial BEM (P-BEM) [15], [16], which models each tap as a linear combination of a set of polynomials. Its modeling performance is rather sensitive to the Doppler spread though it has a better fit for low Doppler spreads than for high Doppler spreads. Note that it is also possible to combine the aforementioned BEMs for different purposes [17]–[19].

Apart from the BEM, another modeling approach is to use a Gauss–Markov process to simulate the channel dynamics [20]. Such a model is interesting for sequential time-domain processing. When we deal with block transmission/precoding schemes, such as orthogonal frequency-division multiplexing (OFDM), it is often more convenient to use a block-based channel model such as a BEM.

Focusing on the estimation of channels that are modeled by a BEM, we basically only need to estimate the BEM coefficients, which are usually smaller in number than the total number of channel unknowns. References [8], [12], and [21] belong to the few works that focus on blind BEM channel estimation. References [6], [10], [15], and [22] propose pilot-assisted channel estimators based on different BEM assumptions, where commonly, pilots are clustered in the time domain such that the channel estimation can be realized without interference from neighboring data symbols. For frequency-domain communication systems such as OFDM, it is not clear what is

Manuscript received November 1, 2005; revised July 27, 2006. The associate editor coordinating the review of this manucsript and approving it for publication was Dr. Mounir Ghogho. This work was supported in part by the NWO-STW foundation under the VICI Program DTC.5893 and the VIDI Program DTC.6577.

Z. Tang and G. Leus are with the Faculty of Electrical Engineering, Mathematics and Computer Science (EEMCS), Delft University of Technology, 2628 CD Delft, The Netherlands (e-mail: tang@cas.et.tudelft.nl; leus@cas.et.tudelft.nl).

R. C. Cannizzaro and P. Banelli are with the Dipartimento di Ingegneria Elettronica e dell'Informazione (DIEI), the University of Perugia, Perugia 06125, Italy (e-mail: banelli@diei.unipg.it).

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.2007.893198

the "optimum" strategy to place the pilots. This is due to the Doppler spread, which corrupts the orthogonality among the subcarriers and induces intercarrier interference (ICI). The receiver can find no subcarrier that solely depends on pilots and thus is not contaminated by data symbols. For this reason, many existing works view the frequency-domain channel matrix either as diagonal [6], [23], [24] thus ignoring the ICI completely, or strictly banded as in [13] that relies on a CE-BEM assumption. Apparently, these approaches suffer from a large estimation error for channels with a high Doppler spread, but admit a clustered pilot scheme in the frequency domain, which is adopted in many OFDM standards, e.g., terrestrial digital video broadcasting (DVB-T) [25]. A different approach is to view the frequency-domain channel matrix as a full matrix [17], [26], [27], which reflects the true situation, but generally requires the pilots to occupy a whole OFDM symbol.

In this paper, we study the clustered pilot scheme in the frequency domain and at the same time view the frequency-domain channel matrix as approximately banded, which implies that we consider yet a full channel matrix but with most of its power concentrated around the main diagonal and degrading rapidly from the diagonal to the margin. This view complies with the observations made in [18] and [28] and can be represented by most BEMs except for the CE-BEM, which results in a strictly banded frequency-domain channel matrix. However, the bandwidth of such an approximately banded channel matrix is hard to define and if we artificially select a clear-cut bandwidth, the out-of-band entries will give rise to interference. This paper will show that by taking this interference wisely into account in traditional estimator designs, such as the linear minimum mean square error (LMMSE) estimator or the best linear unbiased estimator (BLUE), we can improve the estimation accuracy. This is in contrast to the least squares (LS) estimator, which requires the interference to be as small as possible. In other words, the amount of interference we take into account has a significant impact on each estimator. This effect will be analyzed in this paper and a criterion to select the optimal amount of interference for different types of channel estimators will be proposed.

Recently, an increased interest has emerged in low-complexity equalization of TV channels for OFDM systems, many of which rely on the strictly banded assumption of the channel matrix in the frequency domain [29]–[31]. Undoubtedly, for the data model introduced in this paper, the out-of-band interference can compromise the performance of such equalizers. [29], [30] proposed receiver windowing to suppress this interference, but they have not considered how this will affect the channel estimation performance. In a nutshell, now that the channel is reshaped by the window, both the channel estimator and the BEM design must be adapted accordingly. To motivate the latter briefly, it is not hard to imagine that if the channel changes due to the windowing, the traditional BEM design needs to be adapted also to maintain a tight fit. Notably, the CE-BEM forms an exception and should remain unaltered irrespective of the existence of windowing.

The rest of this paper is organized as follows. Section II presents how a windowed TV channel can be approximated by a proper BEM and accordingly establishes the OFDM system model. Section III formulates the data model that we will apply to channel estimation. In Section IV, we propose three channel estimators and their optimization. Simulation results are exhibited in Section V, and we conclude the paper with Section VI.

Notation: We use upper (lower) bold face letters to denote matrices (column vectors). $(\cdot)^*$, $(\cdot)^T$, and $(\cdot)^H$ represent conjugate, transpose, and complex conjugate transpose (Hermitian), respectively. $\mathcal{E}_x\{\cdot\}$ stands for the expected value with respect to the random variable x. $\otimes$ and $\odot$ represent the Kronecker product and the Schur–Hadamard (elementwise) product, respectively. $\dagger$ represents the pseudoinverse. We denote the $N \times N$ identity matrix as $\mathbf{I}_N$, the $M \times N$ all-zero matrix as $\mathbf{0}_{M\times N}$, and the $M \times N$ all-one matrix as $\mathbf{1}_{M\times N}$. $\mathbf{e}_k$ stands for a unit vector with a 1 at the $(k+1)$th position. Further, we use $[\mathbf{x}]_p$ to indicate the $(p+1)$th element of the vector $\mathbf{x}$, $[\mathbf{X}]_{p,q}$ to indicate the $(p+1, q+1)$th entry of the matrix $\mathbf{X}$, and $\mathrm{diag}\{\mathbf{x}\}$ to indicate a diagonal matrix with $\mathbf{x}$ as its diagonal.

II. System Model

A. OFDM System Model

Let us consider an OFDM system with N subcarriers, as illustrated in Fig. 1. The kth OFDM symbol $\mathbf{s}(k)$ is used to modulate N carriers as $\mathbf{s}^{(t)}(k) = \mathbf{F}^H\mathbf{s}(k)$, where $\mathbf{F}$ stands for the N point unitary discrete Fourier transform (DFT) matrix with $[\mathbf{F}]_{p,q} = 1/\sqrt{N}\exp(-j2\pi pq/N)$. Making abstraction of the digital-to-analog and analog-to-digital conversions, $\mathbf{s}^{(t)}(k)$ is next concatenated by a cyclic prefix, sent over the channel, stripped from the cyclic prefix, and reshaped by a windowing filter. The resulting data stream can then be summarized as

$$\begin{aligned}\mathbf{y}^{(t)}(k) &= \mathrm{diag}\{\mathbf{w}\}\tilde{\mathbf{H}}^{(t)}(k)\mathbf{F}^H\mathbf{s}(k) + \mathrm{diag}(\mathbf{w})\tilde{\mathbf{n}}^{(t)}(k)\\ &= \mathbf{H}^{(t)}(k)\mathbf{F}^H\mathbf{s}(k) + \mathbf{n}^{(t)}(k)\end{aligned} \tag{1}$$

where $\mathbf{w} = [w_0, \ldots, w_{N-1}]^T$ represents the time-domain window and $\tilde{\mathbf{H}}^{(t)}(k)$ ($\tilde{\mathbf{n}}^{(t)}(k)$) and $\mathbf{H}^{(t)}(k)$ ($\mathbf{n}^{(t)}(k)$) denote the channel matrix (noise) in the time domain without and with windowing, respectively. Let us now define $\tilde{h}_{n,l}^{(t)}$ and $h_{n,l}^{(t)}$ as the lth channel tap at the nth time instant without and with windowing, respectively, and let us assume that $\tilde{h}_{n,l}^{(t)}$ and $h_{n,l}^{(t)}$ have finite order L, i.e., $\tilde{h}_{n,l}^{(t)} = 0$ for $l < 0$ or $l > L$ and $h_{n,l}^{(t)} = 0$ for $l < 0$ or $l > L$. If we then assume that the length of the cyclic prefix L_{cp} satisfies $L_{cp} \geq L$, both $\tilde{\mathbf{H}}^{(t)}(k)$ and $\mathbf{H}^{(t)}(k)$ are "pseudocirculant" with

$$[\tilde{\mathbf{H}}^{(t)}(k)]_{p,q} = \tilde{h}^{(t)}_{k(N+L_{cp})+L_{cp}+p,\,\mathrm{mod}(p-q,N)}$$

and

$$[\mathbf{H}^{(t)}(k)]_{p,q} = h^{(t)}_{k(N+L_{cp})+L_{cp}+p,\,\mathrm{mod}(p-q,N)}.$$

Note that by means of an appropriate window $\mathbf{w}$, a banded approximation of $\mathbf{H}^{(t)}(k) = \mathrm{diag}\{\mathbf{w}\}\tilde{\mathbf{H}}^{(t)}(k)$ is more accurate than the corresponding banded approximation of $\tilde{\mathbf{H}}^{(t)}(k)$. Therefore, low-complexity equalization schemes that exploit the banded assumption of the channel matrix in the frequency domain [29]–[31] will have an improved performance when appropriate windowing is employed. In the sequel, we will refer

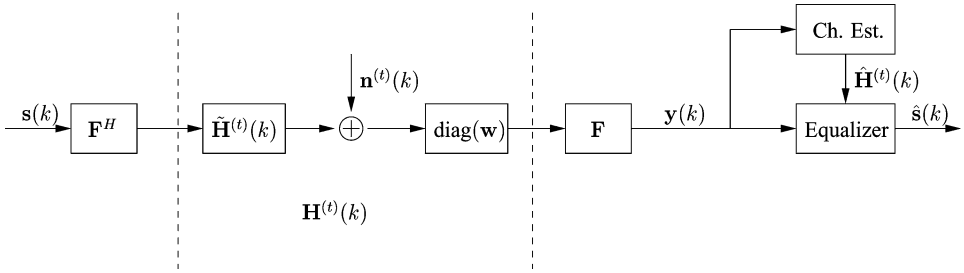

Fig. 1. Transceiver block diagram.

to the windowed channel as "the channel," unless otherwise stated.

The received samples $\mathbf{y}^{(t)}(k)$ are afterwards demodulated by the DFT matrix $\mathbf{F}$, resulting in

$$\begin{aligned}\mathbf{y}(k) &= \mathbf{F}\mathbf{y}^{(t)}(k) = \mathbf{F}\mathbf{H}^{(t)}(k)\mathbf{F}^H\mathbf{s}(k) + \mathbf{F}\mathbf{n}^{(t)}(k) \\ &= \mathbf{H}(k)\mathbf{s}(k) + \mathbf{n}(k) \end{aligned} \tag{2}$$

with $\mathbf{H}(k) := \mathbf{F}\mathbf{H}^{(t)}(k)\mathbf{F}^H$ ($\mathbf{n}(k) := \mathbf{F}\mathbf{n}^{(t)}(k)$) denoting the channel matrix (noise) in the frequency domain.

