

Identifying Time-Varying Channels with Aid of Pilots for MIMO-OFDM

Zijian Tang¹, and Geert Leus²

¹ TNO Defense, Security and Safety
P.O. Box 96864, 2509 JG The Hague, The Netherlands
zijian.tang@tno.nl

² Delft University of Technology - Fac. EEMCS
Mekelweg 4, 2628 CD Delft, The Netherlands
g.j.t.leus@tudelft.nl

Abstract

In this paper, we consider pilot-aided channel estimation for orthogonal frequency division multiplexing (OFDM) systems with a multiple-input multiple-output (MIMO) setup. The channel is time-varying due to Doppler effects, and can be approximated by an oversampled complex-exponential basis expansion model [(O)CE-BEM]. We use a best linear unbiased estimator (BLUE) to estimate the channel with the aid of frequency-multiplexed pilots. The applicability of the BLUE, which is referred to as the channel identifiability in this paper, relies upon a proper pilot structure. Depending on whether the channel is estimated within a single OFDM symbol or multiple OFDM symbols, we propose simple pilot structures that guarantee channel identifiability. Further, it is shown that by employing more receive antennas, the BLUE can combat more effectively the Doppler-induced interference and therefore improve the channel estimation performance.

Keywords

MIMO, OFDM, BLUE, time-varying channel, pilot-aided channel estimation, BEM

I. INTRODUCTION

Orthogonal frequency division multiplexing (OFDM) systems have attracted enormous attention recently, and have been adopted in numerous existing communication systems. OFDM gains most of its popularity thanks to its ability to transmit signals on separate subcarriers without mutual interference. To further enhance the capacity of the transmission link, OFDM systems can be combined with multiple-input multiple-output (MIMO) features.

The fact that OFDM can transmit signals on separate subcarriers can be mathematically represented in the frequency domain by a diagonal channel matrix. This property holds only in a situation where the channel stays (almost) constant for at least one OFDM symbol interval. In practice, a time-invariant channel assumption can become invalid due to, e.g., Doppler effects resulting from the motion between the transmitter and receiver. In such a case, the frequency-domain channel matrix is not diagonal but generally full with the non-zero off-diagonal elements leading to inter-carrier interference (ICI).

To equalize such channels, the knowledge of all the elements in the channel matrix is required. In order to reduce the number of unknown channel parameters, a widely adopted approach is approximating the variation of the channel in the time domain with a parsimonious model, e.g., a basis expansion model (BEM). Consequently, channel estimation boils down to estimating the corresponding BEM coefficients. Among the various BEMs that have been proposed, this paper

will concentrate on the so-called oversampled complex exponential BEM [(O)CE-BEM] [1]. By tuning the oversampling factor, the (O)CE-BEM is reported in [2] to fit time-varying channels much tighter than its variant, the critically-sampled complex exponential BEM [(C)CE-BEM] [3, 4], and it has a steady modelling performance for a wide range of Doppler spreads [5].

Based on a general BEM assumption, the OFDM channel is estimated in [6] utilizing pilots that are multiplexed with data in the frequency domain. The same paper shows that the channel estimators that view the frequency-domain channel matrix as full, such as the (O)CE-BEM, render a better performance than those that view the channel matrix as diagonal [5], or strictly banded [4], such as the (C)CE-BEM. In this paper, the results of [6] will be extended from a single-input single-output (SISO) scenario to MIMO, with a focus on channel identifiability issues.

Estimating time-varying channels in a MIMO-OFDM system gives rise to a number of additional challenges. In the first place, due to multiple transmit-receive links, more channel unknowns need to be estimated, which requires more pilots and thus imposes a higher pressure on the bandwidth efficiency. To alleviate this problem, we will employ more pilot-carrying OFDM symbols to leverage the channel correlation along the time axis as in [7, 8]. Although this comes at a penalty of a larger BEM modeling error, the overall channel estimation performance can still be improved.

Another challenge in a MIMO-OFDM system is how to distribute pilots in the time, frequency, and spatial domains. [9, 10] proposes optimal pilot schemes but only for time-invariant channels or systems for which the time-variation of the channel within one OFDM symbol can be neglected. Except for [7, 11], much less attention has been paid to systems dealing with channels varying faster. In this paper, we will use the channel identifiability criterion as a guideline to design pilot schemes. It is noteworthy that the proposed pilot structures can be independent of the oversampling factor of the (O)CE-BEM, which endows the receiver with the freedom to choose the most suitable oversampling factor.

Pilot structures can have a great impact on both channel identifiability and estimation performance. The latter is, however, difficult to tackle analytically for time-varying channels. In this paper, we will try to establish, by means of simulations, a guideline for designing pilots that render a satisfactory channel estimation performance for different channel situations.

The MIMO feature brings not only design challenges, but also performance benefits. Due to

the ICI, the contribution of the pilots is always mixed with the contribution of the unknown data in the received samples. By taking this interference explicitly into account in the channel estimator design, [6] shows that the resulting best linear unbiased estimator (BLUE) can cope with the interference reasonably well, producing a performance close to the Crámer-Rao bound (CRB). When multiple receive antennas are deployed, we observe that the channel estimation performance can even be further improved. This is attributed to the fact that each receive antenna gets a different copy of the same transmitted data. The interference is therefore correlated across the receive antennas, which can be exploited by the BLUE to suppress the interference more effectively than in the single receive antenna case. To our best knowledge, this effect has not been reported before.

The remainder of the paper is organized as follows. In Section II, we present a general MIMO-OFDM system model. In Section III, we describe how the BLUE can be used to estimate the BEM coefficients. Channel identifiability is discussed in Section IV, based on which we propose a variety of pilot structures. The simulation results are given in Section V, where we discuss the impact of the various pilot structures on the performance. Conclusions are given in 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. $[\mathbf{x}]_p$ indicates the p th element of the vector \mathbf{x} and $[\mathbf{X}]_{p,q}$ indicates the (p, q) th entry of the matrix \mathbf{X} . $\mathcal{D}\{\mathbf{x}\}$ is used to denote a diagonal matrix with \mathbf{x} on the diagonal, and $\mathcal{D}\{\mathbf{A}_0, \dots, \mathbf{A}_{N-1}\}$ is used to denote a block-wise diagonal matrix with the matrices $\mathbf{A}_0, \dots, \mathbf{A}_{N-1}$ on the diagonal. \otimes and \dagger represent the Kronecker product and the pseudo inverse, respectively. \mathbf{I}_N stands for the $N \times N$ identity matrix; $\mathbf{1}_{M \times N}$ for the $M \times N$ all-one matrix, and \mathbf{W}_K for a K -point normalized discrete Fourier transform (DFT) matrix. We use $\mathbf{X}^{\{\mathcal{R}, \mathcal{C}\}}$ to denote the submatrix of \mathbf{X} , whose row and column indices are collected in the sets \mathcal{R} and \mathcal{C} , respectively; Similarly, we use $\mathbf{X}^{\{\mathcal{R}, \cdot\}}$ ($\mathbf{X}^{\{\cdot, \mathcal{C}\}}$) to denote the rows (columns) of \mathbf{X} , whose indices are collected in \mathcal{R} (\mathcal{C}). The cardinality of the set \mathcal{S} is denoted by $|\mathcal{S}|$.

II. SYSTEM MODEL

Let us consider a MIMO-OFDM system with N_T transmit antennas and N_R receive antennas, where the channel in the time domain is assumed to be a time-varying causal finite impulse response (FIR) filter with a maximum order L . Using $h_{p,l}^{(m,n)}$ to denote the time-domain channel gain of the l th lag at the p th time instant for the channel between the m th transmit antenna

and n th receive antenna, we can assume that $h_{p,l}^{(m,n)} = 0$ for $l < 0$ or $l > L$. Note that this channel model can take the transmit/receiver filter, the propagation environment, and the possible synchronization errors among different transmission links into account.

For the j th OFDM symbol that is transmitted via the m th transmit antenna, the data symbols $\mathbf{s}^{(m)}[j]$ are first modulated on K subcarriers by means of the inverse DFT (IDFT) matrix \mathbf{W}_K^H , then concatenated by a cyclic prefix (CP) of length $L_{\text{cp}} \geq L$, and finally sent over the channel. At the receiver, the received samples corresponding to the CP are discarded, and the remaining samples are demodulated by means of the DFT matrix \mathbf{W}_K . Mathematically, we can express the received samples during the j th OFDM symbol as

$$\mathbf{y}^{(n)}[j] = \sum_{m=1}^{N_T} \underbrace{\mathbf{W}_K \mathbf{H}_c^{(m,n)}[j] \mathbf{W}_K^H}_{\mathbf{H}_d^{(m,n)}[j]} \mathbf{s}^{(m)}[j] + \mathbf{z}^{(n)}[j], \quad (1)$$

where $\mathbf{z}^{(n)}[j]$ represents the additive noise related to the n th receive antenna; $\mathbf{H}_c^{(m,n)}[j]$ denotes the channel matrix between the m th transmit antenna and n th receive antenna in the time domain, and $\mathbf{H}_d^{(m,n)}[j] := \mathbf{W}_K \mathbf{H}_c^{(m,n)}[j] \mathbf{W}_K^H$ represents its counterpart in the frequency domain. Under the FIR assumption of the channel and letting $L_{\text{cp}} = L$ without loss of generality, we can express the entries of $\mathbf{H}_c^{(m,n)}[j]$ as $[\mathbf{H}_c^{(m,n)}[j]]_{p,q} = h_{j(K+L)+p+L, \text{mod}(p-q, K)}^{(m,n)}$ with $\text{mod}(a, b)$ standing for the remainder of a divided by b .

Obviously, if the channel stays constant within an OFDM symbol, $\mathbf{H}_c^{(m,n)}[j]$ will be a circulant matrix (hence the subscript c). This results in a diagonal matrix $\mathbf{H}_d^{(m,n)}[j]$ (hence the subscript d), which means that the subcarriers are orthogonal to each other. This property is however corrupted if the time variation within an OFDM symbol is not negligible.

III. CHANNEL ESTIMATION

For the ease of analysis, we will differentiate between two cases throughout the whole paper. The first case is based on a single OFDM symbol, which means that the channel will be estimated for each OFDM symbol individually. The other case employs multiple OFDM symbols. Because these two cases are characterized by some unique properties, we treat them separately.

III-A. Single OFDM Symbol

A.1 Data Model and BEM Based on a Single OFDM Symbol

Let us use a BEM to model the time variation of the channel within one OFDM symbol: for the channel between the m th transmit antenna and the n th receive antenna, the l th lag during the j th OFDM symbol can be approximated as

$$\begin{bmatrix} h_{j(K+L),l}^{(m,n)} \\ \vdots \\ h_{j(K+L)+K-1,l}^{(m,n)} \end{bmatrix} \approx \underbrace{[\mathbf{u}_0, \dots, \mathbf{u}_Q]}_{\mathbf{U}} \begin{bmatrix} c_{0,l}^{(m,n)}[j] \\ \vdots \\ c_{Q,l}^{(m,n)}[j] \end{bmatrix}, \quad (2)$$

where \mathbf{u}_q denotes the q th basis function of a BEM, and $c_{q,l}^{(m,n)}[j]$ the corresponding BEM coefficient. Under a CE-BEM assumption,

$$\mathbf{u}_q := [1, e^{-j\frac{2\pi}{\kappa(K+L)}q}, \dots, e^{-j\frac{2\pi}{\kappa(K+L)}q(K-1)}]^T, \quad (3)$$

where κ stands for the oversampling factor with $\kappa = \frac{K}{K+L}$ used for the (C)CE-BEM and $\kappa > \frac{K}{K+L}$ for the (O)CE-BEM.

Assuming that the BEM inflicts a negligible modeling error, the $K(L+1)$ channel taps within the j th OFDM symbol will be uniquely represented by the $(L+1)(Q+1)$ BEM coefficients $c_{q,l}^{(m,n)}[j]$. As a result, the frequency-domain channel matrix $\mathbf{H}_d^{(m,n)}[j]$ given in (1) can be rewritten in terms of the BEM as

$$\mathbf{H}_d^{(m,n)}[j] = \sum_{q=0}^Q \mathbf{W}_K \mathcal{D}\{\mathbf{u}_q\} \mathbf{C}_q^{(m,n)}[j] \mathbf{W}_K^H,$$

where $\mathbf{C}_q^{(m,n)}[j]$ is a circulant matrix with $[c_q^{(m,n)T}[j], \mathbf{0}_{1 \times (K-L-1)}]^T$ as its first column. Here, $\mathbf{c}_q^{(m,n)}[j] := [c_{q,0}^{(m,n)}[j], \dots, c_{q,L}^{(m,n)}[j]]^T$. Due to its circularity, we can express $\mathbf{C}_q^{(m,n)}[j]$ as

$$\mathbf{C}_q^{(m,n)}[j] = \mathbf{W}_K^H \mathcal{D}\{\mathbf{V}_L \mathbf{c}_q^{(m,n)}[j]\} \mathbf{W}_K, \quad (4)$$

where \mathbf{V}_L denotes the matrix that consists of the first $L+1$ columns of $\sqrt{K} \mathbf{W}_K$. Accordingly, $\mathbf{H}_d^{(m,n)}[j]$ can be written as

$$\mathbf{H}_d^{(m,n)}[j] = \sum_{q=0}^Q \mathbf{W}_K \mathcal{D}\{\mathbf{u}_q\} \mathbf{W}_K^H \mathcal{D}\{\mathbf{V}_L \mathbf{c}_q^{(m,n)}[j]\}. \quad (5)$$

Because we will only concentrate on a single OFDM symbol in this section, we drop the index j for the sake of simplicity.

