On Optimum Pilot Design for Comb-Type OFDM Transmission over Doubly-Selective Channels

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

***Abstract*—We consider comb-type OFDM transmission over doubly-selective channels. Given a fixed number and total power of the pilot subcarriers, we show that the MMSE-optimum pilot design consists of identical equally-spaced clusters where each cluster is zero-correlation-zone sequence.**

***Index Terms*—Pilot optimization, Doppler, ICI, OFDM, ZCZ sequence.**

I. INTRODUCTION

UNDER high mobility, the subcarriers of an orthogonal frequency-division multiplexing (OFDM) symbol lose their orthogonality resulting in performance-limiting Inter-Carrier Interference (ICI). ICI makes channel estimation more challenging since both the sub-carrier frequency responses and the interference caused by each sub-carrier into other subcarriers in each OFDM symbol have to be estimated.

Recently, we proposed in [1] a frequency-domain high-performance computationally-efficient OFDM channel estimation algorithm in the presence of severe ICI. We exploited 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). In most OFDM-based wireless systems, pilot subcarriers are inserted in each OFDM symbol for channel estimation and tracking. When the channel is fixed over each OFDM symbol, the optimum pilot structure consists of equally-spaced individual pilot subcarriers [2], [3]. On the other hand, when the channel varies within the OFDM symbol, [4] argued that the pilot subcarriers should be grouped into equally-spaced clusters. However, [4] did not optimize the pilot subcarrier clusters which is the subject of this paper.

The main contributions of this Letter are

- Proving that the MMSE-optimum pilot design for OFDM over doubly-selective channels consists of identical equally-spaced frequncy-domain pilot clusters.
- Proving that ZCZ sequences (see [5] and references therein) are MMSE-optimal designs for the frequency-domain pilot clusters (see Fig. 1 and Appendix B).
- A new proof (more rigorous than the one in [1]) that the MMSE-optimal OFDM channel estimation error covariance matrix over doubly-selective channels is diagonal (see Appendix A).

Reference [6] proposed a frequency-domain clustered pilot pattern where each cluster has an impulsive structure made of a single pilot subcarrier padded with zero subcarriers as guard band on both sides to eliminate the ICI. This impulsive pilot design ignores signal energy dispersed into the adjacent subcarriers. The novelty of the pilot designs we propose in this paper lies in designing MMSE-optimal non-impulsive periodic pilot clusters which exploit the banded structure of the CFR matrix to increase the accuracy of channel estimation.

This paper is organized as follows. In Section II, we present the doubly-selective channel model and assumptions and briefly review the channel estimation algorithm in [1]. The formulation and solution of the pilot cluster optimization problem are given in Section III. Performance comparisons of our proposed pilot design with the impulsive design are given in Section IV followed by conclusions. For the convenience of the reader, we summarized the key variables used in the paper in Table I.

II. PRELIMINARIES AND BACKGROUND

A. System Model

We start with the following frequency-domain representation of an OFDM system with N subcarriers over a doubly-selective channel

$$\boldsymbol{\mathcal{Y}} \triangleq \mathbf{Q}\mathbf{H}\mathbf{Q}^H\boldsymbol{\mathcal{X}} + \boldsymbol{\mathcal{Z}} = \mathbf{G}\boldsymbol{\mathcal{X}} + \boldsymbol{\mathcal{Z}} \tag{1}$$

where $\mathbf{Q}$ is the N-point FFT matrix and $(.)^H$ is the Hermitian operator. $\boldsymbol{\mathcal{X}}$ is a pilot-data-multiplexed OFDM symbol where certain subcarriers are allocated as pilots surrounded by data subcarriers. We refer to such a multiplexed OFDM symbol structure as comb-type OFDM symbol hereafter. $\mathbf{H}$ is the $N \times N$ time-domain channel matrix which corresponds to convolution with the time-varying CIR coefficients $h_n(l)$ at lag l (for $0 \leq l \leq L-1$) and time instant n and $\boldsymbol{\mathcal{Z}}$ is the frequency-domain noise vector. Over doubly-selective channels, the CFR matrix $\mathbf{G} \triangleq \mathbf{Q}\mathbf{H}\mathbf{Q}^H$ is not diagonal as in time-invariant channels. Rather, the energy of the main diagonal is dispersed into adjacent diagonals depending on the severity of the Doppler spread. We approximate $\mathbf{G}$ as a banded matrix and set all elements of $\mathbf{G}$ outside of M main diagonals to zero where M is odd integer.

Paper approved by H. Arslan, the Editor for Cognitive Radio and OFDM of the IEEE Communications Society. Manuscript received June 24, 2010; no revision.

K. M. Z. Islam and N. Al-Dhahir are with the Department of Electrical Engineering, The University of Texas at Dallas, Richardson, TX, USA (e-mail: zislam@student.utdallas.edu, aldhahir@utdallas.edu).

T. Y. Al-Naffouri is with the Electrical Engineering Department, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia (e-mail: naffouri@kfupm.edu.sa).

This work was supported by King Abdulaziz City for Science and Technology (KACST), Saudi Arabia, Project no. AR 27-98.