B. Windowed BEM Channel Model

In this section, we will try to accurately model the time-domain channel $h_{n,l}^{(t)}$ by using a BEM. Collecting the TV behavior of the lth channel tap within the kth OFDM symbol in the $N \times 1$ vector $\mathbf{h}_l^{(t)}(k) := \left[h_{k(N+L_{cp})+L_{cp},l}^{(t)}, \ldots, h_{(k+1)(N+L_{cp})-1,l}^{(t)}\right]^T$, we can express $\mathbf{h}_l^{(t)}(k)$ as

$$\mathbf{h}_l^{(t)}(k) = \mathbf{B}\mathbf{h}_l(k) + \boldsymbol{\epsilon}_l(k) \tag{3}$$

where $\mathbf{B} := [\mathbf{b}_0, \ldots, \mathbf{b}_Q]$ is an $N \times (Q+1)$ matrix that collects $Q+1$ orthonormal basis functions $\mathbf{b}_q$ as columns, $\mathbf{h}_l(k) := [h_{0,l}(k), \ldots, h_{Q,l}(k)]^T$ represents the BEM coefficients for the lth tap of the kth OFDM symbol, and $\boldsymbol{\epsilon}_l(k) := [\epsilon_{0,l}(k), \ldots, \epsilon_{N-1,l}(k)]^T$ represents the corresponding BEM modeling error, which is assumed to be minimized in the MSE sense. Stacking all the channel taps within the kth observation block in one vector

$$\begin{aligned}\mathbf{h}^{(t)}(k) := \Big[&h_{k(N+L_{cp})+L_{cp},0}^{(t)}, \ldots, h_{k(N+L_{cp})+L_{cp},L}^{(t)}, \\ &\ldots, h_{(k+1)(N+L_{cp})-1,0}^{(t)}, \ldots, h_{(k+1)(N+L_{cp})-1,L}^{(t)}\Big]^T\end{aligned}$$

we obtain

$$\mathbf{h}^{(t)}(k) = (\mathbf{B} \otimes \mathbf{I}_{L+1})\mathbf{h}(k) + \boldsymbol{\epsilon}(k) \tag{4}$$

where $\mathbf{h}(k) := [h_{0,0}(k), \ldots, h_{0,L}(k), \ldots, h_{Q,0}(k), \ldots, h_{Q,L}(k)]^T$ and $\boldsymbol{\epsilon}(k) := [\epsilon_{0,0}(k), \ldots, \epsilon_{0,L}(k), \ldots, \epsilon_{N-1,0}(k), \ldots, \epsilon_{N-1,L}(k)]^T$.

Denoted as $\tilde{\mathbf{B}}$, various traditional BEM designs have been reported to model the unwindowed TV channels, e.g., the CE-BEM $[\tilde{\mathbf{B}}]_{p,q} := e^{(j2\pi p/N)(q-Q/2)}$ [13]; the GCE-BEM $[\tilde{\mathbf{B}}]_{p,q} := e^{(j2\pi p/KN)(q-Q/2)}$ with $K > 1$ [14]; the P-BEM $[\tilde{\mathbf{B}}]_{p,q} := (p+1)^q$[15]; and the DKL-BEM, which employs basis sequences that corresponds to the most significant eigenvectors of the autocorrelation matrix $\tilde{\mathbf{R}}_{h,l}^{(t)}$ [5] with

$$\left[\tilde{\mathbf{R}}_{h,l}^{(t)}\right]_{p,q} := \mathcal{E}\left\{\tilde{h}_{p,l}^{(t)}\tilde{h}_{q,l}^{(t)*}\right\} \tag{5}$$

where we assume that $\tilde{\mathbf{R}}_{h,l}^{(t)}$ is identical for each channel tap l up to a scaling. The DPS-BEM is constructed in a similar fashion, but based on a special $\tilde{\mathbf{R}}_{h,l}^{(t)}$ that is associated with a rectangular Doppler spectrum [6].

The question arises as how to choose an appropriate BEM $\mathbf{B}$ in case of windowing. Basically, we have the following two options:

$$\mathbf{B} = \begin{cases} \tilde{\mathbf{B}}\mathbf{Q} & \text{i)} \\ \mathrm{diag}\{\mathbf{w}\}\tilde{\mathbf{B}}\mathbf{Q} & \text{ii).} \end{cases} \tag{6}$$

In (6), we add a square matrix $\mathbf{Q}$ to ensure that the columns of $\mathbf{B}$ are orthonormal. Its usage is in general not mandatory but can simplify the analysis as will be clear later on. We observe in (6) that option i) ignores the window in the BEM design and sticks to the traditional "unwindowed" BEMs to model the windowed channels whereas option ii) includes the knowledge of the window in the BEM design. We will show in Section V that option ii) generally yields a tighter fit. This is because the window itself brings some additional time-variation on top of the unwindowed channel, which probably requires more basis functions for the traditional "unwindowed" BEM to maintain a tight fit. However, the time-variation due to the window is totally predictable and, hence, can be counteracted by simply absorbing it in our BEM design. The CE-BEM is an exception especially for channels with high Doppler spreads, in which case it is better to use option i). This is because the window is usually designed to make the frequency-domain channel matrix as banded as possible, and the CE-BEM yields itself a perfectly banded frequency-domain channel matrix. As a result, it is not necessary to include the window in the CE-BEM design.

C. OFDM System Model Based on BEM

From now on, we can describe the OFDM system model derived previously in light of the BEM. Since all the algorithms in this paper will be based on a single OFDM symbol, we drop the block index k in the sequel. Substituting (4) in (2), we obtain after some algebra

$$\mathbf{y} = \sum_{q=0}^{Q} \mathbf{D}_q \boldsymbol{\Delta}_q \mathbf{s} + \mathbf{n} + \boldsymbol{\delta} \tag{7}$$

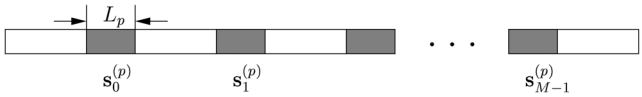

Fig. 2. Pilot placement pattern.

where $\boldsymbol{\delta}$ denotes the error vector induced by the BEM fitting error $\boldsymbol{\epsilon}$, $\mathbf{D}_q$ is a circulant matrix whose first column is the frequency response of the qth basis function

$$\mathbf{D}_q := \mathbf{F}\mathrm{diag}\{\mathbf{b}_q\}\mathbf{F}^H \tag{8}$$

and $\boldsymbol{\Delta}_q$ is a diagonal matrix whose diagonal is the frequency response of the BEM coefficients corresponding to the qth basis function

$$\boldsymbol{\Delta}_q := \mathrm{diag}\left\{\mathbf{F}_L[h_{q,0},\ldots,h_{q,L}]^T\right\}. \tag{9}$$

Here, $\mathbf{F}_L$ stands for the first $L+1$ columns of the matrix $\sqrt{N}\mathbf{F}$.

Note that (7) subsumes the expression for time-invariant (TI) channels, in which case $Q=0$ and $\mathbf{D}_q$ becomes a scaled identity matrix. For TV channels, the nondiagonal entries of $\mathbf{D}_q$ are in general not zero any more. This leads to a loss of orthogonality among the subcarriers known as ICI.

Before going on further, we will make the following assumption:

a1) the BEM approximation holds perfectly, i.e., $\boldsymbol{\epsilon}=\mathbf{0}$ or $\boldsymbol{\delta}=\mathbf{0}$.

This assumption is motivated by the fact that we will mainly focus on BEMs that allow for a very good fit.[1] For other BEMs that fail to capture the time-variation adequately, such as the CE-BEM for instance, we should actually take the modeling error into account. This topic is partly treated in [32]. However, even if we include this error term to derive the best estimator possible, we still do not have a reliable channel estimate simply because the BEM itself is not capable of fitting the true channel. This suggests that it makes not much sense to take the modeling error into account and explains why we apply a1) for all possible BEMs.

III. Data Model for Channel Estimation

Instead of estimating the true bulky channel matrix $\mathbf{H}$, we will estimate the $(L+1)(Q+1)$ BEM coefficients in $\mathbf{h}$ with the aid of pilots. We assume there are M pilot clusters of length L_p denoted as $\mathbf{s}_m^{(p)}$, $m=0,1,\ldots,M-1$, as indicated in Fig. 2. All these pilot clusters stacked together form the pilot vector $\mathbf{s}^{(p)} := \left[\mathbf{s}_0^{(p)T},\ldots,\mathbf{s}_{M-1}^{(p)T}\right]^T$. The pilot clusters are interleaved with information symbols, which can be collected in the information symbol vector $\mathbf{s}^{(d)}$.

It is not clear what is the optimal pilot placement for TV OFDM systems. References [10] and [13] claim that equidistant pulse-shaped pilot clusters are optimal based on a CE-BEM channel assumption, while equidistant pilot clusters also find their practical advantage in [18] though the channel follows the bathtub-shaped Doppler spectrum in that case. Despite its significance, we will not discuss this issue here but allow our receiver design to be applicable for any frequency-multiplexed pilot placement scheme.

For a certain frequency-multiplexed pilot placement scheme, it is up to the receiver to decide which of the received samples must be used for channel estimation. This is crucial for a TV OFDM system since, due to the ICI (or in other words, the nondiagonal entries of $\mathbf{D}_q$), the pilots' power is spread out over the whole frequency band. A judicious choice of the observation samples will enhance the channel estimation performance.

Generally speaking, $\mathbf{D}_q$ is approximately banded suggesting that the ICI primarily comes from adjacent subcarriers [18], [28]. An extreme case occurs with the CE-BEM channel model, where the corresponding $\mathbf{D}_q$ is an identity matrix but circularly shifted over $q-(Q/2)$ columns, which implies that only the Q neighboring subcarriers give rise to interference.

To clarify the notations that will come forth, we plot the structure of $\mathbf{D}_q$ in Fig. 3, where the columns of $\mathbf{D}_q$ are related to the positions of the pilots and data, which operate on $\mathbf{D}_q$ through the diagonal matrix $\boldsymbol{\Delta}_q$. The rows of $\mathbf{D}_q$ are related to the observation samples. For the mth pilot cluster $\mathbf{s}_m^{(p)} = [[\mathbf{s}]_{P_m},\ldots,[\mathbf{s}]_{P_m+L_p-1}]^T$, where P_m stands for its begin position, let us consider the following vector consisting of L_p-2B_c observation samples:

$$\mathbf{y}_m := [[\mathbf{y}]_{P_m+B_c},\ldots,[\mathbf{y}]_{P_m+L_p-B_c-1}]^T. \tag{10}$$

It can be imagined that if $\mathbf{D}_q$ were strictly banded with only $2B_c+1$ nonzero diagonals, $\mathbf{y}_m$ would be the vector of maximal length that exclusively depends on the pilot cluster $\mathbf{s}_m^{(p)}$. In this sense, B_c could be interpreted as the assumed bandwidth of $\mathbf{D}_q$, as suggested in Fig. 3. However, we must be cautious with this interpretation, because $\mathbf{D}_q$ is not strictly banded for most BEMs. As a matter of fact, B_c actually provides a handle on the amount of interference that we want to take into account. More importantly, B_c is not confined to positive values as we will see later on, in which case the bandwidth physical interpretation cannot be directly accounted for.

To formulate the previous discussion in mathematical expressions with notations indicated in Fig. 3, we obtain

$$\mathbf{y}_m = \sum_{q=0}^{Q}\mathbf{D}_{q,m}^{(p)}\boldsymbol{\Delta}_q^{(p)}\mathbf{s}^{(p)} + \underbrace{\sum_{q=0}^{Q}\mathbf{D}_{q,m}^{(d)}\boldsymbol{\Delta}_q^{(d)}\mathbf{s}^{(d)}}_{\mathbf{d}_m} + \mathbf{n}_m \tag{11}$$

where $\mathbf{D}_{q,m}^{(p)}$ is an $(L_p-2B_c)\times ML_p$ matrix, representing the hatched parts of $\mathbf{D}_q$ in Fig. 3; $\boldsymbol{\Delta}_q^{(p)}$ is an $ML_p\times ML_p$ diagonal matrix, which is carved out of $\boldsymbol{\Delta}_q$ corresponding to the pilot-carrying subcarriers; $\mathbf{D}_{q,m}^{(d)}$ is an $(L_p-2B_c)\times(N-ML_p)$ matrix, representing the shaded parts of $\mathbf{D}_q$ in Fig. 3; $\boldsymbol{\Delta}_q^{(d)}$ is an $(N-ML_p)\times(N-ML_p)$ diagonal matrix, which is carved out of $\boldsymbol{\Delta}_q$ corresponding to the data-carrying subcarriers; finally, $\mathbf{n}_m$ stands for the noise related to $\mathbf{y}_m$. In (11), we have thus uncoupled the effect of the data from the pilots, and put it in a separate term $\mathbf{d}_m$. This term, which poses a nuisance to channel estimation, is in general not zero since $\mathbf{D}_q$ is not strictly banded. Let us rewrite (11) as a function of $\mathbf{h}$

$$\mathbf{y}_m = \mathbf{D}_m^{(p)}\boldsymbol{\mathcal{S}}^{(p)}\mathbf{h} + \mathbf{d}_m + \mathbf{n}_m \tag{12}$$

[1] A modeling performance comparison (except for the DKL-BEM) can be found in [6].

with

$$\mathbf{D}_m^{(p)} := \left[\mathbf{D}_{0,m}^{(p)}, \ldots, \mathbf{D}_{Q,m}^{(p)}\right]$$

and

$$\boldsymbol{\mathcal{S}}^{(p)} := \mathbf{I}_{Q+1} \otimes \left(\mathrm{diag}\{\mathbf{s}^{(p)}\}\mathbf{F}_L^{(p)}\right).$$

Here, $\mathbf{F}_L^{(p)}$ collects the rows of $\mathbf{F}_L$ corresponding to the positions of the pilots. Further, we want to underline that the interference term $\mathbf{d}_m$ carries also channel information $\mathbf{h}$ as can be seen from

$$\mathbf{d}_m = \mathbf{D}_m^{(d)}\boldsymbol{\mathcal{S}}^{(d)}\mathbf{h} \tag{13}$$

with $\mathbf{D}_m^{(d)} := \left[\mathbf{D}_{0,m}^{(d)}, \ldots, \mathbf{D}_{Q,m}^{(d)}\right]$ and $\boldsymbol{\mathcal{S}}^{(d)} := \mathbf{I}_{Q+1} \otimes \left(\mathrm{diag}\{\mathbf{s}^{(d)}\}\mathbf{F}_L^{(d)}\right)$. Here, $\mathbf{F}_L^{(d)}$ collects the rows of $\mathbf{F}_L$ corresponding to the positions of the information symbols.

