MIMO-OFDM Channel Estimation in the Presence of Frequency Offsets

Zhongshan Zhang, *Member, IEEE*, Wei Zhang, *Student Member, IEEE,*
and Chintha Tellambura, *Senior Member, IEEE*

***Abstract*—Optimal pilot design and placement for channel estimation in Multiple-input Multiple-output (MIMO) Orthogonal Frequency-Division Multiplexing (OFDM) systems with frequency offsets are considered. Both the single-frequency-offset case and the multiple-frequency-offset case are treated. We show that the Constant-Envelope (CE) condition is sufficient but not necessary for pilot design, and that pilots with multiple envelopes can also achieve the optimal performance in terms of the Mean Square Error (MSE) minimization, provided that an additional constraint on the pilot placement is satisfied simultaneously. New pilot designs, which take into account the multiple-frequency-offset case, are proposed to eliminate Inter-Pilot-Interference (IPI) and to optimize the MSE performance. The Least-Squares (LS) and Linear Minimum Mean Square Error (LMMSE) channel estimators for the multiple-frequency-offset case are designed for uncorrelated and correlated MIMO-OFDM channels, respectively. The LMMSE estimator requires the channel covariance matrix. Both optimal adaptive pilot power allocation and suboptimal uniform pilot power allocation are developed for the proposed LMMSE estimator. The adaptive allocation performs 4 dB better than the uniform allocation in the high noise region, but they both perform identically in the low noise region. Performance comparisons are made against several previous pilot designs due to [1], [2]. The proposed LMMSE estimator significantly outperforms the LS estimator.**

***Index Terms*—Frequency offset, channel estimation, MIMO, OFDM.**

I. Introduction

CONSIDERABLE research has focused on Multiple-input Multiple-output (MIMO) technology for increasing the wireless system capacity. Compared to a single-input single-output (SISO) system, a MIMO system can improve the capacity by a factor of the minimum number of transmit and receive antennas. Space-time coding, including space-time block codes (STBC) and space-time trellis codes (STTC) [3], can extract transmit diversity in MIMO systems.

The frequency-selective fading MIMO channel can be transformed into a set of flat-fading MIMO channels by using Orthogonal Frequency-Division Multiplexing (OFDM). This transformation achieves a high capacity at a low cost of equalization and demodulation [4], [5]. However, just as with the SISO-OFDM, MIMO-OFDM systems too are sensitive to frequency offset. Many SISO-OFDM frequency offset estimators have been proposed [6]–[8]. A synchronization algorithm for MIMO-OFDM systems is proposed in [9], where identical timing offset and frequency offset with respect to each transmit-receive antenna pair are assumed. Parameters estimation of MIMO flat-fading channels is discussed in [10], where frequency offsets for different transmit-receive antennas are assumed to be different, and the Cramer-Rao Lower Bound (CRLB) for either the frequency offsets or channel estimation variance error is derived.

Another prevalent impairment is the channel estimation error, which degrades the bit error rate of MIMO-OFDM systems. Robust channel estimation for OFDM systems is discussed in [11]. Optimal pilot design and placement for channel estimation is developed in terms of minimization of the CRLB [12]. Optimal training signal design for frequency-selective block fading channel estimation in MIMO-OFDM systems is discussed in [13]. Some other constraints for training signal designs, such as low peak-to-average energy ratio (PAR) and robustness to frequency offsets, are also considered in [13]. Barhumi, Leus and Moonen (BLM) propose a high-quality channel estimator for a MIMO-OFDM channel based on frequency-domain uniformly placed pilots (the frequency offset and channel correlation are not considered) [1]. Hu improves the BLM pilots with a nonuniform placement design, which mitigates the performance loss due to the non-modulated pilots in virtual subcarriers [14]. MIMO-OFDM channel estimation for a correlated channel is discussed in [15], [16]. However, neither [15] nor [16] considers the effect of frequency offset. Training sequence design for MIMO channel estimation in the presence of a single frequency offset is discussed in [17], [18]. In [17], optimal training signal design for MIMO-OFDM channel estimation while considering a single frequency offset and phase noise is proposed, and its performance is much more robust than the algorithms that consider only the effect of frequency offset. In [18], the estimation of frequency-selective channel and frequency offset in either SISO or MIMO systems is discussed. By using the exact CRLB as a metric, a power efficient training preamble is designed in [18] to reduce the complexity of estimation. Joint frequency offset and channel estimation with either single or multiple frequency offsets for MIMO frequency selective fading channels is discussed in [2], where pilots for different transmit antennas are orthogonal in the time-domain, and pilot optimization in terms of MSE reduction is also studied.

This paper considers optimal pilot design and placement for channel estimation in Multiple-input Multiple-output (MIMO) Orthogonal Frequency-Division Multiplexing (OFDM) systems in the presence of frequency offsets. Both the single-frequency-offset case and the multiple-frequency-offset case

Manuscript received January 14, 2007; revised July 17, 2007; accepted September 16, 2007. The associate editor coordinating the review of this letter and approving it for publication was C. Xiao.

The authors are with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2V4, Canada (e-mail: {zszhang, wzhang, chintha}@ece.ualberta.ca). The corresponding author is Dr. Chintha Tellambura.

Digital Object Identifier 10.1109/TWC.2008.070044.

1536-1276/08$25.00 © 2008 IEEE

are treated[1]. Each frequency offset is modeled as an Independent and Identically Distributed (i.i.d.) random variable (RV). In our proposed scheme, several subcarriers are allocated to each transmit antenna to transmit either pilots or data symbols. The pilots of different transmit antennas may be either orthogonal (in the frequency-domain or time-domain) or not. We show that the Constant-Envelope (CE) condition is sufficient but not necessary for pilot design, and that pilots with multiple envelopes can also achieve the optimal performance in terms of the Mean Square Error (MSE) minimization, provided that an additional constraint on the pilot placement is satisfied simultaneously. New pilot designs, which take into account the multiple-frequency-offset case, are proposed to eliminate Inter-Pilot-Interference (IPI) and to optimize the MSE performance. The Least-Squares (LS) and Linear Minimum Mean Square Error (LMMSE) channel estimators for the multiple-frequency-offset case are designed for uncorrelated and correlated MIMO-OFDM channels, respectively. The LMMSE estimator requires the channel covariance matrix. Both optimal adaptive pilot power allocation and suboptimal uniform pilot power allocation are developed for the proposed LMMSE estimator. Performance comparisons are made against several previous pilot designs due to [1], [2]. The estimator of [1] is found to exhibit high error floors for the multiple-frequency-offset case. The proposed LMMSE estimator significantly outperforms the LS estimator.

The remainder of this paper is organized as follows. A MIMO-OFDM system model with multiple frequency offsets is discussed in Section II, and the LS frequency offset and channel estimation are analyzed in Section III. The optimal pilot design and placement for LS channel estimation in the presence of multiple frequency offsets are discussed in Section IV, and an LMMSE estimator for a correlated MIMO-OFDM channel is proposed in Section V. Numerical results are given in Section VI, followed by conclusions in Section VII[2].

II. MIMO-OFDM Signal Model

Input data bits of MIMO-OFDM are mapped to complex symbols drawn from a typical signal constellation, e.g., phase-shift keying (PSK) or quadrature amplitude modulation (QAM). An OFDM symbol is generated by taking the *Inverse Discrete Fourier Transform* (IDFT) of N input sub-symbols, where N is the IDFT size, and the IDFT matrix $\mathbf{F}$ is defined as $[\mathbf{F}]_{nk} = \frac{1}{\sqrt{N}} e^{\frac{\jmath 2\pi nk}{N}}$ for $0 \leq n, k \leq N-1$. We also assume that each OFDM symbol has a useful part of duration T_s seconds and a cyclic prefix of length T_g seconds to mitigate Inter-Symbol-Interference (ISI), where T_g is longer than the channel-response duration. For a MIMO-OFDM system with N_t transmit antennas and N_r receive antennas, a $N \times 1$ vector $\mathbf{x}_i(z)$ is used to represent the z-th block of frequency-domain symbols sent by the i-th transmit antenna, where $i \in \{1, 2, \cdots, N_t\}$. In the following sections, the temporal index z will be omitted for the sake of simplicity. Without loss of generality, each entry of $\mathbf{x}_i$ is assumed to be an i.i.d. RV with mean zero and variance $\sigma_x^2 = \frac{E_s}{NN_t}$, where $E_s = \sum_{i=1}^{N_t} \mathbb{E}\left\{\|\mathbf{x}_i\|_2^2\right\}$ is the total transmit power.

By using $h_{k,i}(z)$ to represent the discrete-time impulse response of the z-th tap channel between the i-th transmit and the k-th receive antenna, the related channel response vector can be represented as $\mathbf{h}_{k,i} = [\tilde{\mathbf{h}}_{k,i}^T, \mathbf{0}_{N-L_{max}}^T]^T = [h_{k,i}(0), h_{k,i}(1), \cdots, h_{k,i}(L_{max}-1), \mathbf{0}_{N-L_{max}}^T]^T$ with L_{max} representing the maximum length of all channels. The corresponding frequency-domain channel attenuation matrix is given by $\mathbf{H}_{k,i} = \text{diag}\left\{H_{k,i}^{(0)}, H_{k,i}^{(1)}, \cdots, H_{k,i}^{(N-1)}\right\}$ with $H_{k,i}^{(n)} = \sum_{d=0}^{L_{max}-1} h_{k,i}(d) e^{-\frac{\jmath 2\pi nd}{N}}$ representing the channel attenuation at the n-th subcarrier.

In the following sections, $\psi_{k,i}$ and $\varepsilon_{k,i}$ are used to represent the initial phase and the normalized frequency offset (frequency offset normalized to a subcarrier spacing of OFDM symbols) between the i-th transmit and the k-th receive antennas. Since the initial phase is independent of the frequency offsets and channel attenuation and also because the estimation of the initial phase is beyond the discussion of this paper, without loss of generality, we assume that $\psi_{k,i}$ for each (k,i) has been estimated and compensated for and that, therefore, $\psi_{k,i} = 0$. By considering the channel attenuations and frequency offsets, the q-th received vector can be represented as

$$\mathbf{y}(q) = [\mathbf{y}_1^T(q), \mathbf{y}_2^T(q), \cdots, \mathbf{y}_{N_r}^T(q)]^T, \tag{1}$$

where $\mathbf{y}_k(q) = \sum_{i=1}^{N_t} \mathbf{E}_{k,i}\mathbf{F}\mathbf{H}_{k,i}\mathbf{x}_i(q) + \mathbf{w}_k(q)$, $\mathbf{E}_{k,i} = \text{diag}\left\{1, e^{\frac{\jmath 2\pi\varepsilon_{k,i}}{N}}, \cdots, e^{\frac{\jmath 2\pi\varepsilon_{k,i}(N-1)}{N}}\right\}$, and $\mathbf{w}_k(q)$ is a vector of additive white Gaussian noise (AWGN) with $\mathbf{w}_k(q)[i] \sim \mathcal{CN}(0, \sigma_w^2)$.

At the k-th receive antenna, by taking the DFT operation to the received vector, we obtain

$$\begin{aligned}\mathbf{r}_k(q) = \mathbf{F}^H\mathbf{y}_k(q) &= \sum_{i=1}^{N_t} \underbrace{\sqrt{N}\mathbf{E}_{k,i}^{\text{cir}}\mathbf{X}_i^p(q)\mathbf{F}_{(L_{max})}^H}_{\mathbf{P}_{k,i}(q)\ (N\times L_{max})} \tilde{\mathbf{h}}_{k,i} \\ &+ \sum_{i=1}^{N_t} \underbrace{\sqrt{N}\mathbf{E}_{k,i}^{\text{cir}}\mathbf{X}_i^d(q)\mathbf{F}_{(L_{max})}^H}_{\mathbf{D}_{k,i}(q)\ (N\times L_{max})} \tilde{\mathbf{h}}_{k,i} + \underbrace{\mathbf{F}^H\mathbf{w}_k(q)}_{\boldsymbol{\eta}_k(q)\ (N\times 1)},\end{aligned} \tag{2}$$

where $\mathbf{E}_{k,i}^{\text{cir}} = \mathbf{F}^H\mathbf{E}_{k,i}\mathbf{F}$ is a circulant matrix, $\mathbf{F}_{(L_{max})}$ is the first L_{max} rows of $\mathbf{F}$, and $\mathbf{X}_i(q) = \mathbf{X}_i^d(q) + \mathbf{X}_i^p(q) = \text{diag}\{\mathbf{x}_i^d(q)\} + \text{diag}\{\mathbf{x}_i^p(q)\}$ with $\mathbf{x}_i^d(q)$ and $\mathbf{x}_i^p(q)$ being some $N \times 1$ data and pilot vectors, respectively. Note that the averaged power of the data and pilot may be different. With the pilots modulated over consecutive M symbols ($M \geq 1$), the received vector becomes

$$\mathbf{r}_k = \left[\mathbf{r}_k^T(0), \cdots, \mathbf{r}_k^T(M-1)\right]^T = \mathbf{P}_k\mathbf{h}_k + \mathbf{D}_k\mathbf{h}_k + \boldsymbol{\eta}_k, \tag{3}$$

[1]This case can occur when, for example, MIMO OFDM is used in a multiple user scenario. This case can also occur in cooperative relaying, which can be seen as virtual MIMO systems. The single frequency offset case is a special case of this case.