Digital Object Identifier 10.1109/TCOMM.2011.020411.100151

TABLE I
LIST OF KEY VARIABLES

Variable	Description
N	FFT size
f_d	Doppler frequency
L	Number of channel impulse response taps
N_T	Total number of pilot subcarriers
N_p	Number of subcarriers in each pilot cluster
$\mathbf{H}$	$(N \times N)$ Time-domain channel matrix
$\mathbf{G}$	$(N \times N)$ Frequency-domain channel matrix
N_d	Number of dominant $\mathbf{R}_H$ eigenvalues for each tap
M	Number of diagonals in banded $\mathbf{G}$
$\boldsymbol{\mathcal{X}}$	$(N \times 1)$ Frequency-domain comb-type input vector
$\mathbf{C}_\epsilon$	$(N_dL \times N_dL)$ Channel estimation error-covariance matrix
L_c	Period of the pilot clusters in $\boldsymbol{\mathcal{X}}$
N_c	Total number of pilot clusters in $\boldsymbol{\mathcal{X}}$

B. Reduced-Complexity Frequency-Domain MMSE OFDM Channel Estimation

In [1], we derived a relation between the eigen-decompositions of $\mathbf{R}_G \triangleq E[\text{vec}(\mathbf{G})\text{vec}(\mathbf{G})^H]$ and $\mathbf{R}_H \triangleq E[\text{vec}(\mathbf{H})\text{vec}(\mathbf{H})^H]$. Assuming Jakes's model with $E[h_m(l)h_n^*(l)] = J_0(2\pi f_d(m-n)T_s) \triangleq J(m-n)$ where f_d is the Doppler frequency and $J_0(\cdot)$ is the zero-order Bessel function of the first kind, we derived the eigen-decomposition of $\mathbf{R}_G$ in closed form in terms of the $N \times N$ symmetric Toeplitz Bessel function matrix $\mathbf{J}$ whose (m,n)-th element is given by $J(m,n) = J(|m-n|) = J_0(2\pi f_d|m-n|T_s)$. Let $\mathbf{G}_p$ denote the matrices formed by un-vectorizing the NL eigenvectors of $\mathbf{R}_G$. We showed in [1] that $\mathbf{G}_p$ can be expressed in terms of the eigenvectors of $\mathbf{J}$ as follows

$$\mathbf{G}_p = \mathbf{Q}\text{diag}(\mathbf{v}_n)\mathbf{B}^l\mathbf{Q}^H; \quad 0 \le l \le (L-1) \text{ and } 1 \le p \le NL \tag{2}$$

where $\mathbf{v}_n, n = 1, 2, \cdots, N$ are the dominant eigenvectors of $\mathbf{J}$ and $\mathbf{B}$ is a circulant shift matrix whose first column is $\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \end{bmatrix}^T$. Considering the N_dL dominant eigenvectors of $\mathbf{R}_G$, (1) can be approximated as follows

$$\boldsymbol{\mathcal{Y}} = \mathbf{G}\boldsymbol{\mathcal{X}} + \boldsymbol{\mathcal{Z}} \approx \sum_{p=1}^{N_dL} \alpha_p \underbrace{\mathbf{G}_p\boldsymbol{\mathcal{X}}}_{\boldsymbol{\mathcal{E}}_p} + \boldsymbol{\mathcal{Z}} = \sum_{p=1}^{N_dL} \alpha_p \boldsymbol{\mathcal{E}}_p + \boldsymbol{\mathcal{Z}} \tag{3}$$

where the α_p's are unknown independent random variables. Considering only the T output subcarriers that result in input-output equations free of unknown data subcarriers in (3), we arrive at the following linear system of T equations in N_dL unknowns

$$\underline{\boldsymbol{\mathcal{Y}}} = \sum_{p=1}^{N_dL} \alpha_p \underline{\boldsymbol{\mathcal{E}}}_p + \underline{\boldsymbol{\mathcal{Z}}} = \underline{\mathbf{E}}_p\boldsymbol{\alpha} + \underline{\boldsymbol{\mathcal{Z}}} \tag{4}$$

where $\underline{\mathbf{E}}_p = \begin{bmatrix} \underline{\boldsymbol{\mathcal{E}}}_1 \cdots \underline{\boldsymbol{\mathcal{E}}}_{N_dL} \end{bmatrix}$ and $\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \cdots \alpha_{N_dL} \end{bmatrix}^T$. This is a Bayesian estimation model since the unknown random vector $\boldsymbol{\alpha}$ is assumed zero mean with covariance matrix $\mathbf{R}_\alpha = \text{diag}([\gamma_1\lambda_1, \ldots, \gamma_{N_dL}\lambda_{N_dL}])$ where γ_p and $\lambda_p, p = 1, 2 \cdots, N_dL$ are the channel power-delay profile (PDP) path variances and the dominant eigenvalues of $\mathbf{R}_G$, respectively. Hence, we can estimate $\boldsymbol{\alpha}$ using the following linear minimum mean square error (LMMSE) estimator [7]

$$\hat{\boldsymbol{\alpha}} = \underbrace{\frac{1}{\sigma_z^2}\left[\mathbf{R}_\alpha^{-1} + \frac{1}{\sigma_z^2}\underline{\mathbf{E}}_p^H\underline{\mathbf{E}}_p\right]^{-1}\underline{\mathbf{E}}_p^H}_{\triangleq \mathbf{W}}\underline{\boldsymbol{\mathcal{Y}}} = \mathbf{W}\underline{\boldsymbol{\mathcal{Y}}} \tag{5}$$

where σ_z^2 is the noise variance (assuming the $\mathcal{Z}(k)$'s in (4) are i.i.d. samples). Given N, f_d and σ_z^2 and the PDP, $\mathbf{W}$ in (5) can be pre-computed and stored in look-up tables to reduce the real-time implementation complexity significantly. The performance of this channel estimator is measured by the error vector $\boldsymbol{\epsilon} = \boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}$ which has zero mean with the following covariance matrix

$$\mathbf{C}_\epsilon = \left[\mathbf{R}_\alpha^{-1} + \frac{1}{\sigma_z^2}\underbrace{\underline{\mathbf{E}}_p^H\underline{\mathbf{E}}_p}_{\triangleq \mathbf{R}_E}\right]^{-1} = \left[\mathbf{R}_\alpha^{-1} + \frac{1}{\sigma_z^2}\mathbf{R}_E\right]^{-1} \tag{6}$$

Hence, the MSE in estimating α_i is $\text{MSE}(\hat{\alpha}_i) = \mathbf{C}_\epsilon(i,i)$.