Let us now use $\mathbf{p}^{(m)}$ to denote the pilots sent by the m th transmit antenna, whose subcarrier positions are contained in the set $\mathcal{P}^{(m)}$, and $\mathbf{d}^{(m)}$ to denote the data sent by the m th transmit antenna, whose subcarrier positions are contained in the set $\mathcal{D}^{(m)}$. Because in this paper we focus on frequency-domain multiplexed pilots, this implies that $\mathcal{P}^{(m)} \cap \mathcal{D}^{(m)} = \emptyset$ and $\mathcal{P}^{(m)} \cup \mathcal{D}^{(m)} = \{0, \dots, K-1\}$. Further, we assume that the pilots are grouped in G clusters, each of length $P+1$: $\mathbf{p}^{(m)} = [\mathbf{p}_0^{(m)T}, \dots, \mathbf{p}_{G-1}^{(m)T}]^T$. For the g th pilot cluster $\mathbf{p}_g^{(m)}$, the positions of its elements are collected in the set $\mathcal{P}_g^{(m)} = \{P_g^{(m)}, \dots, P_g^{(m)} + P\}$ with $P_g^{(m)}$ standing for its starting position. Corresponding to the positions of $\mathbf{p}_g^{(m)}$, let us consider the observation samples at the receiver, whose indices are collected in the set

$$\mathcal{O}_g^{(m)} = \{P_g^{(m)} + \frac{D}{2} - \ell, \dots, P_g^{(m)} + P - \frac{D}{2} + \ell\}. \quad (6)$$

It can be seen from the above that the number of observation samples in $\mathcal{O}_g^{(m)}$, given by $P - D + 2\ell + 1$, is controlled by the two parameters D and ℓ . To understand the physical meaning of D , we know that for a small Doppler spread, the ICI is mostly limited to the neighbouring subcarriers, which is equivalent to the assumption that the frequency-domain channel matrix has most of its power located on the main diagonal, the $D/2$ sub- and $D/2$ super-diagonals for an appropriate value of D . In an ideal case where the channel matrix is strictly banded, we should choose

$$\mathcal{O}_g^{(m)} = \{P_g^{(m)} + \frac{D}{2}, \dots, P_g^{(m)} + P - \frac{D}{2}\}, \quad (7)$$

such that the resulting observation samples will depend exclusively on the pilots $\mathbf{p}_g^{(m)}$. However, such a strictly banded assumption is not true, and the channel matrix is full in nature especially at high Doppler spreads. This implies that there is always a power leakage outside the band, which is accounted for in (6) by adding an additional parameter ℓ . The relationship between $\mathbf{p}_g^{(m)}$ and the corresponding observation samples is illustrated in Fig. 1. As shown in [6], the choice of ℓ can have a great impact on the channel estimation performance.

The above analysis is based on a single transmit antenna. For a MIMO scenario, every receiver ‘sees’ a superposition of OFDM symbols from all the transmit antennas. This implies that the g th observation cluster \mathcal{O}_g must be a union of all the individual observation clusters related to the transmit antennas:

$$\mathcal{O}_g = \mathcal{O}_g^{(0)} \cup \dots \cup \mathcal{O}_g^{(N_T-1)}. \quad (8)$$

As a result, we can use the input-output relationship given in (1) to express $\mathbf{y}^{(n)\{\mathcal{O}_g\}}$ as

$$\mathbf{y}^{(n)\{\mathcal{O}_g\}} = \sum_{m=0}^{N_T-1} \left(\mathbf{H}_d^{(m,n)\{\mathcal{O}_g, \mathcal{P}^{(m)}\}} \mathbf{p}^{(m)} + \mathbf{H}_d^{(m,n)\{\mathcal{O}_g, \mathcal{D}^{(m)}\}} \mathbf{d}^{(m)} \right) + \mathbf{z}^{(n)\{\mathcal{O}_g\}}. \quad (9)$$

where $\mathbf{H}_d^{(m,n)\{\mathcal{O}_g, \mathcal{P}^{(m)}\}}$ and $\mathbf{H}_d^{(m,n)\{\mathcal{O}_g, \mathcal{D}^{(m)}\}}$ represent submatrices of $\mathbf{H}_d^{(m,n)}$, which are schematically depicted in Fig. 1. As a consequence of the full matrix $\mathbf{H}_d^{(m,n)}$, we can see from (9) that $\mathbf{y}^{(n)\{\mathcal{O}_g\}}$ does not only depend on $\mathbf{p}_g^{(m)}$, but also on the data $\mathbf{d}^{(m)}$ as well as the other pilot clusters.

We repeat the relationship in (9) for each cluster $g = 0, \dots, G-1$, and for each receive antenna $n = 0, \dots, N_R - 1$, and stack the results in one vector $\check{\mathbf{y}} = [\mathbf{y}^{(0)\{\mathcal{O}\}T}, \dots, \mathbf{y}^{(N_R-1)\{\mathcal{O}\}T}]^T$, with $\mathcal{O} = \mathcal{O}_0 \cup \dots \cup \mathcal{O}_{G-1}$. This leads to

$$\check{\mathbf{y}} = \underbrace{\mathcal{D}\{\mathbf{A}^{(0)}, \dots, \mathbf{A}^{(N_R-1)}\}}_{\mathbf{A}} \mathbf{c} + \mathbf{i} + \mathbf{z}, \quad (10)$$

where \mathbf{z} is similarly defined as \mathbf{y} , and $\mathbf{c} = [\mathbf{c}_0^{(0,0)T}, \dots, \mathbf{c}_Q^{(0,0)T}, \dots, \mathbf{c}_Q^{(N_T-1, N_R-1)T}]^T$. From (5), it can be shown that each diagonal block of \mathbf{A} can be expressed as

$$\mathbf{A}^{(n)} = [\mathbf{A}_c^{(0,n)}, \dots, \mathbf{A}_c^{(N_T-1,n)}] \mathcal{D}\{\mathbf{A}_d^{(0)}, \dots, \mathbf{A}_d^{(N_T-1)}\}, \quad (11)$$

with

$$\begin{aligned} \mathbf{A}_c^{(m,n)} &= \mathbf{W}_K^{\{\mathcal{O},:\}} [\mathcal{D}\{\mathbf{u}_0\}, \dots, \mathcal{D}\{\mathbf{u}_Q\}] (\mathbf{I}_{Q+1} \otimes \mathbf{W}_K^{\{\mathcal{P}^{(m)},:\}H}), \\ \mathbf{A}_d^{(m)} &= \mathbf{I}_{Q+1} \otimes \mathcal{D}\{\mathbf{p}^{(m)}\} \mathbf{V}_L^{\{\mathcal{P}^{(m)},:\}}. \end{aligned} \quad (12)$$

The interference due to the data is represented in (10) by \mathbf{i} , which can be expressed as $\mathbf{i} = \mathbf{B}\mathbf{d}$ with

$$\begin{aligned} \mathbf{B} &= \begin{bmatrix} \mathbf{H}_d^{(0,0)\{\mathcal{O}, \mathcal{D}^{(0)}\}} & \dots & \mathbf{H}_d^{(N_T-1,0)\{\mathcal{O}, \mathcal{D}^{(N_T-1)}\}} \\ \vdots & \ddots & \vdots \\ \mathbf{H}_d^{(0, N_R-1)\{\mathcal{O}, \mathcal{D}^{(0)}\}} & \dots & \mathbf{H}_d^{(N_T-1, N_R-1)\{\mathcal{O}, \mathcal{D}^{(N_T-1)}\}} \end{bmatrix}, \\ \mathbf{d} &= [\mathbf{d}^{(0)T}, \dots, \mathbf{d}^{(N_T-1)T}]^T. \end{aligned} \quad (13)$$

A detailed derivation of (11)-(13) for the SISO case can be found in [6]. The extension to the MIMO case is rather straightforward.

A.2 Best Linear Unbiased Estimator Based on a Single OFDM Symbol

Based on (10), \mathbf{c} can be estimated by diverse channel estimators. Due to space restrictions, this paper will not list all the possible channel estimators, but will only focus on the BLUE.

The BLUE is a compromise between the linear minimum mean-square error (LMMSE) and the least-squares (LS) estimator: it treats \mathbf{c} as a deterministic variable, thus avoiding a possible error in calculating channel statistics, that are necessary for the LMMSE estimator; at the same time, it leverages the statistics of the data symbols and noise, which are easier to attain, such that the interference and the noise can still be better suppressed than with the LS estimator. Simulation results in [6] show that the BLUE is able to yield a performance close to that of the LMMSE estimator, even if the latter is equipped with perfect knowledge of the channel statistics.

In a nutshell, the BLUE uses a linear filter \mathbf{F} to produce an unbiased estimate $\hat{\mathbf{c}} = \mathbf{F}\mathbf{y}$, whose mean-square error (MSE) w.r.t. \mathbf{c} is minimized:

$$\mathbf{F}_{\text{BLUE}} = \arg \min_{\{\mathbf{F}\}} \mathbb{E}_{\mathbf{d}, \mathbf{z}} \{\|\mathbf{F}\mathbf{y} - \mathbf{c}\|^2\}, \text{ s.t. } \mathbb{E}_{\mathbf{d}, \mathbf{z}} \{\mathbf{F}\mathbf{y}\} = \mathbf{c}.$$

Let us assume that the data sent from all the transmit antennas are zero-mean white with variance σ_d^2 , and the noise perceived by all the receive antennas is zero-mean white with variance σ_z^2 . By comprising the interference \mathbf{i} and noise \mathbf{z} in a single disturbance term, we can follow the steps given in [12, Appendix 6B] to derive the BLUE as:

$$\mathbf{F}_{\text{BLUE}} = (\mathbf{A}^H \mathbf{R}^{-1}(\mathbf{c}) \mathbf{A})^{-1} \mathbf{A}^H \mathbf{R}^{-1}(\mathbf{c}), \quad (14)$$

where $\mathbf{R}(\mathbf{c})$ denotes the covariance matrix of the disturbance with \mathbf{c} taken as a deterministic variable. Conform the assumptions on the data and noise statistics and taking (13) into account, we can show that:

$$\begin{aligned} \mathbf{R}(\mathbf{c}) &= \mathbb{E}_{\mathbf{d}} \{\mathbf{i}\mathbf{i}^H\} + \mathbb{E}_{\mathbf{z}} \{\mathbf{z}\mathbf{z}^H\}, \\ &= \sigma_d^2 \mathbf{B}\mathbf{B}^H + \sigma_z^2 \mathbf{I}_{N_{\text{R}}|\mathcal{O}|}. \end{aligned} \quad (15)$$

Clearly, (14) cannot be resolved in closed-form since the computation of $\mathbf{R}(\mathbf{c})$ entails the knowledge of \mathbf{c} itself (contained in \mathbf{B}). As a remedy, we apply a recursive approach. Suppose at the k th iteration, an estimate of \mathbf{c} has been attained, which is denoted as $\hat{\mathbf{c}}^{[k]}$. Next, we utilize this intermediate estimate to update the covariance matrix $\mathbf{R}(\mathbf{c})$, which in turn is used to

produce the BLUE for the subsequent iteration and so on:

$$\begin{aligned}\mathbf{F}_{\text{BLUE}}^{[k+1]} &= (\mathbf{A}^H \mathbf{R}^{-1}(\hat{\mathbf{c}}^{[k]}) \mathbf{A})^{-1} \mathbf{A}^H \mathbf{R}^{-1}(\hat{\mathbf{c}}^{[k]}), \\ \hat{\mathbf{c}}^{[k+1]} &= \mathbf{F}_{\text{BLUE}}^{[k+1]} \mathbf{y}.\end{aligned}\quad (16)$$

Note that a similar idea is adopted in [13] though in a different context. To initialize the iteration, we can set $\hat{\mathbf{c}}^{[0]} = \mathbf{0}$, which results in the following expression for the first iteration:

$$\mathbf{F}_{\text{MLE}} = \mathbf{F}_{\text{BLUE}}^{[1]} = (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H. \quad (17)$$

The above expression is actually the maximum likelihood estimator [12] that is obtained by ignoring the interference \mathbf{i} .

Using the symbol $\Gamma^{[k]}$ to denote the normalized difference in energy between the estimates from the present and previous iterations:

$$\Gamma^{[k]} := \frac{|\mathbf{c}^{[k]} - \mathbf{c}^{[k-1]}|^2}{|\mathbf{c}^{[k-1]}|^2}, \quad (18)$$

we can halt the iterative BLUE if $\Gamma^{[k]}$ is smaller than a pre-defined value or the number of iterations K is higher than a pre-defined value.

In the previous section, we have mentioned that a different choice of ℓ in (9) will have an impact on the channel estimator. For the BLUE in the SISO scenario, it is shown in [6] that the best performance is attained when the whole OFDM symbol is employed for channel estimation.

III-B. Multiple OFDM Symbols

In the previous section, the channel is estimated for each block separately. To improve the performance, we will exploit more observation samples in this section. It is nonetheless noteworthy that in the context of time-varying channels, the channel coherence time is rather short, which means that we cannot utilize an infinite number of OFDM symbols to enhance the estimation precision.

Considering J consecutive OFDM symbols, out of which there are V OFDM symbols carrying pilots, we use the symbol \mathcal{V} to denote the set that contains the indexes of all the pilot OFDM symbols:

$$\mathcal{V} = \{j_0, \dots, j_{V-1}\}, \quad (19)$$

where j_v stands for the position of the v th pilot OFDM symbol. Further, the symbol $\mathcal{P}^{(m)}[j_v]$, as analogously introduced in the previous section, represents the set of pilot subcarriers within

the v th pilot OFDM symbol that is used by the m th transmit antenna. Similar extensions hold for $\mathcal{D}^{(m)}[j_v]$, $\mathcal{O}^{(m)}[j_v]$ and $\mathcal{O}[j_v]$. An interesting topic when utilizing multiple OFDM symbols is how to distribute the pilots along the time as well as frequency axis. To differentiate between various pilot patterns, let us borrow the terms used in [14] to categorize two pilot placement scenarios¹.