[2]*Notation*: $(\cdot)^{-1}$, $(\cdot)^T$ and $(\cdot)^H$ are the inverse, transpose and complex conjugate transpose of a matrix. The imaginary unit is $\jmath = \sqrt{-1}$. $\Re\{x\}$, and $\Im\{x\}$ are the real and imaginary part of x, respectively. A circularly symmetric complex Gaussian RV w with mean m and variance σ^2 is denoted by $w \sim \mathcal{CN}(m, \sigma^2)$. A real Gaussian RV x with mean a and variance σ^2 is denoted by $x \sim \mathcal{N}(a, \sigma^2)$. $\mathbf{I}_N$ is the $N \times N$ identity matrix, and $\mathbf{O}_N$ is the $N \times N$ all-zero matrix. $\mathbf{0}_N$ is the $N \times 1$ all-zero vector. $\mathbf{a}[i]$ is the i-th entry of vector $\mathbf{a}$, and $\|\mathbf{a}\|_2^2 = \sum_i |\mathbf{a}[i]|^2$. $[\mathbf{B}]_{mn}$ is the mn-th entry of matrix $\mathbf{B}$. $(x)_n$ represents the remainder after division of x by n. $\mathbb{E}\{x\}$ and $\text{Var}\{x\}$ are the mean and variance of x.

where $\mathbf{P}_k = \begin{bmatrix} \mathbf{P}_{k,1}(0) & \cdots & \mathbf{P}_{k,N_t}(0) \\ \vdots & \ddots & \vdots \\ \mathbf{P}_{k,1}(M-1) & \cdots & \mathbf{P}_{k,N_t}(M-1) \end{bmatrix}$, $\mathbf{D}_k = \begin{bmatrix} \mathbf{D}_{k,1}(0) & \cdots & \mathbf{D}_{k,N_t}(0) \\ \vdots & \ddots & \vdots \\ \mathbf{D}_{k,1}(M-1) & \cdots & \mathbf{D}_{k,N_t}(M-1) \end{bmatrix}$, $\mathbf{h}_k = \left[\tilde{\mathbf{h}}_{k,1}^T, \cdots, \tilde{\mathbf{h}}_{k,N_t}^T\right]^T$ and $\boldsymbol{\eta}_k = \left[\boldsymbol{\eta}_k^T(0), \cdots, \boldsymbol{\eta}_k^T(M-1)\right]^T$.

III. LS Frequency Offset and Channel Estimation in MIMO-OFDM Systems

This section and the next assume that the channel coefficients for the different transmit and receive antennas independent. Section V considers correlated MIMO-OFDM channels.

Using $\mathbf{r}_k$, the frequency offsets and channel coefficients are jointly estimated as

$$\left\{\hat{\varepsilon}_{k,1}, \cdots, \hat{\varepsilon}_{k,N_t}; \hat{\mathbf{h}}_k\right\} = \arg \min_{\hat{\varepsilon}_{k,1},\cdots,\hat{\varepsilon}_{k,N_t};\hat{\mathbf{h}}_k} \left\|\mathbf{r}_k - \hat{\mathbf{P}}_k \hat{\mathbf{h}}_k\right\|_2^2, \tag{4}$$

where $\hat{\mathbf{P}}_k = \sqrt{N} \begin{bmatrix} \boldsymbol{\Gamma}_{k,1}(0) & \cdots & \boldsymbol{\Gamma}_{k,N_t}(0) \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Gamma}_{k,1}(M-1) & \cdots & \boldsymbol{\Gamma}_{k,N_t}(M-1) \end{bmatrix}$, $\boldsymbol{\Gamma}_{k,i}(m) = \hat{\mathbf{E}}_{k,i}^{\text{cir}} \mathbf{X}_i^p(m) \mathbf{F}_{(L_{max})}^H$, and $[\hat{\mathbf{E}}_{k,i}^{\text{cir}}]_{ls} = \dfrac{\sin\left(\pi\left((s-l)_N + \hat{\varepsilon}_{k,i}\right)\right)}{N \sin\left(\frac{\pi}{N}\left((s-l)_N + \hat{\varepsilon}_{k,i}\right)\right)} \cdot e^{\frac{\jmath(N-1)\left((s-l)_N + \hat{\varepsilon}_{k,i}\right)}{N}}$, and $\hat{\varepsilon}_{k,i}$ represents the estimate of $\varepsilon_{k,i}$.

We perform an estimation in two steps: first, we keep each channel attenuation unchanged and design a robust frequency offset estimator whose performance is insensitive to the wireless channel; second, based on the frequency offset estimation result, $\mathbf{P}_k$ can be achieved and used for channel estimation. The frequency offset can be estimated as

$$\begin{aligned} &\{\hat{\varepsilon}_{k,1}, \cdots, \hat{\varepsilon}_{k,N_t}\} \\ &= \arg \min_{\hat{\varepsilon}_{k,1},\cdots,\hat{\varepsilon}_{k,N_t}} \sum_{n=0}^{M-1} \left\|\mathbf{y}_k(n) - \sum_{i=1}^{N_t} \hat{\mathbf{E}}_{k,i} \mathbf{F} \mathbf{H}_{k,i} \mathbf{x}_i^p(n)\right\|_2^2, \end{aligned} \tag{5}$$

where $\hat{\mathbf{E}}_{k,i} = \text{diag}\left\{1, e^{\frac{\jmath 2\pi \hat{\varepsilon}_{k,i}}{N}}, \cdots, e^{\frac{\jmath 2\pi \hat{\varepsilon}_{k,i}(N-1)}{N}}\right\}$. By transmitting the pilot vector $\mathbf{m}_k(n) = \sum_{i=1}^{N_t} \mathbf{E}_{k,i} \mathbf{F} \mathbf{H}_{k,i} \mathbf{x}_i^p(n)$, from [19, page 926], for an unbiased estimator, the CRLB for a variance error of $\hat{\varepsilon}_{k,i}$ is given by

$$\begin{aligned} \text{Var}\{\hat{\varepsilon}_{k,i}\} &\geq \frac{\sigma_w^2}{\sum_{n=0}^{M-1} \Re\left\{\frac{\partial \mathbf{m}_k^H(n)}{\partial \varepsilon_{k,i}} \frac{\partial \mathbf{m}_k(n)}{\partial \varepsilon_{k,i}}\right\}} \\ &= \frac{\sigma_w^2}{\sum_{n=0}^{M-1} \|\boldsymbol{\Lambda} \mathbf{F} \mathbf{H}_{k,i} \mathbf{x}_i^p(n)\|_2^2} \\ &= \frac{3 N N_t \sigma_w^2}{2\pi^2 (N-1)(2N-1) E_p}, \end{aligned} \tag{6}$$

where $\boldsymbol{\Lambda} = \text{diag}\left\{0, \dfrac{2\pi}{N}, \cdots, \dfrac{2\pi(N-1)}{N}\right\}$, and E_p is the total transmit power of the pilots. Evidently, $\text{Var}\{\hat{\varepsilon}_{k,i}\}$ is inversely proportional to the Pilot-to-Noise Ratio (PNR), as defined by PNR $= E_p/\sigma_w^2$. Note that we will mainly focus on the channel estimation in the following sections, and the design of a frequency offset estimator that satisfies (5) is beyond the scope of this paper. Many existing algorithms, e.g., that proposed in [20]–[22], can be used to perform the frequency offset estimation (although these algorithms were originally designed for OFDMA systems, they can be easily tailored to MIMO-OFDM with each user in OFDMA being virtually seen as a transmit antenna in MIMO-OFDM). In Section IV, when the variance error of the frequency offset estimator is smaller than 10^{-2} (this accuracy is easy to achieve in [20]–[22]), a robust channel estimation can be performed by using the proposed pilots.

When performing LS channel estimation, $\hat{\mathbf{P}}_k$, the estimate of $\mathbf{P}_k$, should be full column rank, so that $MN \geq L_{max} N_t$. Since the frequency offsets have been estimated with negligible errors, $\mathbf{P}_k$ instead of $\hat{\mathbf{P}}_k$ is used in channel estimation (the effect of a large frequency offset estimation error on the performance of the proposed channel estimation will be analyzed in the following sections, where $\hat{\mathbf{P}}_k \neq \mathbf{P}_k$ is considered). By defining the pseudo-inverse of $\mathbf{P}_k$ as $\mathbf{P}_k^\dagger = \left(\mathbf{P}_k^H \mathbf{P}_k\right)^{-1} \mathbf{P}_k^H$, the LS estimation of $\mathbf{h}_k$ is given by

$$\hat{\mathbf{h}}_{k|\varepsilon_{k,1},\cdots,\varepsilon_{k,N_t}} = \mathbf{P}_k^\dagger \mathbf{r}_k = \mathbf{h}_k + \mathbf{P}_k^\dagger \mathbf{D}_k \mathbf{h}_k + \mathbf{P}_k^\dagger \boldsymbol{\eta}_k, \tag{7}$$

where the subscript of $\hat{\mathbf{h}}_{k|\varepsilon_{k,1},\cdots,\varepsilon_{k,N_t}}$ means the LS estimator is a function of frequency offsets $\varepsilon_{k,1}, \cdots, \varepsilon_{k,N_t}$. The MSE of LS estimator $\hat{\mathbf{h}}_{k|\varepsilon_{k,1},\cdots,\varepsilon_{k,N_t}}$ is given by

$$\begin{aligned} \text{MSE}\left(\hat{\mathbf{h}}_{k|\text{LS}}\right) &= \frac{1}{L_{max} N_t} \mathbb{E}\left\{\left\|\hat{\mathbf{h}}_{k|\varepsilon_{k,1},\cdots,\varepsilon_{k,N_t}} - \mathbf{h}_k\right\|_2^2\right\} \\ &= \frac{\text{trace}\left\{\mathbf{V}_k^H \left(\mathbf{P}_k^H \mathbf{P}_k\right)^{-2} \mathbf{V}_k \boldsymbol{\Phi}_k\right\}}{L_{max} N_t} \\ &\quad + \frac{\sigma_w^2 \text{trace}\left\{\left(\mathbf{P}_k^H \mathbf{P}_k\right)^{-1}\right\}}{L_{max} N_t}, \end{aligned} \tag{8}$$

where $\boldsymbol{\Phi}_k = \mathbb{E}\left\{\mathbf{h}_k \mathbf{h}_k^H\right\}$, and $\mathbf{V}_k = \mathbf{P}_k^H \mathbf{D}_k$ ($\mathbf{V}_k$ represents the power spread of $\mathbf{D}_k$ to the signal space of $\mathbf{P}_k$). When in the presence of frequency offsets, $\text{trace}\left\{\mathbf{V}_k^H \left(\mathbf{P}_k^H \mathbf{P}_k\right)^{-2} \mathbf{V}_k \boldsymbol{\Phi}_k\right\} = 0$ is achieved if and only if the following two conditions are satisfied simultaneously:

Preposition 1:

1. $\mathbf{x}_i^d(m) \mathbf{x}_k^{pH}(q) = 0$ for each $1 \leq i, k \leq N_t$, $0 \leq m, q \leq M-1$; i.e., the subcarrier spaces allocated to data symbols are orthogonal to each pilot subcarrier space of each transmit antenna.
2. $\mathbf{P}_k^\dagger \mathbf{D}_k = \left(\mathbf{P}_k^H \mathbf{P}_k\right)^{-1} \mathbf{P}_k^H \mathbf{D}_k = \mathbf{O}_{L_{max} N_t}$.