III. Main Results

A. Problem Formulation

Consider $\boldsymbol{\mathcal{X}}$ to be a comb-type OFDM symbol with data subcarriers masked out by zeros. Our objective is to design a frequency-domain pilot structure for the LMMSE channel estimator in (5) to minimize the trace of $\mathbf{C}_\epsilon$ in (6). In Appendix A, we show that this is achieved by making $\mathbf{R}_E$ a diagonal matrix. Using (3), (4) and (6), it is clear that making $\mathbf{R}_E$ diagonal is equivalent to designing $\boldsymbol{\mathcal{X}}$ such that

$$\boldsymbol{\mathcal{X}}^H\mathbf{G}_i^H\mathbf{G}_j\boldsymbol{\mathcal{X}} = 0, \quad \text{for } i \ne j; \quad i, j = 1, 2, \cdots, N_dL \tag{7}$$

and $\boldsymbol{\mathcal{X}}^H\boldsymbol{\mathcal{X}} = c$ where c is a constant which depends on the total pilot power constraint.

B. Asymptotic Analysis

Using (2), the (m,n)-th element of $\mathbf{R}_E$ can be written as follows

$$\begin{aligned} R_E(m,n) &= \boldsymbol{\mathcal{X}}^H\mathbf{G}_m^H\mathbf{G}_n\boldsymbol{\mathcal{X}} \\ &= \boldsymbol{\mathcal{X}}^H\mathbf{Q}\mathbf{B}^{H(j1)}\underbrace{\text{diag}(\mathbf{v}_{i1})^H\text{diag}(\mathbf{v}_{i2})}_{\Lambda_{i_1i_2}}\mathbf{B}^{(j2)}\mathbf{Q}^H\boldsymbol{\mathcal{X}} \\ &= \mathbf{x}^H\underbrace{\mathbf{B}^{H(j1)}\Lambda_{i_1i_2}\mathbf{B}^{(j2)}}_{\mathbf{I}_c(i_1,i_2,j_1,j_2)}\mathbf{x} = \mathbf{x}^H\mathbf{I}_c(i_1,i_2,j_1,j_2)\mathbf{x} \end{aligned} \tag{8}$$

where $m = (i_1-1)N_d + j_1$, $n = (i_2-1)N_d + j_2$ for $i_1, i_2 = 1, 2, \cdots, N_d$ and $j_1, j_2 = 1, 2, \cdots, L$. We can gain further insight into the pilot optimization problem by approximating the Toeplitz matrix $\mathbf{J}$ defined in Section II-B by a circulant matrix for large N using Szego's theorem [8]. Hence, the eigenvectors and eigenvalues of $\mathbf{J}$ converge to the FFT columns and FFT transform of the first column of $\mathbf{J}$, respectively. Based on this circulant approximation of $\mathbf{J}$, there are 4 possible values of $R_E(m,n)$ in (8) as listed in Table II. Now, as long as $j_1 = j_2$, $\mathbf{I}_c(i_1,i_2,j_1,j_2)$ is a diagonal matrix whose entries are real if $i_1 = i_2$ or complex otherwise. If $j_1 \ne j_2$, $\mathbf{I}_c(i_1,i_2,j_1,j_2)$ has zero diagonal elements and a non-zero d_j-th super-diagonal or sub-diagonal

TABLE II
OFF-DIAGONAL ELEMENTS OF $\mathbf{R}_E$

Case	$i_1 = i_2$	$j_1 = j_2$	$R_E(m,n)$ and $m \neq n$	Comments
1	Yes	Yes	0	$\mathbf{R}_E = \|\boldsymbol{\mathcal{X}}\|^2 \mathbf{I}_{N_d l}$
2	Yes	No	0 (assuming pilot structure in Fig. 1)	$\mathbf{I}_c(\cdot)$ is upper/lower shifted diagonal matrix
3	No	Yes	$c_i' \boldsymbol{\mathcal{X}}_p^H \mathbf{Z}_u^{d_i} \boldsymbol{\mathcal{X}}_p$ or $c_i' \boldsymbol{\mathcal{X}}_p^H \mathbf{Z}_l^{d_i} \boldsymbol{\mathcal{X}}_p$	$\mathbf{Z}_u$ is linear upper-shift matrix, $\mathbf{Z}_l$ is linear lower-shift matrix
4	No	No	0 (assuming pilot structure in Fig. 1)	$\mathbf{I}_c(\cdot)$ is upper/lower shifted diagonal matrix

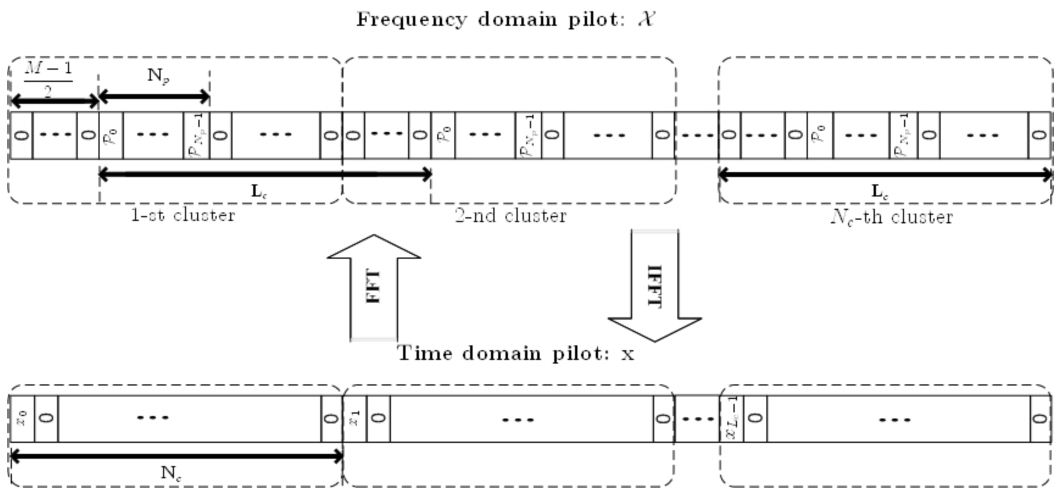

Fig. 1. Optimized pilot structure for our channel estimation algorithm in [1].

where $j_1 - j_2 = +d_j$ or $-d_j$, respectively. Since, our objective is to make $\mathbf{R}_E$ diagonal, we have to force the off-diagonal elements of $\mathbf{R}_E$ in Table II to zero.