Comb-Type. This scheme is adopted in [16–18], in which pilots occupy only a fraction of the subcarriers, but such pilots are carried by each OFDM symbol. In other words, we have $|\mathcal{V}| = J$ and $|\mathcal{P}^{(m)}[j_v]| < K$. This is equivalent to the pilot scheme that we discussed in the previous section, but now extended to multiple OFDM symbols. An example of the comb-type scheme with two transmit antennas is sketched in the left and middle plot of Fig. 2.

Block-Type. This scheme is considered in [19–21], in which the pilots occupy the entire OFDM symbol, and such pilot OFDM symbols are interleaved along the time axis with pure data OFDM symbols. In mathematics, $|\mathcal{V}| < J$ and $|\mathcal{P}^{(m)}[j_v]| = K$. An example of the block-type scheme with two transmit antennas is sketched in the right plot of Fig. 2.

B.1 Data Model and BEM Based on Multiple OFDM Symbols

The biggest difference between the multiple OFDM symbol and the single OFDM symbol case is that we need here to use a larger BEM to approximate the time-varying channel that spans several OFDM symbol intervals. More specifically, we need to model $J(K + L)$ consecutive samples of the l th channel tap between the m th transmit antenna and the n th receive antenna, i.e., $[h_{0,l}^{(m,n)}, \dots, h_{(J-1)(K+L)-1,l}^{(m,n)}]^T$ as

$$\begin{bmatrix} h_{0,l}^{(m,n)} \\ \vdots \\ h_{(J-1)(K+L)-1,l}^{(m,n)} \end{bmatrix} = \underbrace{[\mathbf{u}_0 \cdots \mathbf{u}_Q]}_{\mathbf{U}} \begin{bmatrix} c_{0,l}^{(m,n)} \\ \vdots \\ c_{Q,l}^{(m,n)} \end{bmatrix}. \quad (20)$$

Here, \mathbf{u}_q stands for the q th BEM function that spans $J(K + L)$ time instants, and $c_{q,l}^{(m,n)}$ for the corresponding BEM coefficient. In comparison with (3), we design the CE-BEM as

$$\mathbf{u}_q := [1, e^{-j\frac{2\pi}{\kappa J(K+L)}q}, \dots, e^{-j\frac{2\pi}{\kappa J(K+L)}q(J(K+L)-1)}]^T. \quad (21)$$

¹A third pilot placement scenario, referred to as the mixed-type, is considered in [8, 15]. It can be succinctly described by $|\mathcal{V}| < J$ and $|\mathcal{P}^{(m)}[j_v]| < K$. Because the channel estimation and identifiability condition based on this pilot scheme will be exactly identical to the comb-type, we will not treat it separately in this paper.

Hence, for the j th OFDM symbol in particular, we obtain

$$\begin{bmatrix} h_{j(K+L)+L,l}^{(m,n)} \\ \vdots \\ h_{(j+1)(K+L)-1,l}^{(m,n)} \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{u}_0[j] \cdots \mathbf{u}_Q[j] \end{bmatrix}}_{\mathbf{U}[j]} \begin{bmatrix} c_{0,l}^{(m,n)} \\ \vdots \\ c_{Q,l}^{(m,n)} \end{bmatrix}, \quad (22)$$

where $\mathbf{u}_q[j]$ is a selection of rows $j(K+L)+L$ through $(j+1)(K+L)-1$ from \mathbf{u}_q . By defining the BEM in this way, the resulting channel matrix of the j th OFDM symbol in the frequency domain will admit a slightly different expression than in (5) defined for the single OFDM symbol case:

$$\mathbf{H}_d^{(m,n)}[j] = \sum_{q=0}^Q \mathbf{W}_K \mathcal{D}\{\mathbf{u}_q[j]\} \mathbf{W}_K^H \mathcal{D}\{\mathbf{V}_L \mathbf{c}_q^{(m,n)}\}. \quad (23)$$

where $\mathbf{c}_q^{(m,n)} := [c_{q,0}^{(m,n)}, \dots, c_{q,L}^{(m,n)}]^T$. Note that in (23), each OFDM symbol is associated with a different BEM sequence $\mathbf{u}_q[j]$, but with common BEM coefficients $\mathbf{c}_q^{(m,n)}$. This is in contrast to (5), where each OFDM symbol is associated with a common BEM, but with different BEM coefficients.

For each pilot OFDM symbol, we will follow the same strategy for choosing the observation samples as in the single OFDM symbol case. By iterating the I/O relationship in (10) for each pilot OFDM symbol $j_v = j_0, \dots, j_{V-1}$, and stacking the results in one vector, we obtain

$$\tilde{\mathbf{y}} := [\mathbf{y}^{(0)\{\mathcal{O}[j_0]\}^T}[j_0], \dots, \mathbf{y}^{(N_R-1)\{\mathcal{O}[j_0]\}^T}[j_0], \dots, \mathbf{y}^{(0)\{\mathcal{O}[j_{V-1}]\}^T}[j_{V-1}], \dots, \mathbf{y}^{(N_R-1)\{\mathcal{O}[j_{V-1}]\}^T}[j_{V-1}]]^T \quad (24)$$

which can also be concisely expressed as

$$\tilde{\mathbf{y}} = \underbrace{[\mathbf{A}^T[j_0], \dots, \mathbf{A}^T[j_{V-1}]]^T}_{\tilde{\mathbf{A}}} \mathbf{c} + \tilde{\mathbf{i}} + \tilde{\mathbf{z}}, \quad (25)$$

where $\mathbf{A}[j_v]$ is defined as in (11) with the OFDM symbol index added, and $\tilde{\mathbf{i}}$ and $\tilde{\mathbf{z}}$ are similarly defined as $\tilde{\mathbf{y}}$. Further, the interference term $\tilde{\mathbf{i}}$ in (25) can be written as

$$\begin{aligned} \tilde{\mathbf{i}} &:= [\mathbf{i}^T[j_0], \dots, \mathbf{i}^T[j_{V-1}]]^T, \\ &= \begin{bmatrix} \mathbf{B}[j_0] & & \\ & \ddots & \\ & & \mathbf{B}[j_{V-1}] \end{bmatrix} \begin{bmatrix} \mathbf{d}[j_0] \\ \vdots \\ \mathbf{d}[j_{V-1}] \end{bmatrix}, \end{aligned} \quad (26)$$

where $\mathbf{B}[j_v]$ and $\mathbf{d}[j_v]$ are defined as in (13) with the OFDM symbol index added.

B.2 Best Linear Unbiased Estimator Based on Multiple OFDM Symbols

We notice that (25) admits an expression analogous to (10). Hence, it is not difficult to understand that a similar iterative BLUE can be applied for channel estimation based on multiple pilot OFDM symbols. The BLUE at the $(k + 1)$ st iteration can thus be expressed as

$$\tilde{\mathbf{F}}_{\text{BLUE}}^{[k+1]} = (\tilde{\mathbf{A}}^H \tilde{\mathbf{R}}^{-1}(\hat{\mathbf{c}}^{[k]}) \tilde{\mathbf{A}})^{-1} \tilde{\mathbf{A}}^H \tilde{\mathbf{R}}^{-1}(\hat{\mathbf{c}}^{[k]}), \quad (27)$$

where $\tilde{\mathbf{R}}(\mathbf{c})$ denotes the covariance matrix of the disturbance based on multiple pilot OFDM symbols. Assuming further that the data and noise from different OFDM symbol intervals are uncorrelated, we can show that

$$\begin{aligned} \tilde{\mathbf{R}}(\mathbf{c}) &= \mathbf{E}_{\mathbf{d}[j_0], \dots, \mathbf{d}[j_{V-1}]} \{\tilde{\mathbf{ii}}^H\} + \mathbf{E}_{\tilde{\mathbf{z}}} \{\tilde{\mathbf{z}}\tilde{\mathbf{z}}^H\}, \\ &= \mathcal{D}\{\mathbf{R}[j_0], \dots, \mathbf{R}[j_{V-1}]\}, \end{aligned} \quad (28)$$

where $\mathbf{R}[j_v]$ is defined as in (15) with the OFDM symbol index added.

The above derivations can be directly applied for the comb-type pilots. For the block-type pilots which occupy the entire OFDM symbol, the corresponding channel estimators are not subject to data interference, i.e., $\tilde{\mathbf{i}} = \mathbf{0}$. In this case, the BLUE in (27) reduces to an LS estimator:

$$\tilde{\mathbf{F}}_{\text{BLUE}} = (\tilde{\mathbf{A}}^H \tilde{\mathbf{A}})^{-1} \tilde{\mathbf{A}}^H, \quad (29)$$

which can be attained in just one shot.

IV. CHANNEL IDENTIFIABILITY

In this paper, we define channel identifiability in terms of the uniqueness of the BLUE. From (16) and (27), we understand that the BLUE is unique when \mathbf{A} or $\tilde{\mathbf{A}}$ is of full column-rank, and \mathbf{R} or $\tilde{\mathbf{R}}$ is non-singular.

Normally speaking, the non-singularity of \mathbf{R} or $\tilde{\mathbf{R}}$ can be easily satisfied in a noisy channel. In contrast, the rank condition of \mathbf{A} or $\tilde{\mathbf{A}}$ is often difficult to examine, because its composition depends on the choice of the BEM and the pilot structure. Especially for the latter, it turns out to be very hard to give an analytical formulation for a general pilot structure. In this paper, we will adopt a specific pilot structure for each pilot OFDM symbol, which is similar to the frequency-domain Kronecker Delta (FDKD) scheme proposed in [7]. Note that for a general BEM assumption as taken in [6], the FDKD scheme always yields a good performance experimentally.

The basic pilot structure adopted in this paper can be summarized as follows:

Pilot Design Criterion 1: We group the pilots from one transmit antenna into G (cyclically) equi-distant clusters, where each cluster contains only one non-zero pilot. The entire set of pilots sent by the m th transmit antenna during the v th pilot OFDM symbol can therefore be expressed in a Kronecker form as

$$\mathbf{p}^{(m)}[j_v] = \bar{\mathbf{p}}^{(m)}[j_v] \otimes [\mathbf{0}_{1 \times (\Delta^{(m)}[j_v]-1)}, \mathbf{1}, \mathbf{0}_{1 \times (P-\Delta^{(m)}[j_v])}]^T, \quad (30)$$

where $\bar{\mathbf{p}}^{(m)}[j_v]$ contains all the non-zero pilots sent by the m th transmit antenna during the v th pilot OFDM symbol, and $\Delta^{(m)}[j_v]$ contains the position of the non-zero pilot within the cluster.

Further, the following assumption is adopted throughout the remainder of the paper.

Assumption 1: All the subcarriers of the pilot OFDM symbol will be used for channel estimation, i.e.,

$$|\mathcal{O}^{(m)}[j_v]| = K. \quad (31)$$

This assumption is shown in [6] to maximize the performance of the BLUE. In addition, it will greatly simplify the derivation of the channel identifiability conditions.

As in the previous sections, in order to derive the channel identifiability conditions, we find it instrumental to first explore the rank condition on \mathbf{A} for the single OFDM symbol case, and then extend the results to multiple pilot OFDM symbols.

IV-A. Single OFDM Symbol

The full column-rank condition of \mathbf{A} is related to the full column-rank condition of $\mathbf{A}^{(n)}$ defined in (10) for an arbitrary receive antenna n . Hence, we need to examine if

$$\text{Rank}\{\mathbf{A}^{(n)}\} = N_T(L+1)(Q+1). \quad (32)$$

Following Pilot Design Criterion 1, [7] shows conditions to ensure that the columns of $\mathbf{A}^{(n)}$ are orthonormal under a (C)CE-BEM assumption. However, these conditions are not suitable for an (O)CE-BEM assumption as adopted in this paper, and we need to impose more restrictions, especially on the pilot design across the transmit antennas. They are summarized in the following theorem (see Appendix A for a proof).

Theorem 1: With the pilots following Pilot Design Criterion 1, the channel will be identifiable under an (O)CE-BEM assumption and Assumption 1 if

$$\frac{K}{N_T(Q+1)} \geq G \geq L+1, \quad (33)$$

and

$$|\mu^{(m')} - \mu^{(m)}| > \frac{KQ}{\kappa(K+L)} \text{ for } m' \neq m, \quad (34)$$

where $\mu^{(m)}$ denotes the position of the first non-zero pilot sent by the m th transmit antenna.

The following remarks are in order at this stage.

Remark 1: For the ‘optimal’ pilot structure proposed in [7], each OFDM symbol contains $G = L + 1$ pilot clusters, with each pilot cluster satisfying (up to a scale)

$$\mathbf{p}_{\text{opt}}^{(m)} = \mathbf{1}_{(L+1) \times 1} \otimes [\mathbf{0}_{1 \times [m(Q+1)-1]}, \mathbf{1}, \mathbf{0}_{1 \times [(N_T+1-m)(Q+1)-1]}]^T. \quad (35)$$

Such a pilot structure complies with (33) and (34) with a (C)CE-BEM assumption, i.e., $\kappa = \frac{K}{K+L}$.

We observe in (35) that the FDKD pilot structure contains a certain number of zeros, which are not specified in Theorem 1. These zeros are beneficial to combat the ICI, but not necessary for the rank condition. Later on, we will show that the total number of zeros within the pilot cluster plays a more significant role at high SNR where the ICI becomes more pronounced.

Remark 2: Viewing a time-invariant channel as a special case of a time-varying channel with a trivial $Q = 0$, we can establish the relationship between the conditions given in (33) and (34), and the conditions given for time-invariant channels. For instance, the pilot structure given in [9] requires the number of non-zero pilots per transmit antenna to be no fewer than $L + 1$. Further, the non-zero pilots from different transmit antennas must occupy different subcarriers, i.e., $\mu^{(m')} - \mu^{(m)} > 0$ for $m' \neq m$.