The second condition in **Preposition 1** can be derived as follows:

$$\begin{aligned} &\text{trace}\left\{\mathbf{V}_k^H \left(\mathbf{P}_k^H \mathbf{P}_k\right)^{-2} \mathbf{V}_k \boldsymbol{\Phi}_k\right\} = 0 \\ &\Leftrightarrow \text{trace}\left\{\left(\left(\mathbf{P}_k^H \mathbf{P}_k\right)^{-1} \mathbf{V}_k \boldsymbol{\Phi}_k^{\frac{1}{2}}\right) \left(\left(\mathbf{P}_k^H \mathbf{P}_k\right)^{-1} \mathbf{V}_k \boldsymbol{\Phi}_k^{\frac{1}{2}}\right)^H\right\} = 0 \\ &\Leftrightarrow \left(\mathbf{P}_k^H \mathbf{P}_k\right)^{-1} \mathbf{V}_k \boldsymbol{\Phi}_k^{\frac{1}{2}} = \mathbf{O}_{L_{max} N_t}. \end{aligned} \tag{9}$$

Since both $\mathbf{P}_k^H\mathbf{P}_k$ and $\mathbf{\Phi}_k$ are non-zero Hermitian matrices, the only solution of (9) is $\mathbf{V}_k = \mathbf{O}_{L_{max}N_t}$, which proves the second condition in **Preposition 1**.

IV. Optimal Pilot Design and Placement for LS Channel Estimation in The Presence of Multiple Frequency Offsets

In the multiple frequency offsets case, define $\lambda_{p,k,j}^2$, $\lambda_{v,k,j}^2$ and $\lambda_{h,k,j}^2$ as the j-th eigenvalue of $\mathbf{P}_k^H\mathbf{P}_k$, $\mathbf{V}_k\mathbf{V}_k^H$ and $\mathbf{\Phi}_k$, respectively, where both $\lambda_{p,k,j}^2$ and $\lambda_{v,k,j}^2$ are functions of $\varepsilon_{k,1}, \cdots, \varepsilon_{k,N_t}$, and, therefore, (8) can be rewritten as

$$\begin{aligned}&\mathrm{MSE}\left(\hat{\mathbf{h}}_{k|\mathrm{LS}}\right)\\ &= \frac{1}{L_{max}N_t}\left(\sum_{j=0}^{L_{max}N_t-1}\frac{\lambda_{v,k,j}^2\lambda_{h,k,j}^2}{\lambda_{p,k,j}^4} + \sum_{j=0}^{L_{max}N_t-1}\frac{\sigma_w^2}{\lambda_{p,k,j}^2}\right),\\ &s.t. \quad \sum_{j=0}^{L_{max}N_t-1}\lambda_{p,k,j}^2 = L_{max}E_p.\end{aligned} \tag{10}$$

The optimal $\lambda_{p,k,j}^2$ that minimizes $\mathrm{MSE}\left(\hat{\mathbf{h}}_{k|\mathrm{LS}}\right)$ can be derived based on (10). Unfortunately, for a frequency selective fading MIMO-OFDM channel where $\lambda_{h,k,0}^2 \neq \lambda_{h,k,1}^2 \neq \cdots \neq \lambda_{h,k,L_{max}N_t-1}^2$, obtaining a closed-form resolution of $\lambda_{p,k,j}^2$ appears intractable, if not impossible. A closed-form resolution of $\lambda_{p,k,j}^2$ is achievable when $\lambda_{h,k,0}^2 = \lambda_{h,k,1}^2 = \cdots = \lambda_{h,k,L_{max}N_t-1}^2 = \lambda_{h,k}^2$, i.e., $\mathbf{\Phi}_k = \lambda_{h,k}^2\mathbf{I}_{L_{max}N_t}$. From the definition of $\mathbf{V}_k$, we can readily represent its j-th eigenvalue as $\lambda_{v,k,j}^2 = \alpha_{d,k,j}^2\lambda_{p,k,j}^2$, which means that $\lambda_{v,k,j}^2$ is proportional to $\lambda_{p,k,j}^2$ (Note that $\alpha_{d,k,j}^2$ here is a function of the frequency offset with $\alpha_{d,k,j}^2 \leq \lambda_{d,k,j}^2$, and that $\lambda_{d,k,j}^2$ is the n-th eigenvalue of $\mathbf{D}_k\mathbf{D}_k^H$. Equality holds only if the subcarriers allocated to the data and pilots for each transmit antenna are totally overlapped.) For MIMO-OFDM systems with frequency offsets, the interference of the pilots, contributed by the data subcarriers, is noise-like; therefore, it is reasonable to assume that the average interference power of each pilot subcarrier is identical, i.e., $\alpha_{d,k,0}^2 = \alpha_{d,k,1}^2 = \cdots = \alpha_{d,k,L_{max}N_t-1}^2 = \alpha_{d,k}^2$, and the optimal $\lambda_{p,k,j}^2$ can be obtained as

$$\lambda_{p,k,j}^2 = \frac{E_p}{N_t}, \quad 0 \leq j \leq L_{max}N_t - 1. \tag{11}$$

A. The Optimal Pilot Design and Placement

Before designing pilots that satisfy (11), let us first analyze $\mathbf{P}_k^H\mathbf{P}_k$:

$$\mathbf{P}_k^H\mathbf{P}_k = \begin{bmatrix}\mathbf{G}_{k,1,1} & \cdots & \mathbf{G}_{k,1,N_t}\\ \vdots & \ddots & \vdots\\ \mathbf{G}_{k,N_t,1} & \cdots & \mathbf{G}_{k,N_t,N_t}\end{bmatrix}, \tag{12}$$

where

$$\mathbf{G}_{k,m,n} = N\sum_{i=0}^{M-1}\mathbf{F}_{(L_{max})}\mathbf{X}_m^{pH}(i)\mathbf{E}_{k,m,n}^{\mathrm{cir}}\mathbf{X}_n^p(i)\mathbf{F}_{(L_{max})}^H, \tag{13}$$

$1 \leq m, n \leq N_t$, with the ls-th element of $\mathbf{E}_{k,m,n}^{\mathrm{cir}}$ being given by

$$\begin{aligned}[\mathbf{E}_{k,m,n}^{\mathrm{cir}}]_{ls} = &\frac{\sin\left(\pi\left((s-l)_N + (\varepsilon_{k,n} - \varepsilon_{k,m})\right)\right)}{N\sin\left(\frac{\pi}{N}\left((s-l)_N + (\varepsilon_{k,n} - \varepsilon_{k,m})\right)\right)}\\ &\times e^{\frac{\jmath(N-1)\left((s-l)_N + \left(\varepsilon_{k,n} - \varepsilon_{k,m}\right)\right)}{N}}.\end{aligned} \tag{14}$$

Note that $\mathbf{G}_{k,m,n} = \mathbf{G}_{k,n,m}^H$ represents the Inter-Pilot-Interference (IPI). To satisfy the requirement of (11), the following conditions should be satisfied simultaneously:

Preposition 2:

1. $\mathbf{G}_{k,n,n} = \frac{E_p}{N_t}\mathbf{I}_{L_{max}}$, $1 \leq n \leq N_t$.
2. $\mathbf{G}_{k,m,n\neq m} = \mathbf{O}_{L_{max}}$ for each $(m, n \neq m)$.

The first condition of **Preposition 2** is easy to explain: since (11) requires that all the eigenvalues of $\mathbf{P}_k^H\mathbf{P}_k$ to be identical, each $\mathbf{G}_{k,n,n}$ should be a diagonal matrix with an identical diagonal element. Since the first item in (8) is noise-like, which cannot be minimized by optimizing the pilots, so that

$$\arg\min_{\mathbf{P}_k}\mathrm{MSE}\left(\hat{\mathbf{h}}_{k|\mathrm{LS}}\right) \Leftrightarrow \arg\min_{\mathbf{P}_k}\mathrm{trace}\left\{\left(\mathbf{P}_k^H\mathbf{P}_k\right)^{-1}\right\} \tag{15}$$

For a given power constraint $\mathrm{trace}\left\{\mathbf{P}_k^H\mathbf{P}_k\right\} = L_{max}E_p$, minimizing $\mathrm{trace}\left\{\left(\mathbf{P}_k^H\mathbf{P}_k\right)^{-1}\right\}$ requires the minimization of the off-diagonal power of $\mathbf{P}_k^H\mathbf{P}_k$, and when the second condition of **Preposition 2** is satisfied, (15) is met.

In order to find the pilots and their placements that satisfy **Preposition 2**, let us first assume that a total of $\mathcal{N}_p$ pilots are allocated to each transmit antenna, and that the frequency-domain indexes of the pilots for the n-th transmit antenna is $(\theta_{n,1}, \cdots, \theta_{n,\mathcal{N}_p})$, where $0 \leq \theta_{n,1} < \cdots < \theta_{n,\mathcal{N}_p} \leq N-1$. To make $\mathbf{P}_k^H\mathbf{P}_k$ full-rank, $\mathcal{N}_p \geq L_{max}$ should be satisfied. Note that $\theta_{n,z}$ for each $1 \leq n \leq N_t$ may be either identical or not. When $M \geq N_t$, the pilots and data transmitted by different transmit antennas can also be orthogonal in the time-domain.

For each transmit antenna, to satisfy condition 1 of **Preposition 2**, the total transmit power of the consecutive M symbols in each pilot subcarrier should be identical (i.e., $\sum_{i=0}^{M-1}\mathbf{X}_n^{pH}(i)\mathbf{X}_n^p(i) = \frac{E_p}{\mathcal{N}_pN_t}\mathrm{diag}\{0, \cdots, 1, \cdots, 0\}$), and the pilots' placement requires

$$\begin{aligned}\left(\theta_{n,2}\cdot l - \theta_{n,1}\cdot l\right)_N = \cdots \left(\theta_{n,\mathcal{N}_p}\cdot l - \theta_{n,\mathcal{N}_p-1}\cdot l\right)_N\\ = \left(\theta_{n,1}\cdot l - \theta_{n,\mathcal{N}_p}\cdot l\right)_N\end{aligned} \tag{16}$$

for each $1 \leq n \leq N_t$, $1 \leq l \leq L_{max} - 1$. Note that $\sum_{i=0}^{M-1}\mathbf{X}_n^{pH}(i)\mathbf{X}_n^p(i) = \frac{E_p}{N_t\mathcal{N}_p}\mathrm{diag}\{0, \cdots, 1, \cdots, 0\}$ should not be understood as indicating that each non-zero element of $\mathbf{X}_n^p(i)$ must be modulated as a constant-envelope (CE) pilot, although a CE is a perfect resolution (e.g., CE pilots are proposed in [1]). A CE is required only when $M = 1$. If $M > 1$, we need neither to modulate each pilot subcarrier for each symbol nor to modulate these non-zero pilots as a CE. For example, each pilot shown in Table I to Table II satisfies this requirement.

Since the single frequency offset (SFO) case is a special condition of the multiple frequency offsets (MFO) case, the optimal pilot for a multiple frequency offsets case should also be optimal in a single frequency offset case. Therefore, we can simplify the condition 2 of **Preposition 2** as

$$\sum_{i=0}^{M-1} \mathbf{F}_{(L_{max})} \mathbf{X}_m^{pH}(i) \mathbf{X}_n^p(i) \mathbf{F}_{(L_{max})}^H = \mathbf{O}_{L_{max}}, \text{ (SFO)}; \tag{17a}$$

$$\mathbf{F}_{(L_{max})} \mathbf{\Pi}_{mn} \mathbf{F}_{(L_{max})}^H = \mathbf{O}_{L_{max}}, \text{ (MFO)}, \tag{17b}$$

where $\mathbf{\Pi}_{mn} = \sum_{i=0}^{M-1} \mathbf{X}_m^{pH}(i) \mathbf{E}_{k,m,n}^{\text{cir}} \mathbf{X}_n^p(i)$, and (17b) implies (17a). Also define a $\mathcal{N}_p \times \mathcal{N}_p$ matrix $\underline{\mathbf{\Pi}}_{mn}$, which is generated by deleting all the zero rows and columns of $\mathbf{\Pi}_{mn}$. In Appendix A, it is shown that when pilot subcarriers allocated to the different transmit antennas are orthogonal in the frequency-domain, $\mathbf{G}_{k,m,n\neq m} = \mathbf{O}_{L_{max}}$ for each $m \neq n$, and that, therefore, all pilots and their placements that satisfy (18) are optimal pilots in terms of the minimum MSE:

$$\begin{aligned}
&\arg\left\{[\mathbf{X}_m^p(i)]_{\theta_{m,z}\theta_{m,z}}\right\} = \frac{2\pi\theta_{m,z}K_p m}{N}, \\
&\quad s.t.\ L_{max} \leq \mathcal{N}_p \leq N, \quad \frac{N}{\mathcal{N}_p} = \text{integer}; \\
&\qquad K_p \geq L_{max}, \quad \frac{\theta_{m,z}K_p(n-m)}{N} \neq \text{integer}; \\
&\qquad \left|[\mathbf{X}_m^p(i)]_{\theta_{m,z}\theta_{m,z}}\right|^2 = \left|[\mathbf{X}_m^p(i)]_{\left(\theta_{m,z}+\frac{N}{2}\right)_N \left(\theta_{m,z}+\frac{N}{2}\right)_N}\right|^2; \\
&\qquad \sum_{i=1}^{M} \left|[\mathbf{X}_m^p(i)]_{\theta_{m,z}\theta_{m,z}}\right|^2 = \frac{E_p}{\mathcal{N}_p N_t}; \\
&\qquad \sum_{i=1}^{M} \sum_{z=1}^{\mathcal{N}_p} \left|[\mathbf{X}_m^p(i)]_{\theta_{m,z}\theta_{m,z}}\right|^2 = \frac{E_p}{N_t}.
\end{aligned} \tag{18}$$

where $1 \leq z \leq \mathcal{N}_p$, $1 \leq m \leq N_t$ and $0 \leq i \leq M-1$.