Proposition: If the frequency-domain pilot vector $\boldsymbol{\mathcal{X}}$ has the periodic clustered structure shown at the top of Fig. 1 with N_p adjacent subcarriers in each pilot cluster and the number N_c and period L_c of the pilot clusters satisfy the relation $N = N_c L_c$, then the time-domain pilot vector $\mathbf{x}$ will be sparse as shown at the bottom of Fig. 1. The proof is given in Appendix B.

Consider a sparse $\mathbf{x}$ of the form shown at the bottom of the Fig. 1, then $\mathbf{x}^H \mathbf{I}_c(.)\mathbf{x}$ will be a weighted sum of the elements of $\mathbf{I}_c(.)$ that correspond to the positions of '1'-s in the puncturing matrix $\mathbf{P}_{\mathbf{I}_c}$ whose (m,n)-th element is given by

$$P_{\mathbf{I}_c}(m,n) = \begin{cases} 1, & m = k_1 N_c + 1,\ n = k_2 N_c + 1 \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

where, $k_1, k_2 = 0, 1, \cdots, (L_c - 1)$ and $(L-1)$ is the highest index of the super or sub diagonal of $\mathbf{I}_c(.)$ that is non-zero. Hence, the $\mathbf{I}_c(.)$'s in Case 2 and Case 4 of Table II can be at most $(L-1)$ shifted upper or lower diagonal matrices. If the sparse $\mathbf{x}$ has at least L zeros between any two of its non-zero elements, all Case 2 and Case 4 off-diagonal elements of $\mathbf{R}_E$ will be zero. In other words, we have to design the number of pilot clusters in the frequency domain to be greater than the length of the CIR vector, i.e.

$$N_c > L \tag{10}$$

In [1], we showed that the pilot cluster size must satisfy

$$M \leq N_p \leq 2M - 1; \quad M = 3, 5, \cdots \tag{11}$$

In addition, the periodic clustered structure of $\boldsymbol{\mathcal{X}}$, as shown in Fig. 1, implies that the pilot clusters must be equally spaced. Hence, the period of pilot clusters L_c is given by

$$L_c = \frac{N N_p}{N_T} \tag{12}$$

where $N_T = N_c N_p$ is the total number of pilot subcarriers. Since $\mathbf{G}_p$ is assumed to be a banded matrix with M diagonals, to include all diagonals in the input-output equations at pilot locations, the first and last $\frac{M-1}{2}$ subcarriers of the comb-type OFDM symbol cannot be assigned as pilots. We can avoid making these edge subcarriers pilots by placing $\frac{M-1}{2}$ zeroes at the start of each period of $\boldsymbol{\mathcal{X}}$ and inserting N_p adjacent pilot subcarriers followed by $L_c - \left(N_p - \frac{M-1}{2}\right)$ zeroes implying the following lower bound

$$L_c \geq (N_p + M - 1) \tag{13}$$

Using (10)-(13), we arrive the following design guideline on N_c

$$\max\left(\frac{N_T}{2M-1}, L\right) \leq N_c \leq \frac{N_T}{M} \tag{14}$$

C. Pilot Cluster Optimization

All we are left with now is the 3rd case in Table II ; i.e. we have to make $\mathbf{x}^H \mathbf{I}_c(i_1, i_2, j_1, j_2)\mathbf{x} = 0$ when $i_1 \neq i_2$

and $j_1 = j_2$. Note that due to the periodic structure of $\boldsymbol{\mathcal{X}}$, all pilot clusters are identical. Hence, we only need to optimize one pilot cluster. Next, we will show how to make the Case 3 $\mathbf{R}_E(m,n)$ elements in Table II equal to zero with periodic clustered pilot designs. From (8), we see that each Case 3 $\mathbf{R}_E(m,n)$ element in Table II corresponds to the case when $i_1 \neq i_2$ and $j_1 = j_2$, i.e. when the eigenvectors are different for the same CIR tap. Under this scenario, $\mathbf{I}_c(i_1,i_2,j_1,j_2)$ becomes a diagonal matrix whose diagonal is a scaled, circularly-shifted FFT vector. Let $\mathbf{a}_i$ contain these modified FFT vectors when $i_1 \neq i_2$ and $j_1 = j_2$ where $i = (-(N_d-1),\cdots,-1,1,\cdots,(N_d-1))_N$ denotes the FFT column index and $(.)_N$ is the $modulo-N$ operation. In [1], we chose N_d dominant eigenvectors of $\mathbf{J}$ to reduce computational complexity. The column indices of the FFT vectors chosen as dominant eigenvectors are given by $\left(-\frac{N_d-1}{2},\cdots,\frac{N_d-1}{2}\right)_N$. For each dominant eigenvector, we have (N_d-1) Case 3 off-diagonal elements resulting in a total of $(N_d-1)N_d$ non-diagonal elements in $\mathbf{R}_E$ to be forced to zero.

