Blind Channel Estimation for OFDM Systems via a Generalized Precoding

Feifei Gao, *Student Member, IEEE*, and Arumugam Nallanathan, *Senior Member, IEEE*

***Abstract*—In this paper, we consider the problem of blind channel estimation for single-input–single-output (SISO) orthogonal frequency-division multiplexing (OFDM) system via second-order statistics only. Based on the assumption that the transmitted symbols are independent and identically distributed, we develop a simple blind channel estimation technique for OFDM systems by utilizing a generalized linear nonredundant block precoding. Instead of using partial information from the signal covariance matrix, as done in previous works where a specific precoder is designed and only one column of the signal covariance matrix is exploited, our work jointly considers all the information contained in the signal covariance matrix. Compared to the popular subspace-based blind channel estimation methods, the proposed algorithm is much more computationally efficient. A design criterion of the precoders by which the performance can be improved is provided, and the closed-form stochastic Cramér–Rao bound is derived. The numerical results clearly show the effectiveness of our proposed algorithm, as well as its improvement over the existing techniques.**

***Index Terms*—Blind channel estimation, block precoding, block transmissions, high-performance local area network (HIPERLAN)/2, IEEE 802.11a, orthogonal frequency-division multiplexing (OFDM), stochastic Cramér–Rao bound (CRB), wireless communications.**

I. Introduction

ORTHOGONAL frequency-division multiplexing (OFDM) [1], [2] is a promising candidate for next-generation high-speed wireless multimedia communication systems due to its high data rate, high spectral efficiency, and robustness to frequency-selective channels. It has been used in European digital audio/video broadcasting [3], [4], high-performance local area network (HIPERLAN) [5], and 802.11a wireless LAN standards [6].

Thanks to the inverse fast Fourier transform (IFFT) and the insertion of a cyclic prefix (CP) with length greater than the channel spread, OFDM is able to convert a frequency-selective channel to multiple flat-fading subchannels, thus providing a simple equalization at the receivers. However, coherent detection in OFDM systems requires reliable channel state information (CSI). Consequently, channel estimation becomes a critical component for most OFDM systems. Several training-based channel estimations are developed in [7] and [8]. However, usage of training sequences decreases the system bandwidth efficiency [9]. Additionally, due to the time-varying nature of the channel in some wireless applications, the training sequence needs to be transmitted periodically, causing further loss of channel throughput. Due to these reasons, reducing the number of training symbols becomes a major concern, and the blind channel estimation algorithms have received considerable attention during the past decade.

A promising family of blind channel estimation work, the so-called subspace-based algorithm is developed in [10]. For OFDM transmissions [11], [12], however, the subspace-based method cannot be directly applied since no freedom of the noise subspace is available. In order to solve this problem, a redundant block precoding technique is used in [13], and virtual carriers are exploited in [14].

A new type of blind channel estimation for OFDM systems has been recently proposed in [15] and [16]. In this method, a nonredundant linear precoder is applied at the transmitter, and the CSI is contained in each entry of the signal covariance matrix. However, the authors focus on a special design of linear precoders and purely extract the CSI from only one column of the covariance matrix, which greatly limits the performance accuracy of the algorithm. Moreover, the authors did not clearly present the way how all the channel estimates from each single-column estimation can be averaged to obtain a better result. Another similar work has been proposed in [17] and [18], where channel is, again, estimated from a single column of the cross correlation matrix of the two consecutive received blocks. Then, a direct average is applied. In fact, a direct average may provide bad performance, as will be explained later. Moreover, the method in [17] and [18] reduces the effective number of OFDM blocks by half and is, thus, less suitable for the case where only a few OFDM blocks are available.