Repeating the previous operations for all the observation vectors and introducing the notations $\mathbf{y}^{(p)} := \left[\mathbf{y}_0^T, \ldots, \mathbf{y}_{M-1}^T\right]^T$, $\mathbf{d} := \left[\mathbf{d}_0^T, \ldots, \mathbf{d}_{M-1}^T\right]^T$, and $\mathbf{n}^{(p)} := \left[\mathbf{n}_0^T, \ldots, \mathbf{n}_{M-1}^T\right]^T$, we obtain

$$\begin{aligned}\mathbf{y}^{(p)} &= \mathbf{D}^{(p)}\boldsymbol{\mathcal{S}}^{(p)}\mathbf{h} + \mathbf{d} + \mathbf{n}^{(p)} \\ &= \boldsymbol{\mathcal{P}}\mathbf{h} + \mathbf{d} + \mathbf{n}^{(p)}\end{aligned} \tag{14}$$

with

$$\begin{aligned}\boldsymbol{\mathcal{P}} &:= \mathbf{D}^{(p)}\boldsymbol{\mathcal{S}}^{(p)} \\ \mathbf{D}^{(p)} &:= \left[\mathbf{D}_0^{(p)T}, \ldots, \mathbf{D}_{M-1}^{(p)T}\right]^T \\ &= \begin{bmatrix} \mathbf{D}_{0,0}^{(p)} & \cdots & \mathbf{D}_{Q,0}^{(p)} \\ \vdots & \ddots & \vdots \\ \mathbf{D}_{0,M-1}^{(p)} & \cdots & \mathbf{D}_{Q,M-1}^{(p)} \end{bmatrix}.\end{aligned}$$

Likewise, we can express the interference term $\mathbf{d}$ as

$$\mathbf{d} = \mathbf{D}^{(d)}\boldsymbol{\mathcal{S}}^{(d)}\mathbf{h} \tag{15}$$

with

$$\begin{aligned}\mathbf{D}^{(d)} &:= \left[\mathbf{D}_0^{(d)T}, \ldots, \mathbf{D}_{M-1}^{(d)T}\right]^T \\ &= \begin{bmatrix} \mathbf{D}_{0,0}^{(d)} & \cdots & \mathbf{D}_{Q,0}^{(d)} \\ \vdots & \ddots & \vdots \\ \mathbf{D}_{0,M-1}^{(d)} & \cdots & \mathbf{D}_{Q,M-1}^{(d)} \end{bmatrix}.\end{aligned}$$

The afore described interference analysis is not restricted to any specific BEM. However, note that for the CE-BEM $\mathbf{D}_q$ is strictly banded, and thus $\mathbf{D}_q^{(d)} = \mathbf{0}$. Apparently, this leaves no vagueness about B_c, which should then be set to $B_c = Q/2$. This case is considered in [13].

IV. Channel Estimation and B_c Optimization

In this section, we will discuss channel estimation based on the data model that has been established in the previous section. We make the following assumptions:

a1) time-domain noise prior to windowing $\tilde{\mathbf{n}}^{(t)}$ is assumed to be zero-mean white Gaussian with variance σ_n^2;

a2) data $\mathbf{s}^{(d)}$ are assumed to be zero-mean white with variance σ_s^2 and uncorrelated with the noise $\mathbf{n}$, i.e., $\mathcal{E}\{\mathbf{s}^{(d)}\mathbf{n}^H\} = \mathbf{0}$.

We will propose three channel estimators in this section: the LMMSE estimator relies on the statistics of $\mathbf{h}$, while the LS estimator and the BLUE treat $\mathbf{h}$ as a deterministic variable. The performance of each channel estimator is sensitive to B_c. This can be understood from (14) and (15), where the pilot-related $\mathbf{D}^{(p)}$ is an $M(L_p - 2B_c) \times (Q+1)ML_p$ matrix, and the interference-related $\mathbf{D}^{(d)}$ is an $M(L_p - 2B_c) \times (Q+1)(N - ML_p)$ matrix. Intuitively, one would reduce the interference term by setting B_c as large as possible. The same idea is adopted in [18] though the authors address the problem from a different point of view. To explain this using the physical interpretation of B_c, a larger B_c corresponds to a more accurate band approximation of $\mathbf{D}_q$, and thus to a smaller interference. On the other hand, a larger B_c, giving rise to a "fatter" $\mathbf{D}^{(p)}$, is often detrimental if we are to deploy a linear channel estimator. We will examine the effect of B_c individually for each estimator.

A. LMMSE Estimator

The LMMSE estimator treats $\mathbf{h}$ as a stochastic variable. To be more specific, we introduce the following assumption:

a4) channel vector $\mathbf{h}$ is assumed to be uncorrelated with the noise $\mathbf{n}$ and the information symbols $\mathbf{s}^{(d)}$, i.e., $\mathcal{E}\{\mathbf{h}\mathbf{n}^H\} = \mathbf{0}$ and $\mathcal{E}\{\mathbf{h}\mathbf{s}^{(d)H}\} = \mathbf{0}$.

We seek a linear filter $\mathbf{W}$ such that the MSE between the estimated BEM coefficients $\hat{\mathbf{h}} = \mathbf{W}\mathbf{y}^{(p)}$ and the true BEM coefficients $\mathbf{h}$ is minimal. In other words, we solve

$$\begin{aligned}\mathbf{W}_{\mathrm{LMMSE}} = \arg\min_{\{\mathbf{W}\}} \mathrm{trace}\{&\mathcal{E}_{\mathbf{h},\mathbf{s}^{(d)},\mathbf{n}^{(p)}}\{(\mathbf{W}\mathbf{y}^{(p)} - \mathbf{h}) \\ &(\mathbf{W}\mathbf{y}^{(p)} - \mathbf{h})^H\}\}.\end{aligned} \tag{16}$$

It can be shown that

$$\begin{aligned}&\mathcal{E}_{\mathbf{h},\mathbf{s}^{(d)},\mathbf{n}^{(p)}}\{(\mathbf{W}\mathbf{y}^{(p)} - \mathbf{h})(\mathbf{W}\mathbf{y}^{(p)} - \mathbf{h})^H\} \\ &\quad= \mathbf{W}\Bigg(\boldsymbol{\mathcal{P}}\mathbf{R}_h\boldsymbol{\mathcal{P}}^H + \mathbf{R}_d + \mathbf{R}_n^{(p)} \\ &\qquad\qquad + 2\Re\left(\mathbf{D}^{(d)}\mathcal{E}_{\mathbf{s}^{(d)}}\{\boldsymbol{\mathcal{S}}^{(d)}\}\mathbf{R}_h\boldsymbol{\mathcal{P}}^H\right)\Bigg)\mathbf{W}^H \\ &\qquad - 2\Re\left(\mathbf{R}_h\boldsymbol{\mathcal{P}}^H\mathbf{W}^H + \mathbf{R}_h\mathcal{E}_{\mathbf{s}^{(d)}}\{\boldsymbol{\mathcal{S}}^{(d)H}\}\mathbf{D}^{(d)H}\mathbf{W}^H\right) \\ &\qquad + \mathbf{R}_h \\ &\quad= \mathbf{W}(\boldsymbol{\mathcal{P}}\mathbf{R}_h\boldsymbol{\mathcal{P}}^H + \mathbf{R}_d + \mathbf{R}_n^{(p)})\mathbf{W}^H \\ &\qquad - 2\Re\left(\mathbf{R}_h\boldsymbol{\mathcal{P}}^H\mathbf{W}^H\right) + \mathbf{R}_h.\end{aligned} \tag{17}$$

In deriving (17), we make use of assumptions a3) and a4), and introduce the following covariance matrix notations $\mathbf{R}_h := \mathcal{E}_{\mathbf{h}}\{\mathbf{h}\mathbf{h}^H\}$, $\mathbf{R}_d := \mathcal{E}_{\mathbf{h},\mathbf{s}^{(d)}}\{\mathbf{d}\mathbf{d}^H\}$, and

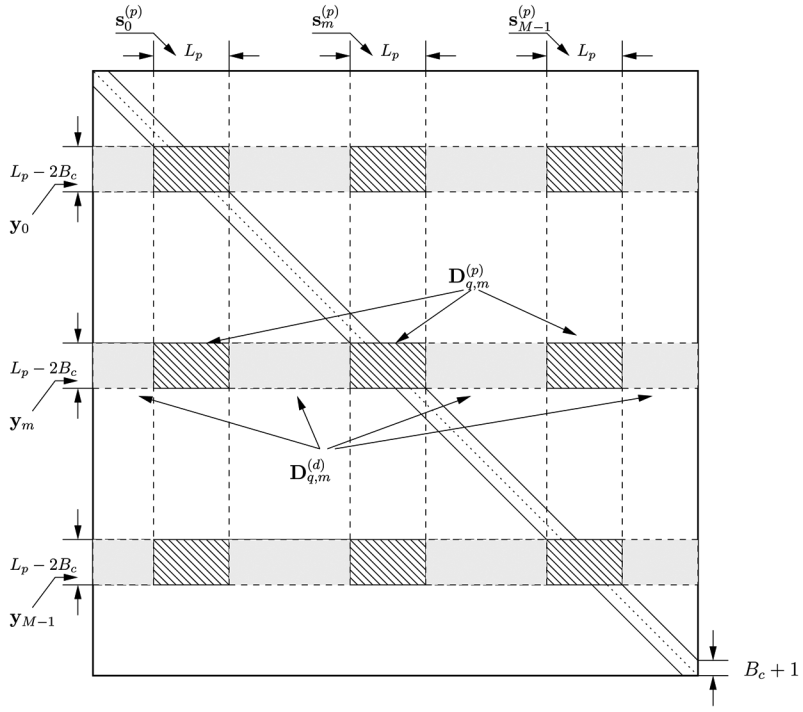

Fig. 3. Relation between the pilots and the received samples.

$\mathbf{R}_n^{(p)} := \mathcal{E}_{\mathbf{n}^{(p)}}\{\mathbf{n}^{(p)}\mathbf{n}^{(p)H}\}$, whose computation is given in Appendix A. Equation (17) leads to

$$\begin{aligned}\hat{\mathbf{h}}_{\mathrm{LMMSE}} &= \mathbf{W}_{\mathrm{LMMSE}}\mathbf{y}^{(p)} \\ \mathbf{W}_{\mathrm{LMMSE}} &= \mathbf{R}_h\boldsymbol{\mathcal{P}}^H\left(\boldsymbol{\mathcal{P}}\mathbf{R}_h\boldsymbol{\mathcal{P}}^H + \mathbf{R}_d + \mathbf{R}_n^{(p)}\right)^{-1} \\ &= \mathbf{R}_h\boldsymbol{\mathcal{P}}^H\left(\boldsymbol{\mathcal{P}}\mathbf{R}_h\boldsymbol{\mathcal{P}}^H + \mathbf{R}_{\mathcal{I}}\right)^{-1} \end{aligned} \tag{18}$$

where $\mathbf{R}_{\mathcal{I}} := \mathbf{R}_d + \mathbf{R}_n^{(p)}$. Note that although (18) bears a similar form to the classical LMMSE estimator [33], it has the extra task to process the interference term $\mathbf{d}$, which contains the information of $\mathbf{h}$ itself. For this purpose, the proposed LMMSE treats $\mathbf{d}$ as stochastic and resorts to the assumed statistics of $\mathbf{h}$ and $\mathbf{s}^{(d)}$. However, these statistics are difficult to retrieve in practice and not always reliable if, for instance, the Doppler spread is only roughly known or the assumed Doppler spectrum deviates from the true value. In such cases, the proposed LMMSE estimator is suboptimal. We will show some examples in Section V.