Towards this objective, the time-domain sparse vector $\mathbf{x}$ is given by

$$\mathbf{x} = \mathbf{Q}^H\tilde{\mathbf{I}}\boldsymbol{\mathcal{X}}_p \tag{15}$$

where $\boldsymbol{\mathcal{X}}_p$ is an individual frequency-domain pilot cluster of length N_p and $\tilde{\mathbf{I}} = \underline{\mathbf{1}}_{N_c} \otimes \tilde{\mathbf{I}}_p$

$$\tilde{\mathbf{I}_p} = \left[\mathbf{0}_{N_p\times\frac{M-1}{2}} \quad \mathbf{I}_{N_p} \quad \mathbf{0}_{N_p\times(L_c-\frac{M-1}{2}-N_p)}\right]^T,$$

$\underline{\mathbf{1}}_{N_c}$ is the length-N_c all-ones column vector and $\otimes$ denotes the Kronecker product. Now, from (8), by using (15) and the sparse structure of $\mathbf{x}$ as shown at the bottom of Fig. 1, we can restate our pilot optimization objective as finding $\boldsymbol{\mathcal{X}}_p$ such that

$$\begin{aligned}&\mathbf{x}^H\mathbf{I}_c(i_1,i_2,j_1,j_2)\mathbf{x}\\ =&\boldsymbol{\mathcal{X}}_p{}^H\tilde{\mathbf{I}}^H\mathbf{Q}\underbrace{\mathbf{B}^{H(j1)}\Lambda_{i_1i_2}\mathbf{B}^{(j1)}}_{\text{diag}(\mathbf{a}_i)}\mathbf{Q}^H\tilde{\mathbf{I}}\boldsymbol{\mathcal{X}}_p = 0\\ \Rightarrow&\boldsymbol{\mathcal{X}}_p{}^H\underbrace{\tilde{\mathbf{I}}^H\mathbf{Q}\ \text{diag}(\mathbf{a}_i)\mathbf{Q}^H\tilde{\mathbf{I}}}_{\mathbf{R}_i}\boldsymbol{\mathcal{X}}_p \triangleq \boldsymbol{\mathcal{X}}_p{}^H\mathbf{R}_i\boldsymbol{\mathcal{X}}_p = 0\\ &i = (-(N_d-1),\cdots,-1,1,\cdots,(N_d-1))_N\end{aligned} \tag{16}$$

Since the $\mathbf{a}_i$'s are scaled, circularly-shifted FFT vectors, $\mathbf{Q}\ \text{diag}(\mathbf{a}_i)\mathbf{Q}^H$ can be written as $c_i\mathbf{Z}_c^i$ where c_i is a complex scalar and $\mathbf{Z}_c$ is the $N\times N$ circular upper-shift matrix whose first column is $\left[\begin{array}{cccc}0 & \cdots & 0 & 1\end{array}\right]^T$. The (m,n)-th element of $\mathbf{R}_i$ will be a weighted sum of the elements of $c_i\mathbf{Z}_c^i$ that correspond to the positions of '1'-s in the puncturing matrix $\mathbf{P}_{\mathbf{R}_i(m,n)}$ given by

$$P_{\mathbf{R}_i(m,n)}(q,r) = \begin{cases}1, & q = k_3N_c+m,\ r = k_4N_c+n\\ 0, & \text{otherwise}\end{cases} \tag{17}$$

where, $k_3,k_4 = 0,1,\cdots,(L_c-1)$. Let d_i denote i, as defined in (16), without the $modulo$-N operation. If d_i is negative, it can be shown that $\mathbf{R}_i = c_i{}'\mathbf{Z}_u^{d_i}$ where $c_i{}'$ is a complex scalar and $\mathbf{Z}_u$ is the $N_p\times N_p$ linear upper-shift matrix whose first row is $\left[\begin{array}{ccccc}0 & 1 & 0 & \cdots & 0\end{array}\right]$. On the other hand, if d_i is positive, $\mathbf{R}_i = c_i{}'\mathbf{Z}_l^{d_i}$ where $\mathbf{Z}_l$ as a linear lower-shift matrix whose first column is $\left[\begin{array}{ccccc}0 & 1 & 0 & \cdots & 0\end{array}\right]^T$. Since the range of d_i is $[-(N_d-1),-1,1,\cdots,(N_d-1)]$, there are (N_d-1) distinct $\mathbf{R}_i$'s associated with $\mathbf{Z}_u$ and another (N_d-1) distinct $\mathbf{R}_i$'s associated with $\mathbf{Z}_l$. Therefore, (16) is equivalent to

$$\boldsymbol{\mathcal{X}}_p{}^H\mathbf{R}_j\boldsymbol{\mathcal{X}}_p = 0 \quad : \quad j = 1,2,\cdots 2(N_d-1) \tag{18}$$

Separating the $\mathbf{R}_i$'s in (18) corresponding to $\mathbf{Z}_u$ and $\mathbf{Z}_l$ yields

$$\boldsymbol{\mathcal{X}}_p{}^H\mathbf{Z}_u^{d_i}\boldsymbol{\mathcal{X}}_p = 0 \quad : \quad d_i = 1,2,\cdots(N_d-1) \tag{19}$$

$$\boldsymbol{\mathcal{X}}_p{}^H\mathbf{Z}_l^{d_i}\boldsymbol{\mathcal{X}}_p = 0 \tag{20}$$

Using the frequency-domain pilot cluster notation shown in Fig. 1, for each d_i, (19) can be written as

$$\sum_{n=\tau}^{N_p-1}\mathcal{P}_n\mathcal{P}_{n-\tau}{}^* = 0 \quad : \quad \tau = 1,2,\cdots(N_d-1) \tag{21}$$

Similarly, for the same d_i, (20) can be written as $\left(\sum_{n=\tau}^{N_p-1}\mathcal{P}_n\mathcal{P}_{n-\tau}{}^*\right)^* = 0$. Since the solution to (19) also satisfies (20), we focus on solving (19) only. From (19), we see that the aperiodic auto-correlation of the optimum pilot cluster sequence must be zero at lag d_i which is the design criterion for a zero-correlation zone (ZCZ) sequence [5] with $Z_c \triangleq N_d-1$ zero lags. To be specific, for $N_p = 5$ and $N_d = 3$, the MMSE-optimal pilot cluster is a ZCZ sequence of length 5 with $Z_c = 2$. In Table III, we present 3 such sequences obtained through numerical search under a total power constraint of $N_p = 5$ with $M = 3$. The aperiodic auto-correlation sequence of these optimized sequences are also given in Table III. The inputs to the numerical optimization algorithm are N_p and N_d. Hence, the optimization can be performed offline and the optimum pilot sequences of different sizes are stored in look-up tables.

