A Model Reduction Approach for OFDM Channel Estimation under High Mobility Conditions

Tareq Y. Al-Naffouri, K.M. Zahidul Islam, *Student Member, IEEE*,
Naofal Al-Dhahir *, *Fellow, IEEE*, and Sili Lu, *Student Member, IEEE*

Abstract—OFDM combines the advantages of high performance and relatively low implementation complexity. However, for reliable coherent detection of the input signal, the OFDM receiver needs accurate channel information. When the channel exhibits fast time variation as it is the case with several recent OFDM-based mobile broadband wireless standards (e.g. WiMAX, LTE, DVB-H), channel estimation at the receiver becomes quite challenging for two main reasons 1) the receiver needs to perform this estimation more frequently and 2) channel time-variations introduce inter-carrier interference among the OFDM sub-carriers which can degrade the performance of conventional channel estimation algorithms significantly. In this paper, we propose a new pilot-aided algorithm for the estimation of fast time-varying channels in OFDM transmission. Unlike many existing OFDM channel estimation algorithms in the literature, we propose to perform channel estimation in the frequency domain, to exploit the structure of the channel response (such as frequency and time correlations and bandedness), optimize the pilot group size and perform most of the computations offline resulting in high performance at substantial complexity reductions.

Index Terms—Channel estimation, Doppler frequency, Model reduction, ICI, OFDM.

I. INTRODUCTION

Orthogonal frequency division multiplexing (OFDM) enables high-speed transmission over frequency-selective channels with simple one-tap equalizers by creating a set of parallel, orthogonal, frequency-flat sub-channels using the computationally-efficient IFFT/FFT modulation/demodulation vectors. OFDM has found widespread applications and is already part of many industry standards including digital audio and video broadcasting (DAB/DVB), high-speed transmission over digital subscriber line (DSL), and wireless local area network (WLAN) standards (e.g., IEEE 802.11a/b/g and HIPER-LAN/2) [1]. OFDM is also the modulation scheme of choice for broadband wireless communications standards including 802.16e (metropolitan area networks), 802.11n (local area networks), and 802.15.3 (personal area networks) and other emerging cellular wireless communications systems like 3GPP evolution and 4G.

However, in a mobile wireless environment (which is the case for many of the applications mentioned above), the channel is continuously changing with time. This represents a burden on the receiver since it has to estimate the channel accurately before coherent detection. A more severe consequence of time variation is that the set of parallel OFDM sub-carriers lose their orthogonality in a phenomenon known as inter-carrier interference (ICI). This ICI phenomenon is the frequency-domain dual of inter-symbol interference (ISI) that plagues single-carrier transmission over frequency-selective channels. The severity of ICI depends on the degree of channel time variation (Doppler frequency). ICI makes data detection less reliable as the OFDM sub-carriers are now coupled and one-tap per sub-channel equalization becomes highly suboptimal. More complex equalizers are thus needed at the receiver to untangle the data streams that are transmitted over the OFDM sub-carriers. Furthermore, this ICI makes the channel estimation task more challenging since not only the individual sub-channel frequency responses need to be estimated but also the interference caused by each sub-channel into the other sub-channels in each OFDM symbol.

In the absence of ICI (which occurs when the channel remains constant within each OFDM symbol), the diagonal of the channel frequency response (CFR) matrix needs to be estimated (i.e., N parameters where N is the FFT size). However, for severe time variation, in principle, one needs to estimate every element of the CFR matrix (i.e., N^2 parameters). In practice, it turns out that it is sufficient to estimate only M ($M \ll N$) diagonals of the CFR matrix and, hence, estimate only MN parameters. Thus, in the presence of ICI, the number of estimated channel parameters increases which, in turn, increases the training overhead and, hence, reduces the useful data throughput of the system [2].

Channel estimation for OFDM can be performed in the frequency or time domains. Conventional frequency-domain channel estimation algorithms ignore ICI and the equalization in an OFDM receiver is implemented as a simple single-tap per subcarrier filter [3], [4], which makes them highly suboptimal under high Doppler. Using pilot OFDM symbols (sometimes also called training sequences or preambles) to estimate the channel and using this channel estimate until the next preamble is another popular solution for OFDM channel estimation [5], however, it assumes the channel to be slowly-varying. Adaptive-filter based channel estimation techniques has also been proposed in [6], [7] with expectation-maximization algorithm.

By performing channel estimation in the time domain for

the time-invariant case, the number of estimated parameters is reduced from N, the number of frequency sub-channels, to L, the number of (time-domain) channel taps which is a drastic reduction since typically $L \ll N$. The reduction in the dimensionality of the parameter estimation space in turn results in improved estimation accuracy. This gain, however, does not come for free. Since OFDM data detection is performed in the frequency domain, even in the time-invariant channel case, estimating the channel in the time-domain makes it necessary to perform an extra FFT operation.

For fast time-varying channels, time-domain channel estimation algorithms typically mark a few rows of the time-domain channel matrix and estimate them using pilot tones embedded within each OFDM symbol [1], [8]. Since the multi-path delay spread can be very large in highly-frequency-selective channels, the number of unknown channel taps for each marker row can be too high to estimate with the limited available pilot tones. Moreover, more marker rows are needed as the Doppler frequency increases which increases complexity and reduces the throughput.

Apart from the computational complexity issue, performing channel estimation in the time domain might be over-solving a problem. For example, in multiple-access OFDM systems like WiMAX, users are not interested in the whole frequency spectrum, but only in that part of the spectrum they are using, something that can be easily achieved by performing channel estimation in the frequency domain. Moreover, even if some users were interested in estimating the whole spectrum, many standards would not be able to support that as there are not enough pilots to do so.

Considering a fast-varying channel, Mostofi proposed in [9] a hybrid frequency/time-domain channel estimation algorithm based on a *linear* approximation of the time variations of each channel impulse response (CIR) coefficient within one OFDM symbol. However, this algorithm introduces a processing delay of at least one OFDM symbol and its performance degrades for long OFDM symbols (necessary to combat severe channel frequency selectivity) at very high Doppler, as we show in Section VI.

The above-mentioned considerations motivate us in this paper to design a high-performance computationally-efficient OFDM channel estimation algorithm in the presence of ICI while reducing the training overhead. The main contributions of this paper are

1) Exploiting the channel correlations in the time and frequency domains to enhance the channel estimation accuracy and reduce its complexity (by performing most of the computations offline).
2) Performing channel estimation and ICI equalization in the frequency domain by exploiting the banded structure [10], [11] of the CFR matrix.
3) Reducing the dimensionality of the parameter estimation space at negligible performance loss.
4) Optimizing the pilot group size.
5) Performing detailed complexity and performance analysis of our proposed algorithm and comparing it with the widely-cited algorithm of [9].

Notation We denote scalars by lower-case letters (e.g., x),

vectors by lower-case boldface letters (e.g., $\mathbf{x}$), and matrices by upper-case boldface letters (e.g., $\mathbf{X}$). Calligraphic notation (e.g. $\mathcal{X}$) is reserved for vectors in the frequency domain. A hat over a variable indicates an estimate of the variable (e.g., $\hat{h}$ is an estimate of h). We use $(\cdot)^*$ to denote complex conjugate, $(\cdot)^T$ transpose, and $(\cdot)^H$ complex conjugate transpose (Hermitian) operation. We use $\mathbf{I}_N$ to denote the $N \times N$ identity matrix and $\mathbf{0}_{m,n}$ to denote an $m \times n$ all-zero matrix. In addition, $\mathrm{Diag}(\mathbf{v})$ denotes the diagonal matrix with diagonal elements given by the vector $\mathbf{v}$, while $\mathrm{vec}(\mathbf{R})$ denotes the operation of vectorizing an $m \times n$ matrix into a $mn \times 1$ vector [12]. The Kronecker product of $\mathbf{R}$ ($m \times n$) and $\mathbf{S}$ ($p \times q$), denoted by $\mathbf{R} \otimes \mathbf{S}$, is a $mp \times nq$ matrix which is defined as follows

$$
\mathbf{R} \otimes \mathbf{S} = \begin{bmatrix} r_{11}\mathbf{S} & \cdots & r_{1n}\mathbf{S} \\ \vdots & \ddots & \vdots \\ r_{m1}\mathbf{S} & \cdots & r_{mn}\mathbf{S} \end{bmatrix}
$$

$$
= \begin{bmatrix} r_{11}s_{11} & \cdots & r_{11}s_{1q} & \cdots & r_{1n}s_{11} & \cdots & r_{1n}s_{1q} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ r_{11}s_{p1} & \cdots & r_{11}s_{pq} & \cdots & r_{1n}s_{p1} & \cdots & r_{1n}s_{pq} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ r_{m1}s_{11} & \cdots & r_{m1}s_{1q} & \cdots & r_{mn}s_{11} & \cdots & r_{mn}s_{1q} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ r_{m1}s_{p1} & \cdots & r_{m1}s_{pq} & \cdots & r_{mn}s_{p1} & \cdots & r_{mn}s_{pq} \end{bmatrix}
$$

Throughout the paper, we use the terms sub-carrier and sub-channel interchangeably. For the convenience of the reader, we have summarized the key system parameters and variables used in the paper in Table I.