We want to use $\hat{\mathbf{h}}$ to reconstruct the BEM channel, and examine how close it is to the best BEM fit of the real channel. To this end, we adopt the MSE criterion as

$$\begin{aligned}\mathrm{MSE} &:= \mathcal{E}_{\mathbf{h},\mathbf{s}^{(d)},\mathbf{n}^{(p)}}\left\{\|(\mathbf{B}\otimes\mathbf{I}_{L+1})\hat{\mathbf{h}} - (\mathbf{B}\otimes\mathbf{I}_{L+1})\mathbf{h}\|^2\right\} \\ &= \mathrm{trace}\Big\{(\mathbf{B}\otimes\mathbf{I}_{L+1})\mathcal{E}_{\mathbf{h},\mathbf{s}^{(d)},\mathbf{n}^{(p)}}\left\{(\hat{\mathbf{h}}-\mathbf{h})(\hat{\mathbf{h}}-\mathbf{h})^H\right\} \\ &\qquad (\mathbf{B}\otimes\mathbf{I}_{L+1})^H\Big\} \\ &= \mathrm{trace}\left\{\mathcal{E}_{\mathbf{h},\mathbf{s}^{(d)},\mathbf{n}^{(p)}}\left\{(\hat{\mathbf{h}}-\mathbf{h})(\hat{\mathbf{h}}-\mathbf{h})^H\right\}\right\}. \end{aligned} \tag{19}$$

The last equality holds since we have designed $\mathbf{B}$ to have orthonormal columns [cf. (6)]. Equation (19) suggests that the channel fitting MSE equals the MSE of the estimated BEM coefficients. Therefore, the MSE resulting from the LMMSE channel estimator can be expressed as

$$\mathrm{MSE}_{\mathrm{LMMSE}} = \mathrm{trace}\left\{\left(\boldsymbol{\mathcal{P}}^H\mathbf{R}_{\mathcal{I}}^{-1}\boldsymbol{\mathcal{P}} + \mathbf{R}_h^{-1}\right)^{-1}\right\}. \tag{20}$$

Obviously, $\mathrm{MSE}_{\mathrm{LMMSE}}$ depends on B_c. Later on, we will show how to find the optimal B_c to minimize $\mathrm{MSE}_{\mathrm{LMMSE}}$.

B. LS Estimator

The LS estimator $\mathbf{W}_{\mathrm{LS}}$ treats $\mathbf{h}$ as a deterministic variable. It is straightforward to obtain $\mathbf{W}_{\mathrm{LS}} := \boldsymbol{\mathcal{P}}^{\dagger}$ such that

$$\hat{\mathbf{h}}_{\mathrm{LS}} = \mathbf{h} + \boldsymbol{\mathcal{P}}^{\dagger}(\mathbf{d} + \mathbf{n}^{(p)}). \tag{21}$$

The LS estimator is the most robust estimator, requiring no knowledge about the channel and noise statistics. However, it performs inferior when the interference is prominent: we show in the simulation part that the LS estimator suffers from a large performance gap in comparison with the Cramer–Rao bound (CRB) (derived in Appendix B). In addition, the performance of the LS estimator relies heavily on the condition number of $\boldsymbol{\mathcal{P}}$ as we can see from the resulting MSE

$$\begin{aligned}\mathrm{MSE}_{\mathrm{LS}} &:= \mathcal{E}_{\mathbf{h},\mathbf{s}^{(d)},\mathbf{n}^{(p)}}\left\{\mathrm{trace}\left\{\boldsymbol{\mathcal{P}}^{\dagger}(\mathbf{d}+\mathbf{n}^{(p)})(\mathbf{d}+\mathbf{n}^{(p)})^H\boldsymbol{\mathcal{P}}^{\dagger H}\right\}\right\} \\ &= \mathrm{trace}\left\{\boldsymbol{\mathcal{P}}^{\dagger}\mathcal{E}_{\mathbf{h},\mathbf{s}^{(d)},\mathbf{n}^{(p)}}\left\{(\mathbf{d}+\mathbf{n}^{(p)})(\mathbf{d}+\mathbf{n}^{(p)})^H\right\}\boldsymbol{\mathcal{P}}^{\dagger H}\right\} \\ &= \mathrm{trace}\left\{\boldsymbol{\mathcal{P}}^{\dagger}\mathbf{R}_{\mathcal{I}}\boldsymbol{\mathcal{P}}^{\dagger H}\right\} \end{aligned} \tag{22}$$

which is again a function of B_c.

C. Iterative BLUE

From (14), we can find an expression for the BLUE following similar steps as in [33, Appendix 6B] by treating the interference $\mathbf{d}$ and noise $\mathbf{n}^{(p)}$ as a single disturbance term such that

$$\begin{aligned}\hat{\mathbf{h}}_{\mathrm{BLUE}} &= \mathbf{W}_{\mathrm{BLUE}}\mathbf{y}^{(p)},\\ \mathbf{W}_{\mathrm{BLUE}} &= \left(\boldsymbol{\mathcal{P}}^H\tilde{\mathbf{R}}_{\mathcal{I}}^{-1}(\mathbf{h})\boldsymbol{\mathcal{P}}\right)^{-1}\boldsymbol{\mathcal{P}}^H\tilde{\mathbf{R}}_{\mathcal{I}}^{-1}(\mathbf{h})\end{aligned} \tag{23}$$

where $\tilde{\mathbf{R}}_{\mathcal{I}}(\mathbf{h})$ denotes the covariance matrix of the disturbance. Here, $\mathbf{h}$ is again viewed as a deterministic variable and, therefore, $\tilde{\mathbf{R}}_{\mathcal{I}}(\mathbf{h}) = \mathcal{E}_{\mathbf{s}^{(d)},\mathbf{n}^{(p)}}\{(\mathbf{d}+\mathbf{n}^{(p)})(\mathbf{d}+\mathbf{n}^{(p)})^H\}$. Due to a3), we have $\tilde{\mathbf{R}}_{\mathcal{I}}(\mathbf{h}) = \tilde{\mathbf{R}}_d(\mathbf{h}) + \mathbf{R}_n^{(p)}$ with $\tilde{\mathbf{R}}_d(\mathbf{h}) := \mathcal{E}_{\mathbf{s}^{(d)}}\{\mathbf{d}\mathbf{d}^H\}$, whose derivation can be found in Appendix A.

However, (23) is not implementable since its computation entails the knowledge of $\mathbf{h}$ itself. A recursive approach can be, therefore, applied: suppose at the kth iteration, an estimate for $\mathbf{h}$ is available denoted as $\hat{\mathbf{h}}_{\mathrm{BLUE}}^{(k)}$. Next, we use this estimate to update the covariance matrix $\tilde{\mathbf{R}}_{\mathcal{I}}(\mathbf{h})$, which in turn is used to produce the BLUE for the next iteration and so on

$$\begin{aligned}\mathbf{W}_{\mathrm{BLUE}}^{(k+1)} &= \left(\boldsymbol{\mathcal{P}}^H\tilde{\mathbf{R}}_{\mathcal{I}}^{-1}\left(\hat{\mathbf{h}}_{\mathrm{BLUE}}^{(k)}\right)\boldsymbol{\mathcal{P}}\right)^{-1}\boldsymbol{\mathcal{P}}^H\tilde{\mathbf{R}}_{\mathcal{I}}^{-1}\left(\hat{\mathbf{h}}_{\mathrm{BLUE}}^{(k)}\right)\\ \hat{\mathbf{h}}_{\mathrm{BLUE}}^{(k+1)} &= \mathbf{W}_{\mathrm{BLUE}}^{(k+1)}\mathbf{y}^{(p)}.\end{aligned}$$

Note that a similar idea is adopted in [34] though applied in a different context. To ensure that this iterative procedure will converge, we can simply initialize with $\hat{\mathbf{h}}_{\mathrm{BLUE}}^{(0)} = \mathbf{0}$, which results in the following expression for the first iteration:

$$\mathbf{W}_{\mathrm{BLUE}}^{(1)} = \left(\boldsymbol{\mathcal{P}}^H\left(\mathbf{R}_n^{(p)}\right)^{-1}\boldsymbol{\mathcal{P}}\right)^{-1}\boldsymbol{\mathcal{P}}^H\left(\mathbf{R}_n^{(p)}\right)^{-1}. \tag{24}$$

From a2), (24) is the maximum-likelihood estimator (MLE) [33] that is obtained by ignoring the interference $\mathbf{d}$. The resulting $\hat{\mathbf{h}}_{\mathrm{BLUE}}^{(1)} = \mathbf{W}_{\mathrm{BLUE}}^{(1)}\mathbf{y}^{(p)}$ is actually the LS fit as obtained in Section IV-Bbut weighted by the noise covariance. If $\mathbf{d}$ is small, which is often the case by carefully selecting B_c, $\hat{\mathbf{h}}_{\mathrm{BLUE}}^{(1)}$ is already close enough to $\mathbf{h}$ to avoid convergence to a local minimum.

Assuming that $\hat{\mathbf{h}}_{\mathrm{BLUE}}^{(k)} \rightarrow \hat{\mathbf{h}}_{\mathrm{BLUE}}$, we use (23) to find the MSE of the channel estimator

$$\begin{aligned}\mathrm{MSE}_{\mathrm{BLUE}} &= \mathcal{E}_{\mathbf{h},\mathbf{s}^{(d)},\mathbf{n}^{(p)}}\Big\{\mathrm{trace}\Big\{\mathbf{W}_{\mathrm{BLUE}}(\mathbf{d}+\mathbf{n}^{(p)})\\ &\qquad (\mathbf{d}+\mathbf{n}^{(p)})^H\mathbf{W}_{\mathrm{BLUE}}^H\Big\}\Big\}\\ &= \mathcal{E}_{\mathbf{h}}\left\{\mathrm{trace}\left(\mathbf{W}_{\mathrm{BLUE}}\tilde{\mathbf{R}}_{\mathcal{I}}(\mathbf{h})\mathbf{W}_{\mathrm{BLUE}}^H\right)\right\}\\ &= \mathcal{E}_{\mathbf{h}}\left\{\mathrm{trace}\left(\left(\boldsymbol{\mathcal{P}}^H\tilde{\mathbf{R}}_{\mathcal{I}}^{-1}(\mathbf{h})\boldsymbol{\mathcal{P}}\right)^{-1}\right)\right\}.\end{aligned} \tag{25}$$

Equation (25) provides a lower bound on the performance of the iterative BLUE. This MSE is, however, difficult to evaluate in closed form due to the inversion of $\tilde{\mathbf{R}}_{\mathcal{I}}(\mathbf{h})$, which forces us to resort to the Monte Carlo method. As will be evident later on, the MSE resulting from the BLUE is also dependent on the choice of B_c.

D. Optimization of B_c

In this section, we will optimize the number of observation samples used for channel estimation, or in other words seek the optimal B_c that will minimize the estimator variance given in (20),(22), and (25) for the LMMSE, LS, and BLUE, respectively

$$B_c = \arg\min_{\{B_c\}} \mathrm{MSE}. \tag{26}$$

It is difficult to find a closed-form solution for (26), especially for the BLUE. An alternative is to evaluate (26) exhaustively, which is feasible since there is only a limited range for the possible values of B_c

$$\frac{L_p}{2} - \frac{N}{2M} \leq B_c \leq \frac{L_p}{2} - \frac{(L+1)(Q+1)}{2M} \tag{27}$$

as we recall that M is the number of pilot clusters and L_p is the size of each pilot cluster. Clearly, the lower bound of B_c is due to the fact that the number of observation samples $M(L_p - 2B_c)$ cannot exceed the number of subcarriers N. The upper bound is associated with the rank condition of $\boldsymbol{\mathcal{P}}$. To have a good performance in the absence of interference and noise, all considered channel estimators require $\boldsymbol{\mathcal{P}}$ to have full column rank. Despite its importance, it is beyond the scope of this paper to discuss sufficient conditions for $\boldsymbol{\mathcal{P}}$ to have full column rank, which will depend on the choice of a specific BEM, the pilot pattern, and B_c. Note that the full column-rank condition can be always checked offline. However, a necessary condition for $\boldsymbol{\mathcal{P}}$ to have a full column rank is that it should be tall or at least square, which posts a lower bound of B_c.