IV. Simulation Results

In our simulations, we assume the SUI-3 channel model with a rate-$\frac{1}{2}$ convolutional code, a high Doppler frequency of 10% (normalized to the subcarrier spacing) with $N = 1024$ and $M = 3$.

Assuming a pilot cluster size $N_p = 2M-1 = 5$, Fig. 2 depicts the BER of our channel estimation algorithm in [1] with the optimized pilot clusters (shown in Table III) along with the BER of perfect CSI under full and banded $\mathbf{G}$ assumptions. While the 3 optimized pilot cluster sequences achieve MMSE and make $\mathbf{R}_E$ diagonal, their BER performance is different at high SNR where ICI dominates noise. It can be seen that sequence 'a' which has a higher aperiodic auto-correlation at lags larger than Z_c, performs worse in ICI-limited (high SNR) scenarios than sequence 'b' and 'c'. As a benchmark, the BER of the impulsive pilot cluster design $\left[\begin{array}{ccccc}0 & 0 & 5 & 0 & 0\end{array}\right]^T$ suffers from an irreducible error floor.

V. Conclusion

In comb-type OFDM transmission over doubly-selective channels, we showed that the channel estimation mean square error is minimized by dividing the available pilot subcarriers into periodic (i.e. identical and equally-spaced) clusters. Under a fixed total pilot power budget, we exploited the banded structure of the CFR matrix to show that the optimum pilot cluster is a ZCZ sequence. Simulation results demonstrated significant BER improvement over impulsive pilot designs which ignore the banded CFR structure.

TABLE III
SAMPLE SIZE-5 MMSE-OPTIMUM PILOT CLUSTERS FOR LARGE N

Sequence a	Aperiodic auto-correlation	Sequence b	Aperiodic auto-correlation	Sequence c	Aperiodic auto-correlation
	-0.7790 + 0.3011i		0.0112 - 0.0084i		0.0006 - 0.0037i
	-0.2792 + 0.7106i		-0.0180 + 0.0053i		0.0009 + 0.0040i
0.2008 - 0.8820i	0.0002 - 0.0013i	-0.0083 - 0.1174i	-0.0001 + 0.0001i	-0.0107 + 0.0576i	0.0001 + 0.0001i
-0.4227 - 0.2853i	0.0031 + 0.0002i	-0.0256 + 0.0768i	-0.0000 - 0.0000i	-0.4252 - 0.2802i	-0.0001 - 0.0001i
0.7785 - 1.5159i	5.0 (zero lag)	-2.1629 - 0.5304i	5.0 (zero lag)	1.6952 - 1.2656i	5.0 (zero lag)
-0.3268 + 0.2416i	0.0031 - 0.0002i	-0.0198 + 0.0752i	-0.0000 + 0.0000i	-0.1461 - 0.4865i	-0.0001 + 0.0001i
-0.5157 + 0.7658i	0.0002 + 0.0013i	0.0645 - 0.0997i	-0.0001 - 0.0001i	-0.0638 - 0.0007i	0.0001 - 0.0001i
	-0.2792 - 0.7106i		-0.0180 - 0.0053i		0.0009 - 0.0040i
	-0.7790 - 0.3011i		0.0112 + 0.0084i		0.0006 + 0.0037i

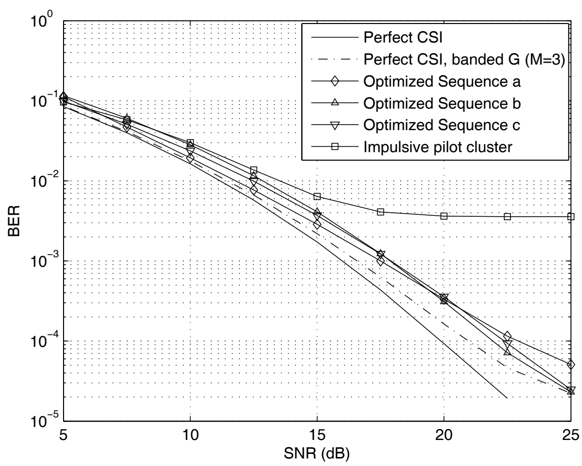

Fig. 2. BER comparison of proposed and impulsive pilot cluster designs with perfect CSI for $N = 1024$, $M = 3$ and $N_p = 5$.