When $\theta_{m,z} = \theta_{n\neq m,z} = \theta_z$ for each $1 \leq m, n \neq m \leq N_t$ and $1 \leq z \leq \mathcal{N}_p$, the N_t transmit antennas share $\mathcal{N}_p$ pilot subcarriers. From (34) in Appendix A, we know that in order to make $\mathbf{G}_{k,m,n} = \mathbf{O}_{L_{max}}$, $\boldsymbol{\mu}_u^T \underline{\mathbf{\Pi}}_{mn} \boldsymbol{\mu}_t^* = 0$ should be satisfied for each $1 \leq u, t \leq L_{max}$. One way to achieve this result is to make $\underline{\mathbf{\Pi}}_{mn} = \mathbf{O}_{\mathcal{N}_p}$, which requires the pilots transmitted by different transmit antennas be orthogonal in the time-domain. These time-domain orthogonal pilots achieve optimal MSE performance, but at the cost of low spectral efficiency.

B. The Suboptimal Pilot Design and Placement to Minimize the Expectation of MSE

Note that the MSE given by (8) is a function of $\varepsilon_{k,1}, \cdots, \varepsilon_{k,N_t}$. In the multiple frequency offsets case, designing the optimal pilots that satisfy $\mathbf{F}_{(L_{max})} \left[\sum_{i=0}^{M-1} \mathbf{X}_m^{pH}(i) \mathbf{E}_{k,m,n}^{\text{cir}} \mathbf{X}_n^p(i)\right] \mathbf{F}_{(L_{max})}^H = \mathbf{O}_{L_{max}}$ is difficult, if the pilots transmitted by the different transmit antennas are neither orthogonal in the time-domain nor orthogonal in the frequency-domain.

However, orthogonal pilot placement considerably reduces the spectral efficiency, and in this subsection, we consider only the case of $\theta_{m,z} = \theta_{n\neq m,z} = \theta_z$ for each ($1 \leq m, n \neq m \leq N_t, 1 \leq z \leq \mathcal{N}_p$) with $M < N_t$. Since the IPI cannot be totally eliminated in this case, we can get

TABLE I
OPTIMAL PILOTS ALLOCATION IN MIMO-OFDM WITH $N = 32$, $N_t = 4$, $L_{max} = 4, \mathcal{N}_p = 4$ AND $M = 2$: CONSTANT ENVELOPE.

Symbol 1				
θ_z	Tx 1	Tx 2	Tx 3	Tx 4
0	$\frac{\rho}{\sqrt{2}}$	$\frac{\rho}{\sqrt{2}}$	$\frac{\rho}{\sqrt{2}}$	$\frac{\rho}{\sqrt{2}}$
8	$\frac{\rho}{\sqrt{2}}e^{\frac{\jmath\pi}{2}}$	$\frac{\rho}{\sqrt{2}}e^{\jmath\pi}$	$\frac{\rho}{\sqrt{2}}e^{\frac{\jmath3\pi}{2}}$	$\frac{\rho}{\sqrt{2}}e^{\jmath2\pi}$
16	$\frac{\rho}{\sqrt{2}}e^{\jmath\pi}$	$\frac{\rho}{\sqrt{2}}e^{\jmath2\pi}$	$\frac{\rho}{\sqrt{2}}e^{\jmath3\pi}$	$\frac{\rho}{\sqrt{2}}e^{\jmath4\pi}$
24	$\frac{\rho}{\sqrt{2}}e^{\frac{\jmath3\pi}{2}}$	$\frac{\rho}{\sqrt{2}}e^{\jmath3\pi}$	$\frac{\rho}{\sqrt{2}}e^{\frac{\jmath9\pi}{2}}$	$\frac{\rho}{\sqrt{2}}e^{\jmath6\pi}$
Symbol 2				
θ_z	Tx 1	Tx 2	Tx 3	Tx 4
0	$\frac{\rho}{\sqrt{2}}$	$\frac{\rho}{\sqrt{2}}$	$\frac{\rho}{\sqrt{2}}$	$\frac{\rho}{\sqrt{2}}$
8	$\frac{\rho}{\sqrt{2}}e^{\frac{\jmath\pi}{2}}$	$\frac{\rho}{\sqrt{2}}e^{\jmath\pi}$	$\frac{\rho}{\sqrt{2}}e^{\frac{\jmath3\pi}{2}}$	$\frac{\rho}{\sqrt{2}}e^{\jmath2\pi}$
16	$\frac{\rho}{\sqrt{2}}e^{\jmath\pi}$	$\frac{\rho}{\sqrt{2}}e^{\jmath2\pi}$	$\frac{\rho}{\sqrt{2}}e^{\jmath3\pi}$	$\frac{\rho}{\sqrt{2}}e^{\jmath4\pi}$
24	$\frac{\rho}{\sqrt{2}}e^{\frac{\jmath3\pi}{2}}$	$\frac{\rho}{\sqrt{2}}e^{\jmath3\pi}$	$\frac{\rho}{\sqrt{2}}e^{\frac{\jmath9\pi}{2}}$	$\frac{\rho}{\sqrt{2}}e^{\jmath6\pi}$

TABLE II
OPTIMAL PILOTS ALLOCATION IN MIMO-OFDM WITH $N = 32$, $N_t = 4$, $L_{max} = 4, \mathcal{N}_p = 4$ AND $M = 3$: MULTIPLE ENVELOPES.

Symbol 1				
θ_z	Tx 1	Tx 2	Tx 3	Tx 4
0	ρ	ρ	0	0
8	$\frac{\rho}{\sqrt{3}}e^{\frac{\jmath\pi}{2}}$	$\frac{\rho}{\sqrt{3}}e^{\jmath\pi}$	$\frac{\rho}{\sqrt{3}}e^{\frac{\jmath3\pi}{2}}$	$\frac{\rho}{\sqrt{3}}e^{\jmath2\pi}$
16	$\rho e^{\jmath\pi}$	$\rho e^{\jmath2\pi}$	0	0
24	$\frac{\rho}{\sqrt{3}}e^{\frac{\jmath3\pi}{2}}$	$\frac{\rho}{\sqrt{3}}e^{\jmath3\pi}$	$\frac{\rho}{\sqrt{3}}e^{\frac{\jmath9\pi}{2}}$	$\frac{\rho}{\sqrt{3}}e^{\jmath6\pi}$
Symbol 2				
θ_z	Tx 1	Tx 2	Tx 3	Tx 4
0	0	0	$\frac{\rho}{\sqrt{2}}$	$\frac{\rho}{\sqrt{2}}$
8	$\frac{\rho}{\sqrt{3}}e^{\frac{\jmath\pi}{2}}$	$\frac{\rho}{\sqrt{3}}e^{\jmath\pi}$	$\frac{\rho}{\sqrt{3}}e^{\frac{\jmath3\pi}{2}}$	$\frac{\rho}{\sqrt{3}}e^{\jmath2\pi}$
16	0	0	$\frac{\rho}{\sqrt{2}}e^{\jmath3\pi}$	$\frac{\rho}{\sqrt{2}}e^{\jmath4\pi}$
24	$\frac{\rho}{\sqrt{3}}e^{\frac{\jmath3\pi}{2}}$	$\frac{\rho}{\sqrt{3}}e^{\jmath3\pi}$	$\frac{\rho}{\sqrt{3}}e^{\frac{\jmath9\pi}{2}}$	$\frac{\rho}{\sqrt{3}}e^{\jmath6\pi}$
Symbol 3				
θ_z	Tx 1	Tx 2	Tx 3	Tx 4
0	0	0	$\frac{\rho}{\sqrt{2}}$	$\frac{\rho}{\sqrt{2}}$
8	$\frac{\rho}{\sqrt{3}}e^{\frac{\jmath\pi}{2}}$	$\frac{\rho}{\sqrt{3}}e^{\jmath\pi}$	$\frac{\rho}{\sqrt{3}}e^{\frac{\jmath3\pi}{2}}$	$\frac{\rho}{\sqrt{3}}e^{\jmath2\pi}$
16	0	0	$\frac{\rho}{\sqrt{2}}e^{\jmath3\pi}$	$\frac{\rho}{\sqrt{2}}e^{\jmath4\pi}$
24	$\frac{\rho}{\sqrt{3}}e^{\frac{\jmath3\pi}{2}}$	$\frac{\rho}{\sqrt{3}}e^{\jmath3\pi}$	$\frac{\rho}{\sqrt{3}}e^{\frac{\jmath9\pi}{2}}$	$\frac{\rho}{\sqrt{3}}e^{\jmath6\pi}$

suboptimal pilots only in terms of the MSE. These suboptimal pilots can be designed by minimizing the expectation of $\text{MSE}\left(\hat{\mathbf{h}}_{k|\varepsilon_{k,1},\cdots,\varepsilon_{k,N_t}}\right)$; i.e.,

$$\begin{aligned}
\mathbf{P}_k^{\text{subopt}} &= \arg\min_{\mathbf{P}_k} \mathbb{E}\left\{\text{MSE}\left(\hat{\mathbf{h}}_{k|\varepsilon_{k,1},\cdots,\varepsilon_{k,N_t}}\right)\right\}, \\
&\quad s.t. \quad \text{trace}\left\{\left(\mathbf{P}_k^H \mathbf{P}_k\right)\right\} = L_{max} E_p,
\end{aligned} \tag{19}$$

where the expectation is with respect to $\varepsilon_{k,1}, \cdots, \varepsilon_{k,N_t}$. When the eigenvalues of $\mathbf{P}_k^H \mathbf{P}_k$ are identical, the minimum MSE is achieved, which requires that

$$\sum_{i=0}^{M-1} \mathbf{P}_{k,m}^H(i) \mathbf{P}_{k,n\neq m}(i) = \mathbf{O}_{L_{max}}, \text{ (SFO)}; \tag{20a}$$

$$\arg\min_{\mathbf{P}_k} \sum_{n\neq m} \left| \mathbb{E}\left\{\text{trace}\left\{\sum_{i=0}^{M-1} \mathbf{P}_{k,m}^H(i) \mathbf{P}_{k,n}(i)\right\}\right\}\right|^2, \text{ (MFO)}. \tag{20b}$$