Fortunately, even the exhaustive search might be avoided as will become evident from the simulations, where the MSE-versus-B_c curves for each channel estimator often exhibit a monotonous trail, although not necessarily in the same direction. For the LS channel estimator, the curve is monotonously descending and thus the optimal B_c must be chosen as large as possible. To explain this, we recall that a larger B_c leads to a more accurate band assumption of the channel matrix and, hence, to a smaller out-of-band interference power as suggested in Fig. 4(a). This is beneficial to the LS estimator, which is not good at suppressing the interference due to the lack of statistical knowledge. Opposed to the LS estimator, the LMMSE estimator and BLUE require the B_c to be as small as possible. For practical setups, this often implies a negative B_c, in which case the observation samples outnumber the pilots as illustrated in Fig. 4(b). We can see that some of the observation samples, e.g., in the two boundary areas, will suffer a very low signal-to-interference-and-noise ratio (SINR) because for these observation samples the unknown data are magnified by the high-power diagonals of the channel matrix and, hence, are much more prominent than the pilots. However, this casts no serious problem to the LMMSE estimator and BLUE since both of them can take the interference into account in a positive way. We will come back to this issue in Section V.

V. Numerical Results

In this section, we consider an OFDM system with $N = 256$ subcarriers, where roughly 80% of the subcarriers are used for transmitting data symbols. The remaining subcarriers are reserved for pilots, which are grouped in $M = 6$ equidistant clusters, each containing $L_p = 9$ pilot tones. Inside each cluster, we adopt the scheme referred to as "frequency-domain Kronecker

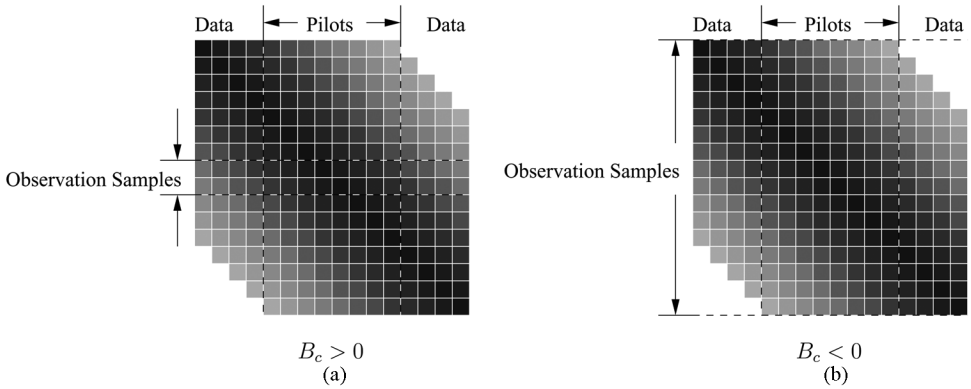

Fig. 4. Relationship between one pilot cluster and its corresponding observation samples via the channel matrix. The tint suggests the power degradation in the channel matrix. (a) $B_c > 0$. (b) $B_c < 0$.

delta" (FDKD) in [13], where a nonzero pilot is located in the middle of the cluster with zero guard bands on both sides.

The proposed channel estimators will be cast to TV channels, whose variation is described by the Doppler frequency normalized to the subcarrier spacing

$$f_D := \frac{v f_c}{c} T_s N \tag{28}$$

where v denotes the "virtual" velocity, f_c the carrier frequency, $T_s N$ the OFDM symbol duration, and c the speed of light. Note that in case of a moving terminal, the "virtual" velocity equals the terminal velocity, whereas in case of a moving scatterer, the "virtual" velocity equals twice the scatterer velocity. To be able to approximate the TV channel by a CE-BEM, we use the standard rule of thumb $Q \leq 2f_D$ to satisfy the Nyquist criterion. In the following test cases, we set $f_D \leq 1$ such that $Q = 2$ could be adequate, but in order to reduce the impact of the BEM modeling error we select $Q = 4$. For other BEMs, a different number of basis functions could lead to a better modeling performance. However, to make a comparison for a fixed estimation complexity, we will fix $Q = 4$ for all BEMs.

In the following test cases, we will restrict ourselves to a moving terminal with many uniformly distributed scatterers in the close vicinity of the terminal, leading to the typical bathtub-shaped Doppler spectrum [1]. We will follow the algorithms given in [35] to generate TV channels conform to such a Doppler spectrum. Further, we assume the channel to be a finite-impulse response (FIR) filter with $L+1 = 6$ taps, which are independent random variables with an exponential power intensity profile. More specifically, we take $\mathbf{R}_{\text{multipath}} = \text{diag}\left(\left[\sigma_0^2, \ldots, \sigma_L^2\right]\right)$ in (34) with $\sigma_l^2 = e^{-l/10}$. In short, we will characterize the TV channel with $(L+1)(Q+1) = 30$ BEM coefficients.

Test Case 1. The BEM Justification: We first list the modeling performance (e.g., the channel fitting performance in the absence of noise) of the DKL-BEM, CE-BEM, GCE-BEM, and P-BEM for a range of f_Ds. The DKL-BEM is constructed based on the bathtub-shaped Doppler spectrum but fixed at $f_D = 0.6$, and is thus suboptimal for other f_Ds. The BEM-modeled channel is compared to the true channel after windowing in terms of the modeling error $\mathcal{E}_{\mathbf{h}^{(t)}}\{\|\boldsymbol{\epsilon}\|^2\}$. For the window design, we adopt the "MBAE-SOE" window presented in [30],

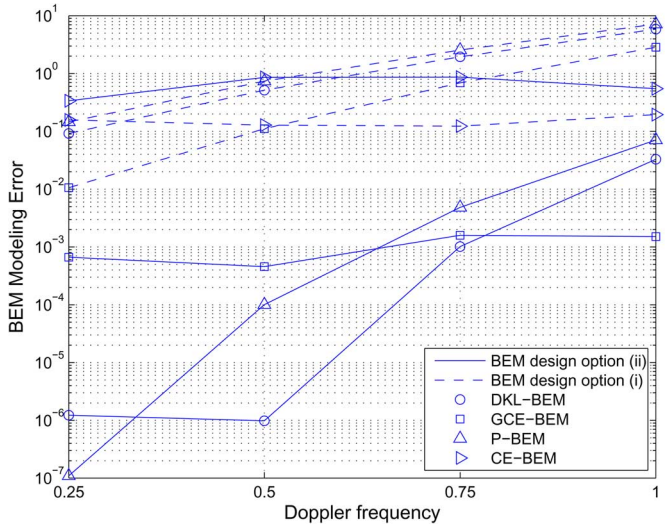

Fig. 5. Justification of (windowed) BEM.

which is a sum of three complex exponentials. We examine the modeling performance for two BEM designs: the first BEM design follows option i) in (6), which is the traditional BEM design ignoring the windowing; the second BEM design follows option ii) in (6), and is adapted to the windowing. From the results that are sketched in Fig. 5, we can observe that by taking the windowing into account, the BEMs following option ii) yield in general a tighter fit with the windowed channel, with the only exception of the CE-BEM, which, by following option i), performs better within the tested Doppler frequency range. Further, it can seen that the DKL-BEM and P-BEM have the smallest modeling error at low Doppler frequencies but lose track if the channel varies faster. Apparently, for the DKL-BEM, the mismatch due to an underestimated Doppler frequency is much more harmful than the mismatch due to an overestimated Doppler frequency. The GCE-BEM that is virtually independent of the Doppler frequency is more robust in this sense.

For the following simulations, we will concentrate on TV channels at the following two Doppler frequencies: 1) $f_D = 0.2$ and 2) $f_D = 1$. For the LMMSE estimator, we will use the DKL-BEM and allow for a mismatch by assuming $f_D = 0.6$, which means that the mismatch will be induced not only in the

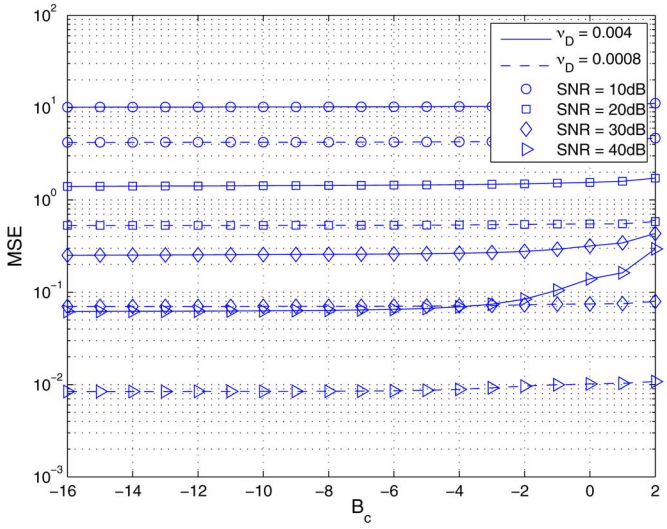

Fig. 6. MSE versus B_c for the LMMSE estimator. Solid curves: $f_D = 1$. Dashed curves: $f_D = 0.2$.

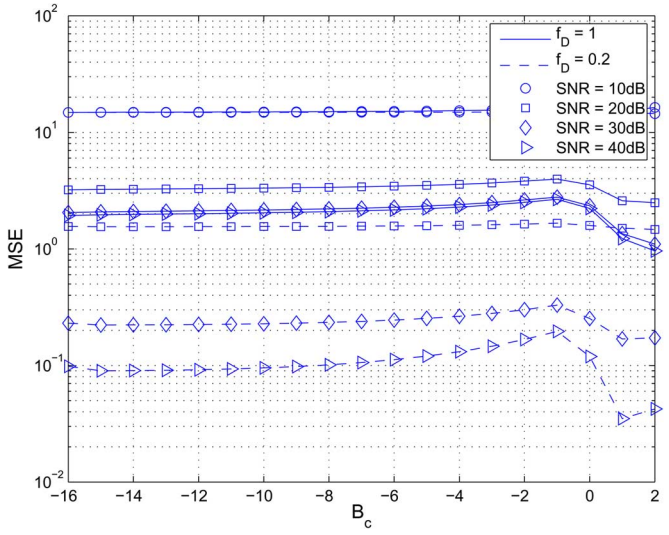

Fig. 7. MSE versus B_c for the LS estimator. Solid curves: $f_D = 1$. Dashed curves: $f_D = 0.2$.

BEM, but also in the estimator design. For the LS estimator and the BLUE, we will just use the GCE-BEM since both the channel estimators and the BEM are independent of the channel statistics. In addition, we will also compare our results with the channel estimation method for the CE-BEM presented in [13]. Note that this method resembles our proposed LMMSE estimator (without mismatch) but uses a data model that is only applicable to the CE-BEM, i.e., the channel matrix is viewed as strictly banded.

Test Case 2. Seeking the Optimal B_c: First, we need to find an optimal B_c for the different channel estimators. From (27), B_c is bounded by

$$-16 \leq B_c \leq 2. \tag{29}$$

For these values, we evaluate the MSE of the LMMSE estimator (20) for signal-to-noise ratio (SNR)= 10, 20, 30, and 40 dB, and depict the results in Fig. 6. We observe that at low SNR, the effect of B_c is not pronounced, whereas at high SNR, a smaller B_c corresponds to a smaller MSE. This is especially true for $f_D = 1$, where the ICI is still severe in spite of windowing. Tuning B_c is of greater importance in that case. Therefore, we choose $B_c = -16$ as the optimal value, which implies that the whole OFDM symbol will be invoked for channel estimation.

The results for the LS estimator are plotted in Fig. 7, where we observe that B_c must be chosen as large as possible, i.e., $B_c = 2$.