APPENDIX A
PROOF THAT OPTIMUM $\mathbf{R}_E$ IS DIAGONAL

We start by presenting the following three useful matrix derivative identities

- **I**.1 $\frac{\partial \text{Tr}(\mathbf{Y}^{-1})}{\partial \mathbf{Y}} = -\mathbf{Y}^{-2T}$ (see (57) in [9])
- **I**.2 $\frac{\partial \text{Tr}(\mathbf{Y}^H\mathbf{Y})}{\partial \mathbf{Y}} = \mathbf{Y}^*$ (see (225) in [9])
- **I**.3 If $\mathbf{A}$ is a function of $\mathbf{F} = \mathbf{C} + \mathbf{B}^H\mathbf{Y}^H\mathbf{Y}\mathbf{B}$ where $\mathbf{C} \geq \mathbf{0}$, then $\frac{\partial \text{Tr}(\mathbf{A})}{\partial \mathbf{Y}} = \mathbf{Y}^*\mathbf{B}^*\frac{\partial \text{Tr}(A)}{\partial \mathbf{F}}\mathbf{B}^T$.
  Proof: Let $\mathbf{D} = \mathbf{B}^H\mathbf{Y}^H$. Hence, $F(i,j) = C(i,j) + \sum_k \sum_m D(i,k)Y(k,m)B(m,j)$. Differentiating $F(i,j)$ with respect to $Y(k,m)$, we get $\frac{\partial F(i,j)}{\partial Y(k,m)} = D(i,k)B(m,j)$. Hence,

$$
\begin{aligned}
\frac{\partial \text{Tr}(\mathbf{A})}{\partial Y(k,m)} &= \sum_j \sum_i \frac{\partial \text{Tr}(\mathbf{A})}{\partial F(i,j)} \frac{\partial F(i,j)}{\partial Y(k,m)} \\
&= \sum_j \sum_i \frac{\partial \text{Tr}(\mathbf{A})}{\partial F(i,j)} D(i,k)B(m,j) \\
\therefore \frac{\partial \text{Tr}(\mathbf{A})}{\partial \mathbf{Y}} &= \mathbf{D}^T \frac{\partial \text{Tr}(\mathbf{A})}{\partial \mathbf{F}}\mathbf{B}^T = \left(\mathbf{B}^H\mathbf{Y}^H\right)^T \frac{\partial \text{Tr}(\mathbf{A})}{\partial \mathbf{F}}\mathbf{B}^T \\
&= \mathbf{Y}^*\mathbf{B}^*\frac{\partial \text{Tr}(\mathbf{A})}{\partial \mathbf{F}}\mathbf{B}^T
\end{aligned} \tag{22}
$$

Our pilot optimization objective is to minimize the trace of $\mathbf{C}_\epsilon = \left(\mathbf{R}_\alpha^{-1} + \frac{1}{\sigma_z^2}\mathbf{R}_E\right)^{-1}$ subject to $\sum_i R_E(i,i) = E_{tot}$ where E_{tot} is the total pilot energy in an OFDM symbol. We form the cost function using Lagrangian multipliers as follows

$$
f_{cost} = \text{Tr}\left(\mathbf{R}_\alpha^{-1} + \frac{1}{\sigma_z^2}\underline{\mathbf{E}}_p^H\underline{\mathbf{E}}_p\right)^{-1} + \lambda\left(\mathbf{e}_i^H\underline{\mathbf{E}}_p^H\underline{\mathbf{E}}_p\mathbf{e}_i\right) - E_{tot} \tag{23}
$$

Next, we compute $\frac{\partial f_{cost}}{\underline{\mathbf{E}}_p}$ and set it to $\mathbf{0}$ to get the optimality (Kuhn-Tucker) condition on $\underline{\mathbf{E}}_p$. From **I**.3 with $\mathbf{F} = \mathbf{R}_\alpha^{-1} + \frac{1}{\sigma_z^2}\underline{\mathbf{E}}_p^H\underline{\mathbf{E}}_p$, $\mathbf{C} = \mathbf{R}_\alpha^{-1}$, $\mathbf{B} = \frac{1}{\sigma_z}\mathbf{I}$ and $\mathbf{A} = \mathbf{F}^{-1}$, we have

$$
\begin{aligned}
\frac{\partial \text{Tr}\left(\mathbf{R}_\alpha^{-1} + \frac{1}{\sigma_z^2}\underline{\mathbf{E}}_p^H\underline{\mathbf{E}}_p\right)}{\partial \underline{\mathbf{E}}_p} &= \frac{1}{\sigma_z^2}\underline{\mathbf{E}}_p^*\frac{\partial}{\partial \mathbf{F}}\text{Tr}\left(\mathbf{F}\right)^{-1} \qquad (24) \\
&= -\frac{1}{\sigma_z^2}\underline{\mathbf{E}}_p^*\mathbf{F}^{-2T} \qquad (\text{Using } \mathbf{I}.1)
\end{aligned}
$$

Moreover, $\frac{\partial}{\partial \underline{\mathbf{E}}_p}\left(\mathbf{e}_i^H\underline{\mathbf{E}}_p^H\underline{\mathbf{E}}_p\mathbf{e}_i\right) = \frac{\partial}{\partial \underline{\mathbf{E}}_p}\left(\underline{\mathbf{E}}_p^H\underline{\mathbf{E}}_p\mathbf{e}_i\mathbf{e}_i^H\right) = \underline{\mathbf{E}}_p^*\mathbf{e}_i^*\mathbf{e}_i^T$ where we used **I**.3 with $\mathbf{B} = \mathbf{e}_i$, $\mathbf{Y} = \underline{\mathbf{E}}_p$, $\mathbf{C} = \mathbf{0}$ and $\frac{\partial \text{tr}(\mathbf{A})}{\partial \mathbf{A}} = \mathbf{I}$. From (24), we get

$$
-\frac{1}{\sigma_z^2}\underline{\mathbf{E}}_p^*\left(\mathbf{R}_\alpha^{-1} + \frac{1}{\sigma_z^2}\underline{\mathbf{E}}_p^H\underline{\mathbf{E}}_p\right)^{-2T} + \lambda\left(\sum_i \underline{\mathbf{E}}_p^*\mathbf{e}_i^*\mathbf{e}_i^T\right) = \mathbf{0}
$$

Transposing both sides yields

$$
\Longrightarrow \left[\mathbf{I} - \lambda\sigma_z^2\left(\mathbf{R}_\alpha^{-1} + \frac{1}{\sigma_z^2}\underline{\mathbf{E}}_p^H\underline{\mathbf{E}}_p\right)^2\left(\sum_i \mathbf{e}_i^*\mathbf{e}_i^T\right)\right]\underline{\mathbf{E}}_p^H = \mathbf{0} \tag{25}
$$