TABLE I
LIST OF KEY PARAMETERS AND VARIABLES

Parameters / Variables	Description
N	FFT size
f_d	Doppler frequency
L	Number of channel impulse response taps
T	Number of training pilot tones in each OFDM symbol
$\mathbf{H}$	Time-domain channel matrix
$\mathbf{G}$	Frequency-domain channel matrix
$\mathbf{R}_H$	Covariance matrix of $\mathrm{vec}(\mathbf{H})$ $\mathbf{R}_H \triangleq E[\mathrm{vec}(\mathbf{H})\mathrm{vec}(\mathbf{H})^H]$
$\mathbf{R}_G$	Covariance matrix of $\mathrm{vec}(\mathbf{G})$ $\mathbf{R}_G \triangleq E[\mathrm{vec}(\mathbf{G})\mathrm{vec}(\mathbf{G})^H]$
N_d	Number of dominant eigenvalues of $\mathbf{R}_H$ for each channel tap
M	Number of diagonals in approximated banded $\mathbf{G}$
$\mathcal{X}$	Frequency-domain comb-type input vector

II. MODEL AND ASSUMPTIONS

We consider an OFDM system with N subcarriers where each OFDM symbol, denoted by $\mathcal{X} = [X_0 \ldots X_{N-1}]^T$, is converted into time-domain samples $\mathbf{x} = [x_0 \ldots x_{N-1}]^T$ using the N-point Inverse Fast Fourier Transform (IFFT) operation $\mathbf{x} = \mathbf{Q}^H \mathcal{X}$ where $\mathbf{Q}^H$ is the N-point IFFT matrix. The (m, n) element of the matrix $\mathbf{Q}$ is defined as

$$
Q_{m,n} = \frac{1}{\sqrt{N}} e^{-\frac{j2\pi mn}{N}} \qquad \text{for} \quad m, n = 0, 1, \cdots N-1
$$

We assume that the cyclic prefix (CP) length is equal to or larger than the CIR memory denoted by L. Then, the received

OFDM symbol $\mathbf{y} = [y_0 \ldots y_{N-1}]^T$ after CP removal is given by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z} \tag{1}$$

where $\mathbf{H}$ is an $N \times N$ time-domain channel matrix given by

$$\mathbf{H} = \begin{bmatrix} h_0(0) & 0 & \cdots & h_0(2) & h_0(1) \\ h_1(1) & h_1(0) & \cdots & h_1(3) & h_1(2) \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ h_{L-1}(L-1) & h_{L-1}(L-2) & \cdots & 0 & 0 \\ 0 & h_L(L-1) & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & h_{N-2}(0) & 0 \\ 0 & 0 & \cdots & h_{N-1}(1) & h_{N-1}(0) \end{bmatrix} \tag{2}$$

where $h_n(l)$ is the complex zero-mean unit-variance CIR tap at lag l (for $0 \le l \le L-1$) and time instant n and $\mathbf{z}$ is the time-domain noise vector. Taking the FFT of (1) yields

$$\boldsymbol{\mathcal{Y}} = \mathbf{Q}\mathbf{y} = \mathbf{Q}\mathbf{H}\mathbf{Q}^H\boldsymbol{\mathcal{X}} + \mathbf{Q}\mathbf{z} = \mathbf{G}\boldsymbol{\mathcal{X}} + \boldsymbol{\mathcal{Z}} \tag{3}$$

where $\mathbf{G} \triangleq \mathbf{Q}\mathbf{H}\mathbf{Q}^H$ is the CFR matrix and $\boldsymbol{\mathcal{Z}}$ is the frequency-domain noise vector. The vectors $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ are the transmit and receive OFDM symbols, respectively, in the frequency domain. The signal-to-noise ratio (SNR) of the system is defined as

$$SNR = \frac{\text{Trace}\left[\text{E}\left(\boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}^H\right)\right]}{\text{Trace}\left[\text{E}\left(\boldsymbol{\mathcal{Z}}\boldsymbol{\mathcal{Z}}^H\right)\right]}$$

where $\text{E}(.)$ denotes the expectation operator. For a time-invariant or a quasi-static fading channel, the CIR remains fixed over the OFDM symbol duration, making $\mathbf{H}$ a circulant matrix and, hence, $\mathbf{G}$ in (3) a diagonal matrix. In this case, the OFDM sub-carriers are decoupled, and a one-tap frequency-domain equalizer (FEQ) is optimal.

However, for a time-variant channel, $\mathbf{H}$ is not circulant, and $\mathbf{G}$ can no longer be assumed diagonal [13]. Some of the signal energy will be dispersed to the off-diagonal elements of $\mathbf{G}$. In this case, the input-output relation for the k-th subcarrier is

$$\mathcal{Y}_k = \underbrace{G_{k,k}\mathcal{X}_k}_{\text{desired}} + \underbrace{\sum_{n=0,n \ne k}^{N-1} G_{k,n}\mathcal{X}_n}_{\text{ICI term}} + \mathcal{Z}_k \tag{4}$$

The first term on the right hand side of (4) is the desired signal term while the second one is the ICI term. When the channel is fast-varying, ICI becomes significant and produces an irreducible error floor.

III. PROPOSED CHANNEL ESTIMATION ALGORITHM

A. Statistics of the Frequency-Domain Channel Coefficients

In this section, we determine the covariance matrix of the CFR matrix coefficients. We start from the following relation between the time-domain and frequency-domain channel matrices

$$\mathbf{G} = \mathbf{Q}\mathbf{H}\mathbf{Q}^H \tag{5}$$

Vectorizing both sides of (5) and using the Kronecker product property in Theorem T2.13 in [12] yields

$$\text{vec}(\mathbf{G}) = ((\mathbf{Q}^H)^T \otimes \mathbf{Q})\text{vec}(\mathbf{H}) = (\mathbf{Q}^* \otimes \mathbf{Q})\text{vec}(\mathbf{H}) \tag{6}$$

We can now use (6) to calculate the covariance matrix of $\text{vec}(\mathbf{G})$ in terms of that of $\text{vec}(\mathbf{H})$

$$\begin{aligned} \mathbf{R}_G &\triangleq E[\text{vec}(\mathbf{G})\text{vec}(\mathbf{G})^H] \\ &= (\mathbf{Q}^* \otimes \mathbf{Q})E[\text{vec}(\mathbf{H})\text{vec}(\mathbf{H})^H](\mathbf{Q}^* \otimes \mathbf{Q})^H \\ &\triangleq (\mathbf{Q}^* \otimes \mathbf{Q})\mathbf{R}_H(\mathbf{Q}^* \otimes \mathbf{Q})^H \end{aligned} \tag{7}$$

Now, we turn our attention to calculating the covariance matrix $\mathbf{R}_H \triangleq E[\text{vec}(\mathbf{H})\text{vec}(\mathbf{H})^H]$. By inspecting (2), we note that we can decompose $\mathbf{H}$ as the sum of L matrices, each of which corresponds to a CIR tap (and its time evolution), i.e.

$$\mathbf{H} = \sum_{l=0}^{L-1} \mathbf{A}_l \tag{8}$$

where $\mathbf{A}_0 = \text{Diag}([h_0(0), h_1(0), \ldots, h_{N-1}(0)])$ and $\mathbf{A}_l$ is the matrix corresponding to the lth CIR tap given by

$$\mathbf{A}_l = \text{Diag}([h_0(l), h_1(l), \ldots, h_{N-1}(l)])\mathbf{B}^l \tag{9}$$

The $N \times N$ matrix $\mathbf{B}$ is a permutation matrix which has the form

$$\mathbf{B} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \tag{10}$$

Thus, post-multiplication of a matrix by $\mathbf{B}$ cyclically shifts its columns to the left. Vectorizing both sides of (8) yields

$$\text{vec}(\mathbf{H}) = \sum_{l=0}^{L-1} \text{vec}(\mathbf{A}_l)$$

Hence, the covariance matrix of $\text{vec}(\mathbf{H})$ is given by

$$\mathbf{R}_H \triangleq E[\text{vec}(\mathbf{H})\text{vec}(\mathbf{H})^H] = \sum_{i=0}^{L-1} E[\text{vec}(\mathbf{A}_l)\text{vec}(\mathbf{A}_l)^H] \triangleq \sum_{l=1}^{L} \mathbf{C}_l \tag{11}$$

where we assumed that the $\mathbf{A}_i$'s are independent, i.e., we assumed that the CIR taps at a given time instant are independent. Moreover, we defined $\mathbf{C}_l$ to be the covariance matrix of $\text{vec}(\mathbf{A}_l)$, i.e., $\mathbf{C}_l = E\left[\text{vec}(\mathbf{A}_l)\text{vec}(\mathbf{A}_l)^H\right]$ which we calculate next. To this end, note that since $\mathbf{A}_0$ is a diagonal matrix, we can write $\text{vec}(\mathbf{A}_0)$ as

$$\text{vec}(\mathbf{A}_0) = [h_0(0), \underbrace{0, \ldots, 0}_{N \text{ zeros}}, h_1(0), \underbrace{0, \ldots, 0}_{N \text{ zeros}}, h_2(0), \ldots, h_{N-1}(0)]^T \tag{12}$$

Assuming Jakes's model for channel time variation [14],[1] we can show that the covariance matrix of $\text{vec}(\mathbf{A}_0)$ is given by

$$\begin{aligned} \mathbf{C}_0 &\triangleq E[\text{vec}(\mathbf{A}_0)\text{vec}(\mathbf{A}_0)^H] \\ &= \begin{bmatrix} J(0) & \mathbf{0}_{1,N} & J(1) & \ldots & \mathbf{0}_{1,N} & J(N-1) \\ \mathbf{0}_{N,1} & \mathbf{0}_{N,N} & \mathbf{0}_{N,1} & \ldots & \mathbf{0}_{1,N} & \mathbf{0}_{N,1} \\ J(1) & \mathbf{0}_{1,N} & J(0) & \ldots & \mathbf{0}_{1,N} & J(N-2) \\ \mathbf{0}_{N,1} & \mathbf{0}_{N,N} & \mathbf{0}_{N,1} & \ldots & \mathbf{0}_{1,N} & \mathbf{0}_{N,1} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ J(N-1) & \mathbf{0}_{1,N} & J(N-2) & \ldots & \mathbf{0}_{1,N} & J(0) \end{bmatrix} \end{aligned} \tag{13}$$

---

[1] Any other CIR time correlation model can be used but we assume Jakes's model in this paper due to its popularity.

which follows from the assumption that

$$E[h_m(l)h_n(l)^*] = J_0(2\pi f_d(m-n)T_s) \triangleq J(m-n) \quad (14)$$

where f_d and T_s are the Doppler frequency and sampling period, respectively, and $J_0(\cdot)$ is the zero-order Bessel function. We can calculate the covariance matrix $\mathbf{C}_l \triangleq E[\text{vec}(\mathbf{A}_l)\text{vec}(\mathbf{A}_l)^H]$ in a similar manner. It is enough though to observe the following properties about $\mathbf{C}_l$

1) $\mathbf{C}_l$ consists of zero entries and non-zero entries all of which belong to the matrix

$$\mathbf{J} \triangleq \begin{bmatrix} J(0) & J(1) & \dots & J(N-1) \\ J(1) & J(0) & \dots & J(N-2) \\ \vdots & \vdots & \ddots & \vdots \\ J(N-1) & J(N-2) & \dots & J(0) \end{bmatrix}$$
$$(15)$$

In fact, $\mathbf{C}_l$ is simply an oversampled version of $\mathbf{J}$.