For the BLUE in Fig. 8, a smaller B_c always yields a lower MSE just like the LMMSE estimator and we should also take $B_c = -16$. However, complexity plays a crucial role in this case, because the BLUE has to be computed recursively and the procedure must be repeated for every OFDM symbol (note that the LMMSE estimator is in essence time-invariant and can thus be precomputed and stored offline). In practice, a smaller B_c often requires more iterations to reach convergence, and during each iteration, it inflicts a larger computational effort because more observation samples have to be processed. Observing that

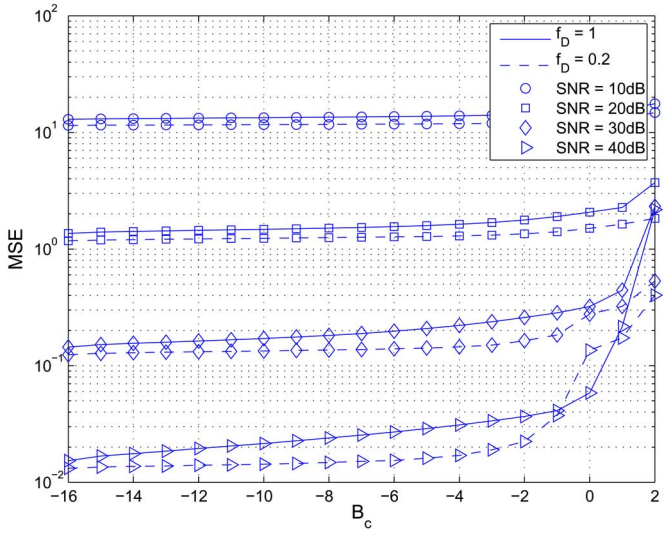

Fig. 8. MSE versus B_c for the BLUE. Solid curves: $f_D = 1$. Dashed curves: $f_D = 0.2$.

the MSE curve descends only slowly for $B_c < -3$, we select $B_c = -3$ as a good compromise between complexity and performance for the BLUE.

Test Case 3. The Estimator Performance: Having determined $B_c = -16$ for the LMMSE estimator, $B_c = 2$ for the LS estimator, and $B_c = -3$ for the BLUE, we inspect their channel estimation performance for a wide range of SNRs. Next to the MSE defined in (19), which we will refer to as the "BEM MSE," we will also look at the so-called "channel MSE" which we define by

$$\text{MSE–CH} := \mathcal{E}_{\mathbf{h}^{(t)}} \left\{ \left\| \mathbf{h}^{(t)} - (\mathbf{B} \otimes \mathbf{I}_{L+1}) \hat{\mathbf{h}} \right\|^2 \right\}. \tag{30}$$

Note that the channel MSE differs from the BEM MSE in that it explicitly takes the BEM modeling error into account, whereas the BEM MSE merely indicates how close the estimated channel is to the best possible BEM fit. Fig. 9 depicts the performance

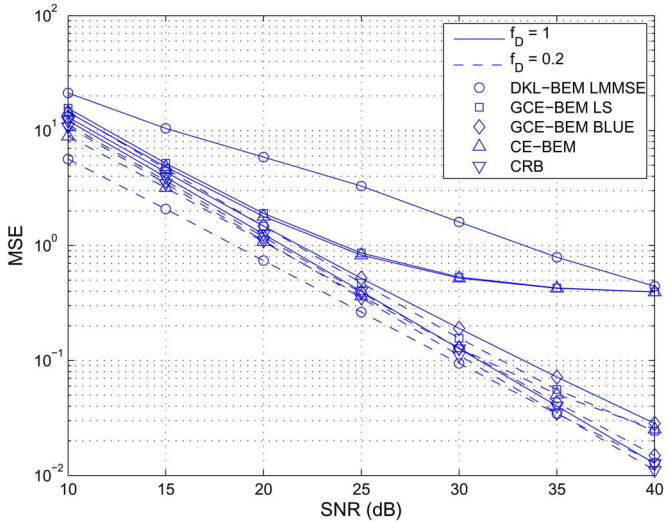

Fig. 9. MSE of the BEM versus SNR. Solid curves: $f_D = 1$. Dashed curves: $f_D = 0.2$.

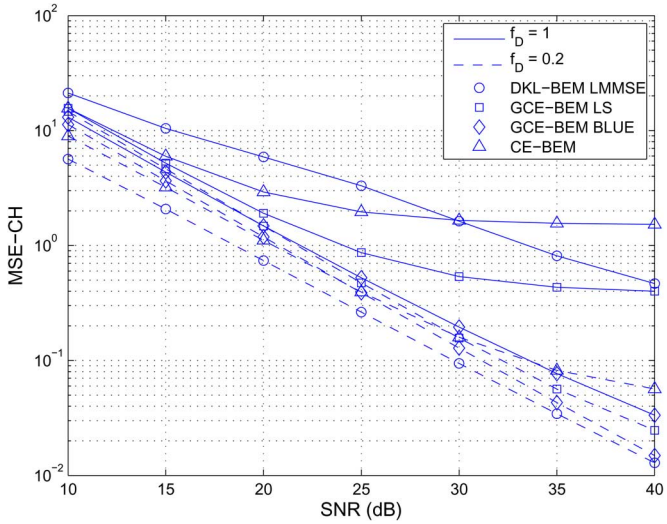

Fig. 10. Channel MSE versus SNR. Solid curves: $f_D = 1$. Dashed curves: $f_D = 0.2$.

in terms of the BEM MSE, whereas Fig. 10 depicts the performance in terms of the channel MSE. We observe that these two performances are in general very close to each other, which suggests that assumption a1) brings no harm to channel estimation. Further, we remark that the LMMSE estimator, which is suboptimal due to the Doppler frequency mismatch, performs much better under TV channels with $f_D = 0.2$ than with $f_D = 1$. This suggests that underestimating the Doppler frequency is more harmful than overestimating it.

The BEM MSE is also compared with the CRB (see the derivation in Appendix B) in Fig. 9. The CRB is based on $B_c = -16$ and obtained using the Monte Carlo method, thereby exploring the channel statistics. We observe that the performance of the BLUE is very close to the CRB.

Test Case 4. Equalization Performance Based on the Estimated Channel: For this test case, we will transmit quadrature phase-shift keying (QPSK)-modulated data symbols, and

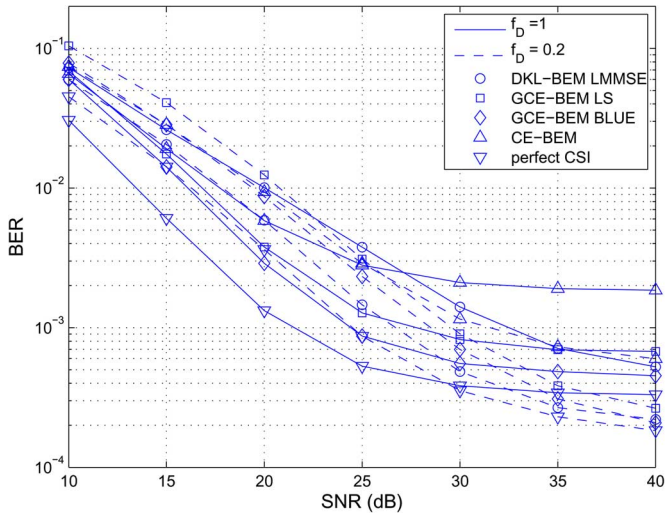

Fig. 11. BER versus SNR. Solid curves: $f_D = 1$. Dashed curves: $f_D = 0.2$.

the channel will be equalized by the banded LMMSE equalizer proposed in [30], although other banded equalizers can be employed as well [29], [31]. In short, the equalizer can be described as

$$\begin{aligned} \hat{\mathbf{s}} &= \bar{\mathbf{H}}_K^H \mathbf{G}^{-1} \mathbf{y} \\ \mathbf{G} &= \left(\bar{\mathbf{H}}_K \bar{\mathbf{H}}_K^H + \mathbf{R}_n + \sigma^2 \mathbf{I} \right) \end{aligned} \tag{31}$$

where $\bar{\mathbf{H}}_K$ is the banded approximation of the original channel matrix, $\bar{\mathbf{H}}_K = \mathbf{H} \odot \mathbf{T}_K$ with $\mathbf{T}_K$ being an $N \times N$ matrix whose main diagonal, K subdiagonals, and K super-diagonals are ones, and the remaining entries are zero. Thanks to the banded approximation and a simple banded LDL decomposition of $\mathbf{G}$, the complexity of the banded LMMSE equalizer is $\mathcal{O}(K^2 N)$. Not disclosed in [30], a regularization term $\sigma^2 \mathbf{I}$ is added to the expression of $\mathbf{G}$, which serves to prevent performance degradation at high SNR. This performance degradation is caused by the following: i) the band approximation error and ii) the fact that the "pseudocirculant" channel matrix $\mathbf{H}$ itself is extremely ill-conditioned. To determine σ in practice, we can first apply (31) by setting $\sigma = 0$, then observe from which SNR the BER curve begins to saturate, and finally choose σ^2 based on that SNR.

To examine the influence of the channel estimation error on the equalization, we construct the banded LMMSE equalizer utilizing the estimated channel obtained from our LMMSE estimator, LS estimator, and BLUE, respectively. We choose $K = 9$ in (31). As a comparison, we also list the equalization performances, which are based on the estimated CE-BEM channel and the perfect CSI. It can be seen that due to the Doppler diversity the equalizer renders a better performance for faster TV channels, but it is plagued by a higher BER floor. Comparing the results in Fig. 10 with those in Fig. 11, we can conclude that the equalization performance is in consistence with the corresponding channel estimation performance for each Doppler frequency. Take the DKL-BEM and the CE-BEM at $f_D = 1$, for instance; the channel estimation for the CE-BEM is better at low SNR but worse at high SNR than that for the DKL-BEM. We can make a corresponding observation in Fig. 11.

VI. Conclusion

In this paper, the TV channel is modeled by the BEM, which takes possible windowing into account. We view the resulting channel to be approximately banded, and, hence, the interference comes primarily from the out-of-band part of the channel matrix, which is combated by an LMMSE estimator, an LS estimator, and an iterative BLUE. It is worth underlining that using these channel estimators, we recover the full frequency-domain channel matrix, instead of its banded approximation (unless we use a CE-BEM assumption).

Appendix A
Derivation of Covariance Matrices $\mathbf{R}_n^{(p)}$, $\mathbf{R}_d$ and $\tilde{\mathbf{R}}_d(\mathbf{h})$

Let us start with the covariance matrix of the noise term $\mathbf{n}^{(p)}$, i.e.,

$$\mathbf{R}_n^{(p)} := \mathcal{E}_{\mathbf{n}}\{\mathbf{n}^{(p)}\mathbf{n}^{(p)H}\}. \tag{32}$$

It is clear that $\mathbf{R}_n^{(p)}$ is extracted from a larger matrix $\mathbf{R}_n = \mathcal{E}_{\mathbf{n}}\{\mathbf{n}\mathbf{n}^H\}$, which can be easily obtained from assumption a2) by taking the windowing and demodulation into account: $\mathbf{R}_n = \sigma_n^2\mathbf{F}\mathrm{diag}\{\mathbf{w}\}\mathrm{diag}\{\mathbf{w}^H\}\mathbf{F}^H$. $\mathbf{R}_n^{(p)}$ is comprised of the rows and columns of $\mathbf{R}_n$ corresponding to the positions of $\mathbf{y}^{(p)}$.

Since the interference term $\mathbf{d}$ depends on the BEM coefficients $\mathbf{h}$, we can make a distinction between two types of covariance matrices. One is based on a stochastic channel assumption, whereas the other is based on a deterministic channel assumption, which lead to $\mathbf{R}_d$ and $\tilde{\mathbf{R}}_d(\mathbf{h})$, respectively. Note that for both cases the information symbols $\mathbf{s}^{(d)}$ are always viewed as stochastic.