IV-B. Multiple OFDM Symbols

In many practical situations, Theorem 1 can be hard to satisfy due to practical constraints. For instance, if the Doppler spread and/or the delay spread of the channel are large, the lower- and upper-bound in (33) will approach each other, making it harder to find a suitable G . Fortunately, these constraints can be loosened by employing multiple pilot OFDM symbols.

One important issue of channel estimation based on multiple pilot OFDM symbols is how to distribute the pilots along the time axis. Prior to proceeding, let us introduce two possible schemes.

Pilot Design Criterion 2: The positions of the equi-distant pilots sent by the same transmit antenna are disparate for each OFDM symbol, i.e.,

$$\mathcal{P}^{(m)}[j_v] \neq \mathcal{P}^{(m)}[j_{v'}] \text{ for } v \neq v'. \quad (36)$$

Adopting the above design criterion leads to the following theorem.

Theorem 2: With the pilots following Pilot Design Criterion 1 and Pilot Design Criterion 2, then for the n th receive antenna, the corresponding $\tilde{\mathbf{A}}^{(n)} = [\mathbf{A}^{(n)T}[j_0], \dots, \mathbf{A}^{(n)T}[j_{V-1}]]^T$ will have a full column-rank under an (O)CE-BEM assumption and Assumption 1 if

$$\frac{K}{N_T(Q+1)} \geq G \geq \frac{L+1}{V}, \quad (37)$$

and

$$|\mu^{(m')} - \mu^{(m)}| > \frac{KQ}{\kappa V(K+L)} \text{ for } m' \neq m. \quad (38)$$

The proof is given in Appendix B.

Remark 3: We observe here again that the right inequality in (37) is identical to the channel identifiability condition in [9] for the time-invariant MIMO channel based on multiple OFDM symbols.

Remark 4: For realistic system parameters, $\frac{KQ}{\kappa V(K+L)} < 1$ in most cases. From (38), it is hence sufficient if $\mu^{(m')} \neq \mu^{(m)}$ for $m' \neq m$: this implies that the transmitter can be transparent to the oversampling factor used by the receiver.

An alternative way of designing the pilots is given by the following construction.

Pilot Design Criterion 3: The values and positions of the equi-distant pilots sent by the same transmit antenna are identical for each OFDM symbol, which implies that

$$\begin{aligned} \bar{\mathbf{p}}^{(m)}[j_0] &= \dots = \bar{\mathbf{p}}^{(m)}[j_{V-1}], \\ \mathcal{P}^{(m)}[j_0] &= \dots = \mathcal{P}^{(m)}[j_{V-1}]. \end{aligned} \quad (39)$$

Adopting the above design criterion leads to the following theorem.

Theorem 3: With the pilots following Pilot Design Criterion 1 and Pilot Design Criterion 3, then for the n th receive antenna, the corresponding $\tilde{\mathbf{A}}^{(n)} = [\mathbf{A}^{(n)T}[j_0], \dots, \mathbf{A}^{(n)T}[j_{V-1}]]^T$ will have a full column-rank under an (O)CE-BEM assumption and Assumption 1 if

$$\begin{aligned} \frac{K}{N_T} &\geq G \geq L+1, \\ V &\geq Q+1, \end{aligned} \quad (40)$$

and

$$|\mu^{(m)} - \mu^{(m')}| > 0 \text{ for } m' \neq m. \quad (41)$$

The proof is given in Appendix C.

Remark 5: Theorem 3 enables the transmitter to be completely transparent to the choice of the oversampling factor at the receiver.

If there is only one transmit antenna, the conditions given in Theorem 3 can be relaxed as stated in the following corollary.

Corollary 1: With the pilots following Pilot Design Criterion 1 and Pilot Design Criterion 3, if there is only one transmit antenna, the matrix $\tilde{\mathbf{A}}^{(n)} = [\mathbf{A}^{(n)T}[j_0], \dots, \mathbf{A}^{(n)T}[j_{V-1}]]^T$ will have full column-rank under an (O)CE-BEM assumption and Assumption 1 if

$$\frac{KV}{(Q+1)} \geq G \geq L+1. \quad (42)$$

The proof is given in the last part of Appendix C. This property has been explored in [22] where a SISO scenario is considered.

V. SIMULATIONS AND DISCUSSIONS

For the simulations, we generate time-varying channels conform Jakes' Doppler profile [23] using the channel generator given in [24]. The channel taps are assumed to be mutually uncorrelated with a variance of $\sigma_l^2 = 1/\sqrt{L+1}$. The variation of the channel is characterized by the normalized Doppler spread $\nu_D = f_c v/c$, where f_c is the carrier frequency; v is the speed of the vehicle parallel to the direction between the transmitter and the receiver, and c is the speed of light.

We consider an OFDM system with 64 subcarriers. The pilots and data symbols are multiplexed in the frequency domain by occupying different subcarriers. The data symbols are modulated by quadrature phase-shift keying (QPSK). Further, we set the average power of the pilots to be the same as the average power of the data symbols.

To qualify the channel estimation performance, we use the normalized mean square error (NMSE), which is defined as

$$\text{NMSE} = \frac{1}{N_T N_R K J} \sum_{j=0}^{J-1} \sum_{m=1}^{N_T} \sum_{n=1}^{N_R} \sum_{l=0}^L \left\| \begin{bmatrix} h_{j(K+L),l}^{(m,n)} \\ \vdots \\ h_{j(K+L)+K-1,l}^{(m,n)} \end{bmatrix} - \mathbf{U}[j] \begin{bmatrix} c_{0,l}^{(m,n)}[j] \\ \vdots \\ c_{Q,l}^{(m,n)}[j] \end{bmatrix} \right\|^2. \quad (43)$$

Note that in the above criterion the true channel $h_{k,l}^{(m,n)}$ is used, which implies that we actually take also the BEM modeling error into account.

For all the numerical examples below, we adopt the stop criterion that halts the iterative BLUE if either $\Gamma^{[k]}$, which is defined in (18) as the normalized difference in energy between the current

TABLE I
PILOT STRUCTURE

	G	$P + 1$	V_a	V_b	J
Comb-type I	4	8	6	1	6
Comb-type II	16	2	1	6	6
Block-type	16	4	1	3	6

estimate and the estimate from the previous iteration, is smaller than 10^{-6} or the number of iterations K is higher than 30.

Study Case 1: Single OFDM Symbol. The pilots used in this study case are grouped in $G = 4$ clusters, each containing seven zero pilots and one non-zero pilot, i.e., $P + 1 = 8$. The non-zero pilot is located within the pilot cluster at the $[3(m + 1) - 1]$ st position, where m corresponds to the transmit antenna index. Because we will use an (O)CE-BEM with $Q = 2$ and $\kappa = 4$ to fit a slower time-varying channel ($\nu_D = 8e^{-4}$) and a faster time-varying channel ($\nu_D = 4e^{-3}$), this pilot structure satisfies the ‘optimal’ pilot structure in (35) as well as Theorem 1 for a channel of length $L = 3$, which is assumed for this study case. The performance of the BLUE is given in Fig. 3. We observe that the performance degrades when the number of transmit antennas is increased from one to two. But more interestingly, this performance degradation can be alleviated by using more receive antennas, especially for the faster channels (the right plot). We will discuss this effect in more detail later on.

In the following study cases, we will focus on pilots carried by multiple OFDM symbols. We compare three different pilot structures as summarized in Table I, where we use V_a to denote the number of pilot OFDM symbols that satisfy Pilot Design Criterion 2, and V_b to denote the number of pilot OFDM symbols that satisfy Pilot Design Criterion 3. The positions of the zero and non-zero pilots and data symbols of the three pilot structures are schematically given in Fig. 2. Note also that the ‘optimal’ pilot structure in (35) is carried by all the OFDM symbols in Comb-type I.

Study Case 2: Short Channels. In this study case, we again examine channels with $\nu_D = 8e^{-4}$ and $\nu_D = 4e^{-3}$. To fit the time variation of the channel for $J = 6$ consecutive OFDM symbols, we use at the receiver an (O)CE-BEM with $Q = 2$ and $\kappa = 3$ if $\nu_D = 8e^{-4}$, and with

$Q = 4$ and $\kappa = 1.5$ if $\nu_D = 4e^{-3}$. Further, we focus on a channel with length $L = 3$, and compare the performance of the pilot structures listed in Table I. The results are given in Fig. 4. We observe that Comb-type I renders a much better performance than the other two, especially when the channel varies faster (the right plot). This can be attributed to the zeros in the pilot cluster that protect the non-zero pilots from the interference much more effectively.

Again, we observe that the channel estimation performance degrades with more transmit antennas, but improves with more receiver antennas especially at high SNR. In contrast, this does not happen to the Block-type scheme. We understand that the interference induced by the Doppler spread to the channel estimator becomes the dominant nuisance at high SNR. At the same time, this interference is a function of the transmitted data, and hence strongly correlated among different receive antennas. The BLUE is able to exploit this correlation to combat the interference better. The following heuristic analysis enables a better insight into this effect.

It can be shown that the variance of the BLUE equals the trace of $(\tilde{\mathbf{A}}^H \tilde{\mathbf{R}}^{-1} \tilde{\mathbf{A}})^{-1}$, where $\tilde{\mathbf{R}}$, as defined in (28), expresses the correlation of the interference as well as the noise. Because $\tilde{\mathbf{R}}$ is a block-diagonal matrix with $\mathbf{R}[j_v]$ as its v th diagonal block, we focus further on $\mathbf{R}[j_v]$. From its definition in (15), and by applying the matrix inversion lemma in [25], its inverse can be written as

$$\begin{aligned} \mathbf{R}^{-1}[j_v] &= \sigma_z^{-2} \mathbf{I} - \sigma_z^{-2} \mathbf{B}[j_v] \left(\frac{\sigma_z^2}{\sigma_d^2} \mathbf{I} + \mathbf{B}^H[j_v] \mathbf{B}[j_v] \right)^{-1} \mathbf{B}^H[j_v], \\ &\longrightarrow \sigma_z^{-2} \mathbf{I} - \sigma_z^{-2} \mathbf{B}[j_v] \left(\mathbf{B}^H[j_v] \mathbf{B}[j_v] \right)^{-1} \mathbf{B}^H[j_v], \end{aligned} \quad (44)$$

where the last is attained at high SNR when $\frac{\sigma_d^2}{\sigma_z^2} \rightarrow \infty$. The presence of $\mathbf{B}[j_v]$ in (44) is associated with the interference. We observe that the $N_R K \times N_R K$ matrix $\mathbf{R}^{-1}[j_v]$ lies in the noise subspace of $\mathbf{B}[j_v]$, i.e., $\mathbf{R}^{-1}[j_v] \mathbf{B}[j_v] = \mathbf{0}$. Suppose the $N_R K \times N_T (K - G(P + 1))$ matrix $\mathbf{B}[j_v]$ has full column-rank $N_T (K - G(P + 1))$. We then have

$$\text{Rank}\{\tilde{\mathbf{R}}^{-1}\} = V \cdot \text{Rank}\{\mathbf{R}^{-1}[j_v]\} = V N_R K - V N_T (K - G(P + 1)). \quad (45)$$

The above suggests that the rank of $\tilde{\mathbf{R}}^{-1}$ increases with the number of receive antennas, the number of pilot OFDM symbols as well as the number of pilots within the OFDM symbol, but decreases with the number of transmit antennas. A higher rank of $\tilde{\mathbf{R}}^{-1}$ is beneficial to the condition of the matrix $\tilde{\mathbf{A}}^H \tilde{\mathbf{R}}^{-1} \tilde{\mathbf{A}}$, which is in turn related to the trace of $(\tilde{\mathbf{A}}^H \tilde{\mathbf{R}}^{-1} \tilde{\mathbf{A}})^{-1}$. Following such a reasoning, it is not difficult to understand that increasing the number of receive

antennas is beneficial to the performance just as increasing the number of pilots or decreasing the number of transmit antennas. To the best of our knowledge, this effect of the number of receive antennas on the channel estimation performance is not widely recognized. The main reason is that most works are based on a scenario where the interference is absent at the receiver, e.g., for time-invariant channels, or in the case of the Block-type scheme, where the pilots occupy the whole OFDM symbol and there is no interference either.

Note that the rank of $\tilde{\mathbf{R}}^{-1}$ also increases with the number of pilot OFDM symbols. Comparing Fig. 4 with Fig. 3, we can indeed observe a performance improvement. However, for faster fading channels, multiple OFDM symbols work only better at low-to-moderate SNR, but suffer from a noise floor at high SNR, where the BEM modeling error plays a dominant role. The BEM modelling error will become larger if more OFDM symbols are considered and/or the channel varies faster. Increasing the BEM order Q can enhance the BEM modelling performance at the penalty that more channel unknowns need to be estimated. An alternative is not to estimate the channel of all the OFDM symbols, but only the middle part, e.g., the 3rd and 4th symbols. This means that the channel estimator will work like an overlapping sliding window, an approach that is adopted in [26].

Study Case 3: Long Channels. We examine now a much longer channel with length $L = 15$, for which the results are given in Fig. 5. Note that in this figure, we do not list the performance of Comb-type I because it failed in the simulation. We will explore the reason later on. Fig. 5 shows that Comb-type II performs in general better than the Block-type, especially when the channel varies faster. Note that the channels where the data are located are not estimated directly in the Block-type scheme, but actually result from an implicit interpolation of the channels estimated at the pilot OFDM symbols. Such an interpolation error gives rise to a performance penalty.

The channel equalization performance based on the estimated channels is given in Fig. 6, where the bit error rate (BER) is used as the performance measure. The results in Fig. 6 follow similar trends as shown in Fig. 5 except for the MISO case with $N_T = 2$ and $N_R = 1$. In this case, the equalizer fails because there are more unknowns than observation samples.