Since $\underline{\mathbf{E}}_p^H$ is a tall full-column rank matrix, we have

$$
\begin{aligned}
\left(\mathbf{R}_\alpha^{-1} + \frac{1}{\sigma_z^2}\underline{\mathbf{E}}_p^H\underline{\mathbf{E}}_p\right)^2 &= \frac{1}{\lambda\sigma_z^2}\left(\mathbf{e}_i^*\mathbf{e}_i^T\right)^{-1} \qquad (26) \\
&= \underbrace{\frac{1}{\sigma_z^2}\text{diag}\left(\frac{1}{\lambda}, \frac{1}{\lambda}, \cdots, \frac{1}{\lambda}\right)}_{\triangleq \Lambda} \\
\Longrightarrow \left(\mathbf{R}_\alpha^{-1} + \frac{1}{\sigma_z^2}\underline{\mathbf{E}}_p^H\underline{\mathbf{E}}_p\right) &= \Lambda^{\frac{1}{2}} \qquad (27)
\end{aligned}
$$

Hence, the trace of $\mathbf{C}_\epsilon$ will be minimized when $\mathbf{R}_E \triangleq \underline{\mathbf{E}}_p^H\underline{\mathbf{E}}_p = \sigma_z^2\left(\Lambda^{\frac{1}{2}} - \mathbf{R}_\alpha^{-1}\right) > \mathbf{0}$. Since Λ and $\mathbf{R}_\alpha$ are diagonal matrices, the optimum $\mathbf{R}_E$ is also a diagonal matrix.

APPENDIX B
PROOF OF PROPOSITION 1

Let L_c and N_c denote the period of the pilot clusters and the total number of pilot clusters in $\boldsymbol{\mathcal{X}}$, respectively, as shown in Fig. 1. Using the DFT relationship, the m-th element of $\mathbf{x}$ is given by

$$x_m = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \mathcal{X}_n e^{j\frac{2\pi mn}{N}} \tag{28}$$

Using the periodic clustered pilot structure shown in Fig. 1, (28) can be written as follows

$$\begin{aligned} x_m =& \frac{1}{\sqrt{N}} \Bigg[\mathcal{P}_0 e^{j\frac{2\pi dm}{N}} \left(1 + e^{j\frac{2\pi L_c m}{N}} + \cdots + e^{j\frac{2\pi L_c (N_c-1)m}{N}} \right) \\ &+ \mathcal{P}_1 e^{j\frac{2\pi (d+1)m}{N}} \left(1 + e^{j\frac{2\pi L_c m}{N}} + \cdots + e^{j\frac{2\pi L_c (N_c-1)m}{N}} \right) + \cdots \\ &+ \mathcal{P}_{N_p-1}\, e^{j\frac{2\pi (d+N_p-1)m}{N}} \left(1 + e^{j\frac{2\pi L_c m}{N}} + \cdots + e^{j\frac{2\pi L_c (N_c-1)m}{N}} \right) \Bigg] \end{aligned} \tag{29}$$

Using the relation $N = N_c L_c$, Equation (29) can be compactly written as follows

$$x_m = \frac{1}{\sqrt{N}} \sum_{i=0}^{N_p-1} \mathcal{P}_i e^{j\frac{2\pi (d+i)}{N}} \left(\sum_{k=0}^{N_c-1} e^{j\frac{2\pi mk}{N_c}} \right) \tag{30}$$

If the index m is not an integer multiple of N_c, (30) is given by

$$x_m = \frac{1}{\sqrt{N}} \sum_{i=0}^{N_p-1} \mathcal{P}_i e^{j\frac{2\pi (d+i)}{N}} \left(\sum_{k=0}^{N_c-1} e^{j\frac{2\pi k}{N_c}} \right)^m \tag{31}$$

Now, $\left(\sum_{k=0}^{N_c-1} e^{j\frac{2\pi k}{N_c}} \right)$ is the sum of the geometric series of the N_c-th roots of unity which equals zero. Thus, $\mathbf{x}$ will have only L_c non-zero elements separated by $N_c - 1$ zeros.

REFERENCES

[1] T. Y. Al-Naffouri, K. M. Zahidul Islam, N. Al-Dhahir, and S. Lu, “A model reduction approach for OFDM channel estimation under high mobility conditions," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2181-2193, Apr. 2010.

[2] L. Tong, B. Sadler, and M. Dong, “Pilot-assisted wireless transmissions: general model, design criteria, and signal processing," *IEEE Signal Process. Mag.*, vol. 21, no. 6, pp. 12-25, Nov. 2004.

[3] I. Barhumi, G. Leus, and M. Moonen, “Optimal training design for MIMO OFDM systems in mobile wireless channels," *IEEE Trans. Signal Process.*, vol. 51, no. 6, pp. 1615-1624, June 2003.

[4] A. Stamoulis, S. Diggavi, and N. Al-Dhahir, “Intercarrier interference in MIMO OFDM," *IEEE Trans. Signal Process.*, vol. 50, pp. 2451-2464, Oct. 2002.

[5] M. Perez, J. Urena, A. Hernandez, A. Jimenez, and C. De Marziani, “Efficient generation and correlation of sequence pairs with three zero-correlation zones," *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3450-3465, Sep. 2009.

[6] A. Kannu and P. Schniter, “Design and analysis of MMSE pilot-aided cyclic-prefixed block transmissions for doubly selective channels," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 1148-1160, Mar. 2008.

[7] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, 1993.

[8] R. M. Gray, “Toeplitz and circulant matrices: a review," Now Publishers Inc., tech. rep., 2002.

[9] K. B. Petersen and M. S. Pedersen, *The Matrix Cookbook*. Available: http://matrixcookbook.com.