We first assume that $\mathbf{h}$ is stochastic with $\mathbf{R}_h = \mathcal{E}_{\mathbf{h}}\{\mathbf{h}\mathbf{h}^H\}$. To derive its expression, let us introduce the $(Q+1)\times(Q+1)$ matrix $\mathbf{R}_{h,l}$, which is defined as the autocorrelation of the BEM coefficients for tap l: $\mathbf{R}_{h,l} := \mathcal{E}\left\{\mathbf{h}_l\mathbf{h}_l^H\right\}\}$. By taking the windowing into account, we can easily derive that

$$\mathbf{R}_{h,l} = \mathbf{B}^{\dagger}\mathrm{diag}\{\mathbf{w}\}\tilde{\mathbf{R}}_{h,l}^{(t)}\mathrm{diag}\{\mathbf{w}^H\}\mathbf{B}^{\dagger H} \tag{33}$$

where $\tilde{\mathbf{R}}_{h,l}^{(t)}$ being defined in (5) is related to the specific Doppler spectrum. In addition, if we use an $(L+1)\times(L+1)$ matrix $\mathbf{R}_{\mathrm{multipath}}$ to describe the correlation due to the channel's frequency selectivity, which depends on the correlation between the channel taps and the power intensity profile, e.g., $[\mathbf{R}_{\mathrm{multipath}}]_{p,q} := \mathcal{E}\left\{\tilde{h}_{n,p}^{(t)}\tilde{h}_{n,q}^{(t)*}\right\}$, we have

$$\mathbf{R}_h = \mathbf{R}_{h,l} \otimes \frac{\mathbf{R}_{\mathrm{multipath}}}{\mathcal{E}\left\{|\tilde{h}_{n,l}^{(t)}|^2\right\}}. \tag{34}$$

With $\mathbf{R}_h$, we can derive the covariance matrix of the interference term $\mathbf{d}$ as

$$\begin{aligned}\mathbf{R}_d &:= \mathcal{E}_{\mathbf{h},\mathbf{s}^{(d)}}\{\mathbf{d}\mathbf{d}^H\}\\ &= \mathcal{E}_{\mathbf{h},\mathbf{s}^{(d)}}\left\{\mathbf{D}^{(d)}\boldsymbol{\mathcal{S}}^{(d)}\mathbf{h}\mathbf{h}^H\boldsymbol{\mathcal{S}}^{(d)H}\mathbf{D}^{(d)H}\right\}\\ &= \mathbf{D}^{(d)}\mathbf{R}_x\mathbf{D}^{(d)H}\end{aligned} \tag{35}$$

where using assumption a4), we have

$$\begin{aligned}\mathbf{R}_x &:= \mathcal{E}_{\mathbf{h},\mathbf{s}^{(d)}}\left\{\boldsymbol{\mathcal{S}}^{(d)}\mathbf{h}\mathbf{h}^H\boldsymbol{\mathcal{S}}^{(d)H}\right\}\\ &= \mathcal{E}_{\mathbf{s}^{(d)}}\left\{\left(\mathbf{I}_{Q+1}\otimes\mathrm{diag}\{\mathbf{s}^{(d)}\}\mathbf{F}_L^{(d)}\right)\right.\\ &\qquad\left.\times\mathbf{R}_h\left(\mathbf{I}_{Q+1}\otimes\mathrm{diag}\left\{\mathbf{s}^{(d)}\right\}\mathbf{F}_L^{(d)}\right)^H\right\}\\ &= \mathcal{E}_{\mathbf{s}^{(d)}}\left\{\left(\mathbf{I}_{Q+1}\otimes\mathrm{diag}\{\mathbf{s}^{(d)}\}\right)\right.\\ &\qquad\left.\times\boldsymbol{\mathcal{X}}\left(\mathbf{I}_{Q+1}\otimes\mathrm{diag}\{\mathbf{s}^{(d)}\}\right)^H\right\}\end{aligned}$$

with $\boldsymbol{\mathcal{X}} := \left(\mathbf{I}_{Q+1}\otimes\mathbf{F}_L^{(d)}\right)\mathbf{R}_h\left(\mathbf{I}_{Q+1}\otimes\mathbf{F}_L^{(d)}\right)^H$. Utilizing assumption a3), we can easily verify that

$$[\mathbf{R}_x]_{m,n} = \begin{cases}\sigma_s^2[\boldsymbol{\mathcal{X}}]_{m,n}, & \text{if } \mathrm{mod}(m-n,N_d)=0\\ 0, & \text{otherwise}\end{cases}$$

where $N_d = N - ML_p$ is the total number of information symbols in $\mathbf{s}^{(d)}$.

In contrast with $\mathbf{R}_d$, $\tilde{\mathbf{R}}_d(\mathbf{h})$ is obtained by assuming that $\mathbf{h}$ is deterministic

$$\begin{aligned}\tilde{\mathbf{R}}_d(\mathbf{h}) &:= \mathcal{E}_{\mathbf{s}^{(d)}}\{\mathbf{d}\mathbf{d}^H\}\\ &= \mathcal{E}_{\mathbf{s}^{(d)}}\{\mathbf{D}^{(d)}\boldsymbol{\mathcal{S}}^{(d)}\mathbf{h}\mathbf{h}^H\boldsymbol{\mathcal{S}}^{(d)H}\mathbf{D}^{(d)H}\}\\ &= \mathbf{D}^{(d)}\tilde{\mathbf{R}}_x(\mathbf{h})\mathbf{D}^{(d)H}\end{aligned} \tag{36}$$

where using assumption a3), we have

$$\begin{aligned}\tilde{\mathbf{R}}_x(\mathbf{h}) &:= \mathcal{E}_{\mathbf{s}^{(d)}}\{\boldsymbol{\mathcal{S}}^{(d)}\mathbf{h}\mathbf{h}^H\boldsymbol{\mathcal{S}}^{(d)H}\}\\ &= \mathrm{diag}\left\{\left(\mathbf{I}_{Q+1}\otimes\mathbf{F}_L^{(d)}\right)\mathbf{h}\right\}\\ &\quad\times\mathcal{E}_{\mathbf{s}^{(d)}}\left\{\left(\mathbf{1}_{Q+1,1}\otimes\mathbf{s}^{(d)}\right)\left(\mathbf{1}_{Q+1,1}\otimes\mathbf{s}^{(d)}\right)^H\right\}\\ &\quad\times\mathrm{diag}\left(\left(\mathbf{I}_{Q+1}\otimes\mathbf{F}_L^{(d)}\right)\mathbf{h}\right)^H\\ &= \sigma_s^2\mathrm{diag}\left(\left(\mathbf{I}_{Q+1}\otimes\mathbf{F}_L^{(d)}\right)\mathbf{h}\right)(\mathbf{1}_{Q+1,Q+1}\otimes\mathbf{I}_{N_d})\\ &\quad\times\mathrm{diag}\left(\left(\mathbf{I}_{Q+1}\otimes\mathbf{F}_L^{(d)}\right)\mathbf{h}\right)^H.\end{aligned} \tag{37}$$

Appendix B
CRB

For the sake of simplicity, we assume that $\mathbf{y}^{(p)}$ is Gaussian distributed with mean $\boldsymbol{\mathcal{P}}\mathbf{h}$ and covariance matrix $\tilde{\mathbf{R}}_{\mathcal{I}}(\mathbf{h})$. This is supported by assumption a2) and by the fact that N is large enough to make $\mathbf{d}$ approximately normal-distributed due to the central limit theorem. The negative Gaussian log-likelihood function $\mathcal{L}$ can hence be written as

$$\begin{aligned}-\mathcal{L} &= \mathcal{C}\ln(\det(\tilde{\mathbf{R}}_{\mathcal{I}}(\mathbf{h})))\\ &\quad+(\mathbf{y}^{(p)}-\boldsymbol{\mathcal{P}}\mathbf{h})^H\tilde{\mathbf{R}}_{\mathcal{I}}^{-1}(\mathbf{h})(\mathbf{y}^{(p)}-\boldsymbol{\mathcal{P}}\mathbf{h})\end{aligned} \tag{38}$$

which leads to the following Fisher information matrix (FIM):

$$\boldsymbol{\mathcal{J}}(\mathbf{h}) := \mathcal{E}_{\mathbf{s}^{(d)},\mathbf{n}} \left\{ \left(\frac{\partial \mathcal{L}}{\partial \mathbf{h}} \right)^T \left(\frac{\partial \mathcal{L}}{\partial \mathbf{h}} \right) \right\}. \tag{39}$$

Adapting the results given in [36], we can formulate the real FIM as

$$\boldsymbol{\mathcal{J}}(\mathbf{h}) = 2 \begin{bmatrix} \mathrm{Re}(\mathbf{J}_{\theta\theta}) & -\mathrm{Im}(\mathbf{J}_{\theta\theta}) \\ \mathrm{Im}(\mathbf{J}_{\theta\theta}) & \mathrm{Re}(\mathbf{J}_{\theta\theta}) \end{bmatrix} + 2 \begin{bmatrix} \mathrm{Re}(\mathbf{J}_{\theta\theta^*}) & -\mathrm{Im}(\mathbf{J}_{\theta\theta^*}) \\ \mathrm{Im}(\mathbf{J}_{\theta\theta^*}) & \mathrm{Re}(\mathbf{J}_{\theta\theta^*}) \end{bmatrix} \tag{40}$$

where

$$[\mathbf{J}_{\theta\theta}]_{i,j} := \left[\boldsymbol{\mathcal{P}}^H \tilde{\mathbf{R}}_{\mathcal{I}}^{-1} \boldsymbol{\mathcal{P}} \right]_{i,j} + \mathrm{trace} \left(\tilde{\mathbf{R}}_{\mathcal{I}}^{-1} \frac{\partial \tilde{\mathbf{R}}_{\mathcal{I}}}{\partial [\mathbf{h}]_i^*} \tilde{\mathbf{R}}_{\mathcal{I}}^{-1} \frac{\partial \tilde{\mathbf{R}}_{\mathcal{I}}}{\partial [\mathbf{h}]_j^*} \right)$$

$$[\mathbf{J}_{\theta\theta^*}]_{i,j} := \mathrm{trace} \left(\tilde{\mathbf{R}}_{\mathcal{I}}^{-1} \frac{\partial \tilde{\mathbf{R}}_{\mathcal{I}}}{\partial [\mathbf{h}]_i^*} \tilde{\mathbf{R}}_{\mathcal{I}}^{-1} \frac{\partial \tilde{\mathbf{R}}_{\mathcal{I}}}{\partial [\mathbf{h}]_j^*} \right).$$

Let us now focus on computing $\left(\partial \tilde{\mathbf{R}}_{\mathcal{I}} / \partial [\mathbf{h}]_j^* \right)$. Since $\left[\partial \tilde{\mathbf{R}}_{\mathcal{I}} / \partial [\mathbf{h}]_j^* \right]_{m,n} = \partial [\tilde{\mathbf{R}}_{\mathcal{I}}]_{m,n} / \partial [\mathbf{h}]_j^*$, we want to formulate $[\tilde{\mathbf{R}}_{\mathcal{I}}]_{m,n}$ as a function of $[\mathbf{h}]_j^*$. This is achieved by realizing that

$$[\tilde{\mathbf{R}}_{\mathcal{I}}]_{m,n} = \mathbf{e}_m^T \left(\mathbf{D}^{(d)} \tilde{\mathbf{R}}_d(\mathbf{h}) \mathbf{D}^{(d)H} + \mathbf{R}_n^{(p)} \right) \mathbf{e}_n \tag{41}$$

with $\tilde{\mathbf{R}}_d(\mathbf{h})$ defined in (37). Following the derivative rules given in [37], we have

$$\frac{\partial [\tilde{\mathbf{R}}_{\mathcal{I}}]_{m,n}}{\partial [\mathbf{h}]_j^*} = \mathbf{e}_m^T \mathbf{D}^{(d)} \frac{\partial \tilde{\mathbf{R}}_d(\mathbf{h})}{\partial [\mathbf{h}]_j^*} \mathbf{D}^{(d)H} \mathbf{e}_n$$

$$\frac{\partial \tilde{\mathbf{R}}_d(\mathbf{h})}{\partial [\mathbf{h}]_j^*} = \sigma_s^2 \left(\left(\mathbf{I}_{Q+1} \otimes \mathbf{F}_L^{(d)} \right) \mathbf{h} \right) \left(\mathbf{1}_{Q+1,Q+1} \otimes \mathbf{I}_{N_d} \right) \times \frac{\partial \mathrm{diag} \left(\left(\mathbf{I}_{Q+1} \otimes \mathbf{F}_L^{(d)} \right) \mathbf{h} \right)^H}{\partial [\mathbf{h}]_j^*}.$$