Study Case 4: Why Comb-type I Fails for Long Channels. For channels with a long delay spread, it is not possible for Comb-type I to satisfy Theorem 1. Although by using multiple symbols, Theorem 2 can still be met, the condition number of $\tilde{\mathbf{A}}$ drastically increases once the

channel order $L + 1$ supersedes the number of pilot clusters G . Here, we define the condition number of a non-square matrix $\tilde{\mathbf{A}}$ as

$$\frac{\max_{0 \leq n \leq (L+1)(Q+1)-1} |\lambda_n(\tilde{\mathbf{A}})|}{\min_{0 \leq n \leq (L+1)(Q+1)-1} |\lambda_n(\tilde{\mathbf{A}})|},$$

where $\lambda_n(\tilde{\mathbf{A}})$ stands for the n th singular value of $\tilde{\mathbf{A}}$. A condition number equal to infinity means that the matrix is rank deficient. In Fig. 7, we depict the condition number as a function of the channel length for the SISO case $N_T = 1$ and $N_R = 1$, where one can observe that the condition number of $\tilde{\mathbf{A}}$ rapidly increases for Comb-type I, once the channel length becomes larger than the number of pilot clusters. In contrast, with a sufficient number of pilot clusters, the condition number for Comb-type II and Block-type stays constant.

The condition number of $\tilde{\mathbf{A}}$ is important to the variance of the BLUE, which is given before as the trace of $(\tilde{\mathbf{A}}^H \tilde{\mathbf{R}}^{-1} \tilde{\mathbf{A}})^{-1}$. An ill-conditioned $\tilde{\mathbf{A}}$ makes the BLUE more sensitive to the interference and noise. In the worst case, the adaptive BLUE will be unable to even converge to a local minimum. In Fig. 8, we show the channel estimation performance of all three pilot structures at SNR = 40dB for the SISO case, where the results exhibit the same tendency as the corresponding condition numbers.

Study Case 5: Convergence performance. As mentioned at the beginning of this section, we have adopted a stopping criterion that halts the BLUE if either $\Gamma^{[k]} < 10^{-6}$ or $K \geq 31$. The actual number of iterations is dependent on several factors such as the channel, the SNR, the number of transmit/receive antennas, etc. As an example, we show in this case the convergence performance for the comb-type II pilots over the channel with $L = 15$ and $\nu_D = 4e^{-3}$. Fig. 9 shows the average number of iterations versus SNR required for different MIMO setups, where the MISO case $N_T = 2$ and $N_R = 1$ requires the most iterations especially at high SNR. In this case, $\tilde{\mathbf{A}}$ has obviously a larger condition number than the other two cases, although it still retains full column-rank. We have learned from **Study Case 4** that when the SNR increase, the condition number of $\tilde{\mathbf{A}}$ plays a more pronounced role in the trace of $(\tilde{\mathbf{A}}^H \tilde{\mathbf{R}}^{-1} \tilde{\mathbf{A}})^{-1}$, which in turn influences the convergence behavior of the BLUE, and thus explains the large discrepancy in the number of iterations at high SNR. Fig. 10 shows the average value of $\Gamma^{[k]}$ during each iteration. With the adopted stopping criterion, we can conclude from this figure that the BLUE halts after around six iterations in most cases.

VI. CONCLUSIONS

In this paper, we have discussed how to design pilots to estimate time-varying channels in a MIMO-OFDM system. We underline that the proposed pilot design criteria can be made (almost) independent of the oversampling factor of the (O)CE-BEM such that each receiver can independently choose the best (O)CE-BEM.

We have compared the performance of three different pilot structures, all conform the proposed design criteria. By means of simulations, we have shown that

- Each pilot OFDM symbol should contain as few pilot clusters as possible provided there are more than the channel order.
- Comb-type pilots can estimate the time-varying channel better than the block-type pilots because they suffer from a smaller interpolation error.
- For comb-type pilots, it is possible to improve the channel estimation performance by employing more receive antennas, which combats the interference more effectively.

APPENDICES

A. PROOF OF THEOREM 1

Because each pilot cluster now contains only one non-zero pilot, we can express the positions of the equi-spaced non-zero pilots sent by transmit antenna m as

$$\bar{\mathcal{P}}^{(m)} = \mu^{(m)} + \{0, X, \dots, X(G-1)\}, \quad (46)$$

with $X = K/G$. Since the zero pilots have no contribution, we can rewrite $\mathbf{A}^{(n)}$, defined in (11), in the following form

$$\begin{aligned} \mathbf{A}^{(n)} &= \bar{\mathbf{A}}_c^{(n)} \bar{\mathbf{A}}_d^{(n)}, \\ \bar{\mathbf{A}}_c^{(n)} &= \mathbf{W}_K[\mathcal{D}\{\mathbf{u}_0\}, \dots, \mathcal{D}\{\mathbf{u}_Q\}][\mathbf{I}_{Q+1} \otimes \mathbf{W}_K^{\{\bar{\mathcal{P}}^{(0)},:\}H}, \dots, \mathbf{I}_{Q+1} \otimes \mathbf{W}_K^{\{\bar{\mathcal{P}}^{(N_T-1)},:\}H}], \\ \bar{\mathbf{A}}_d^{(n)} &= \begin{bmatrix} \mathbf{I}_{Q+1} \otimes \mathcal{D}\{\bar{\mathbf{p}}^{(0)}\} \mathbf{V}_L^{\{\bar{\mathcal{P}}^{(0)},:\}} & & \\ & \ddots & \\ & & \mathbf{I}_{Q+1} \otimes \mathcal{D}\{\bar{\mathbf{p}}^{(N_T-1)}\} \mathbf{V}_L^{\{\bar{\mathcal{P}}^{(N_T-1)},:\}} \end{bmatrix}. \end{aligned} \quad (47)$$

Compared to (12), we keep here only the rows/columns that correspond to the positions of the non-zero pilots, which are represented by $\bar{\mathcal{P}}^{(m)}$. In addition, we have dropped the observation sample index \mathcal{O} in the above as a result of Assumption 1.

The following two lemmas determine the rank of $\bar{\mathbf{A}}_c^{(n)}$ and $\bar{\mathbf{A}}_d^{(n)}$.

Lemma 1: If $K/[N_T(Q+1)] \geq G$, and $\mu^{(m+1)} - \mu^{(m)} > \frac{KQ}{\kappa(K+L)}$, the matrix $\bar{\mathbf{A}}_c^{(n)}$ has full column-rank $N_T G(Q+1)$.

Proof: Let us first examine the m th submatrix of $\bar{\mathbf{A}}_c^{(n)}$:

$$\bar{\mathbf{A}}_c^{(m,n)} := \mathbf{W}_K [\mathcal{D}\{\mathbf{u}_0\}, \dots, \mathcal{D}\{\mathbf{u}_Q\}] \mathbf{I}_{Q+1} \otimes \mathbf{W}_K^{\{\bar{\mathcal{P}}^{(m)}, :\}H}. \quad (48)$$

Given the property that $\bar{\mathcal{P}}^{(m)}$ contains equi-distant elements, we can express $\mathbf{W}_K^{\{\bar{\mathcal{P}}^{(m)}, :\}}$ as

$$\mathbf{W}_K^{\{\bar{\mathcal{P}}^{(m)}, :\}} = \frac{1}{\sqrt{X}} \boldsymbol{\theta}^{\mu^{(m)T}} \otimes \mathbf{W}_G \mathcal{D}\{\boldsymbol{\xi}^{\mu^{(m)}}\}, \quad (49)$$

with $\boldsymbol{\theta} := [e^{-j\frac{2\pi}{X}0}, \dots, e^{-j\frac{2\pi}{X}(X-1)}]^T$ and $\boldsymbol{\xi} := [e^{-j\frac{2\pi}{GX}0}, \dots, e^{-j\frac{2\pi}{GX}(G-1)}]^T$. The rank of $\bar{\mathbf{A}}_c^{(m,n)}$ will not change if we left-multiply it with \mathbf{W}_K^H , and right-multiply it with $(\mathbf{I}_{Q+1} \otimes \mathbf{W}_G^H)$, which leads to

$$\mathbf{W}_K^H \bar{\mathbf{A}}_c^{(m,n)} (\mathbf{I}_{Q+1} \otimes \mathbf{W}_G^H) = \frac{1}{\sqrt{X}} [\mathcal{D}\{\mathbf{u}_0\}, \dots, \mathcal{D}\{\mathbf{u}_Q\}] (\mathbf{I}_{Q+1} \otimes \boldsymbol{\theta}^{-\mu^{(m)}} \otimes \mathcal{D}\{\boldsymbol{\xi}^{-\mu^{(m)}}\}). \quad (50)$$

The above matrix is obviously a stack of $X \times (Q+1)$ sub-matrices, each being diagonal of size G . To be more specific, the (x, q) th sub-matrix $[\mathbf{W}_K^H \bar{\mathbf{A}}_c^{(m,n)} (\mathbf{I}_{Q+1} \otimes \mathbf{W}_G^H)]_{\{xG:(x+1)G-1, qG:(q+1)G-1\}}$ admits an expression as

$$[\mathbf{W}_K^H \bar{\mathbf{A}}_c^{(m,n)} (\mathbf{I}_{Q+1} \otimes \mathbf{W}_G^H)]_{\{xG:(x+1)G-1, qG:(q+1)G-1\}} = e^{j\frac{2\pi}{X}\mu^{(m)}x} \mathcal{D}\{\mathbf{u}_{q,x}\} \mathcal{D}\{\boldsymbol{\xi}^{-\mu^{(m)}}\}. \quad (51)$$

In the above, we have down-sampled the BEM sequence \mathbf{u}_q into length- G subsequences with the x th subsequence being $\mathbf{u}_{q,x} := [[\mathbf{u}_q]_{xG}, \dots, [\mathbf{u}_q]_{(x+1)G-1}]^T$ for $x = 0, \dots, X-1$.

In order to obtain a better perception of its rank, we apply an row-permutation and column-permutation on $\mathbf{W}_K^H \bar{\mathbf{A}}_c^{(m,n)} (\mathbf{I}_{Q+1} \otimes \mathbf{W}_G^H)$, which renders a new-block diagonal matrix

$$\boldsymbol{\Pi}_G \mathbf{W}_K^H \bar{\mathbf{A}}_c^{(m,n)} (\mathbf{I}_{Q+1} \otimes \mathbf{W}_G^H) \bar{\boldsymbol{\Pi}}_G^H = \begin{bmatrix} \boldsymbol{\Phi}_0^{(m)} & & \\ & \ddots & \\ & & \boldsymbol{\Phi}_{G-1}^{(m)} \end{bmatrix}, \quad (52)$$

where $\boldsymbol{\Pi}_G$ and $\bar{\boldsymbol{\Pi}}_G$ are both depth- G interleave matrices with appropriate dimensions²; and

$$\boldsymbol{\Phi}_g^{(m)} := \frac{1}{\sqrt{X}} e^{j\frac{2\pi}{GX}\mu^{(m)}g} \mathcal{D}\{\boldsymbol{\theta}^{-\mu^{(m)}}\} \begin{bmatrix} [\mathbf{u}_0]_g & \cdots & [\mathbf{u}_Q]_g \\ [\mathbf{u}_0]_{G+g} & \cdots & [\mathbf{u}_Q]_{G+g} \\ \vdots & \ddots & \vdots \\ [\mathbf{u}_0]_{(X-1)G+g} & \cdots & [\mathbf{u}_Q]_{(X-1)G+g} \end{bmatrix}. \quad (53)$$

²For instance, with a vector $\mathbf{a} = [a_0, a_1, \dots]^T$, we have $\boldsymbol{\Pi}_G \mathbf{a} = [a_0, a_G, \dots]^T$, and with a vector $\mathbf{b} = [b_0, b_1, \dots]^T$, $\mathbf{b}^T \bar{\boldsymbol{\Pi}}_G^H = [b_0, b_G, \dots]$.

With \mathbf{u}_q defined as the q th basis of the (O)CE-BEM given in (3), we can rewrite $\Phi_g^{(m)}$ after some algebra as

$$\begin{aligned} \Phi_g^{(m)} &= \frac{1}{\sqrt{X}} \bar{\Phi}^{(m)} e^{j\frac{2\pi}{K}g\mu^{(m)}} \mathcal{D}\{[e^{-j\frac{2\pi}{\kappa(K+L)}g^0}, \dots, e^{-j\frac{2\pi}{\kappa(K+L)}gQ}]^T\}, \\ \bar{\Phi}^{(m)} &:= \begin{bmatrix} e^{-j\frac{2\pi}{\kappa X(K+L)}0(K0-\mu^{(m)}\kappa(K+L))} & \dots & e^{-j\frac{2\pi}{\kappa X(K+L)}0(KQ-\mu^{(m)}\kappa(K+L))} \\ \vdots & \ddots & \vdots \\ e^{-j\frac{2\pi}{\kappa X(K+L)}(X-1)(K0-\mu^{(m)}\kappa(K+L))} & \dots & e^{-j\frac{2\pi}{\kappa X(K+L)}(X-1)(KQ-\mu^{(m)}\kappa(K+L))} \end{bmatrix}. \end{aligned} \quad (54)$$

With $\bar{\mathbf{A}}_c^{(n)} = [\bar{\mathbf{A}}_c^{(0,n)}, \dots, \bar{\mathbf{A}}_c^{(N_T-1,n)}]$, we apply the procedure from (48) until (54) on all the submatrices $\bar{\mathbf{A}}_c^{(m,n)}$ for $m = 0, \dots, N_T - 1$. It is not difficult to realize the rank of $\bar{\mathbf{A}}_c^{(n)}$ is determined by the rank of the matrix $[\bar{\Phi}^{(0)}, \dots, \bar{\Phi}^{(N_T-1)}]$ multiplied by G . It is tall if $X = K/G \geq N_T(Q + 1)$. Besides, it contains distinctive columns of a larger $\kappa X(K + L)$ -point DFT matrix if $\mu^{(m+1)}\kappa(K + L) > \mu^{(m)}\kappa(K + L) + KQ$, which is hence of full column-rank. ■

Lemma 2: If $G \geq (L + 1)$, the matrix $\bar{\mathbf{A}}_d^{(n)}$ has full column-rank $N_T(L + 1)(Q + 1)$.