2) From 1), it follows that the rank of $\mathbf{C}_l$ is N.

3) If $C_l(m,n) \neq 0$, then $C_{l'}(m,n) = 0$ for $l \neq l'$, i.e. the matrices C_l and $C_{l'}$ are never non-zero at the same entry.

4) From the above properties and (11), it follows that the rank of $\mathbf{R}_H$ is NL.

These properties allow us to calculate the eigenvalue decomposition of $\mathbf{R}_H$ which we undertake next.

B. Eigenvalue Decomposition of $\mathbf{R}_H$

Let $\lambda_1, \ldots, \lambda_N$ be the eigenvalues of $\mathbf{J}$ and let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N$ be the corresponding eigenvectors. These quantities can be used to construct the eigenvalues and eigenvectors of $\mathbf{C}_l$'s and subsequently those of $\mathbf{R}_H$ and $\mathbf{R}_G$. To this end, define $\bar{\mathbf{v}}_n$ as an oversampled version of eigenvector $\mathbf{v}_n$; i.e.

$$\bar{\mathbf{v}}_n = [v_n(0), \underbrace{0, \ldots, 0}_{N\,\text{zeros}}, v_n(1), \underbrace{0, \ldots, 0}_{N\,\text{zeros}}, v_n(2), \ldots, v_n(N-1)]^T$$
$$(16)$$

Starting from the relation $\mathbf{J}\mathbf{v}_i = \lambda_i\mathbf{v}_i$, we note that this relationship continues to hold when we replace $\mathbf{J}$ by its oversampled version $\mathbf{C}_0$ and $\mathbf{v}_n$ by its over-sampled version $\bar{\mathbf{v}}_n$, i.e.

$$\mathbf{C}_0\bar{\mathbf{v}}_n = \lambda_n\bar{\mathbf{v}}_n \qquad \text{for } n = 1, \ldots, N \qquad (17)$$

From the above, and since $\mathbf{C}_0$ is of rank N, it follows that $\lambda_1, \lambda_2, \ldots, \lambda_N$ are the nonzero eigenvalues of C_0 and $\bar{\mathbf{v}}_1, \bar{\mathbf{v}}_2, \ldots, \bar{\mathbf{v}}_N$ are the corresponding eigenvectors. We now turn our attention to the eigenvalue decomposition (EVD) of $\mathbf{C}_1, \mathbf{C}_2, \ldots, \mathbf{C}_{L-1}$. From (9), we note that $\mathbf{A}_l$ is a cyclicly-shifted version of a diagonal matrix, and so we can easily vectorize $\mathbf{A}_i$ as follows using T2.13 from [12]

$$\text{vec}(\mathbf{A}_l) = (\mathbf{B}^{l^T} \otimes \boldsymbol{I}_N)\text{vec}(\text{Diag}([h_0(l), h_1(l), \ldots, h_{N-1}(l)]))$$
$$= \mathbf{D}^l\text{vec}(\text{Diag}([h_0(l), h_1(l), \ldots, h_{N-1}(l)]))$$
$$(18)$$

where $\mathbf{D}$ is the N^2-by-N^2 permutation matrix defined by

$$\mathbf{D} \triangleq \mathbf{B}^T \otimes \mathbf{I}_N \qquad (19)$$

Now, by definition, we have that

$$\begin{aligned}\mathbf{C}_l &= E[\text{vec}(\mathbf{A}_l)\text{vec}(\mathbf{A}_l)^H] \\ &= \mathbf{D}^l E[\text{vec}(\text{Diag}([h_0(l), h_1(l), \ldots, h_{N-1}(l)])) \\ &\quad .\text{vec}(\text{Diag}([h_0(l), h_1(l), \ldots, h_{N-1}(l)]))^H]\mathbf{D}^{l^H} \\ &= \mathbf{D}^l\mathbf{C}_0\mathbf{D}^{l^H} \end{aligned}$$
$$(20)$$

The last line above follows from the fact that the CIR taps are assumed identically distributed [2] and so the expectation in the second equality remains the same if we replace the lth tap with the zero-th tap. From this relation, it is easy to see that $\mathbf{D}^l\bar{\mathbf{v}}_n$, $n = 1, 2, \ldots, N$ are the eigenvectors of $\mathbf{C}_l$ as follows

$$\mathbf{C}_l\mathbf{D}^l\bar{\mathbf{v}}_n = \mathbf{D}^l\mathbf{C}_0\mathbf{D}^{l^H}\mathbf{D}^l\bar{\mathbf{v}}_n = \mathbf{D}^l\mathbf{C}_0\bar{\mathbf{v}}_n = \lambda_n\mathbf{D}^l\bar{\mathbf{v}}_n \quad (21)$$

where the second equality above follows from the fact that $\mathbf{D}$ is a permutation matrix and, hence, $\mathbf{D}\mathbf{D}^H = \mathbf{I}_{N^2}$.

Now, we construct the eigenvectors of $\mathbf{R}_H = E[\text{vec}(\mathbf{H})\text{vec}(\mathbf{H})^H]$. As shown in (21), the $\mathbf{C}_l$'s are modified versions of $\mathbf{C}_0$ where the first Nl rows are cyclically shifted downwards and the first Nl columns are cyclically shifted to the left. On the other hand, the $\bar{\mathbf{v}}_n$'s are over-sampled versions of eigenvectors $\mathbf{v}_n$, where N consecutive zeros are inserted between adjacent elements of $\mathbf{v}_n$. Therefore, as long as $l \neq 0$, the j-th element of each row of $\mathbf{C}_l$ satisfies $\mathbf{C}_l(i,j) = 0$ where $\bar{\mathbf{v}}_n(j) \neq 0$ and $i, j = 0, 1, \cdots, N^2 - 1$. Hence, $\mathbf{C}_l\bar{\mathbf{v}}_n = \mathbf{0}$ for $l \neq 0$. Following the above argument and using (11) we can write

$$\mathbf{R}_H\bar{\mathbf{v}}_n = \left(\sum_{i=0}^{L-1}\mathbf{C}_i\right)\bar{\mathbf{v}}_n = \mathbf{C}_0\bar{\mathbf{v}}_n + \sum_{i=1}^{L-1}\mathbf{C}_i\bar{\mathbf{v}}_n = \lambda_n\bar{\mathbf{v}}_n \quad (22)$$

Hence, the eigenvectors of $\mathbf{C}_0$ are also eigenvectors of $\mathbf{R}_H$. In a similar manner, for every eigenvector of $\mathbf{C}_l$ we can show that

$$\begin{aligned}\mathbf{R}_H\mathbf{D}^l\bar{\mathbf{v}}_n &= \left(\sum_{i=0}^{L-1}\mathbf{C}_i\right)\mathbf{D}^l\bar{\mathbf{v}}_n \\ &= \mathbf{C}_l\mathbf{D}^l\bar{\mathbf{v}}_n + \sum_{\substack{i=0 \\ i\neq l}}^{N-1}\mathbf{C}_i\mathbf{D}^l\bar{\mathbf{v}}_n = \lambda_n\mathbf{D}^l\bar{\mathbf{v}}_n \end{aligned}$$
$$(23)$$

Equation (23) follows from the observation that the permutation matrix $\mathbf{D}^l$ cyclically shifts the first Nl elements of $\bar{\mathbf{v}}_n$ downwards. Unless $i = l$, the q-th element of each row of $\mathbf{C}_i$ is $\mathbf{C}_i(p,q) = 0$ where $\mathbf{D}^l\bar{\mathbf{v}}_n(q) \neq 0$ and $p, q = 0, 1, \cdots, N^2 - 1$ and, hence, $\mathbf{C}_i\mathbf{D}^l\bar{\mathbf{v}}_n = \mathbf{0}$. Thus, every eigenvector of $\mathbf{C}_l$ is also an eigenvector of $\mathbf{R}_H$. This produces the NL eigenvalues and the NL corresponding eigenvectors of $\mathbf{R}_H$. Since $\text{rank}(\mathbf{R}_H) = NL$, those eigenvalues are all the nonzero eigenvalues of $\mathbf{R}_H$.

Proposition 1

Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$ denote the eigenvalues of the matrix $\mathbf{J}$ defined in (15), and let $\mathbf{v}_1, \ldots, \mathbf{v}_N$ denote the corresponding eigenvectors. Then, the covariance matrix $\mathbf{R}_H$, defined in (11), has the NL eigenvalues

---

[2]See Remark 4 below for the generalization to the case of an arbitrary power delay profile.

$$\underbrace{\lambda_1, \ldots, \lambda_1}_{L \text{ times}}, \ldots, \underbrace{\lambda_N, \ldots, \lambda_N}_{L \text{ times}} \qquad (24)$$

with corresponding eigenvectors

$$\bar{\mathbf{v}}_1, \mathbf{D}\bar{\mathbf{v}}_1, \ldots, \mathbf{D}^{L-1}\bar{\mathbf{v}}_1, \ldots, \bar{\mathbf{v}}_N, \mathbf{D}\bar{\mathbf{v}}_N, \ldots, \mathbf{D}^{L-1}\bar{\mathbf{v}}_N \quad (25)$$

where $\bar{\mathbf{v}}_n$ is the over-sampled version of $\mathbf{v}_n$ defined in (16) and $\mathbf{D}$ is the permutation matrix defined in (19).

Remarks

1) Proposition 1 shows that the EVD of the N^2-by-N^2 matrix $\mathbf{R}_H$ can be obtained from the EVD of the N-by-N matrix $\mathbf{J}$.

2) The EVD of $\mathbf{J}$ can be performed off-line as it is only a function of the number of sub-carriers N and the Doppler frequency f_d.