To work out the last equation, we realize that in the matrix $\mathrm{diag} \left((\mathbf{I}_{Q+1} \otimes \mathbf{F}_L^{(d)}) \mathbf{h} \right)^H$, only the $(1 + \lfloor j - 1/L + 1 \rfloor N_d)$th until the $(N_d + \lfloor j - 1/L + 1 \rfloor N_d)$th diagonal entries are associated with $[\mathbf{h}]_j^*$, with the coefficients $\mathbf{F}_L^{(d)*} \mathbf{e}_{\mathrm{mod}(j-1,L+1)+1}$. Therefore

$$\frac{\partial \mathrm{diag} \left(\left(\mathbf{I}_{Q+1} \otimes \mathbf{F}_L^{(d)} \right) \mathbf{h} \right)^H}{\partial [\mathbf{h}]_j^*} = \mathrm{diag} \left(\left[\mathbf{0}_{1,\lfloor j-1/L+1 \rfloor N_d}, \mathbf{e}_{\mathrm{mod}(j-1,L+1)+1}^T \mathbf{F}_l^{(d)H}, \mathbf{0}_{1,(Q-\lfloor j-1/L+1 \rfloor)N_d} \right]^T \right). \tag{42}$$

Coining some new matrix definitions

$$\boldsymbol{\mathcal{H}} := \mathrm{diag} \left(\left(\mathbf{I}_{Q+1} \otimes \mathbf{F}_L^{(d)} \right) \mathbf{h} \right)$$

$$\boldsymbol{\mathcal{F}}_j := \mathbf{1}_{Q+1,1} \otimes \mathrm{diag} \left\{ \mathbf{F}_l^{(d)*} \mathbf{e}_{\mathrm{mod}(j-1,L+1)+1} \right\}$$

$$\mathbf{D}_j^{(d)} := \mathbf{D}^{(d)} \left[\mathbf{0}_{N_d,\lfloor j-1/L+1 \rfloor N_d} \mathbf{I}_{N_d} \mathbf{0}_{N_d,(Q-\lfloor j-1/L+1 \rfloor)N_d} \right]^T \tag{43}$$

we can easily show

$$\frac{\partial \tilde{\mathbf{R}}_{\mathcal{I}}}{\partial [\mathbf{h}]_j^*} = \sigma_s^2 \mathbf{D}^{(d)} \boldsymbol{\mathcal{H}} \boldsymbol{\mathcal{F}}_j \mathbf{D}_j^{(d)H}. \tag{44}$$

With the obtained FIM, we find a lower bound on the channel estimator's variance [33]

$$\mathrm{CRB} = \mathrm{trace} \left\{ (\boldsymbol{\mathcal{J}}(\mathbf{h}))^{-1} \right\} \leq \mathrm{trace} \left\{ \mathcal{E}_{\mathbf{s}^{(d)},\mathbf{n}} \{ (\hat{\mathbf{h}} - \mathbf{h})(\hat{\mathbf{h}} - \mathbf{h})^H \} \right\} \tag{45}$$

which is also a lower bound on the MSE of the BEM channel, as can be seen from (19).

REFERENCES

[1] W. C. Jakes, *Microwave Mobile Channels*. New York: Wiley, 1974.

[2] S. Thoen, L. V. d. Perre, and M. Engles, "Modeling the channel time-variance for fixed wireless communications," *IEEE Commun. Lett.*, vol. 6, no. 8, pp. 331–333, Aug. 2002.

[3] M. Visintin, "Karhunen-Loeve expansion of a fast Rayleigh fading process," *IEEE Electron. Lett.*, vol. 32, no. 8, pp. 1712–1713, Aug. 1996.

[4] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1996.

[5] K. D. Teo and S. Ohno, "Optimal MMSE finite parameter model for doubly-selective channels," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, 2005, pp. 3503–3507.

[6] 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.

[7] M. K. Tsatsanis and G. B. Giannakis, "Modeling and equalization of rapidly fading channels," *Int. J. Adapt. Control Signal Process.*, vol. 10, pp. 159–176, Mar. 1996.

[8] H. A. Cirpan and M. K. Tsatsanis, "Maximum likelihood blind channel estimation in the presence of Doppler shifts," *IEEE Trans. Signal Process.*, vol. 47, no. 6, pp. 1559–1569, Jun. 1999.

[9] M. Guillaud and D. T. M. Slock, "Channel modeling and associated inter-carrier interference equalization for OFDM systems with high Doppler spread," in *Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Apr. 2003, vol. IV, pp. 237–240.

[10] X. Ma, G. Giannakis, and S. Ohno, "Optimal training for block transmissions over doubly-selective fading channels," *IEEE Trans. Signal Process.*, vol. 51, no. 5, pp. 1351–1366, May 2003.

[11] G. B. Giannakis and C. Tepedelenlioğlu, "Basis expansion models and diversity techniques for blind identification and equalization of time-varying channels," *Proc. IEEE*, vol. 86, no. 10, pp. 1969–1986, Oct. 1998.

[12] G. Leus and M. Moonen, "Deterministic subspace based blind channel estimation for doubly-selective channels," in *Proc. IEEE Signal Process. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2003, pp. 210–214.

[13] A. P. Kannu and P. Schniter, "MSE-optimal training for linear time-varying channels," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Mar. 2005, pp. iii/789–iii/792.

[14] G. Leus, "On the estimation of rapidly time-varying channels," in *Euro. Signal Process. Conf. (EUSIPCO)*, Sep. 2004.

[15] D. K. Borah and B. D. Hart, "Frequency-selective fading channel estimation with a polynomial time-varying channel model," *IEEE Trans. Commun.*, vol. 47, no. 6, pp. 862–873, Jun. 1999.
[16] S. Tomasin, A. Gorokhov, H. Yang, and J.-P. Linnartz, "Iterative interference cancellation and channel estimation for mobile OFDM," *IEEE Trans. Wireless Commun.*, vol. 4, no. 1, pp. 238–245, Jan. 2005.
[17] A. Gorokhov and J.-P. Linnartz, "Robust OFDM receivers for dispersive time-varying channels: Equalization and channel acquisition," *IEEE Trans. Commun.*, vol. 52, no. 4, pp. 572–583, Apr. 2004.
[18] A. Stamoulis, S. N. Diggavi, and N. Al-Dhahir, "Intercarrier interference in MIMO OFDM," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2451–2464, Oct. 2002.
[19] M. Nicoli, O. Simeone, and U. Spagnolini, "Multislot estimation of frequency-selective fast-varying channels," *IEEE Trans. Commun.*, vol. 51, no. 8, pp. 1337–1347, Aug. 2003.
[20] M. Dong, L. Tong, and B. Sadler, "Optimal insertion of pilot symbols for transmissions over time-varying flat fading channels," *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1403–1418, May 2004.
[21] J. K. Tugnait and W. Luo, "Linear prediction error method for blind identification of periodically time-varying channel," *IEEE Trans. Signal Process.*, vol. 50, no. 12, pp. 3070–3082, Dec. 2002.
[22] O. Simeone and U. Spagnolini, "Lower bound on training-based channel estimation error for frequency-selective block-fading Rayleigh MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 11, pp. 3265–3277, Nov. 2004.
[23] R. Negi and J. Cioffi, "Pilot tone selection for channel estimation in a mobile OFDM system," *IEEE Trans. Consum. Electron.*, vol. 44, no. 8, pp. 1122–1128, Aug. 1998.
[24] C. R. N. Athaudage and A. D. S. Jayalath, "Enhanced MMSE channel estimation using timing error statistics for wireless OFDM systems," *IEEE Trans. Broadcast.*, vol. 50, no. 4, pp. 369–376, Dec. 2004.
[25] *Implementation Guidelines for DVB Terrestrial Services; Transmission Aspects*, Euro. Telecommun. Standards Inst. (ETSI), 1998.
[26] Y.-S. Choi, P. J. Voltz, and F. A. Cassara, "On channel estimation and detection for multicarrier signals in fast and selective Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 49, no. 8, pp. 1375–1387, Aug. 2001.
[27] T. Cui, C. Tellambura, and Y. Wu, "Low-complexity pilot-aided channel estimation for OFDM systems over doubly-selective channels," in *IEEE Int. Conf. Commun. (ICC)*, May 2005, vol. 3, pp. 1980–1984.
[28] X. Cai and G. B. Giannakis, "Bounding performance and suppressing intercarrier interference in wireless mobile OFDM," *IEEE Trans. Commun.*, vol. 51, no. 12, pp. 2047–2056, Dec. 2003.
[29] P. Schniter, "Low-complexity equalization of OFDM in doubly-selective channels," *IEEE Trans. Signal Process.*, vol. 52, no. 4, pp. 1002–1011, Apr. 2004.
[30] L. Rugini, P. Banelli, and G. Leus, "Block DFE and windowing for Doppler-affected OFDM systems," in *IEEE Signal Process. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2005, pp. 470–474.
[31] S. Ohno, "Maximum likelihood inter-carrier interference suppression for wireless OFDM with null subcarriers," in *Int.Conf. Acoust., Speech, Signal Process. (ICASSP)*, Mar. 2005, vol. 3, pp. 849–852.
[32] L. Rugini and P. Banelli, "Windowing techniques for ICI mitigation in multicarrier systems," in *Euro. Signal Process. Conf. (EUSIPCO)*, 2005.
[33] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
[34] M. Ghogho and A. Swami, "Improved channel estimation using superimposed training," in *IEEE Signal Process. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jul. 2004, pp. 110–114.
[35] 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.
[36] E. d. Carvalho and D. T. M. Slock, "Cramer-Rao bounds for semi-blind, blind and training sequence based channel estimation," in *Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Apr. 1997, pp. 3593–3596.
[37] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 2000.

Zijian Tang (S'06) received the M.S. degree in electrical engineering from the Delft University of Technology, Delft, The Netherlands in 2003, where he is now working towards the Ph.D. degree in signal processing for communication.

From June to September 2003, he took an internship at the Philips Research Center (NATLAB), Eindhoven, the Netherlands. Since October 2003, he has been with the Circuit and Systems group at the Electrical Engineering Department, Delft University of Technology. His research interests are in the area of signal processing for communication over time-varying channels.

Rocco Claudio Cannizzaro received the B.S. degree in information technology engineering and the M.S. degree in computer science and telecommunications engineering from the University of Perugia, Perugia, Italy, in 2003 and 2005, respectively.

His research interests include time varying channel estimation and equalization techniques and their applications to wireless communication.

Geert Leus (SM'95) was born in Leuven, Belgium, in 1973. He received the electrical engineering degree and the Ph.D. degree in applied sciences from the Katholieke Universiteit Leuven, Leuven, Belgium, in June 1996 and May 2000, respectively.

He has been a Research Assistant and a Postdoctoral Fellow of the Fund for Scientific Research—Flanders, Belgium, from October 1996 till September 2003. During that period, he was with the Electrical Engineering Department, the Katholieke Universiteit Leuven. Currently, he is an Assistant Professor at the Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Delft, The Netherlands. During the summer of 1998, he visited Stanford University, and from March 2001 until May 2002 he was a Visiting Researcher and Lecturer at the University of Minnesota. His research interests are in the area of signal processing for communications.

Dr. Leus received a 2002 IEEE Signal Processing Society Young Author Best Paper Award and a 2005 IEEE Signal Processing Society Best Paper Award. He is a member of the IEEE Signal Processing for Communications Technical Committee and an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the *EURASIP Journal on Applied Signal Processing*. In the past, he has served on the Editorial Board of the IEEE SIGNAL PROCESSING LETTERS and the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.

Paolo Banelli (S'90–M'99) received the Laurea degree in electronics engineering and the Ph.D. degree in telecommunications from the University of Perugia, Perugia, Italy, in 1993 and 1998, respectively.

In 2005, he was appointed Associate Professor at the Department of Electronic and Information Engineering (DIEI), University of Perugia, where he has been an Assistant Professor since 1998. In 2001, he joined the SpinComm group at the Electrical and Computer Engineering Department, University of Minnesota, Minneapolis, as a Visiting Researcher. His research interests include nonlinear distortions, broadcasting, time-varying channels estimation and equalization, and block-transmission techniques for wireless communications. He has been serving as a Reviewer for several technical journals and as technical program committee member of leading international conferences on signal processing and telecommunications.