Proof: Expressing $\bar{\mathbf{A}}_d^{(n)}$ in the form of $\bar{\mathbf{A}}_d^{(n)} = \mathcal{D}\{\bar{\mathbf{A}}_d^{(0,n)}, \dots, \bar{\mathbf{A}}_d^{(N_T-1,n)}\}$, we can see that the rank of its m th sub-matrix

$$\bar{\mathbf{A}}_d^{(m,n)} := \mathbf{I}_{Q+1} \otimes \mathcal{D}\{\bar{\mathbf{p}}^{(m)}\} \mathbf{V}_L^{\{\bar{\mathcal{P}}^{(m)},:\}} \quad (55)$$

is determined by the rank of $\mathbf{V}_L^{\{\bar{\mathcal{P}}^{(m)},:\}}$. The latter is a submatrix of the Vandermonde matrix \mathbf{W}_K , and is thus of full column-rank $L + 1$ if $G \geq L + 1$. In this case, the matrix $\bar{\mathbf{A}}_d^{(n)}$ is of full column-rank $N_T(L + 1)$. ■

For the matrix product $\mathbf{A}^{(n)} = \bar{\mathbf{A}}_c^{(n)} \bar{\mathbf{A}}_d^{(n)}$, the rank inequality [25] reads

$$\text{Rank}\{\bar{\mathbf{A}}_c^{(n)}\} + \text{Rank}\{\bar{\mathbf{A}}_d^{(n)}\} - N_T G(Q + 1) \leq \text{Rank}\{\mathbf{A}^{(n)}\} \leq \min\{\text{Rank}\{\bar{\mathbf{A}}_c^{(n)}\}, \text{Rank}\{\bar{\mathbf{A}}_d^{(n)}\}\}. \quad (56)$$

Combining Lemma 1 and Lemma 2 concludes the proof.

B. PROOF OF THEOREM 2

Similar to (47), we can express $\tilde{\mathbf{A}}^{(n)}$ as $\tilde{\mathbf{A}}^{(n)} := \tilde{\tilde{\mathbf{A}}}_c^{(n)} \tilde{\tilde{\mathbf{A}}}_d^{(n)}$ with

$$\begin{aligned} \tilde{\tilde{\mathbf{A}}}_c^{(n)} &:= \begin{bmatrix} \bar{\mathbf{A}}_c^{(n)}[j_0] & & \\ & \ddots & \\ & & \bar{\mathbf{A}}_c^{(n)}[j_{V-1}] \end{bmatrix}, \\ \tilde{\tilde{\mathbf{A}}}_d^{(n)} &:= [\bar{\mathbf{A}}_d^{(n)T}[j_0] \cdots \bar{\mathbf{A}}_d^{(n)T}[j_{V-1}]]^T, \end{aligned} \quad (57)$$

where $\bar{\mathbf{A}}_c^{(n)T}[j_v]$ and $\bar{\mathbf{A}}_d^{(n)T}[j_v]$ are defined in (47) but with the symbol index j_v added.

We first prove the full column-rank condition of $\bar{\mathbf{A}}_c^{(n)}[j_v]$ by following the same steps as in Lemma 1 except for (54), where we need to plug in the (O)CE-BEM that is based on multiple blocks as defined in (21). As a result, we obtain after some algebra that

$$\begin{aligned} \Phi_g^{(m)}[j_v] &= \frac{1}{\sqrt{X}} \bar{\Phi}^{(m)} e^{j \frac{2\pi}{K} g \mu^{(m)}} \mathcal{D}\{[e^{-j \frac{2\pi}{K'}(j_v(K+L)+L+g)0}, \dots, e^{-j \frac{2\pi}{K'}(j_v(K+L)+L+g)Q}]^T\}, \\ \bar{\Phi}^{(m)} &:= \begin{bmatrix} e^{-j \frac{2\pi}{XK'} 0(K0-\mu^{(m)}K')} & \dots & e^{-j \frac{2\pi}{XK'} 0(KQ-\mu^{(m)}K')} \\ \vdots & \ddots & \vdots \\ e^{-j \frac{2\pi}{XK'} (X-1)(K0-\mu^{(m)}K')} & \dots & e^{-j \frac{2\pi}{XK'} (X-1)(KQ-\mu^{(m)}K')} \end{bmatrix}. \end{aligned} \quad (58)$$

with $K' := \kappa V(K+L)$. Like in Lemma 1, the rank of $\bar{\mathbf{A}}_c^{(n)}[j_v]$ is determined by the rank of $[\bar{\Phi}^{(0)}, \dots, \bar{\Phi}^{(N_T-1)}]$ multiplied by G . It is tall if $X = K/G \geq N_T(Q+1)$. Besides, if $\mu^{(m+1)}K' > \mu^{(m)}K' + KQ$, this matrix contains distinctive columns of a larger XK' -point DFT matrix, and is in that case of full column-rank.

To check the rank of $\tilde{\mathbf{A}}_d^{(n)}$, we permute its rows, which admits an expression as

$$(\mathbf{\Pi}_{N_T(Q+1)} \otimes \mathbf{I}_G) \tilde{\mathbf{A}}_d^{(n)} = \mathcal{D}\{\mathbf{I}_{Q+1} \otimes \mathcal{D}\{\tilde{\mathbf{p}}^{(0)}\}\mathbf{V}_L^{\{\tilde{\mathcal{P}}^{(0)},:\}}, \dots, \mathbf{I}_{Q+1} \otimes \mathcal{D}\{\tilde{\mathbf{p}}^{(N_T-1)}\}\mathbf{V}_L^{\{\tilde{\mathcal{P}}^{(N_T-1)},:\}}\}, \quad (59)$$

where

$$\begin{aligned} \tilde{\mathbf{p}}^{(m)} &:= [\bar{\mathbf{p}}^{(m)T}[j_0], \dots, \bar{\mathbf{p}}^{(m)T}[j_{V-1}]]^T, \\ \tilde{\mathcal{P}}^{(m)} &:= \bar{\mathcal{P}}^{(m)}[j_0] \cup \dots \cup \bar{\mathcal{P}}^{(m)}[j_{V-1}]. \end{aligned}$$

Because $\tilde{\mathcal{P}}^{(m)}$ contains VG distinctive elements, $\mathbf{V}_L^{\{\tilde{\mathcal{P}}^{(m)},:\}}$ is a tall Vandermonde matrix if $VG \geq L+1$.

Since $\tilde{\mathbf{A}}_c^{(n)}$ and $\tilde{\mathbf{A}}_d^{(n)}$ are both of full column-rank, we can utilize the rank inequality in [25] to conclude the proof.

C. PROOF OF THEOREM 3 AND COROLLARY 1

The identical pilot assumption implies that $\bar{\mathbf{A}}_d^{(n)}[j_0] = \dots = \bar{\mathbf{A}}_d^{(n)}[j_{V-1}]$ and therefore

$$\tilde{\mathbf{A}}^{(n)} = [\bar{\mathbf{A}}_c^{(n)T}[j_0], \dots, \bar{\mathbf{A}}_c^{(n)T}[j_{V-1}]]^T \bar{\mathbf{A}}_d^{(n)}[j_0],$$

where $\bar{\mathbf{A}}_c^{(n)T}[j_v]$ and $\bar{\mathbf{A}}_d^{(n)T}[j_v]$ are defined in (47) with the symbol index j_v added. Obviously, $\bar{\mathbf{A}}_d^{(n)}[j_0]$ is of full column-rank if $G \geq N_T(L+1)$. To prove the full column-rank condition of

$[\bar{\mathbf{A}}_c^{(n)T}[j_0], \dots, \bar{\mathbf{A}}_c^{(n)T}[j_{V-1}]]^T$, we can follow similar steps as in Appendix A and Appendix B, which lead eventually to the full column-rank condition of a larger matrix

$$\begin{bmatrix} \Phi_g^{(0)}[j_0] & \cdots & \Phi_g^{(N_T-1)}[j_0] \\ \vdots & \ddots & \vdots \\ \Phi_g^{(0)}[j_{V-1}] & \cdots & \Phi_g^{(N_T-1)}[j_{V-1}] \end{bmatrix}, \quad (60)$$

where $\Phi_g^{(m)}[j_v]$ is defined in (58). If the above is not of full column-rank, then there should exist a vector $[\mathbf{a}_0^T, \dots, \mathbf{a}_{N_T-1}^T]^T$, which contains at least one non-zero element, such that

$$\begin{bmatrix} \Phi_g^{(0)}[j_0] & \cdots & \Phi_g^{(N_T-1)}[j_0] \\ \vdots & \ddots & \vdots \\ \Phi_g^{(0)}[j_{V-1}] & \cdots & \Phi_g^{(N_T-1)}[j_{V-1}] \end{bmatrix} \begin{bmatrix} \mathbf{a}_0 \\ \vdots \\ \mathbf{a}_{N_T-1} \end{bmatrix} = \mathbf{0}_{V \times 1}. \quad (61)$$

With (58) taken into account, the above can be equivalently rewritten as $\Theta \Gamma = \mathbf{0}_{X \times V}$, with

$$\Theta = \sum_m \bar{\Phi}^{(m)} \mathcal{D}\{e^{j\frac{2\pi}{K}g\mu^{(m)}} \mathbf{a}_m\},$$

$$\Gamma = \begin{bmatrix} e^{-j\frac{2\pi}{K'}0(j_0(K+L)+L+g)} \cdots e^{-j\frac{2\pi}{K'}0(j_{V-1}(K+L)+L+g)} \\ \vdots & \ddots & \vdots \\ e^{-j\frac{2\pi}{K'}Q(j_0(K+L)+L+g)} \cdots e^{-j\frac{2\pi}{K'}Q(j_{V-1}(K+L)+L+g)} \end{bmatrix}.$$

Γ has a full row-rank if $V \geq Q + 1$, and therefore in order for (61) to hold, Θ must be an all-zero matrix. Based on the definition of $\bar{\Phi}^{(m)}$ in (58), the q th column of Θ can be expressed as

$$\Theta^{\{:,q\}} = \begin{bmatrix} e^{-j\frac{2\pi}{XK'}0(Kq-\mu^{(0)}K')} \cdots e^{-j\frac{2\pi}{XK'}0(Kq-\mu^{(N_T-1)}K')} \\ \vdots & \ddots & \vdots \\ e^{-j\frac{2\pi}{XK'}(X-1)(Kq-\mu^{(0)}K')} \cdots e^{-j\frac{2\pi}{XK'}(X-1)(Kq-\mu^{(N_T-1)}K')} \end{bmatrix} \begin{bmatrix} e^{j\frac{2\pi}{K}g\mu^{(0)}} [\mathbf{a}_0]_q \\ \vdots \\ e^{j\frac{2\pi}{K}g\mu^{(N_T-1)}} [\mathbf{a}_{N_T-1}]_q \end{bmatrix}. \quad (62)$$

The first matrix on the right-hand-side of the above will have a full column-rank if $X \geq N_T$ and $\mu^{(m)} \neq \mu^{(m')}$. This means that there exists at least one column in Θ that is not all-zero. Hence, (61) cannot hold, and the matrix in (60) has a full column-rank. This concludes the proof of Theorem 3.

If there is only one transmit antenna N_T , we only need to prove the full column-rank condition

of the following matrix [c.f. (60)]

$$\left[\Phi_g^{(0)T}[j_0] \cdots \Phi_g^{(0)T}[j_{V-1}] \right]^T = \frac{1}{\sqrt{X}} e^{j \frac{2\pi}{GX} \mu^{(m)} g} \mathcal{D} \{ \mathbf{1}_{V \times 1} \otimes \boldsymbol{\theta}^{-\mu^{(m)}} \}$$

$$\begin{bmatrix} e^{-j \frac{2\pi}{\kappa V(K+L)} 0[j_0(K+L)+0G+g]} & \cdots & e^{-j \frac{2\pi}{\kappa V(K+L)} Q[j_0(K+L)+0G+g]} \\ \vdots & \ddots & \vdots \\ e^{-j \frac{2\pi}{\kappa V(K+L)} 0[j_0(K+L)+(X-1)G+g]} & \cdots & e^{-j \frac{2\pi}{\kappa V(K+L)} Q[j_0(K+L)+(X-1)G+g]} \\ \vdots & \ddots & \vdots \\ e^{-j \frac{2\pi}{\kappa V(K+L)} 0[j_{V-1}(K+L)+0G+g]} & \cdots & e^{-j \frac{2\pi}{\kappa V(K+L)} Q[j_{V-1}(K+L)+0G+g]} \\ \vdots & \ddots & \vdots \\ e^{-j \frac{2\pi}{\kappa V(K+L)} 0[j_{V-1}(K+L)+(X-1)G+g]} & \cdots & e^{-j \frac{2\pi}{\kappa V(K+L)} Q[j_{V-1}(K+L)+(X-1)G+g]} \end{bmatrix}$$

The rightmost matrix on the RHS of the above is of dimension $VX \times (Q+1)$, and is obviously a submatrix carved out of a larger $\kappa V(K+L)$ -point DFT matrix. Hence, it will have full column rank if $VX \geq (Q+1)$. This completes the proof of Corollary 1.