3) We can derive an approximate analytical closed form for the EVD of $\mathbf{J}$ by noting that the matrix $\mathbf{J}$ is a Toeplitz Hermitian matrix and, hence, can be approximated by a circulant matrix for large N [15]. Therefore, we can write $\mathbf{J} \approx \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^H$, where $\mathbf{\Lambda}$ is a diagonal matrix whose (m, m) element is given by $\boldsymbol{\lambda}(m,m) \triangleq \sum_{k=0}^{N-1} J_0\left(2\pi k f_d T_s\right) e^{\frac{-j2\pi km}{N}}$.

4) So far, we have assumed that the CIR power delay profile (PDP) is uniform, meaning that all CIR taps have the same unit power, i.e. $E[| h_n(l) |^2] = E[| h_n(l') |^2]$ for $l \neq l'$. We can relax this assumption to an arbitrary PDP as follows. Assuming that the l-th CIR tap has a variance of $\gamma_l = E[| h_n(l) |^2]$, it is easy to show that $E[h_m(l)h_n(l)^*] = \gamma_l J(m - n)$, hence, the covariance matrix of $\text{vec}(\mathbf{A}_l)$, which we have denoted by $\mathbf{C}_l$, is scaled by a factor of γ_l compared to the uniform PDP case. Therefore, the NL eigenvectors of the covariance matrix $\mathbf{R}_H$ remain the same $(\bar{\mathbf{v}}_1, \mathbf{D}\bar{\mathbf{v}}_1, \ldots, \mathbf{D}^{L-1}\bar{\mathbf{v}}_1, \ldots, \bar{\mathbf{v}}_N, \mathbf{D}\bar{\mathbf{v}}_N, \ldots, \mathbf{D}^{L}\bar{\mathbf{v}}_N)$ while the eigenvalues are scaled by the individual variance of each tap, i.e. $\underbrace{\gamma_0 \lambda_1, \ldots, \gamma_{L-1} \lambda_1}_{L \text{ times}}, \ldots, \underbrace{\gamma_0 \lambda_N, \ldots, \gamma_{L-1} \lambda_N}_{L \text{ times}}$.

C. Eigenvalue Decomposition of $\mathbf{R}_G$

From (7), we can deduce the EVD of $\mathbf{R}_G$ from that of $\mathbf{R}_H$ since the matrix $(\mathbf{Q}^* \otimes \mathbf{Q})$ is orthonormal. This observation is the basis for the following proposition.

Proposition 2

Consider the setting of Proposition 1. The covariance matrix $\mathbf{R}_G$ of the CFR matrix $\mathbf{G}$ has the NL nonzero eigenvalues

$$\underbrace{\lambda_1, \ldots, \lambda_1}_{L \text{ times}}, \ldots, \underbrace{\lambda_N, \ldots, \lambda_N}_{L \text{ times}} \qquad (26)$$

with corresponding eigenvectors $\boldsymbol{\mathcal{G}}_p$ $(1 \leq p \leq NL)$ given by

$$(\mathbf{Q}^* \otimes \mathbf{Q})\bar{\mathbf{v}}_1, \ldots, (\mathbf{Q}^* \otimes \mathbf{Q})\mathbf{D}^{L-1}\bar{\mathbf{v}}_1, \ldots,$$
$$(\mathbf{Q}^* \otimes \mathbf{Q})\bar{\mathbf{v}}_N, \ldots, (\mathbf{Q}^* \otimes \mathbf{Q})\mathbf{D}^{L-1}\bar{\mathbf{v}}_N \qquad (27)$$

IV. MODEL REDUCTION FOR BAYESIAN MMSE CHANNEL ESTIMATION

In the previous section, we found the EVD of $\mathbf{R}_G$ in terms of that of $\mathbf{J}$. In this section, we show how to estimate $\mathbf{G}$ or equivalently $\text{vec}(\mathbf{G})$ efficiently using the EVD of $\mathbf{R}_G$.

A. Estimating $\mathbf{G}$

We can write $\text{vec}(\mathbf{G})$ in terms of its EVD as follows

$$\text{vec}(\mathbf{G}) = \sum_{p=1}^{NL} \alpha_p \boldsymbol{\mathcal{G}}_p \qquad (28)$$

where the α_p's are independent random variables each with zero mean and variance equal to the eigenvalue λ_p. Equation (28) can be justified as given in Equation (29), where $\lambda_1, \lambda_2, \cdots, \lambda_{NL}$ are the NL eigenvalues of the covariance matrix $\mathbf{R}_G$ and $\boldsymbol{\mathcal{G}}_1, \boldsymbol{\mathcal{G}}_2, \cdots, \boldsymbol{\mathcal{G}}_{NL}$ are the corresponding eigenvectors. We can reduce the parameter estimation space for $\mathbf{G}$ by retaining only those α_i's with large variance and considering the rest as modeling noise. Thus, let N_d denote the number of dominant eigenvalues of $\mathbf{J}$ and let $N_s = N - N_d$ denote the negligible ones. This results in $N_d L$ dominant parameters α_p in (28) while the rest can be considered as noise. In other words, we can approximate (28) as follows

$$\text{vec}(\mathbf{G}) = \sum_{p=1}^{N_d L} \alpha_p \boldsymbol{\mathcal{G}}_p + \widetilde{\boldsymbol{\mathcal{Z}}} \approx \sum_{p=1}^{N_d L} \alpha_p \boldsymbol{\mathcal{G}}_p \qquad (30)$$

where the term $\widetilde{\boldsymbol{\mathcal{Z}}} = \sum_{p=N_d L+1}^{NL} \alpha_p \boldsymbol{\mathcal{G}}_p$ is ignored. Figure 1 shows the percentage of energy of the most dominant eigenvalues of $\mathbf{J}$ normalized to the total energy of all the eigenvalues for $N = 256$ and 1024. It illustrates that for practical choices of normalized Doppler, f_d, up to 20%, the 3 most dominant eigenvalues capture more than 99.99% of the total energy of the all the eigenvalues. Hence, we set $N_d = 3$. For a given Doppler spread, it is possible to calculate $\mathbf{J}$ off-line and, thus, we can select N_d based on a judiciously-determined energy threshold for the eigenvalues. Now since $\mathbf{D} = \mathbf{B}^T \otimes \mathbf{I}_N$ and $\bar{\mathbf{v}}_n = \text{vec}(\text{diag}(\mathbf{v}_n))$, using Proposition 2 and Theorem T2.4 from [12] we can rewrite $\boldsymbol{\mathcal{G}}_p$ as

$$\begin{aligned}
\boldsymbol{\mathcal{G}}_p &= (\mathbf{Q}^* \otimes \mathbf{Q})(\mathbf{B}^{l^T} \otimes \mathbf{I})\text{vec}(\text{Diag}(\mathbf{v}_n)) \\
&= (\mathbf{Q}^* \mathbf{B}^{l^T} \otimes \mathbf{Q})\text{vec}(\text{Diag}(\mathbf{v}_n))
\end{aligned} \qquad (31)$$

By un-vectorizing both sides of (31) and using Theorem T2.13 from [12] we get

$$\begin{aligned}
\mathbf{G}_p &= \mathbf{Q}\text{diag}(\mathbf{v}_n)\mathbf{B}^l \mathbf{Q}^H \\
&= \underbrace{\mathbf{Q}\text{diag}(\mathbf{v}_n)\mathbf{Q}^H}_{\mathbf{C}_{\mathbf{v}_n}} \cdot \underbrace{\mathbf{Q}\mathbf{B}^l \mathbf{Q}^H}_{\mathbf{D}_w^l} \\
&= \mathbf{C}_{\mathbf{v}_n} \cdot \mathbf{D}_w^l
\end{aligned} \qquad (32)$$

where $\mathbf{C}_{\mathbf{v}_n}$ is a circulant matrix and $\mathbf{G}_p$ is such that $\boldsymbol{\mathcal{G}}_p = \text{vec}(\mathbf{G}_p)$ and n and l are chosen such that $p = N_d l + n$. More specifically, we have

$$l = \left\lfloor \frac{p-1}{N_d} \right\rfloor \quad \text{and} \quad n = (p-1)_{N_d} + 1 \qquad (33)$$

$$\mathbf{R}_G \triangleq E\left[(\text{vec}(\mathbf{G})\text{vec}(\mathbf{G})^H\right] = E\left(\sum_{p=1}^{NL}\alpha_p\alpha_p^H\boldsymbol{\mathcal{G}}_p\boldsymbol{\mathcal{G}}_p^H\right)$$
$$= [\boldsymbol{\mathcal{G}}_1\ \boldsymbol{\mathcal{G}}_2\ \cdots\ \boldsymbol{\mathcal{G}}_{NL}]\,E\left(\text{Diag}\left(\mid\alpha_1\mid^2, \mid\alpha_2\mid^2, \cdots, \mid\alpha_{NL}\mid^2\right)\right)[\boldsymbol{\mathcal{G}}_1\ \boldsymbol{\mathcal{G}}_2\ \cdots\ \boldsymbol{\mathcal{G}}_{NL}]^H \qquad (29)$$
$$= [\boldsymbol{\mathcal{G}}_1\ \boldsymbol{\mathcal{G}}_2\ \cdots\ \boldsymbol{\mathcal{G}}_{NL}]\,\text{Diag}\left(\lambda_1, \lambda_2, \cdots, \lambda_{NL}\right)[\boldsymbol{\mathcal{G}}_1\ \boldsymbol{\mathcal{G}}_2\ \cdots\ \boldsymbol{\mathcal{G}}_{NL}]^H$$
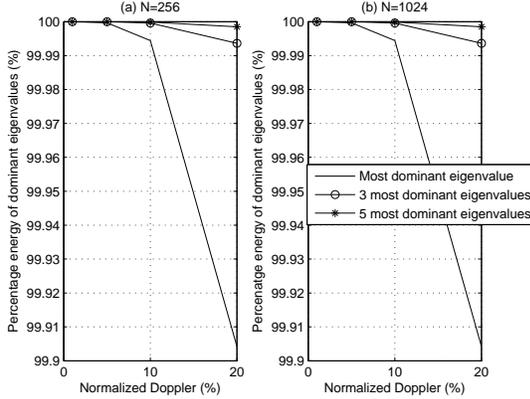