In this paper, we propose a blind channel estimation method that can overcome the aforementioned shortcomings. We focus on the scenario in [15] and [16], where the covariance matrix is constructed. Nevertheless, the proposed algorithm can be readily applied to the transmission scenarios of [17] and [18], where the cross correlation matrix is considered. The proposed method utilizes a general precoder and jointly obtain the channel estimation from all the entries of the signal covariance matrix. This joint estimation algorithm does not increase the computational complexity compared to the existing method but yields better performance. A precoder design criterion is also provided by properly considering both the channel estimation error and data detection error. The proposed method, as well as the proposed design on precoders, is shown to be better than those in [15] and [16] through numerical simulations.

Manuscript received April 21, 2005; revised March 24, 2006, June 27, 2006, and June 28, 2006. The review of this paper was coordinated by Prof. Z. Wang.

The authors are with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore 119260 (e-mail: feifeigao@nus.edu.sg; elena@nus.edu.sg).

Digital Object Identifier 10.1109/TVT.2007.895562

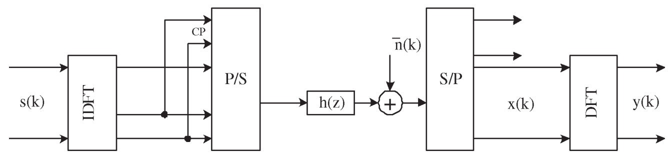


Fig. 1. Typical baseband OFDM system.

This paper is organized as follows. Section II presents the system model of OFDM transmissions. Section III provides our proposed algorithm, together with the design criteria for the linear precoders. Section IV derives the stochastic Cramér–Rao bound (CRB) of the proposed estimator. In Section V, we provide various simulations to show the effectiveness of our algorithm. Finally, conclusions are made in Section VI, and the related proof is given in the Appendix.

The following notations are used in this paper: The transpose, complex conjugate, Hermitian, inverse, and pseudoinverse of the matrix $\mathbf{A}$ are denoted by $\mathbf{A}^T$, $\mathbf{A}^*$, $\mathbf{A}^H$, $\mathbf{A}^{-1}$, and $\mathbf{A}^\dagger$, respectively; $\mathbf{a}_R$ and $\mathbf{a}_I$ represent the real and imaginary parts of a vector; $\mathrm{tr}(\mathbf{A})$ stands for the trace operation; $\mathrm{diag}\{\mathbf{a}\}$ denotes the diagonal matrix with the diagonal element constructed from $\mathbf{a}$; $\otimes$ and $\odot$ stand for the Kronecker and Hadamard products, respectively; $\mathrm{vec}(\mathbf{A})$ represents the vectorization of $\mathbf{A}$; and $\mathrm{E}\{\cdot\}$ denotes the statistical expectation. MATLAB notations are used. For example, $\mathbf{A}(m,q)$ and $\mathbf{A}(:,q)$ represent the (m, q)th entry and the qth column of matrix $\mathbf{A}$, respectively.

II. Problem Formulation

Fig. 1 shows a baseband single-input–single-output (SISO) OFDM system that we consider. In each OFDM block, M symbols are transmitted. The discrete channel response is supposed to have a memory upper bounded by L and is represented by the $(L+1)\times 1$ vector $\mathbf{h} = [h_0, \ldots, h_L]^T$. The CP is added at the front of each transmitted block and is discarded at each received block. As long as the length of the CP is greater than or equal to L, the remaining signal at the receiver for the kth block can be represented as [11]

$$\mathbf{x}(k) = \bar{\mathbf{H}}\mathbf{F}^H\mathbf{s}(k) + \bar{\mathbf{n}}(k) \tag{1}$$

where $\mathbf{s}(k)$ is the $M \times 1$ transmitted symbol vector in the frequency domain, $\bar{\mathbf{n}}(k)$ is the $M\times 1$ vector of the unknown white Gaussian noise with equivalent variance σ_n^2 on each sampling time. The $M \times M$ normalized discrete Fourier transform (DFT) matrix $\mathbf{F}$ has the form