REFERENCES

- [1] T. A. Thomas and F. W. Vook, "Multi-user frequency-domain channel identification, interference suppression, and equalization for time-varying broadband wireless communications," *Proceedings of the 2000 IEEE Sensor Array and Multichannel Signal Processing Workshop*, pp. 444–448, Mar. 2000.
- [2] G. Leus, "On the estimation of rapidly time-varying channels," *European Signal Processing Conference, EUSIPCO*, pp. 2227–2230, Sept. 2004.
- [3] X. Ma, G. Giannakis, and S. Ohno, "Optimal training for block transmissions over doubly-selective fading channels," *IEEE Transactions on Signal Processing*, vol. 51, pp. 1351–1366, May 2003.
- [4] A. P. Kannu and P. Schniter, "MSE-optimal training for linear time-varying channels," *International Conference on Acoustics, Speech, and Signal Processing, ICASSP*, Mar. 2005.
- [5] T. Zemen and C. F. Mecklenbräuker, "Time-variant channel estimation using discrete prolate spheroidal sequences," *IEEE Transactions on Signal Processing*, vol. 53, pp. 3597–3607, Sept. 2005.
- [6] Z. Tang, R. C. Cannizzaro, P. Banelli, and G. Leus, "Pilot-assisted time-varying channel estimation for OFDM systems," *IEEE Transactions on Signal Processing*, vol. 55, pp. 2226–2238, May 2007.
- [7] A. P. Kannu and P. Schniter, "Minimum mean-squared error pilot-aided transmission for MIMO doubly selective channels," *Proc. Conference on Information Sciences and Systems, (Princeton, NJ)*, Mar. 2006.
- [8] Z. Tang and G. Leus, "Pilot schemes for time-varying channel estimation in OFDM systems," *IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications, SPAWC*, June 2007.
- [9] I. Barhumi, G. Leus, and M. Moonen, "Optimal training design for MIMO OFDM systems in mobile wireless channels," *IEEE Transactions on Signal Processing*, vol. 51, pp. 1615–1624, June 2003.
- [10] H. Minn and N. Al-Dhahir, "Optimal training signals for MIMO OFDM channel estimation," *IEEE Transactions on Wireless Communications*, vol. 5, pp. 1158–1168, May 2006.

- [11] X. Dai, "Optimal training design for linearly time-varying MIMO/OFDM channels modelled by a complex exponential basis expansion," *IET Communications*, vol. 1, no. 5, pp. 945–953, 2007.
- [12] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. New Jersey, USA, 1993.
- [13] M. Ghogho and A. Swami, "Improved channel estimation using superimposed training," *IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications, SPAWC*, pp. 110–114, July 2004.
- [14] S. Coleri, M. Ergen, A. Puri, and A. Bahai, "Channel estimation techniques based on pilot arrangement in OFDM systems," *IEEE Transactions on Broadcasting*, vol. 48, pp. 223–229, Sept. 2002.
- [15] J.-W. Choi and Y.-H. Lee, "Optimum pilot pattern for channel estimation in OFDM systems," *IEEE Transactions on Wireless Communications*, vol. 4, pp. 2083–2088, Sept. 2005.
- [16] Y. Li, L. J. Cimini, and N. R. Sollenberger, "Robust channel estimation for OFDM systems with rapid dispersive fading channels," *IEEE Transactions on Communications*, vol. 46, pp. 1146–1162, July 1998.
- [17] Y. Mostofi and D. C. Cox, "ICI mitigation for pilot-aided OFDM mobile systems," *IEEE Transactions on Wireless Communications*, vol. 4, pp. 765–774, Mar. 2005.
- [18] D. Schafhuber and G. Matz, "MMSE and adaptive prediction of time-varying channels for OFDM systems," *IEEE Transactions on Wireless Communications*, vol. 4, pp. 593–602, Mar. 2005.
- [19] 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 Transactions on Communications*, vol. 49, pp. 1375–1387, Aug. 2001.
- [20] T. Cui, C. Tellambura, and Y. Wu, "Low-complexity pilot-aided channel estimation for OFDM systems over doubly-selective channels," *IEEE International Conference on Communications, ICC*, vol. 3, pp. 1980–1984, May 2005.
- [21] P. Schniter, "On doubly dispersive channel estimation for pilot-aided pulse-shaped multicarrier modulation," *Proc. Conference on Information Sciences and Systems, (Princeton, NJ)*, Mar. 2006.
- [22] Z. Tang, *OFDM Transmission over Rapidly Changing Channels*. Ph.D Dissertation at Delft University of Technology, the Netherlands, 2007.
- [23] W. C. Jakes, *Microwave Mobile Channels*. New York: Wiley, 1974.
- [24] Y. R. Zheng and C. Xiao, "Simulation models with correct statistical properties for Rayleigh fading channels," *IEEE Transactions on Communications*, vol. 51, pp. 920–928, June 2003.
- [25] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1999.
- [26] Z. Tang, G. Leus, and P. Banelli, "Pilot-assisted time-varying OFDM channel estimation based on multiple OFDM symbols," *IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications, SPAWC*, June 2006.

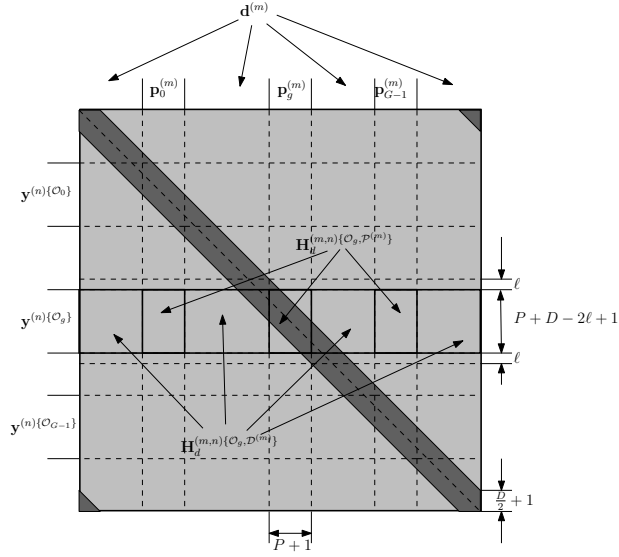


Fig. 1. The partitioning of the frequency-domain channel matrix $\mathbf{H}_d^{(m,n)}$. Its rows correspond to the positions of the received samples; its columns to the positions of the pilots and data. Note that $\mathbf{H}_d^{(m,n)}$ is in principle a full matrix, but with most of its energy concentrated around the diagonal. This effect is represented in the figure by the different shades.

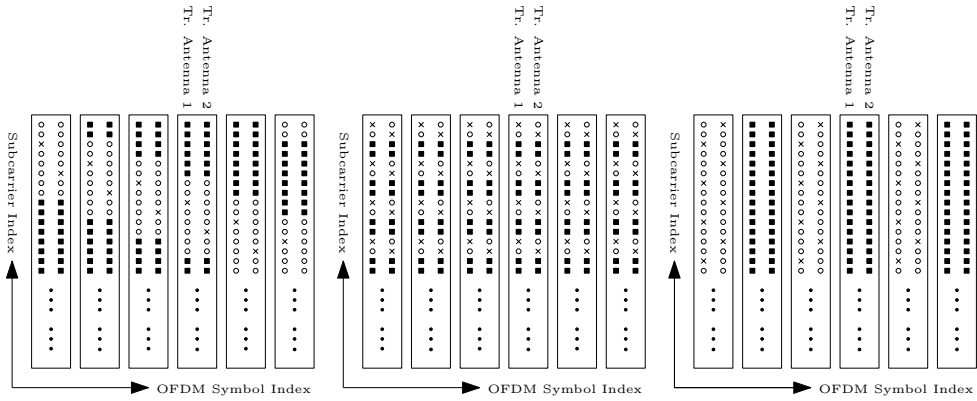


Fig. 2. Overview of the pilot schemes studied. The left subplot depicts the Comb-type I pilot structure; the middle subplot the Comb-type II pilot structure, and the right subplot the block-type pilot structure. Each rectangle corresponds to one OFDM symbol interval and contains OFDM symbols from each transmit antenna. Inside the rectangle, the zero pilots are represented by circles; the non-zero pilots by crosses, and the data symbols by squares.

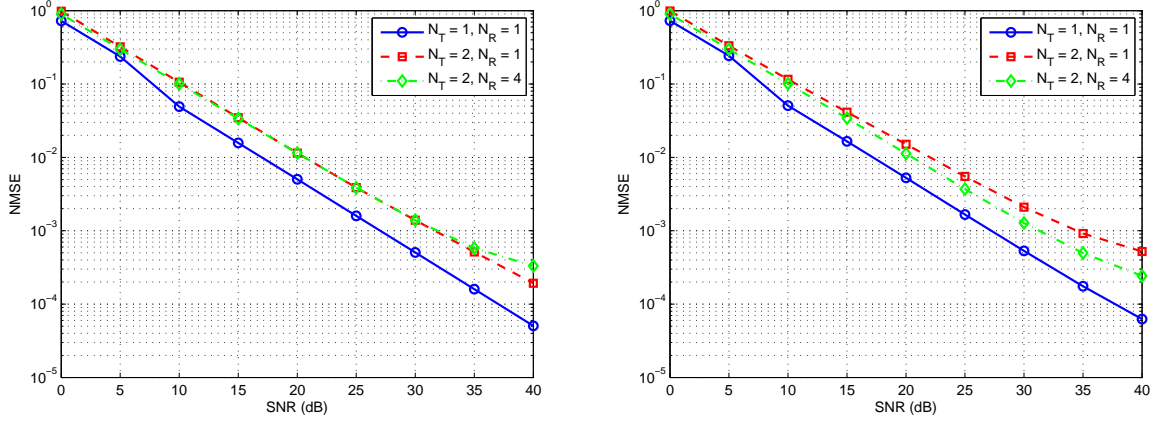


Fig. 3. Channel estimation performance based on a single OFDM symbol for a short channel $L = 3$. Left plot $\nu_D = 8e^{-4}$; right plot $\nu_D = 4e^{-3}$.

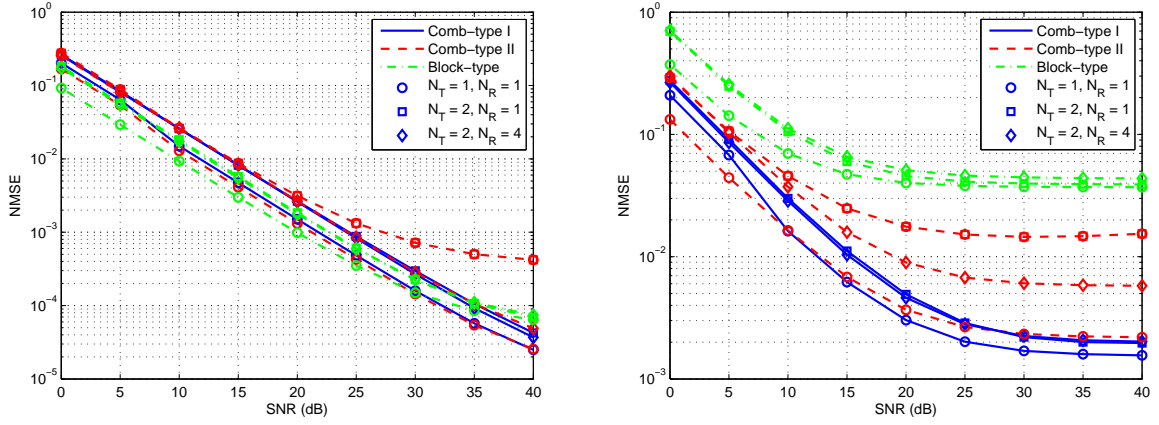


Fig. 4. Channel estimation performance based on multiple OFDM symbols for a short channel $L = 3$. Left plot $\nu_D = 8e^{-4}$; right plot $\nu_D = 4e^{-3}$.

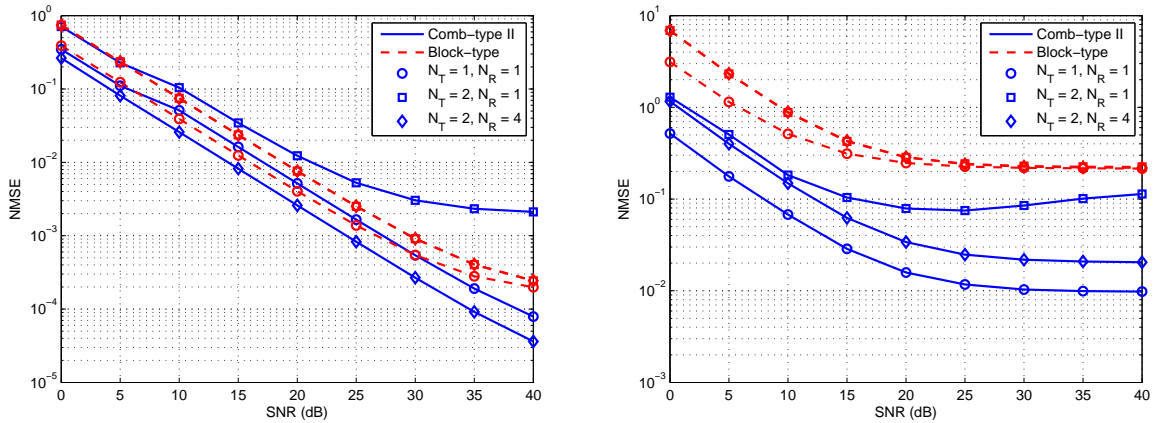


Fig. 5. Channel estimation performance based on multiple OFDM symbols for a long channel $L = 15$. Left plot $\nu_D = 8e^{-4}$; right plot $\nu_D = 4e^{-3}$.