Fig. 1. Variation of percentage energy of the dominant eigenvalues of $\mathbf{J}$ with normalized Doppler.

where $(.)_N$ is the *modulo-N* operation and $\lfloor.\rfloor$ denotes the floor function which maps a real number to the integer less than or equal to it. Now, from (30) and (32), we have

$$\mathbf{G} \approx \sum_{p=1}^{N_dL}\alpha_p\mathbf{G}_p \approx \sum_{p=1}^{N_dL}\alpha_p\mathbf{Q}\text{diag}(\mathbf{v}_n)\mathbf{B}^l\mathbf{Q}^H \qquad (34)$$

where l and n are related to p through (33). Alternatively, we can write (34) as

$$\mathbf{G} \approx \sum_{l=0}^{L-1}\sum_{n=1}^{N_d}\alpha_{N_dl+n}\mathbf{Q}\text{diag}(\mathbf{v}_n)\mathbf{B}^l\mathbf{Q}^H \qquad (35)$$

B. Bandedness Constraint

Now, we invoke another constraint on the $\mathbf{G}_p$'s in (34), and hence on $\mathbf{G}$, namely, the constraint that they are approximately banded matrices with, say, M main diagonals (where M is an odd integer). To impose this constraint, we set all elements of $\mathbf{G}_p$ outside these M diagonals to zero. Analytically, $\mathbf{B}$, given in (10), is a shifting matrix which is circulant as well. Depending on the channel tap index, i.e. l, $\text{diag}(\mathbf{v}_n)\mathbf{B}^l$ is either a diagonal or a shifted diagonal matrix. The matrix $\mathbf{C}_{\mathbf{v}_n}$ in (32) is formed by the cyclic shift of a column with only few leading dominant entries. Hence, we can approximate $\mathbf{C}_{\mathbf{v}_n}$ as a banded matrix. Note that multiplying $\mathbf{C}_{\mathbf{v}_n}$ by $\mathbf{D}_w^l$ only introduces a weighted phase change in the columns of the $\mathbf{C}_{\mathbf{v}_n}$, leaving the absolute value of the columns unchanged. Hence, multiplying $\mathbf{C}_{\mathbf{v}_n}$ by $\mathbf{D}_w^l$ preserves the banded property of $\mathbf{C}_{\mathbf{v}_n}$ and $\mathbf{G}_p$ as well. Therefore, we can approximate $\mathbf{G}_p$ as a banded matrix. Under the banded assumption on $\mathbf{G}_p$ and the circulant approximation of $\mathbf{J}$ for large N, the (r,c)-th element of $\mathbf{G}_p$ as given by (32) can be expressed as in Equation (36)

where $v_n(i)$ is the i-th element of $\mathbf{v}_n$. Under the banded constraint, $\mathbf{G}$ in (34) can be further simplified using (36). We are now ready to formulate the input-output equations that are used to estimate the dominant parameters α_p for $1 \le p \le N_dL$ in Equation (35) as follows

$$\boldsymbol{\mathcal{Y}} = \mathbf{G}\boldsymbol{\mathcal{X}} + \boldsymbol{\mathcal{Z}} \approx \sum_{p=1}^{N_dL}\alpha_p\mathbf{G}_p\boldsymbol{\mathcal{X}} + \boldsymbol{\mathcal{Z}} \qquad (37)$$

$$= \sum_{p=1}^{N_dL}\alpha_p\boldsymbol{\mathcal{E}}_p + \boldsymbol{\mathcal{Z}} \qquad (38)$$

where $\boldsymbol{\mathcal{Y}} = [\mathcal{Y}(0), \ldots, \mathcal{Y}(N-1)]^T$, $\boldsymbol{\mathcal{X}} = [\mathcal{X}(0), \ldots, \mathcal{X}(N-1)]^T$, and $\boldsymbol{\mathcal{Z}} = [\mathcal{Z}(0), \ldots, \mathcal{Z}(N-1)]^T$ are the frequency-domain receive, transmit, and noise vectors, respectively, and $\boldsymbol{\mathcal{E}}_p \triangleq \mathbf{G}_p\boldsymbol{\mathcal{X}}$. Now, some entries of $\boldsymbol{\mathcal{X}}$ are pilot sub-carriers while the rest are data sub-carriers. The presence of ICI forces us to group enough pilot sub-carriers together to create an output signal at the k-th sub-carrier that is independent of unknown data sub-carriers. Specifically, since each row of $\mathbf{G}_p$ in (37) has M non-zero elements, we need to group the pilots in clusters of size greater than or equal to M to generate a data-independent output vector that can be used to solve for the α_p's. Since M is odd in general, from (37), the input-output equation at the k-th sub-carrier is given by

$$\mathcal{Y}(k) = \sum_{p=1}^{N_dL}\alpha_p\underbrace{\sum_{m=-\frac{M-1}{2}}^{\frac{M-1}{2}}G_p(k, k+m)\mathcal{X}(k+m)}_{\mathcal{E}_p(k)} + \mathcal{Z}(k)$$

$$= \sum_{p=1}^{N_dL}\alpha_p\mathcal{E}_p(k) + \mathcal{Z}(k) \qquad (39)$$

where $\frac{M-1}{2} \le k \le N - \frac{M-1}{2}$. Thus, for $\mathcal{E}_p(k)$ to be known (i.e. independent of the data), the sub-carriers $\mathcal{X}(k - \frac{M-1}{2}), \cdots, \mathcal{X}(k + \frac{M-1}{2})$ must be pilot sub-carriers resulting in pilot grouping.

Let $\{k_1, k_2, \ldots, k_T\}$ denote the set of sub-carrier indices that can be used for training. We can prune (37) of all sub-carriers that do not belong to the training set, resulting in the following linear system of T equations in N_dL unknowns

$$\underline{\boldsymbol{\mathcal{Y}}} = \sum_{p=1}^{N_dL}\alpha_p\underline{\boldsymbol{\mathcal{E}}}_p + \underline{\boldsymbol{\mathcal{Z}}}$$

$$G_p(r,c) = \begin{cases} \frac{1}{N}\sum_{m=0}^{N-1} e^{\frac{i2\pi}{N}[mc-r(m+l)_N]} v_n(m+l)_N & |r-c| \leq \frac{M-1}{2} \;\;\&\;\; |r-c| \geq N - \frac{M-1}{2} \\ 0 & \frac{M-1}{2} < |r-c| < N - \frac{M-1}{2} \end{cases} \quad (36)$$

where

$$\underline{\mathcal{Y}} = \begin{bmatrix} \mathcal{Y}(k_1) \\ \mathcal{Y}(k_2) \\ \vdots \\ \mathcal{Y}(k_T) \end{bmatrix} \quad \underline{\mathcal{E}}_p = \begin{bmatrix} \mathcal{E}_p(k_1) \\ \mathcal{E}_p(k_2) \\ \vdots \\ \mathcal{E}_p(k_T) \end{bmatrix} \quad (40)$$

In matrix form, we can write

$$\underline{\mathcal{Y}} = \underline{\mathbf{E}}_p \boldsymbol{\alpha} + \underline{\mathcal{Z}} \quad (41)$$

where

$$\underline{\mathbf{E}}_p = \begin{bmatrix} \underline{\mathcal{E}}_1 \cdots \underline{\mathcal{E}}_{N_dL} \end{bmatrix} \quad \text{and} \quad \boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \cdots \alpha_{N_dL} \end{bmatrix}^T$$

This is a Bayesian estimation model since the unknown random vector $\boldsymbol{\alpha}$ is assumed zero mean with covariance matrix $\mathbf{R}_\alpha = \mathrm{diag}([\lambda_1, \ldots, \lambda_{N_dL}])$. Hence, we can compute $\boldsymbol{\alpha}$ for a generalized noise covariance matrix $\mathbf{R}_z$ using the following linear minimum mean square error (LMMSE) estimator [16]

$$\begin{aligned} \hat{\boldsymbol{\alpha}} &= \mathbf{R}_\alpha \underline{\mathbf{E}}_p^H \left[\underline{\mathbf{E}}_p \mathbf{R}_\alpha \underline{\mathbf{E}}_p^H + \mathbf{R}_z \right]^{-1} \underline{\mathcal{Y}} \\ &= \left[\mathbf{R}_\alpha^{-1} + \underline{\mathbf{E}}_p^H \mathbf{R}_z^{-1} \underline{\mathbf{E}}_p \right]^{-1} \underline{\mathbf{E}}_p^H \mathbf{R}_z^{-1} \underline{\mathcal{Y}} \triangleq \mathbf{W} \underline{\mathcal{Y}} \quad (42) \end{aligned}$$

For white noise, the second form in (42) is preferred since it reduces the size of the inverted matrix from T to N_dL where $T \geq N_dL$. The performance of this estimator is measured by the error vector $\boldsymbol{\epsilon} = \boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}$ whose mean is zero and whose covariance matrix is [16]

$$\mathbf{C}_\epsilon = \left[\mathbf{R}_\alpha^{-1} + \underline{\mathbf{E}}_p^H \mathbf{R}_z^{-1} \underline{\mathbf{E}}_p \right]^{-1} \quad (43)$$

Hence, the MSE in estimating α_i is $\mathrm{MSE}(\hat{\alpha}_i) = \mathbf{C}_\epsilon(i,i)$. In Appendix A, we show how to select the pilot group size to minimize the MSE of the $\hat{\alpha}_i$'s .

Algorithm Summary:

Input : N, N_d, f_d, power delay profile (otherwise assumed uniform) and noise covariance matrix (otherwise assumed white).

1) Determine the N_d dominant eigenvectors and eigenvalues of $\mathbf{J}$ exactly off-line (e.g. using a look-up table) or approximately using the circulant approximation (for large N).
2) Determine the N_dL dominant eigenvalues and eigenvectors of $\mathbf{R}_H$ off-line using (24) and (25) in Proposition 1.
3) Determine the N_dL dominant eigenvalues and eigenvectors of $\mathbf{R}_G$ off-line using Equations (26) and (27) in Proposition 2.
4) Compute $\hat{\boldsymbol{\alpha}}$ using (42). Note that the $N_dL \times T$ matrix $\mathbf{W}$ can be computed off-line. Hence, the on-line computation consists of multiplying $\mathbf{W}$ by the received data vector $\underline{\mathcal{Y}}$ only !
5) Approximate $\mathbf{G}$ as in (34) under the banded constraint and the circulant approximation (for large N).