$$\mathbf{F} = \frac{1}{\sqrt{M}}\begin{bmatrix} 1 & 1 & \ldots & 1 \\ 1 & f_M^1 & \ldots & f_M^{M-1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & f_M^{M-1} & \ldots & f_M^{(M-1)(M-1)} \end{bmatrix} \tag{2}$$

where $f_M = \exp(-j2\pi/M)$, and $\bar{\mathbf{H}}$ is the circulant channel matrix with its (m,q)th entry given by $\bar{\mathbf{h}}((m-q)\,\mathrm{mod}\,M)$. The $M\times 1$ vector $\bar{\mathbf{h}}$ is obtained by inserting $M-L-1$ zeros at the end of $\mathbf{h}$.

Let

$$\begin{aligned} \mathbf{H} &= \mathrm{DFT}(\mathbf{h}) = \sqrt{M}\mathbf{F}(:, 1:L+1)\mathbf{h} \\ &= [H_0, H_1, \ldots, H_{M-1}]^T \end{aligned} \tag{3}$$

denote the M-point DFT of the channel vector $\mathbf{h}$. The normalized DFT of the received signal vector is then represented as

$$\begin{aligned} \mathbf{y}(k) &= \mathbf{F}\mathbf{x}(k) = \mathbf{F}\bar{\mathbf{H}}\mathbf{F}^H\mathbf{s}(k) + \mathbf{F}\bar{\mathbf{n}}(k) \\ &= \tilde{\mathbf{H}}\mathbf{s}(k) + \mathbf{n}(k) \end{aligned} \tag{4}$$

where $\tilde{\mathbf{H}}$ is a diagonal matrix with the form

$$\tilde{\mathbf{H}} = \mathrm{diag}\{\mathbf{H}\}. \tag{5}$$

The new random noise vector $\mathbf{n}(k)$ has the same statistical distribution as $\bar{\mathbf{n}}(k)$. Therefore, the frequency-selective channel is converted into M parallel flat-fading subchannels, which greatly reduces the complexity and enhances the speed of the equalization at the receiver.

The signal covariance matrix $\mathbf{R}$ is defined as

$$\begin{aligned} \mathbf{R} &= \mathrm{E}\left\{\mathbf{y}(k)\mathbf{y}(k)^H\right\} = \tilde{\mathbf{H}}\mathbf{R}_\mathbf{s}\tilde{\mathbf{H}}^H + \sigma_n^2\mathbf{I} \\ &= \mathrm{diag}\left\{\sigma_s^2|H_0|^2 + \sigma_n^2, \sigma_s^2|H_1|^2\right. \\ &\qquad \left. +\ \sigma_n^2, \ldots, \sigma_s^2|H_{M-1}|^2 + \sigma_n^2\right\} \end{aligned} \tag{6}$$

where the source covariance matrix

$$\mathbf{R}_\mathbf{s} = \mathrm{E}\left\{\mathbf{s}(k)\mathbf{s}(k)^H\right\} = \sigma_s^2\mathbf{I} \tag{7}$$

is a scaling of the identity matrix, and σ_s^2 is the signal power on each subcarrier. It is known from (6) that the phase information of $\mathbf{H}$ is lost. Hence, it is impossible to recover $\mathbf{H}$ purely relying on $\mathbf{R}$.

III. Generalized Precoding Technique

A. Algorithm

Suppose the symbol vector $\mathbf{s}(k)$ is precoded by a predefined matrix $\mathbf{W}$ before the IFFT operation. The received signal vector can be rewritten as

$$\mathbf{y} = \tilde{\mathbf{H}}\mathbf{W}\mathbf{s}(k) + \mathbf{n}(k). \tag{8}$$