(20a) can be achieved by all the pilots that satisfy (18). To resolve (20b), we have

$$\mathbb{E}\left\{\text{trace}\left\{\sum_{i=0}^{M-1}\mathbf{P}_{k,m}^{H}(i)\mathbf{P}_{k,n}(i)\right\}\right\} = \text{trace}\left\{\sum_{i=0}^{M-1}\mathbf{F}_{(L_{max})}\mathbf{X}_{m}^{pH}(i)\mathbf{F}^{H}\boldsymbol{\Lambda}_{p}\mathbf{F}\mathbf{X}_{n}^{p}(i)\mathbf{F}_{(L_{max})}^{H}\right\}, \tag{21}$$

where $\boldsymbol{\Lambda}_p \cong \text{diag}\left\{0, -\frac{4\pi^2\times 4}{N^2}\sigma_\varepsilon^2, \cdots, -\frac{4\pi^2(N-1)^2}{N^2}\sigma_\varepsilon^2\right\}$. To minimize (21), it requires

$$\sum_{m=1}^{N_t}\sum_{i=0}^{M-1}\mathbf{F}_{(L_{max})}\mathbf{X}_{m}^{pH}(i)\mathbf{F}^{H} = \frac{E_p}{N_t}\left[\underbrace{\mathbf{B}_L,\cdots,\mathbf{B}_L}_{N/\mathcal{N}_p \text{ times } \mathbf{B}_L}\right], \tag{22}$$

where $\mathbf{B}_L = \left[\underbrace{\mathbf{I}_{L_{max}},\cdots,\mathbf{I}_{L_{max}}}_{L_{max}\times(L_{max}N_t)}, \mathbf{O}_{L_{max}\times(\mathcal{N}_p - L_{max}N_t)}\right]$, and $\mathcal{N}_p \geq L_{max}N_t$ should be satisfied. By resolving (20) to (22), the suboptimal pilots and their placements are

$$\begin{aligned}
&\arg\left\{[\mathbf{X}_m^p(i)]_{\theta_z\theta_z}\right\} = \frac{2\pi\theta_z(m-1)L_{max}}{N},\\
&\quad s.t.\ L_{max}N_t \leq \mathcal{N}_p \leq N, \quad \frac{N}{\mathcal{N}_p} = \text{integer};\\
&\quad\quad (\theta_2\cdot l - \theta_1\cdot l)_N = \cdots \left(\theta_{\mathcal{N}_p}\cdot l - \theta_{\mathcal{N}_p - 1}\cdot l\right)_N\\
&\quad\quad\quad = \left(\theta_1 \cdot l - \theta_{\mathcal{N}_p}\cdot l\right)_N, l = 1,2,\cdots,L_{max}-1;\\
&\quad\quad \frac{\theta_z(n-m)L_{max}}{N} \neq \text{integer};\\
&\quad\quad [\mathbf{X}_m^p(i)]_{\theta_1\theta_1} = \cdots = [\mathbf{X}_m^p(i)]_{\theta_{\mathcal{N}_p}\theta_{\mathcal{N}_p}} \geq 0;\\
&\quad\quad \sum_{i=1}^{M}\left|[\mathbf{X}_m^p(i)]_{\theta_z\theta_z}\right|^2 = \frac{E_p}{\mathcal{N}_p N_t};\\
&\quad\quad \sum_{i=1}^{M}\sum_{z=1}^{\mathcal{N}_p}\left|[\mathbf{X}_m^p(i)]_{\theta_z\theta_z}\right|^2 = \frac{E_p}{N_t}.
\end{aligned} \tag{23}$$

When pilots for different transmit antennas are orthogonal in the frequency-domain, pilots and their placements that satisfy (18) are the optimal; if pilots for different transmit antennas are overlapped in the frequency-domain, pilots and their placements that satisfy (23) are sub-optimal in terms of LS MSE minimization. For a given $(N, N_t, L_{max}, \mathcal{N}_p, M)$, the number of the optimal pilots is so large that numerically exemplifying all the optimal pilots as well as their placements would be trivial. Table I to Table II present some examples of the optimal pilots and their placements that satisfy both (18) and (23) when $N = 32$, $N_t = 4$, $L_{max} = 4$, $\mathcal{N}_p = 4$ and $M = 2, 3$.

Note that both the optimal pilots and the suboptimal pilots are uniformly placed in the frequency-domain, and may suffer a performance degrade due to the pilots losses at the virtual subcarriers. One way to mitigate this performance loss here is to make the pilot distance larger than the virtual-subcarrier bandwidth and to guarantee that $\mathcal{N}_p \geq L_{max}N_t$ simultaneously.

V. Linear Minimum Mean Square Error (LMMSE) Estimator by Exploiting The Channel Correlation

Sections III to IV considered pilot design for an uncorrelated MIMO-OFDM channel. However, channel correlation is usually exploited to improve the performance, as discussed in [15], [16].

A. LMMSE Estimator

Define the received vector on the N_r receive antennas as $\mathbf{r} = \left[\mathbf{r}_1^T,\cdots,\mathbf{r}_{N_r}^T\right]^T = \mathbf{Ph} + \mathbf{Dh} + \boldsymbol{\eta}$, where $\mathbf{P} = \text{diag}\{\mathbf{P}_1,\cdots,\mathbf{P}_{N_r}\}$, $\mathbf{D} = \text{diag}\{\mathbf{D}_1,\cdots,\mathbf{D}_{N_r}\}$,$\mathbf{h} = \left[\mathbf{h}_1^T,\cdots,\mathbf{h}_{N_r}^T\right]^T$, $\boldsymbol{\eta} = \left[\boldsymbol{\eta}_1^T,\cdots,\boldsymbol{\eta}_{N_r}^T\right]^T$. Based on it, we define

$$\tilde{\mathbf{h}} = \mathbf{P}^\dagger\mathbf{r} = \mathbf{h} + \mathbf{P}^\dagger\mathbf{D}\mathbf{h} + \mathbf{P}^\dagger\boldsymbol{\eta}, \tag{24}$$

where $\mathbf{P}^\dagger = \left(\mathbf{P}^H\mathbf{P}\right)^{-1}\mathbf{P}^H$. The channel correlation matrix is given by $\boldsymbol{\Phi} = \mathbb{E}\left\{\mathbf{h}\mathbf{h}^H\right\} = \begin{bmatrix}\boldsymbol{\Phi}_{1,1} & \boldsymbol{\Phi}_{1,2} & \cdots & \boldsymbol{\Phi}_{1,N_r}\\ \boldsymbol{\Phi}_{2,1} & \boldsymbol{\Phi}_{2,2} & \cdots & \boldsymbol{\Phi}_{2,N_r}\\ \vdots & \vdots & \ddots & \vdots\\ \boldsymbol{\Phi}_{N_r,1} & \boldsymbol{\Phi}_{N_r,2} & \cdots & \boldsymbol{\Phi}_{N_r,N_r}\end{bmatrix}$, where $\boldsymbol{\Phi}_{k,i} = \mathbb{E}\left\{\mathbf{h}_k\mathbf{h}_i^H\right\} = \begin{bmatrix}\boldsymbol{\Phi}_{k,i,1,1} & \boldsymbol{\Phi}_{k,i,1,2} & \cdots & \boldsymbol{\Phi}_{k,i,1,N_t}\\ \boldsymbol{\Phi}_{k,i,2,1} & \boldsymbol{\Phi}_{k,i,2,2} & \cdots & \boldsymbol{\Phi}_{k,i,2,N_t}\\ \vdots & \vdots & \ddots & \vdots\\ \boldsymbol{\Phi}_{k,i,N_t,1} & \boldsymbol{\Phi}_{k,i,N_t,2} & \cdots & \boldsymbol{\Phi}_{k,i,N_t,N_t}\end{bmatrix}$ and $\boldsymbol{\Phi}_{k,i,m,n} = \mathbb{E}\left\{\tilde{\mathbf{h}}_{k,m}\tilde{\mathbf{h}}_{i,n}^H\right\} = \varrho_{k,i,m,n}\cdot \text{diag}\left\{\sigma_{h,0}^2,\sigma_{h,1}^2,\cdots,\sigma_{h,L_{max}-1}^2\right\}$ with $\sigma_{h,l}^2 = \mathbb{E}\left\{|h_{k,i}(l)|^2\right\}$ for each (k,i). In the following sections, we assume that $\sum_{l=0}^{L_{max}-1}\sigma_{h,l}^2 = 1$. Note that $\varrho_{k,k,m,m} = 1$ and that $0 \leq \varrho_{k,i,m,n} \leq 1$ for each $k \neq i$ or $m \neq n$. The correlation matrix of $\tilde{\mathbf{h}}$ is given by

$$\mathbf{C}_{\tilde{\mathbf{h}}\tilde{\mathbf{h}}} = \boldsymbol{\Omega}\boldsymbol{\Phi}\boldsymbol{\Omega}^H + \sigma_w^2\left(\mathbf{P}^H\mathbf{P}\right)^{-1}, \tag{25}$$

where $\boldsymbol{\Omega} = \left(\mathbf{I}_{L_{max}N_tN_r} + \left(\mathbf{P}^H\mathbf{P}\right)^{-1}\underbrace{\mathbf{P}^H\mathbf{D}}_{\mathbf{V}}\right)$. We also have $\mathbf{C}_{\mathbf{h}\tilde{\mathbf{h}}} = \mathbb{E}\left\{\mathbf{h}\tilde{\mathbf{h}}^H\right\} = \boldsymbol{\Phi}\boldsymbol{\Omega}^H$. From [23], an LMMSE estimator can be designed as

$$\hat{\mathbf{h}}_{\text{LMMSE}} = \left(\sigma_w^2\boldsymbol{\Phi}^{-1} + \boldsymbol{\Omega}^H\left(\mathbf{P}^H\mathbf{P}\right)\boldsymbol{\Omega}\right)^{-1}\boldsymbol{\Omega}^H\left(\mathbf{P}^H\mathbf{P}\right)\tilde{\mathbf{h}} \tag{26}$$

with its covariance matrix of estimation error being given by

$$\mathbf{C}_{\mathbf{e}} = \mathbb{E}\left\{\left(\hat{\mathbf{h}}_{\text{LMMSE}} - \mathbf{h}\right)\left(\hat{\mathbf{h}}_{\text{LMMSE}} - \mathbf{h}\right)^H\right\}. \tag{27}$$

The MSE of $\hat{\mathbf{h}}_{\text{LMMSE}}$ is $\text{MSE}\left(\hat{\mathbf{h}}_{\text{LMMSE}}\right) = \frac{\text{trace}\{\mathbf{C}_{\mathbf{e}}\}}{L_{max}N_tN_r}$.

B. Adaptive Power Allocation to Minimize MSE

Using the eigenvalue decomposition of $\mathbf{C_e}$, we have

$$\begin{aligned}&\mathrm{MSE}\left(\hat{\mathbf{h}}_{\mathrm{LMMSE}}\right)\\&=\sum_{j=0}^{L_{max}N_tN_r-1}\frac{\sigma_w^2\lambda_{\phi,j}^2\lambda_{p,j}^2}{L_{max}N_tN_r\left(\sigma_w^2\lambda_{p,j}^2+\lambda_{\phi,j}^2\left(\lambda_{p,j}^2+\lambda_{v,j}\right)^2\right)},\end{aligned} \tag{28}$$

where $\lambda_{\phi,j}^2$ represents the j-th eigenvalue of $\mathbf{\Phi}$, $\lambda_{p,j}^2$ and $\lambda_{v,j}^2$ represent the j-th eigenvalue of $\mathbf{P}^H\mathbf{P}$ and $\mathbf{V}\mathbf{V}^H$, respectively. By resolving (36) in Appendix B, the optimal $\lambda_{p,j}^2$ to minimize the MSE $\left(\hat{\mathbf{h}}_{\mathrm{LMMSE}}\right)$ can be derived as

$$\lambda_{p,j}^2=\frac{\left(N_rL_{max}E_p+\sum\limits_{m=0}^{L_{max}N_tN_r-1}\frac{\sigma_w^2}{\lambda_{\phi,m}^2}\right)}{L_{max}N_tN_r-1}-\frac{\sigma_w^2}{\lambda_{\phi,j}^2}, \tag{29}$$

where $0\leq j\leq L_{max}N_tN_r-1$. By considering $\sum_{j=0}^{L_{max}N_tN_r-1}\lambda_{p,j}^2=N_rL_{max}E_p$, the minimum MSE $\left(\hat{\mathbf{h}}_{\mathrm{LMMSE}}\right)$ is given by

$$\begin{aligned}&\min\left\{\mathrm{MSE}\left(\hat{\mathbf{h}}_{\mathrm{LMMSE}}\right)\right\}\\&\quad=\sum_{j=0}^{L_{max}N_tN_r-1}\frac{\sigma_w^2}{N_rL_{max}E_p+\sum\limits_{m=0}^{L_{max}N_tN_r-1}\frac{\sigma_w^2}{\lambda_{\phi,m}^2}}.\end{aligned} \tag{30}$$