C. Complexity Comparison

In this section, we compare the complexity of our proposed algorithm with the widely-cited algorithm proposed by Mostofi and Cox in [9] (referred to henceforth as the MC algorithm). Both algorithms estimate the CFR matrix $\mathbf{G}$ and then detect information symbols using the estimate of $\mathbf{G}$. In this section, we will compare the complexity of both algorithms in estimating $\mathbf{G}$. Their performance will be compared in Section VI.

1) Complexity of Our Proposed Algorithm: For given N and f_d, Steps 1-3 of the algorithm summary need to be calculated only once offline. Hence, we consider the complexity of Steps 4 and 5 only as described in Table II. For given N, f_d and SNR, Step 4 except $\mathbf{S}_6$ can be computed offline to reduce processing complexity and delay. In our analysis, we assumed $\underline{\mathcal{Z}}$ to be white Gaussian noise and, therefore, used the second form of (42) to estimate $\boldsymbol{\alpha}$.

2) Complexity of MC Algorithm: Next, in Table III, we evaluate the complexity of estimating $\mathbf{G}$ as proposed in [9]. For concreteness, we have quantified the complexity of both algorithms in terms of floating point operations (FLOP) where each FLOP represents a real-valued addition or multiplication operation. Conventional processors require more than four clock cycles for multiplication of real numbers. Pipeline arrangements speed up the process and make it possible to perform one FLOP per clock cycle. In Figure 2 we have compared the number of required FLOPs of the two algorithms to process each OFDM symbol for different FFT sizes and CIR lengths $L = 3$ and 6. For example, to process one OFDM symbol (real-time complexity only) of size $N=1024$ assuming a 3-tap channel, a TI TMS320C6455-1000 DSP with a processing capability of 8000 MIPS, takes around 25.42 μs and 9.79 μs for the exact and approximate circulant versions of our proposed algorithm, respectively, compared to 19.42 μs required by the MC algorithm. Although the complexity of our algorithm increases linearly with L, significant complexity reductions are achieved by performing most of the computations offline. Hence, the real-time complexity of our algorithm is comparable to that of the MC algorithm for small to moderate CIR lengths and is less than the MC algorithm under the circulant approximation which becomes more accurate for large N. The MC algorithm suffers from higher latency and requires larger buffer because it processes three adjacent OFDM symbols together while our algorithm processes one OFDM symbol at a time.

V. REDUCED-COMPLEXITY ICI EQUALIZATION

Based on the banded structure of $\mathbf{G}$, a reduced-tap FEQ in the form of an FIR filter designed using the MMSE criterion, hence the name FIR-MMSE FEQ, was proposed in [11] and is based only on the following $M \times (2M-1)$ sub-matrix of $\mathbf{G}$ which is shown in Figure 3.

TABLE II
COMPLEXITY OF OUR PROPOSED ALGORITHM

	Operation	Complexity	Remarks
Step 4	$\mathbf{S}_1$: Forming $\underline{\mathbf{E}}_p$	TN_dLM complex multiplications $TN_dL(M-1)$ complex additions	Can be computed offline for given N, f_d and SNR
	$\mathbf{S}_2$: Inverse of $\mathbf{R}_\alpha$	N_dL real divisions	
	$\mathbf{S}_3$: Calculate $\mathbf{S}_1^H\mathbf{R}_z^{-1}\mathbf{S}_1$	$\frac{1}{2}(N_dL)^2T$ complex multiplications $\frac{1}{2}(N_dL)^2(T-1)$ complex additions $(N_dL)^2$ real divisions	
	$\mathbf{S}_4$: Inverse of $(\mathbf{S}_2+\mathbf{S}_3)$	$N_dL(N_dL-1)(N_dL+1)$ complex divisions $N_dL(N_dL-1)(N_dL+1)$ complex multiplications $N_dL(N_dL-1)(N_dL+1)$ complex subtractions	
	$\mathbf{S}_5$: Calculate $\mathbf{S}_4\mathbf{S}_1^H\mathbf{R}_z^{-1}$	$T(N_dL)^2$ complex multiplications $T(N_dL)(N_dL-1)$ complex additions N_dL real divisions	
	$\mathbf{S}_6$: Calculate $\mathbf{S}_5\boldsymbol{\mathcal{Y}}$	TN_dL complex multiplications $(T-1)N_dL$ complex additions	Real-time
Step 5	Approximate $\mathbf{G}$ using (34)	$NMN_d(L-1)+MN_d$ complex multiplications $NM(N_dL-1)$ complex additions	Real-time

TABLE III
COMPLEXITY OF MC ALGORITHM

| Operation	Complexity	Remarks
Estimate $H_{l_{i,0}}$, $0 \le i \le (T-1)$	$T(2M-1)$ complex divisions	Real-time
Estimate h_k^{ave}, $0 \le k \le (L-1)$	$3LT(2M-1)$ complex multiplications $3L\left[(2M-1)T-1\right]$complex additions $3L$ real divisions	
Calculating slope	$2L$ complex subtractions $2L$ real divisions	
FFT	$3NM$ complex multiplications $3N(M-1)$ complex additions	
Estimate $\mathbf{G}$	$2NM$ complex multiplications $2NM$ complex additions	
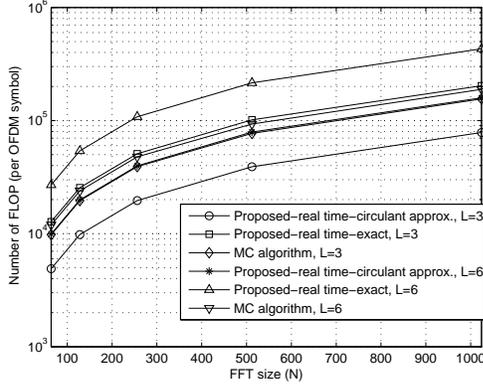
Fig. 2. Real-time complexity comparison of our proposed and MC algorithms, $T = \lfloor \frac{\lfloor 0.15N \rfloor}{2M-1} \rfloor (N_p - M + 1)$, $N_d = 3$ and $M = 3$.

Fig. 3. Banded Structure of the $\mathbf{G}$ matrix.

The M-tap FIR-MMSE FEQ is given by

$$\boldsymbol{\mathcal{W}}_m = \boldsymbol{\mathcal{C}}_m^H(\mathbf{G}_m\mathbf{G}_m^H + \sigma^2\mathbf{I}_M)^{-1} \qquad (44)$$

where $\boldsymbol{\mathcal{C}}_m = \mathbf{G}_m(:,M)$ is the middle column of $\mathbf{G}_m$ shown in Equation (45). Hence, the M-tap FEQ output for the m-th subcarrier is

$$\hat{\boldsymbol{\mathcal{X}}}_m = \boldsymbol{\mathcal{W}}_m\boldsymbol{\mathcal{Y}}_m \qquad (46)$$

where $\boldsymbol{\mathcal{Y}}_m = [\mathcal{Y}_{\left(m-\frac{M-1}{2}\right)_N} \ldots \mathcal{Y}_{\left(m+\frac{M-1}{2}\right)_N}]^T$. Instead of detecting all sub-carriers of an OFDM symbol simultaneously using an $N \times N$ matrix inversion, our proposed FIR-MMSE
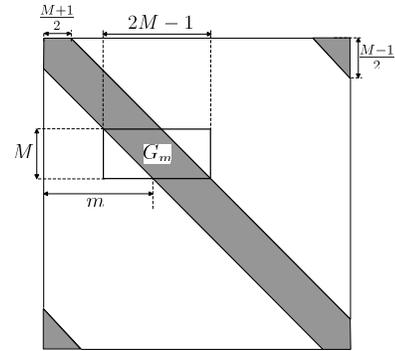
FEQ detects each subcarrier individually, but taking ICI from neighboring sub-carriers into account, requiring N matrix inversions each of size $M \times M$ (where $M \ll N$), which significantly reduces the detection complexity.

VI. SIMULATION RESULTS

In this section, we compare the performance of our proposed algorithm with the MC algorithm for a mobile coded OFDM system for large (N=1024) and small ($N = 256$) FFT sizes. A rate-$\frac{1}{2}$ convolutional code with bit interleaving is implemented. Unless otherwise stated, we assume a high Doppler frequency of 10% (normalized to the subcarrier spacing). Furthermore, we have assumed the 3-tap SUI-3 channel model with $L = 3$ and having path delays of 0, 0.5 and 1 μs with powers of 0,-5

$$\mathbf{G}_m = \begin{bmatrix} G_{\left(m-\frac{M-1}{2}\right)_N,(m-M+1)_N} & \cdots & G_{\left(m-\frac{M-1}{2}\right)_N,(m+M-1)_N} \\ \vdots & \ddots & \vdots \\ G_{m,(m-M+1)_N} & \cdots & G_{m,(m+M-1)_N} \\ \vdots & \ddots & \vdots \\ G_{\left(m+\frac{M-1}{2}\right)_N,(m-M+1)_N} & \cdots & G_{\left(m+\frac{M-1}{2}\right)_N,(m+M-1)_N} \end{bmatrix} \tag{45}$$

and -10 dB, respectively. For a fair comparison with the MC algorithm, we estimated the channel assuming the variances of its CIR taps to be equal to 1 (i.e. uniform PDP) by ignoring any knowledge of the PDP at the receiver.