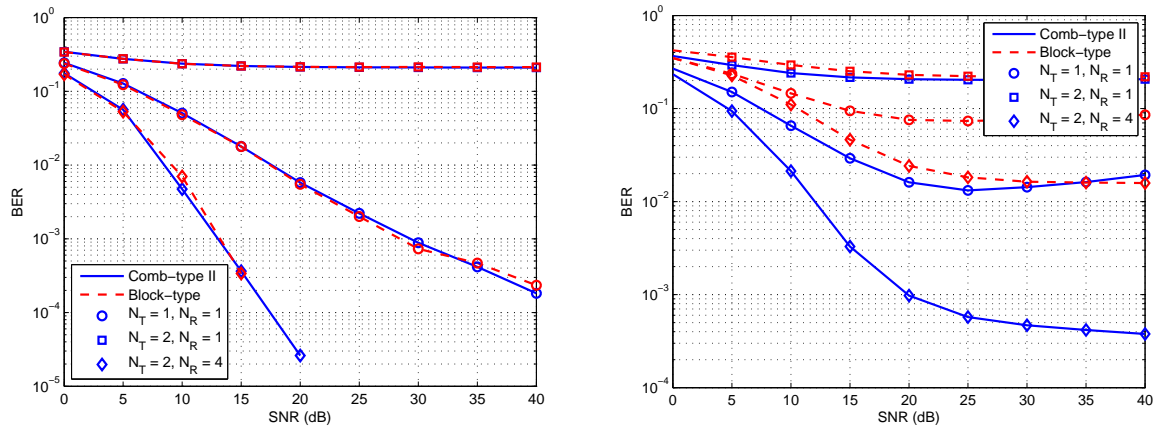


Fig. 6. Channel equalization performance based on multiple OFDM symbols for a long channel $L = 15$. Left plot $\nu_D = 8e^{-4}$; right plot $\nu_D = 4e^{-3}$.

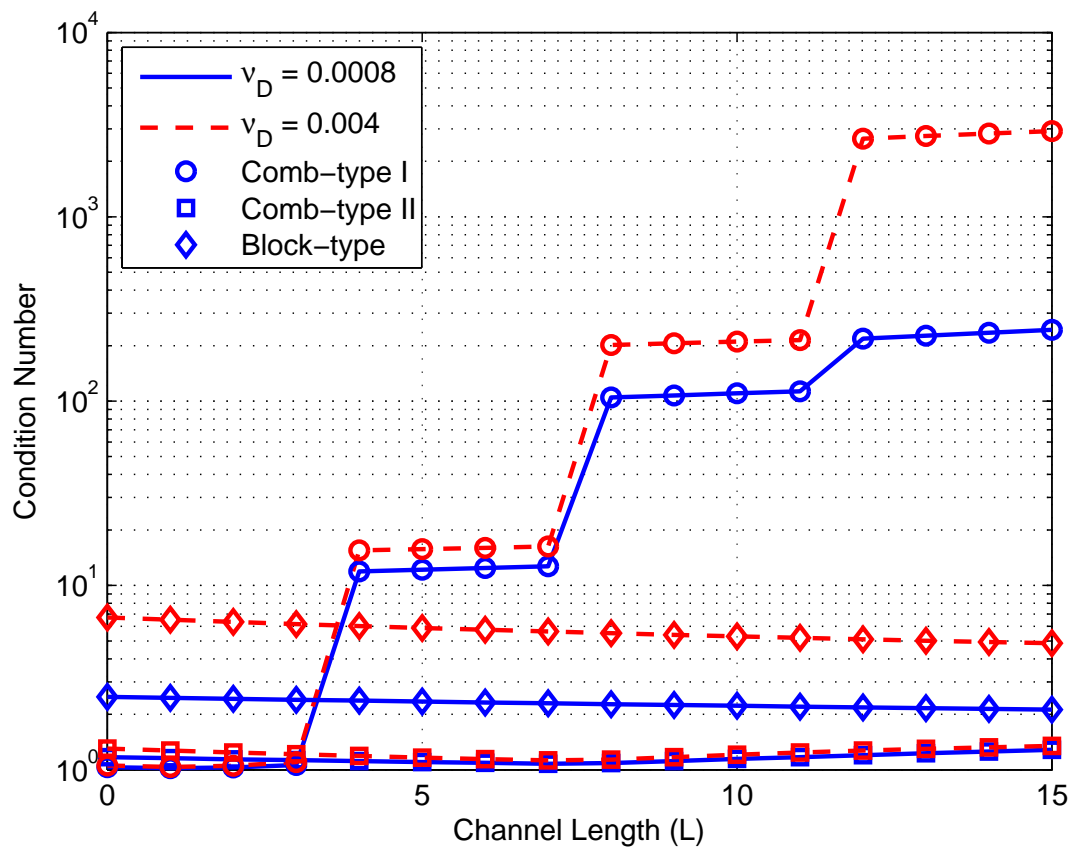


Fig. 7. Condition number vs channel length for different pilot structures.

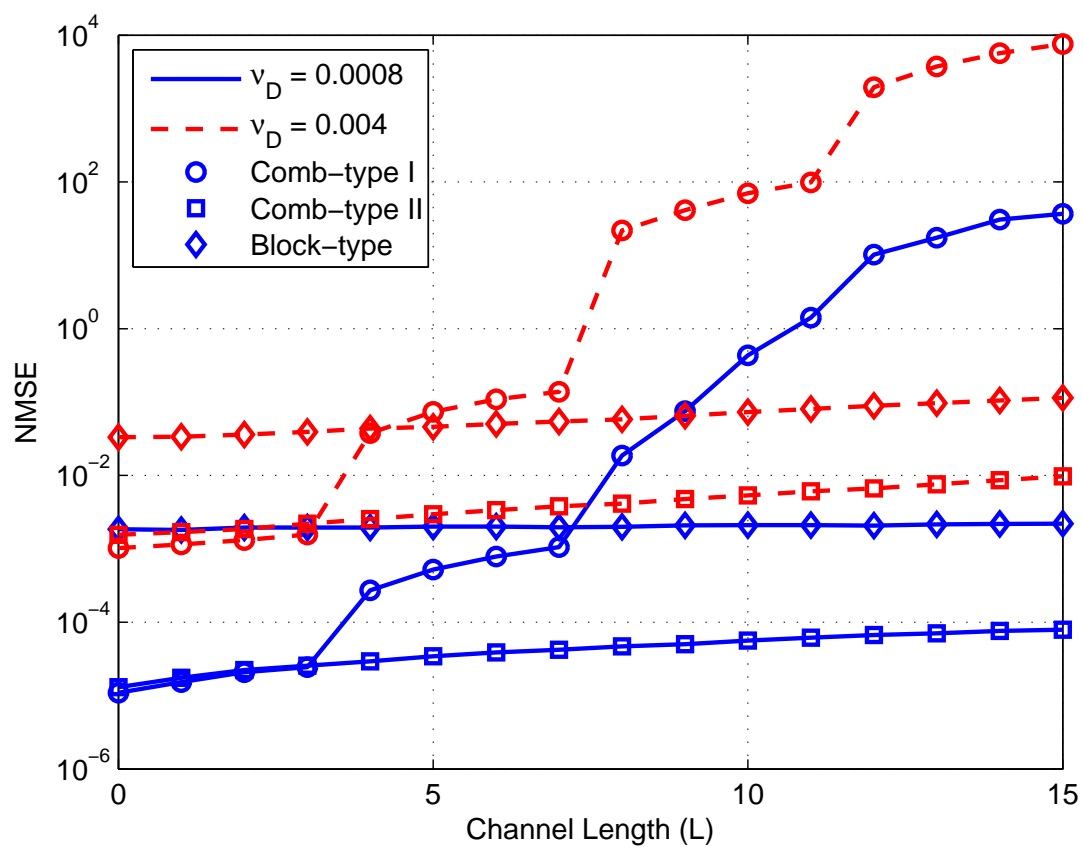


Fig. 8. Channel estimation performance vs channel length for different pilot structures at SNR = 40dB.

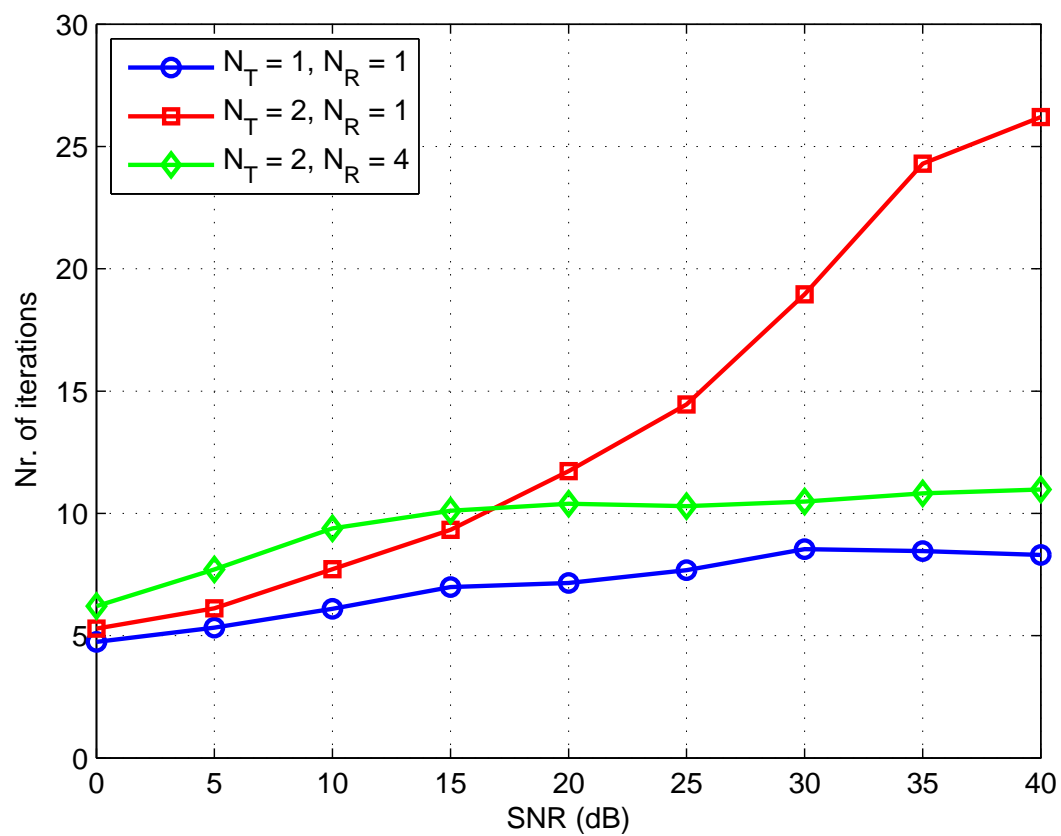


Fig. 9. Average number of iterations vs SNR.

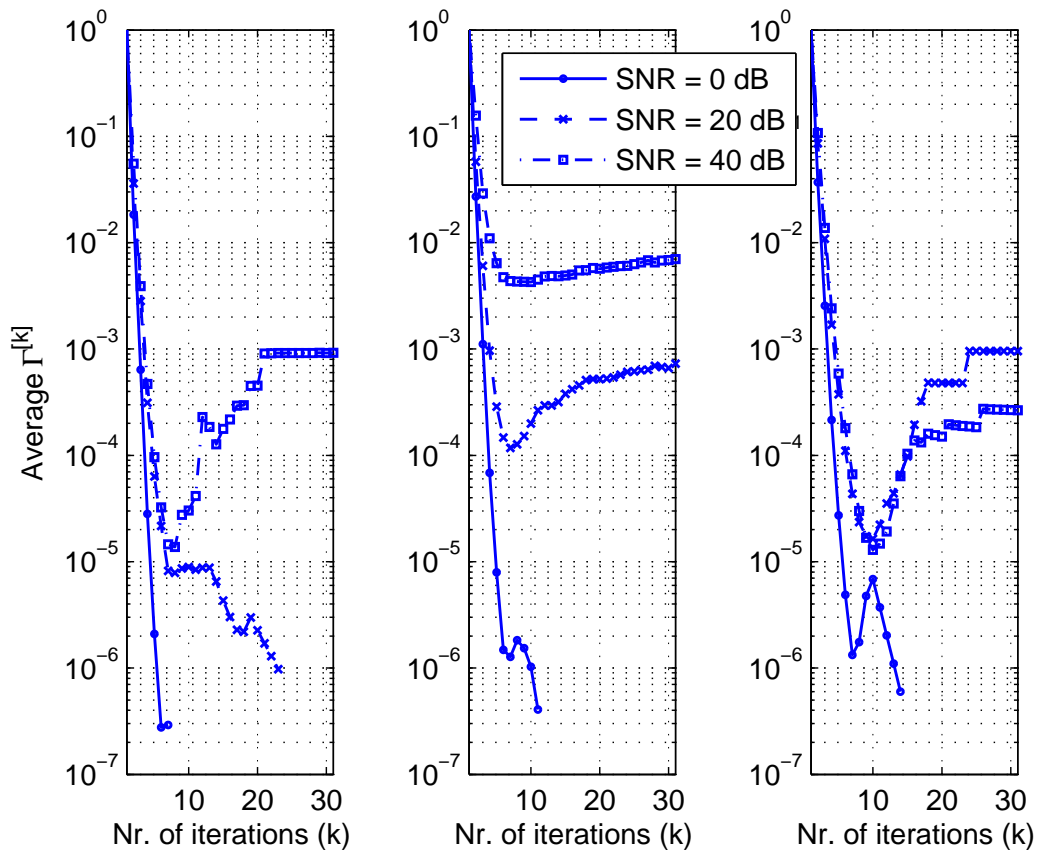


Fig. 10. Average normalized difference in energy over subsequent estimates. Left plot $N_T = 1$ and $N_R = 1$; middle plot $N_T = 2$ and $N_R = 1$; right plot $N_T = 2$ and $N_R = 4$;