As compared to (30), when uniform power allocation to the pilots is applied; i.e., when $\mathbf{P}^H\mathbf{P}=\frac{E_p}{N_t}\mathbf{I}_{L_{max}N_tN_r}$, the suboptimal MSE $\left(\hat{\mathbf{h}}_{\mathrm{LMMSE}}\right)$ is given by

$$\begin{aligned}&\mathrm{MSE}\left(\hat{\mathbf{h}}_{\mathrm{LMMSE}}\middle|\mathbf{P}^H\mathbf{P}=\frac{E_p}{N_t}\mathbf{I}_{L_{max}N_tN_r}\right)\\&\quad=\sum_{j=0}^{L_{max}N_tN_r-1}\frac{\sigma_w^2\lambda_{\phi,j}^2}{N_rL_{max}\left(N_t\sigma_w^2+\lambda_{\phi,j}^2E_p\right)}.\end{aligned} \tag{31}$$

VI. NUMERICAL RESULTS

A multipath-fading channel with path gains $h_{k,i}(l)=e^{-l}$ and the channel correlation coefficient $\varrho_{k,i,m,n\neq n}=0.5$ is considered, where $l=0,1,2,\cdots,L_{max}-1$, $k,i=1,2,\cdots,N_r$ and $m,n=1,2,\cdots,N_t$. We also assume that $N=128$, $M=1,2,4$, $L_{max}=4$, $N_t=2,4$, $N_r=2,4$ and $N_p=L_{max}N_t$. Here, we use the Normalized-Pilot-to-Noise Ratio (NPNR), i.e., $\mathrm{NPNR}=\frac{E_p}{N_t\cdot\mathcal{N}_p\cdot\sigma_w^2}$, instead of the PNR to represent the normalized PNR of each pilot subcarrier for each transmit antenna, and without loss of generality, the average power of each pilot subcarrier is also assumed to be identical to that of each data subcarrier, unless otherwise stated. Multiple frequency offsets are considered for different transmit and receive antennas with $\varepsilon_{k,i}\sim\mathcal{N}(0,\sigma_\varepsilon^2)$.

In a MIMO-OFDM system, pilots transmitted by N_t transmit antennas are modulated into consecutive $M\geq 1$ symbols (Sections III to IV). When $M=1$, the pilot subcarriers for the different transmit antennas can be either orthogonal in the frequency-domain or not. Fig. 1 to Fig. 5 compares the performance of the proposed LS estimator and the BLM estimator with $M=1$. Fig. 1 shows that with a single frequency offset, the BLM estimator exhibits a performance floor at high NPNR, and a larger σ_ε^2 implies a worse MSE performance. However, the proposed estimator eliminates this floor, and for different σ_ε^2, it achieves an identical MSE.

Fig. 2 to Fig. 5 are for the multiple frequency offsets case. Fig. 2 and Fig. 3 present the simulation results with overlapped pilot subcarriers allocated for the different transmit antennas when $N_t=2$ and $N_t=4$, respectively. In the multiple frequency offsets case, both the BLM estimator and the proposed

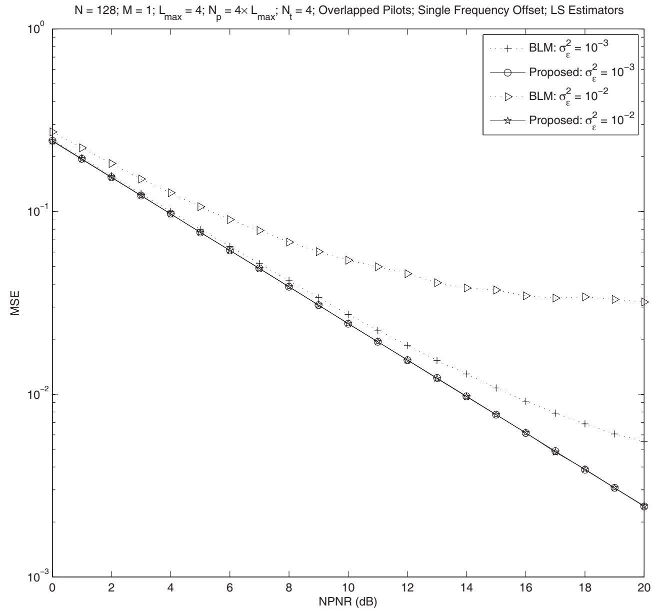

Fig. 1. LS channel estimation in single frequency offset MIMO-OFDM systems with $M=1$, $N_t=4$ and pilot subcarriers of different transmit antennas are overlapped.

3

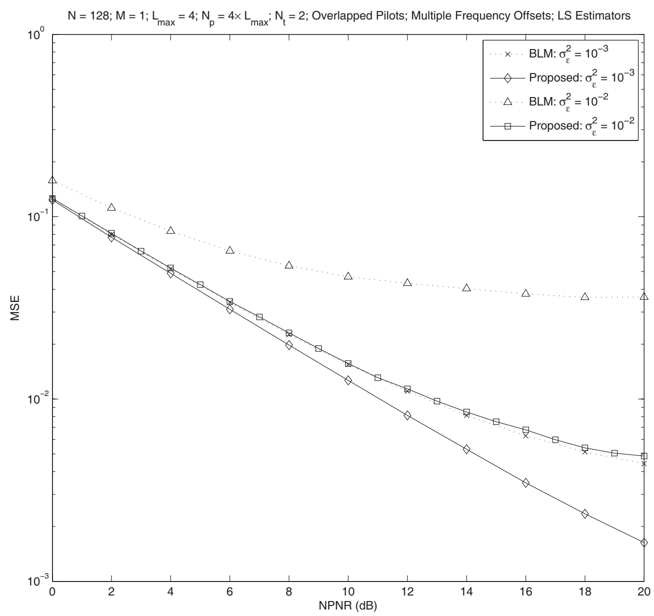

Fig. 2. LS channel estimation in multiple frequency offsets MIMO-OFDM systems with $M=1$, $N_t=2$ and pilot subcarriers of different transmit antennas are overlapped.

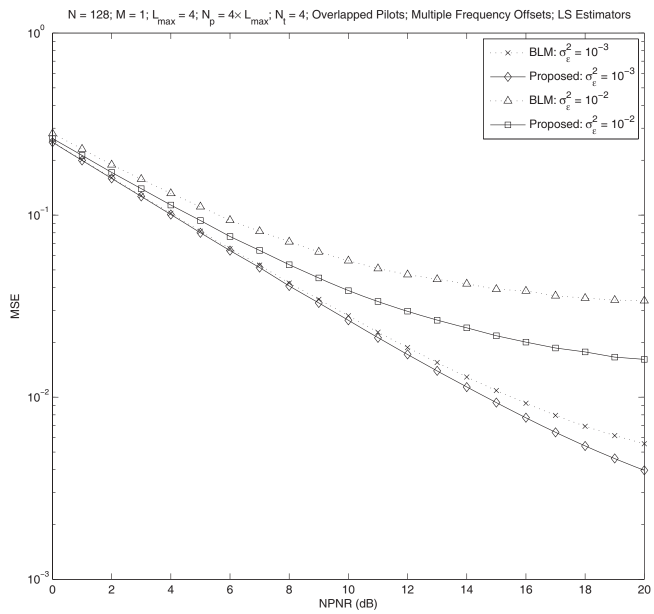

Fig. 3. LS channel estimation in multiple frequency offsets MIMO-OFDM systems with $M = 1$, $N_t = 4$ and pilot subcarriers of different transmit antennas are overlapped.

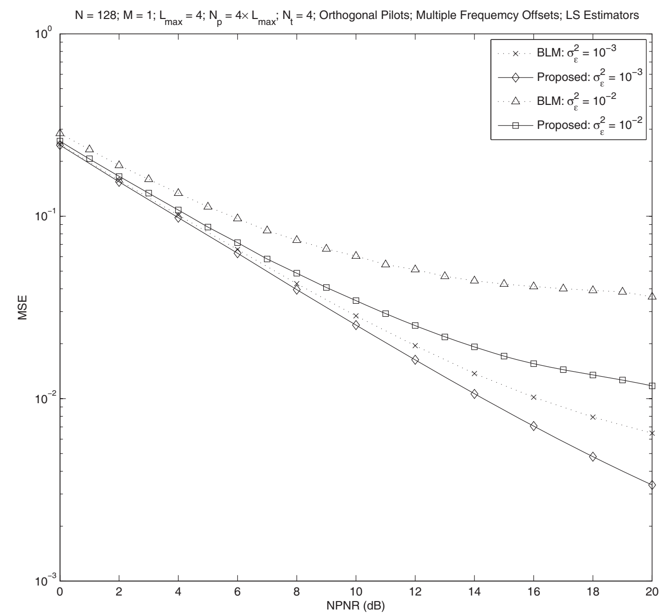

Fig. 5. LS channel estimation in multiple frequency offsets MIMO-OFDM systems with $M = 1$, $N_t = 4$ and pilot subcarriers of different transmit antennas are orthogonal.

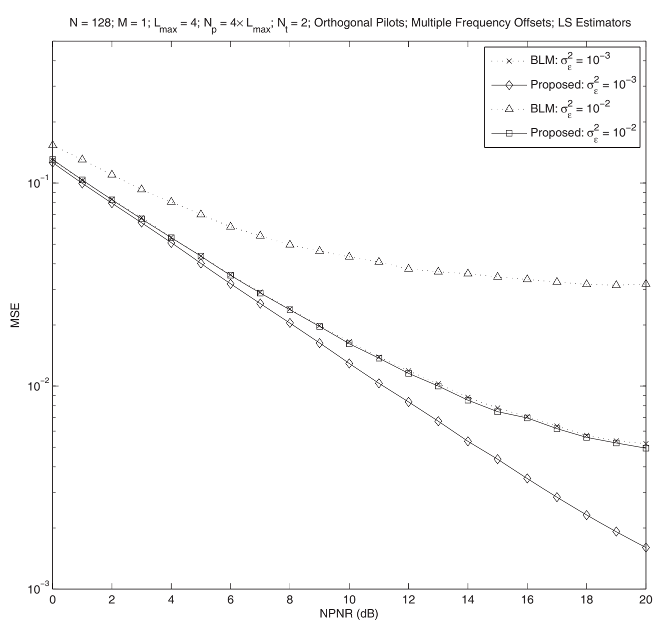

Fig. 4. LS channel estimation in multiple frequency offsets MIMO-OFDM systems with $M = 1$, $N_t = 2$ and pilot subcarriers of different transmit antennas are orthogonal.

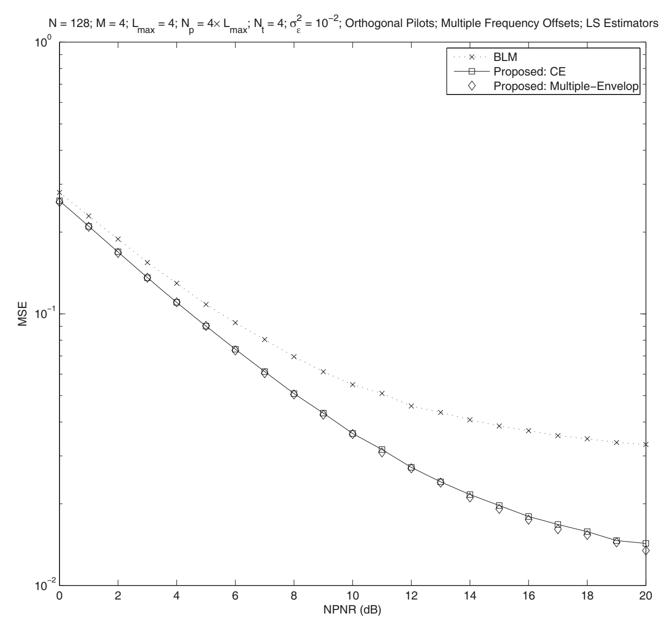

Fig. 6. LS channel estimation in multiple frequency offsets MIMO-OFDM systems with $M = 4$, $N_t = 4$ and pilot subcarriers of different transmit antennas are orthogonal.

one exhibit a performance floor. However, the latter maintains a performance advantage. For example, when $N_t = 2$ and NPNR=20 dB, the MSE of the proposed estimator is about 1.6×10^{-3} (or 5×10^{-3}) for $\sigma_\varepsilon^2 = 10^{-3}$ (or $\sigma_\varepsilon^2 = 10^{-2}$), and that for the BLM estimator is about 5.1×10^{-3} (or 3.1×10^{-2}) for $\sigma_\varepsilon^2 = 10^{-3}$ (or $\sigma_\varepsilon^2 = 10^{-2}$).