We found that for a normalized Doppler of 10% with the afore-mentioned FFT sizes, $\mathbf{G}$ can be well approximated by a banded matrix with $M = 3$ diagonals and the first 3 eigenvalues of $\mathbf{J}$ in (15) are dominant, i.e. $N_d = 3$. Hence, we retained only $N_d L = 9$ eigenvalues of $\mathbf{R}_H$ and $\mathbf{R}_G$ (see (30)) and computed the eigenvalues and eigenvectors off-line. For data detection, we implemented a 3-tap FEQ computed using (44). We assumed a 15% pilot training overhead with equally-spaced pilot clusters. In each cluster, we have N_p pilots generated as random BPSK (± 1) symbols. Using Equations (41) and (42), we estimate $\boldsymbol{\alpha}$ where $T = \lfloor \frac{\lfloor 0.15N \rfloor}{2M-1} \rfloor (N_p - M + 1)$. Figures 4 and 5 show the BER achieved by our proposed algorithm for $N = 256$ and 1024, respectively, with $N_p = 3$ and 5 pilots per cluster. As performance benchmarks, we also show the achievable BER with perfect CSI when $\mathbf{G}$ is constrained to be a banded matrix (with $M = 3$ diagonals) and when it is not. It can be seen from both figures that the performance loss of our channel estimation algorithm with 5 pilots per cluster from the case of perfect CSI with banded G is negligible. Figure 6 depicts the BER results for a very high Doppler of 20%. In this case, we set $M = 5$ to account for the resulting severe ICI. Again, this figure demonstrates that the best estimated CSI performance is achieved with $(2M-1) = 9$ pilots per cluster. This corroborates our conjecture in Appendix A that the optimum number of pilot sub-carriers per cluster is $(2M - 1)$. Figures 4 and 5 also show that the banded constraint which is imposed to reduce channel estimation and ICI equalization complexity results in only 1 dB perform loss at $BER = 10^{-4}$ which is a modest penalty. Figures 7 and 8 demonstrate the significant performance gains achieved by our algorithm (both exact and approximate circulant versions) over the MC algorithm for $N = 256$ and 1024, respectively, where the later has an error floor at $BER = 10^{-3}$. Finally Figure 8 shows that the BER performance gain of our algorithm with perfect PDP knowledge at the receiver is negligible compared to the case of no PDP knowledge. This illustrates that our algorithm is not sensitive to PDP knowledge.

VII. CONCLUSIONS

In this paper, we proposed a novel algorithm for the estimation of the channel and ICI coefficients in OFDM systems under high-mobility conditions. The key ideas in developing this algorithm are exploiting the time and frequency correlations of the channel taps, the (approximately)
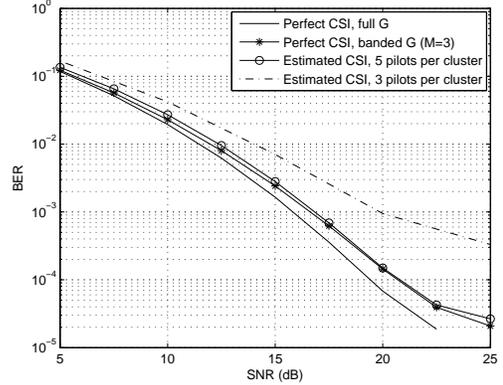

Fig. 4. BER comparison between perfect CSI and estimated CSI using our proposed algorithm for N=256, f_d=10% and $M = 3$.

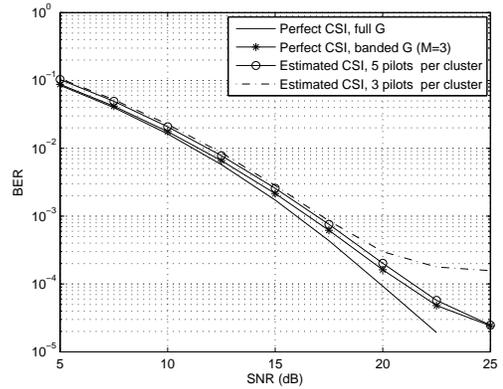
Fig. 5. BER comparison between perfect CSI and estimated CSI using our proposed algorithm for N=1024, f_d=10% and $M = 3$.
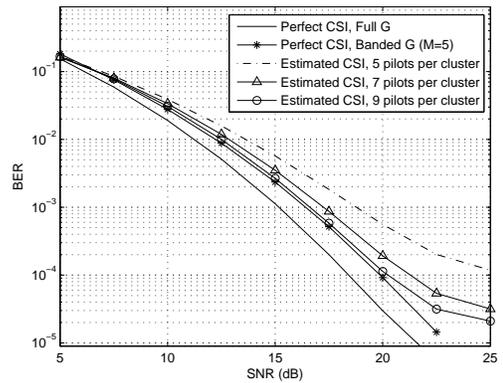

Fig. 6. BER comparison between perfect CSI and estimated CSI using our proposed algorithm for N=1024, f_d=20% and $M = 5$.
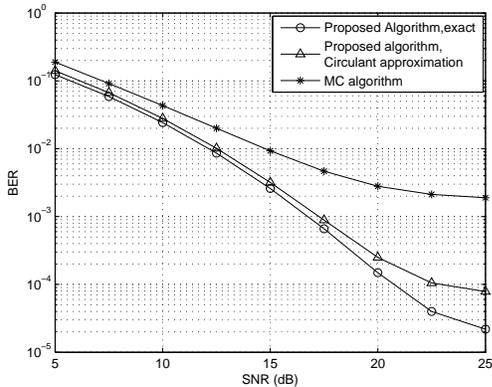
Fig. 7. BER comparison between our proposed algorithm (both exact and approximate versions) and MC algorithm for N=256.
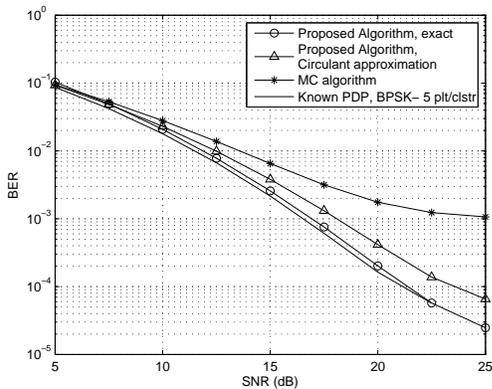

Fig. 8. BER comparison between our proposed algorithm (both exact and approximate versions) and MC algorithm for N=1024.

banded structure of the frequency-domain channel matrix, the asymptotic equivalence of Toeplitz and circulant matrices, and reducing the dimensionality of the parameter estimation space by retaining only the dominant terms in an off-line eigen-decomposition. Additional performance gains are achieved by optimizing the pilot subcarrier group size bringing the performance very close to the case of perfect CSI at practical training overhead ratios. Detailed performance and complexity comparisons of our proposed algorithm with a widely-cited recent algorithm demonstrate a significant performance advantage at comparable real-time complexity levels.

APPENDIX A
OPTIMIZING THE PILOT GROUP SIZE

Assuming that the $\mathcal{Z}(k)$'s are i.i.d samples with zeros mean and variance of σ_z^2, (43) can be written as

$$\mathbf{C}_\epsilon = (\mathbf{R}_\alpha^{-1} + \frac{1}{\sigma_z^2} \underline{\mathbf{E}}_p^H \underline{\mathbf{E}}_p)^{-1} \qquad (47)$$

Now, $\mathbf{C}_\epsilon^{-1} \triangleq \mathbf{R}_\alpha^{-1} + \frac{1}{\sigma_z^2} \underline{\mathbf{E}}_p^H \underline{\mathbf{E}}_p$ is a positive-definite Hermitian matrix of size $N_d L \times N_d L$ for any $\mathcal{X} \in \mathbb{C}^N$. Hence, all the eigenvalues of $\mathbf{C}_\epsilon^{-1}$, and in turn of $\mathbf{C}_\epsilon$, will be real and positive. Let $\lambda_{\mathbf{C}}(i)$, $i = 1, 2, \cdots, N_d L$ be the $N_d L$

eigenvalues of $\mathbf{C}_\epsilon^{-1}$. For positive real numbers, using Jensen's inequality [17], it can be shown that the arithmetic mean is greater than the harmonic mean. Therefore, we have the following inequality

$$\frac{\sum_{i=1}^{N_d L} \lambda_{\mathbf{C}}(i)}{N_d L} \geq \frac{N_d L}{\sum_{i=1}^{N_d L} \frac{1}{\lambda_{\mathbf{C}}(i)}} \qquad (48)$$

$$\Rightarrow \quad \frac{\text{Trace}(\mathbf{C}_\epsilon^{-1})}{N_d L} \geq \frac{N_d L}{\text{Trace}(\mathbf{C}_\epsilon)}$$

$$\Rightarrow \quad \text{Trace}(\mathbf{C}_\epsilon)\text{Trace}(\mathbf{C}_\epsilon^{-1}) \geq (N_d L)^2$$

$$\Rightarrow \quad \text{Trace}(\mathbf{C}_\epsilon) \geq \frac{(N_d L)^2}{\text{Trace}\left(\mathbf{R}_\alpha^{-1}\right) + \text{Trace}\left(\frac{1}{\sigma_z^2} \underline{\mathbf{E}}_p^H \underline{\mathbf{E}}_p\right)}$$

with equality if and only if $\mathbf{C}_\epsilon^{-1} = k\mathbf{I}_{N_d L}$ where k is a real positive constant. From (47) and using the fact that $\mathbf{R}_\alpha$ is a diagonal matrix, we conclude that $\text{Trace}(\mathbf{C}_\epsilon)$ in minimized by making $\mathbf{R}_E \triangleq \underline{\mathbf{E}}_p^H \underline{\mathbf{E}}_p$ a diagonal matrix. In this paper, we assume equally-spaced [3] pilot clusters where each cluster consists of N_p randomly chosen (± 1) BPSK symbols. Under this pilot structure, $\mathbf{R}_E$ can not be made diagonal in general and hence $\mathbf{C}_\epsilon$ will also be non-diagonal. The cost function $\text{Trace}(\mathbf{C}_\epsilon)$ measures only the self-variances of the $\hat{\alpha}_i$'s ignoring the co-variances between the $\hat{\alpha}_i$'s which can have significant effect on the BER performance.

When $\mathbf{C}_\epsilon$ is diagonal, $\mathbf{C}_\epsilon^{-1}$ will also be diagonal and the eigenvalues of $\mathbf{C}_\epsilon^{-1}$ are equal to its diagonal elements, i.e. $\lambda_{\mathbf{C}_\epsilon^{-1}}(i) = \mathbf{C}_\epsilon(i,i)$ where $i = 1, 2, \cdots, N_d L$. Therefore, in general, we want to minimize the difference between the eigenvalues and diagonal elements of $\mathbf{C}_\epsilon^{-1}$, i.e. $|\lambda_{\mathbf{C}_\epsilon^{-1}}(i) - \mathbf{C}_\epsilon^{-1}(i,i)|$. By using the Gershgorin circle theorem [4] [18], an upper bound on this difference is given by

$$|\lambda_{\mathbf{C}_\epsilon^{-1}}(i) - \mathbf{C}_\epsilon^{-1}(i,i)| \leq \sum_{i \neq j} |\mathbf{C}_\epsilon^{-1}(i,j)| \qquad (49)$$

Hence, by minimizing the sum of the magnitudes of the off-diagonal elements of $\mathbf{C}_\epsilon^{-1}$ we can achieve our objective. But $\mathbf{R}_\alpha^{-1}$ is diagonal, hence, the off-diagonal elements of $\mathbf{C}_\epsilon^{-1}$ are the off-diagonal elements of $\mathbf{R}_E$. Therefore, our objective is to choose N_p to minimize the sum of the magnitudes of the off-diagonal elements of $\mathbf{R}_E$ subject to the pilot power constraint. The structure of $\mathbf{R}_E$ is given by Equation (50).

To design the optimum pilot symbols for channel estimation, our objective is to make $\mathbf{R}_E$ diagonal or as close to diagonal as possible, i.e. to minimize the sum of the absolute values of the off-diagonal elements of $\mathbf{C}_\epsilon^{-1}$. We have investigated the minimization of this sum as a function of the clustered pilot symbols in the frequency domain with an average pilot power constraint. This turns out to be a non-linear non-convex optimization problem and is outside the scope of this paper which will be investigated in future work [19]. Hence, we determine the optimum pilot group size only and then use

---

[3] For further motivation and more details on equally-spaced pilot groups, see [2].

[4] This theorem assumes a diagonally-dominant matrix which is the case for $\mathbf{C}_\epsilon^{-1}$.

$$\mathbf{R}_E = \underline{\mathbf{E}}_p^H \underline{\mathbf{E}}_p$$
$$= \begin{bmatrix} \boldsymbol{\mathcal{X}}^H\mathbf{G}_1^H\mathbf{G}_1\boldsymbol{\mathcal{X}} & \boldsymbol{\mathcal{X}}^H\mathbf{G}_1^H\mathbf{G}_2\boldsymbol{\mathcal{X}} & \cdots & \boldsymbol{\mathcal{X}}^H\mathbf{G}_1^H\mathbf{G}_{N_dL}\boldsymbol{\mathcal{X}} \\ \boldsymbol{\mathcal{X}}^H\mathbf{G}_2^H\mathbf{G}_1\boldsymbol{\mathcal{X}} & \boldsymbol{\mathcal{X}}^H\mathbf{G}_2^H\mathbf{G}_2\boldsymbol{\mathcal{X}} & \cdots & \boldsymbol{\mathcal{X}}^H\mathbf{G}_2^H\mathbf{G}_{N_dL}\boldsymbol{\mathcal{X}} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\mathcal{X}}^H\mathbf{G}_{N_dL}^H\mathbf{G}_1\boldsymbol{\mathcal{X}} & \boldsymbol{\mathcal{X}}^H\mathbf{G}_{N_dL}^H\mathbf{G}_2\boldsymbol{\mathcal{X}} & \cdots & \boldsymbol{\mathcal{X}}^H\mathbf{G}_{N_dL}^H\mathbf{G}_{N_dL}\boldsymbol{\mathcal{X}} \end{bmatrix} \tag{50}$$
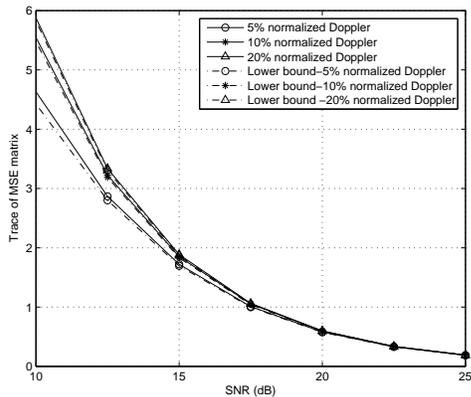

Fig. 10. Comparison of the trace of $\mathbf{C}_\epsilon$ to its lower bound for pilot cluster size of $2M-1$ ($N = 1024$).
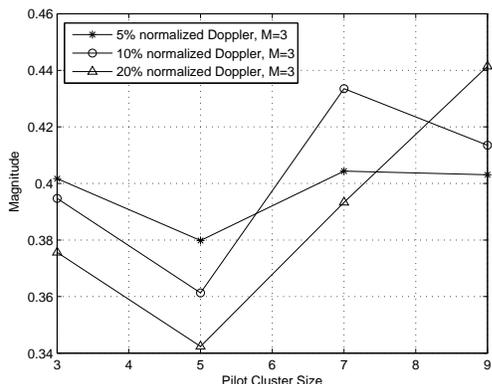

Fig. 9. Sum of the Magnitudes of the off-diagonal elements of $\mathbf{R}_E$ for different pilot cluster sizes ($N = 1024$).

random $BPSK$ symbols as pilot tones which are easy to generate and commonly used in industrial standards.

Since the $\mathbf{G}_i$'s are constrained to be banded matrices with M main diagonals, $\mathbf{G}_i^H\mathbf{G}_j$'s will also be banded but with $(2M-1)$ diagonals. To eliminate ICI effects from adjacent data sub-carriers into pilot sub-carriers we must have at least M pilot sub-carriers in each pilot cluster. An upper-limit on the number of pilot sub-carriers in each cluster, N_p, is derived by observing that if we choose N_p greater than $(2M-1)$ sub-carriers, we will be only adding zeros to the cost function in (49) since the $\mathbf{G}_i^H\mathbf{G}_j$'s are banded with $(2M-1)$ diagonals. On the other hand, for a given pilot overhead ratio, increasing N_p beyond $(2M-1)$ will decrease the number of pilot clusters. Hence, we must have

$$M \le N_p \le 2M-1: \qquad M = 3,5,\cdots \quad \text{and } N_p \text{ odd} \tag{51}$$

Figure 9 shows the variation of the sum of the magnitudes of the off-diagonal elements of $\mathbf{R}_E$ versus N_p as it varies within its range given by (51). It is evident from Figure 9 that the optimum pilot cluster size is $(2M-1)$ pilot sub-carriers for normalized Doppler up to 20%. Figure 10 shows that the difference between the achievable Trace($\mathbf{C}_\epsilon$) with our pilot design and its theoretical lower bound given in (48) is negligible for a pilot cluster size of $(2M-1)$ pilots.

REFERENCES

[1] A. Stamoulis, S. Diggavi, and N. Al-Dhahir, "Intercarrier interference in MIMO OFDM," *IEEE Transactions on Signal Processing*, vol. 50, pp. 2451–2464, Oct. 2002.
[2] L. Tong, B. Sadler, and M. Dong, "Pilot-assisted wireless transmissions: general model, design criteria, and signal processing," *IEEE Signal Processing Magazine*, vol. 21, pp. 12–25, Nov 2004.
[3] Y. Li, "Pilot-symbol-aided channel estimation for OFDM in wireless systems," *IEEE Transactions on Vehicular Technology*, vol. 49, pp. 1207–1215, Jul 2000.
[4] B. Yang, K. B. Letaief, R. Cheng, and Z. Cao, "Channel estimation for OFDM transmission in multipath fading channels based on parametric channel modeling," *IEEE Transactions on Communications*, vol. 49, pp. 467–479, Mar 2001.
[5] J. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels mobile radio," *IEEE Transactions on Vehicular Technology*, vol. 40, pp. 686 – 693, Nov 1991.
[6] T. Al-Naffouri, "An EM-based forward-backward Kalman filter for the estimation of time-variant channels in OFDM," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3924–3930, July 2007.
[7] T. Y. Al-Naffouri and A. A. Quadeer, "A forward-backward Kalman filter-based STBC MIMO OFDM receiver," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, Article ID 158037, doi=10.1155/2008/158037.
[8] W.-G. Song and J.-T. Lim, "Pilot-symbol aided channel estimation for OFDM with fast fading channels," *IEEE Transactions on Broadcasting*, vol. 49, pp. 398–402, Dec 2003.
[9] Y. Mostofi and D. Cox, "ICI mitigation for pilot-aided OFDM mobile systems," *IEEE Transactions on Wireless Communications*, vol. 4, pp. 765–774, Mar. 2005.
[10] W. Jeon, K. Chang, and Y. Cho, "An equalization technique for orthogonal frequency-division multiplexing systems in time-variant multipath channels," *IEEE Transactions on Communications*, vol. 47, pp. 27–32, Jan. 1999.
[11] S. Lu and N. Al-Dhahir, "Coherent and differential ICI cancellation for mobile OFDM with application to DVB-H," *IEEE Transactions on Wireless Communications*, vol. 7, no. 11, pp. 4110–4116, November 2008.
[12] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Transactions on Circuits and Systems*, vol. 25, pp. 772–781, Sep 1978.
[13] S. Chen and T. Yao, "Intercarrier interference suppression and channel estimation for OFDM system in time-varying frequency-selective fading channels," *IEEE Transactions on Consumer Electronics*, vol. 50, pp. 429–435, May 2004.
[14] W. C. Jakes, *Microwave Mobile Communications*. New York: Wiley, 1994.
[15] R. M. Gray, "Toeplitz and circulant matrices: A review," Now Publishers Inc, Tech. Rep., 2002.
[16] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ: Prentice Hall, 1993.

[17] S. G. Krantz, *Handbook of Complex Variables*. Boston, MA: Birkhäuser, 1999.

[18] C. F. V. L. Gene Howard Golub, *Matrix Computations*. Baltimore and London: JHU Press, 1996.

[19] K. Z. Islam, T. Al-Naffouri, and N. Al-Dhahir, "Pilot optimization for comb-type OFDM transmissions over doubly-selective channels," *in preparation.*