Fig. 4 and Fig. 5 compare the proposed estimator and the BLM estimator with orthogonal pilot placement in the frequency-domain. Note that even with orthogonal pilots placement, a performance floor will always appear at the proposed estimator in the presence of multiple frequency offsets, although the proposed estimator still outperforms the BLM estimator. Since in frequency-domain orthogonal pilot placement, the IPI is reduced as compared to that in the overlapped pilot placement, the MSE is smaller than that of the latter, but at the cost of lower spectral efficiency. For example, when $N_t = 4$, $\sigma_\varepsilon^2 = 10^{-2}$ and NPNR=20 dB, the MSE for the proposed estimator with overlapped pilots placement is about 1.7×10^{-2}, whereas that with orthogonal pilots placement is about 1.2×10^{-2}, as shown in Fig. 3 and Fig. 5.

When $N_t > 1$, pilots for each transmit antenna can be modulated into consecutive $M > 1$ symbols, the envelope can either be CE or not. The simulation results with $N_t = M = 4$ are illustrated in Fig. 6, where frequency-domain orthogonal

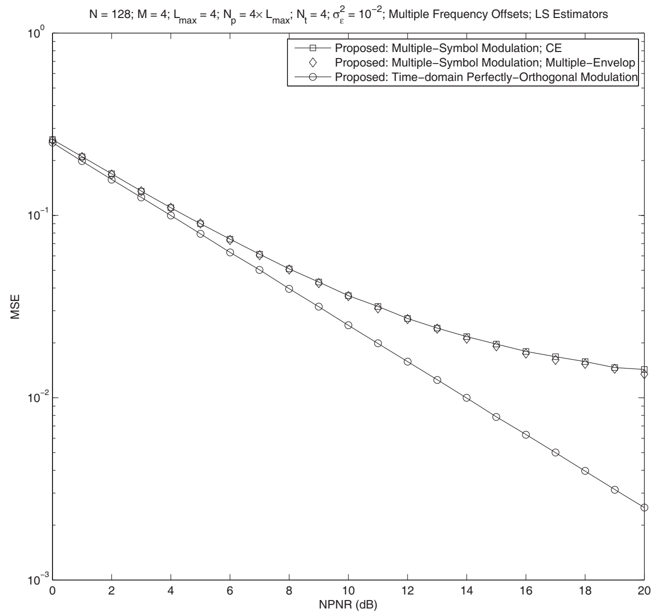

Fig. 7. LS channel estimation in multiple frequency offsets MIMO-OFDM systems with $M = 4$, $N_t = 4$ and pilots modulated with either consecutive multiple symbols or perfect orthogonal in time-domain.

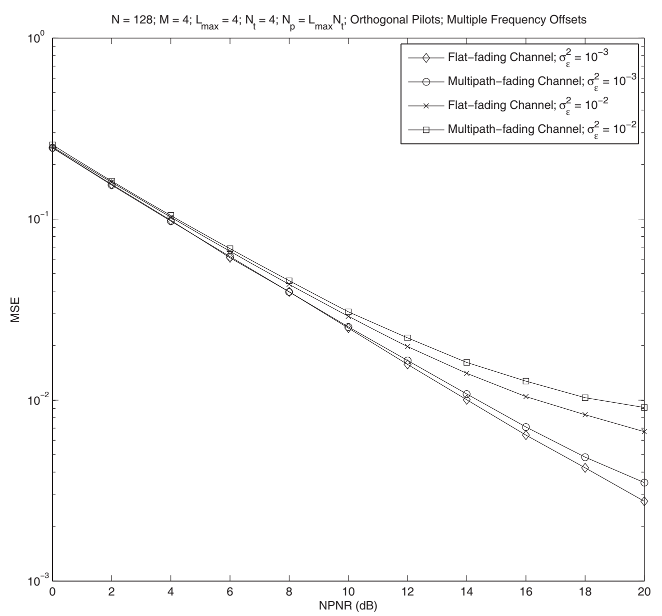

Fig. 8. Performance degradation of LS channel estimation in the multipath-fading channels with $M = 4$, $N_t = 4$ and pilot subcarriers of different transmit antennas are orthogonal.

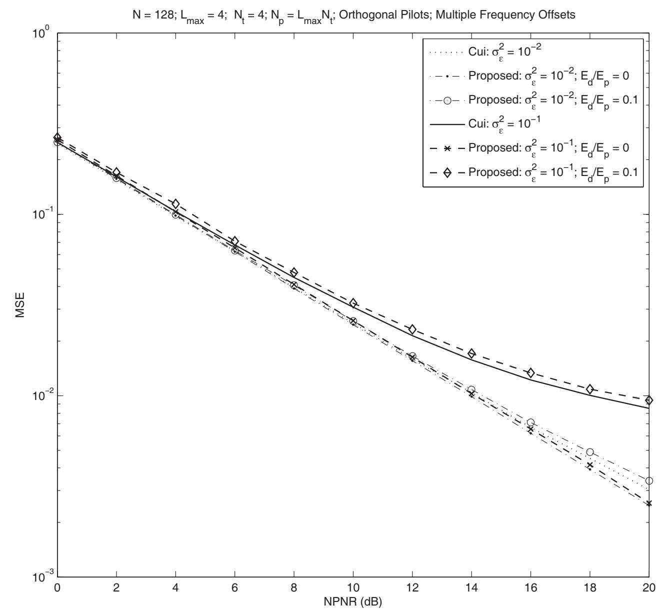

Fig. 9. Performance comparison between the proposed LS estimator and Cui's estimator with $L_{max} = 4$ and $N_t = 4$.

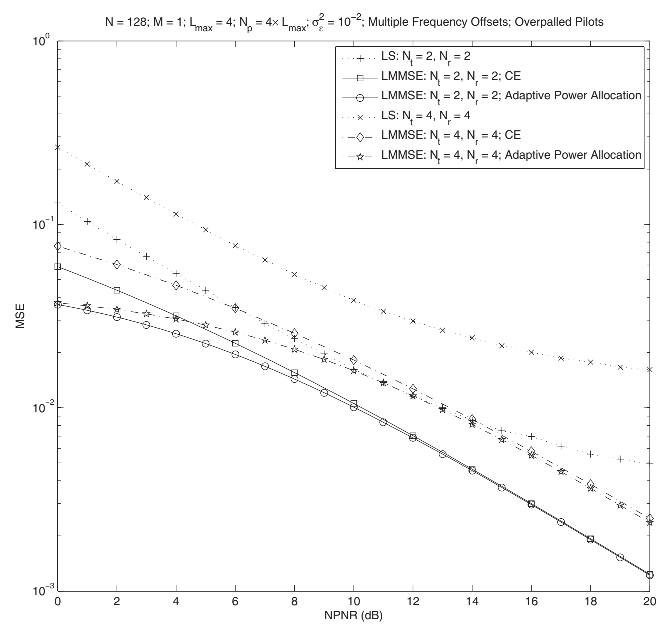

Fig. 10. Performance comparison between LS and LMMSE estimators.

pilot placement is assumed for the different transmit antennas in each symbol (a similar result can be achieved in the case of overlapped pilots placement for different transmit antennas). The proposed pilots, either a CE or Multiple-Envelope, can achieve the same performance advantage over that of the BLM estimator, provided that the total pilots power E_p remains fixed. For a given E_p, we can also conclude that an identical performance can be achieved in the proposed estimator with either $M = 1$ or $M > 1$. For example, when $N_t = 4$, $\sigma_\varepsilon^2 = 10^{-2}$ and NPNR=20 dB, an MSE of about 1.5×10^{-2} can be achieved with either $M = 1$ or $M = 4$, as shown in Fig. 5 and Fig. 6. No matter what pilots are used, multiple symbol pilots have no performance advantage over that of single-symbol pilots in terms of MSE, and more seriously, multiple-symbol pilot modulation has lower spectrum efficiency. Multiple-symbol pilot modulation is nevertheless applied for two main reasons: (1) it reduces the peak power of each pilot subcarrier in each symbol; and (2) it has an advantage over single-symbol pilot modulation in tracking the time-variant channel. In the presence of multiple frequency offsets, non-zero Inter-Antenna-Interference will degrade the channel estimation performance if neither time-domain nor frequency-domain orthogonal pilot placement is utilized. To eliminate this Inter-Antenna-Interference, we can make $M = N_t$ and ensure that the pilots and data transmitted by the different transmit antennas are orthogonal in the time-domain; i.e., when one transmit antenna is transmitting, the other transmit antennas remain silent. Fig. 7 presents the performance analysis of the proposed estimator in two pilot placement cases: the time-domain orthogonal and non-orthogonal pilot placements. The simulation results show that the performance floor at high NPNR can be mitigated in the case of orthogonal pilot placement, but at the cost of lower spectral efficiency than that of non-orthogonal placement.

Note that (11) is optimal in flat fading, but the MSE performance may degrade in frequency selective fading. Fig. 8 shows that a performance degradation of about 0.4 dB (or about 0.9 dB) is achieved in the multipath-fading channel at a high NPNR when $\sigma_\epsilon^2 = 10^{-3}$ (or when $\sigma_\epsilon^2 = 10^{-2}$), as compared to the MSE performance in the flat-fading channels. For a small σ_ϵ^2, this performance degradation is negligible.

We also compare the performance of the proposed LS estimator with Cui's estimator [2] with multiple frequency offsets, and the numerical results are shown in Fig. 9. Note that in Cui's estimator, the total preamble is used for training only. For a fair comparison, we use 16 pilot subcarriers in the proposed estimator and use a length-16 training sequence in Cui's estimator, with an identical total power E_p being allocated to the pilots/training in each estimator. If the data subcarriers in the proposed symbol are not modulated, the proposed estimator always outperforms Cui's estimator at a high NPNR, and a larger performance gain can be achieved as σ_ϵ^2 increases. For example, at a high NPNR, the performance improvement of the proposed estimator with $E_d = 0$ over Cui's estimator is about 0.7 dB (or 2.6 dB) with $\sigma_\epsilon^2 = 10^{-2}$ (or $\sigma_\epsilon^2 = 10^{-1}$). A performance degradation will be achieved in the proposed estimator if $E_d \neq 0$. In this simulation, we assume that $E_d/E_p = 10^{-1}$, and the MSE performance of the proposed estimator is about 0.5 dB worse than that of Cui's estimator. In other words, Cui's estimator outperforms the proposed estimator by sacrificing the spectral efficiency.

When the covariance matrix is available at both the transmitter and the receiver, an LMMSE estimator can be designed to improve the estimation accuracy. Fig. 10 compares the channel estimation performance of the proposed LS and the proposed LMMSE estimators. Multiple frequency offsets are considered with $\sigma_\varepsilon^2 = 10^{-2}$. The LMMSE estimator with CE pilot modulation can considerably reduce the MSE by applying adaptive pilot power allocation at the transmitter. For example, when $N_t = N_r = 2$, the performance advantage of the proposed LMMSE estimator with adaptive pilot power allocation over that with CE pilot modulated at a low NPNR is about 3.1 dB, and its performance advantage over that of the LS estimator is about 5.8 dB.

VII. Conclusions

The optimal pilot design and placement for channel estimation in MIMO-OFDM with multiple frequency offsets were discussed. The IPI was eliminated in the proposed estimator, and, therefore, a performance advantage over that of conventional estimators was achieved. Given a total pilot power E_p, the pilots for the different transmit antennas was modulated into one or consecutive multiple OFDM symbols with each pilot subcarrier in each symbol being modulated as a CE or not (a CE is required in pilot design only when $M = 1$). With the channel covariance matrix known at both the transmitter and the receiver, an LMMSE estimator with CE pilots modulation was designed based on the proposed optimal pilots. This estimator outperformed the proposed LS estimator in terms of the MSE. By applying an adaptive pilot power allocation at the transmitter, a much lower MSE as compared to that achieved with a CE pilot modulation was achieved in the proposed LMMSE estimator.

Appendix A
Analysis of $\mathbf{G}_{k,m,n\neq m}$

Define $\mathbf{G}_{k,m,n} = N\mathbf{F}_{(L_{max})}\mathbf{\Pi}_{mn}\mathbf{F}_{(L_{max})}^H$ where $\mathbf{\Pi}_{mn} = \sum_{i=0}^{M-1}\mathbf{F}_{(L_{max})}\mathbf{X}_m^{pH}(i)\mathbf{E}_{k,m,n}^{\mathrm{cir}}\mathbf{X}_n^p(i)$, where the $\theta_{m,l}\theta_{n,s}$-th element of $\mathbf{\Pi}_{mn}$ is non-zero, as given by

$$[\mathbf{\Pi}_{mn}]_{\theta_{m,l}\theta_{n,s}} = \sum_{i=0}^{M-1}[\mathbf{X}_m^p(i)]^*_{\theta_{m,l}\theta_{m,l}}\left[\mathbf{E}_{k,m,n}^{\mathrm{cir}}\right]_{\theta_{m,l}\theta_{n,s}}[\mathbf{X}_n^p(i)]_{\theta_{n,s}\theta_{n,s}}, \tag{32}$$

where $1 \leq l, s \leq \mathcal{N}_p$. Let us consider the following two cases:

1) Pilot Subcarriers for the m-th and the n-th Transmit Antennas are Orthogonal in the frequency-domain: In this case, $\theta_{m,l} \neq \theta_{n,s}$ for each l and s, and it is easy to show that $\mathbf{F}_{(L_{max})}\mathbf{\Pi}_{mn}\mathbf{F}_{(L_{max})}^H = \mathbf{O}_{L_{max}}$, so that $\mathbf{G}_{k,m,n\neq m} = \mathbf{O}_{L_{max}}$ for each $m \neq n$.

2) Pilot Subcarriers for the m-th and the n-th Transmit Antennas are Overlapped in the frequency-domain ($\theta_{m,z} = \theta_{n,z} = \theta_z$ for $z = 1, 2, \cdots \mathcal{N}_p$): Define $\mathbf{\Pi}_{mn}$ as

$$[\mathbf{\Pi}_{mn}]_{\theta_l\theta_s} = \sum_{i=0}^{M-1}[\mathbf{X}_m^p(i)]^*_{\theta_l\theta_l}\left[\mathbf{E}_{k,m,n}^{\mathrm{cir}}\right]_{\theta_l\theta_s}[\mathbf{X}_n^p(i)]_{\theta_s\theta_s}, \tag{33}$$

where $1 \leq l, s \leq \mathcal{N}_p$. Also define a $\mathcal{N}_p \times \mathcal{N}_p$ matrix $\underline{\mathbf{\Pi}}_{mn}$, which is generated by deleting all the zero rows and columns of $\mathbf{\Pi}_{mn}$. Evidently, $\underline{\mathbf{\Pi}}_{mn} = \underline{\mathbf{\Pi}}_{mn}^H$. $\mathbf{G}_{k,m,n}$ can be rewritten as

$$\begin{aligned}\mathbf{G}_{k,m,n} &= N\underbrace{\left[\mathbf{f}_{\theta_1}, \cdots, \mathbf{f}_{\theta_{\mathcal{N}_p}}\right]}_{\underline{\mathbf{F}}_{(L_{max})}}\underline{\mathbf{\Pi}}_{mn}\underbrace{\left[\mathbf{f}_{\theta_1}, \cdots, \mathbf{f}_{\theta_{\mathcal{N}_p}}\right]^H}_{\underline{\mathbf{F}}_{(L_{max})}^H} \\ &= N\left[\boldsymbol{\mu}_1, \cdots, \boldsymbol{\mu}_{L_{max}}\right]^T\underline{\mathbf{\Pi}}_{mn}\left[\boldsymbol{\mu}_1^*, \cdots, \boldsymbol{\mu}_{L_{max}}^*\right],\end{aligned} \tag{34}$$

where $\boldsymbol{\mu}_z^T$ is the z-th row of $\underline{\mathbf{F}}_{(L_{max})}$. Therefore, the ut-th element of $\mathbf{G}_{k,m,n}$ is $[\mathbf{G}_{k,m,n}]_{ut} = N\boldsymbol{\mu}_u^T\underline{\mathbf{\Pi}}_{mn}\boldsymbol{\mu}_t^*$, and $\mathbf{G}_{k,m,n} = \mathbf{O}_{L_{max}}$ is achieved only when $\boldsymbol{\mu}_u^T\underline{\mathbf{\Pi}}_{mn}\boldsymbol{\mu}_t^* = 0$, $1 \leq u, t \leq L_{max}$.

Appendix B
Optimal Eigenvalues for Adaptive Allocation

Define a cost function,

$$\begin{aligned}&\mathcal{C}\left(\lambda_{p,0}^2, \cdots, \lambda_{p,L_{max}N_tN_r-1}^2\right) \\ &= \sum_{j=0}^{L_{max}N_tN_r-1}\frac{\sigma_w^2\lambda_{\phi,j}^2\lambda_{p,j}^2}{L_{max}N_tN_r\left(\sigma_w^2\lambda_{p,j}^2 + \lambda_{\phi,j}^2\left(\lambda_{p,j}^2 + \lambda_{v,j}\right)^2\right)} \\ &\quad + \beta\left(\sum_{j=0}^{L_{max}N_tN_r-1}\lambda_{p,j}^2 - N_rL_{max}E_p\right),\end{aligned} \tag{35}$$

where β is a positive real coefficient.

To minimize (35), let us take the partial derivative to $\mathcal{C}\left(\lambda_{p,0}^2, \cdots, \lambda_{p,L_{max}N_tN_r-1}^2\right)$ with respect to each $\lambda_{p,j}^2$ $(0 \leq j \leq L_{max}N_tN_r - 1)$ and set the result to zero. The result is

$$\begin{aligned} & L_{max}N_tN_r\left(\sigma_w^2\lambda_{p,j}^2 + \lambda_{\phi,j}^2\left(\lambda_{p,j}^2 + \lambda_{v,j}\right)^2\right)^2\beta \\ & \quad - \sigma_w^2\lambda_{\phi,j}^4\left(\lambda_{p,j}^4 - \lambda_{v,j}^2\right) = 0, \\ & \quad st. \qquad \sum_{j=0}^{L_{max}N_tN_r-1} \lambda_{p,j}^2 = N_rL_{max}E_p. \end{aligned} \tag{36}$$

REFERENCES

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

[2] T. Cui and C. Tellambura, "Channel and frequency offset estimation and training sequence design for MIMO frequency selective channels," *IEEE Trans. Wireless Commun.*, submitted, to appear.

[3] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance analysis and code construction," *IEEE Trans. Inform. Theory*, vol. 44, pp. 744–765, Mar. 1998.

[4] Y. Li, J. Winters, and N. Sollenberger, "MIMO-OFDM for wireless communications: signal detection with enhanced channel estimation," *IEEE Trans. Commun.*, vol. 50, no. 9, pp. 1471–1477, Sept. 2002.

[5] H. Bolcskei, "MIMO-OFDM wireless systems: basics, perspectives, and challenges," *IEEE Wireless Commun.*, vol. 13, no. 4, pp. 31–37, Aug. 2006. (See also *IEEE Personal Commun.*)

[6] T. Schmidl and D. Cox, "Robust frequency and timing synchronization for OFDM," *IEEE Trans. Commun.*, vol. 45, no. 12, pp. 1613–1621, Dec. 1997.

[7] Z. Zhang, M. Zhao, H. Zhou, Y. Liu, and J. Gao, "Frequency offset estimation with fast acquisition in OFDM system," *IEEE Commun. Lett.*, vol. 8, no. 3, pp. 171–173, Mar. 2004.

[8] Z. Zhang, W. Jiang, H. Zhou, Y. Liu, and J. Gao, "High accuracy frequency offset correction with adjustable acquisition range in OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 4, no. 1, pp. 228–237, Jan. 2005.

[9] A. Mody and G. Stuber, "Synchronization for MIMO OFDM systems," in *Proc. IEEE Global Telecommn. Conf. (GLOBECOM)*, vol. 1, San Antonio, TX, 2001, pp. 509–513.

[10] O. Besson and P. Stoica, "On parameter estimation of MIMO flat-fading channels with frequency offsets," *IEEE Trans. Signal Processing*, vol. 51, no. 3, pp. 602–613, Mar. 2003.

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

[12] M. Dong and L. Tong, "Optimal design and placement of pilot symbols for channel estimation," *IEEE Trans. Signal Processing*, vol. 50, no. 12, pp. 3055–3069, Dec. 2002.

[13] H. Minn and N. Al-Dhahir, "Optimal training signals for MIMO OFDM channel estimation," *IEEE Trans. Wireless Commun.*, vol. 5, no. 5, pp. 1158–1168, May 2006.

[14] D. Hu, L. Yang, Y. Shi, and L. He, "Optimal pilot sequence design for channel estimation in MIMO OFDM systems," *IEEE Commun. Lett.*, vol. 10, no. 1, pp. 1–3, Jan. 2006.

[15] J. Kotecha and A. Sayeed, "Transmit signal design for optimal estimation of correlated MIMO channels," *IEEE Trans. Signal Processing*, vol. 52, no. 2, pp. 546–557, Feb. 2004.

[16] H. Zhang, Y. Li, A. Reid, and J. Terry, "Channel estimation for MIMO OFDM in correlated fading channels," in *Proc. IEEE Int. Conf. Communications (ICC)*, vol. 4, May 2005, pp. 2626–2630.

[17] H. Minn, N. Al-Dhahir, and Y. Li, "Optimal training signals for MIMO OFDM channel estimation in the presence of frequency offset and phase noise," *IEEE Trans. Commun.*, vol. 54, no. 6, pp. 1081–1096, June 2006.

[18] M. Ghogho and A. Swami, "Training design for multipath channel and frequency-offset estimation in MIMO systems," *IEEE Trans. Signal Processing*, vol. 54, no. 10, pp. 3957–3965, Oct. 2006.

[19] H. L. V. Trees, *Detection, Estimation, and Modulation Theory, Part IV, Optimum Array Processing*. New York: John Wiley & Sons, 2002.

[20] Z. Zhang and H. Kayama, "Robust uplink carrier frequency offset estimation with interference mitigation in OFDMA systems," in *Proc. IEEE Wireless Commun. and Networking Conf.*, vol. 2, pp. 816–821, Apr. 2006.

[21] X. Fu, H. Minn, and C. Cantrell, "Two novel iterative joint frequency-offset and channel estimation methods for OFDMA uplink," in *Proc. IEEE Global Telecommn. Conf. (GLOBECOM)*, San Francisco, CA, 2006.

[22] Z. Zhang, W. Zhang, and C. Tellambura, "Optimal pilots for OFDMA uplink frequency offsets and channel estimation," *IEEE Trans. Veh. Technol.*, to be submitted.

[23] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice Hall, 1993.

Zhongshan Zhang received the M.S. degree in computer science in 2001 and Ph.D. degree in electrical engineering in 2004 from the Beijing University of Posts and Telecommunications (BUPT). From Aug. 2004 he joined DoCoMo Beijing Laboratories as an associate researcher, and was promoted to be a researcher in Dec. 2005. From Feb. 2006, he joined University of Alberta, Edmonton, AB, Canada, as a postdoctoral fellow. His main research interests include statistical signal processing, synchronization and channel estimation in Multiple-input Multiple-output (MIMO) orthogonal frequency division multiplexing (OFDM) systems, and cooperative communications.

Wei Zhang received her B.E. and M.E. degrees in control engineering from Shandong University, Jinan, P.R. China, in 2000 and 2004., respectively. In Sept. 2005, she joined the Department of Electrical and Computer Engineering at University of Alberta where she is currently pursuing her Ph.D degree. Her current research interests include wireless communications theory, signal detection and estimation, diversity and cooperative communications.

Chintha Tellambura (SM'02) received the B.Sc. degree (with first-class honors) from the University of Moratuwa, Moratuwa, Sri Lanka, in 1986, the M.Sc. degree in electronics from the University of London, London, U.K., in 1988, and the Ph.D. degree in electrical engineering from the University of Victoria, Victoria, BC, Canada, in 1993.

He was a Postdoctoral Research Fellow with the University of Victoria (1993-1994) and the University of Bradford (1995-1996). He was with Monash University, Melbourne, Australia, from 1997 to 2002. Presently, he is a Professor with the Department of Electrical and Computer Engineering, University of Alberta. His research interests include Diversity and Fading Countermeasures, Multiple-Input Multiple-Output (MIMO) Systems and Space-Time Coding, and Orthogonal Frequency Division Multiplexing (OFDM).

Prof. Tellambura is an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS and the Area Editor for Wireless Communications Systems and Theory in the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He was Chair of the Communication Theory Symposium in Globecom'05 held in St. Louis, MO.