Note that the precoding matrix $\mathbf{W}$ should be designed in a way that the signal power in each transmitted block does not change. As a result, the signal covariance matrix can be obtained as

$$\begin{aligned} \mathbf{R} &= \sigma_s^2\tilde{\mathbf{H}}\mathbf{P}\tilde{\mathbf{H}}^H + \sigma_n^2\mathbf{I} \\ &= \sigma_s^2(\mathbf{H}\mathbf{H}^H)\odot\mathbf{P} + \sigma_n^2\mathbf{I} \end{aligned} \tag{9}$$

where

$$\mathbf{P} = \mathbf{W}\mathbf{W}^H \tag{10}$$

is the square of the precoding matrix. On the qth column of $\mathbf{R}$, except the diagonal element $\mathbf{R}(q,q)$, all the other $M-1$ elements are of the form

$$\mathbf{R}(m,q) = \sigma_s^2 \mathbf{P}(m,q) H_{m-1} H_{q-1}^*, \quad m = 1,\ldots,q-1,q+1,\ldots,M. \tag{11}$$

Since $\mathbf{P}$ is known as *a priori*, $\sigma_s^2 H_{m-1} H_{q-1}^*$ can be obtained from

$$\sigma_s^2 H_{m-1} H_{q-1}^* = \frac{\mathbf{R}(m,q)}{\mathbf{P}(m,q)}. \tag{12}$$

However, $\sigma_s^2 H_{q-1} H_{q-1}^*$ cannot be obtained since the (q,q)th entry of $\mathbf{R}$ is corrupted by the unknown noise variance. As long as $M-1 \geq L+1$, which is the usual case, both the $(L+1)\times 1$ time-domain channel vector $\mathbf{h}$ and the $M \times 1$ frequency-domain channel vector $\mathbf{H}$ can be estimated as in (13)–(15), shown at the bottom of the page.

Define new variables as in (13)–(15), where entries of $\tilde{\mathbf{R}}$ can be obtained from (11) and (12). We can derive M estimates of $\mathbf{h}$ as

$$\mathbf{h}_{eq} = \frac{1}{\sqrt{M}} \mathbf{F}_q^\dagger \mathbf{r}_q = \sigma_s^2 H_{q-1}^* \mathbf{h}, \qquad q = 1,\ldots,M, \tag{16}$$

where the vector $\mathbf{h}_{eq}$ is truncated to have a correct length $L+1$. The frequency response of channel vector $\mathbf{H}$ is then obtained from

$$\begin{aligned}\mathbf{H}_{eq} &= \mathrm{DFT}(\mathbf{h}_{eq}) = \sqrt{M}\mathbf{F}(:,1:L+1)\mathbf{h}_{eq} \\ &= \mathbf{F}(:,1:L+1)\mathbf{F}_q^\dagger \mathbf{r}_q = \sigma_s^2 H_{q-1}^* \mathbf{H}.\end{aligned} \tag{17}$$

In other words, the channel length is enforced in the time domain to obtain a truncated channel vector, and then, the M-point DFT is performed to obtain a channel frequency response. This process is known as denoising and can improve the estimation accuracy [19]. The estimated channel vectors from both (16) and (17) are different from the true one by a complex scalar, which is well known for almost all blind channel estimation techniques. Since the qth entry of $\mathbf{H}_{eq}$ is $\sigma_s^2 |\mathbf{H}_{q-1}|^2$, the scaling ambiguity of the estimation can be eliminated if the signal power is known.

Note that our proposed precoding technique can be seen as a generalization of the work in [15] and [16]. For example, if we take $\mathbf{W}$ with the (m,q)th entry as

$$\mathbf{W}(m,q) = \begin{cases} 1, & m = q \neq t \\ 2, & m = q = t \\ 1, & q = t \\ 0, & \text{otherwise} \end{cases} \tag{18}$$

where t is some predefined value, which is usually taken as $M/4$, and obtain the channel response vector purely through the tth column of $\tilde{\mathbf{R}}$, then the proposed method reduces exactly to the one developed in [15].

B. Improved Algorithm via Joint Estimation

From the previous discussion, we know that M estimates of $\mathbf{H}$, denoted as $\mathbf{H}_{e1},\ldots,\mathbf{H}_{eM}$, can be obtained separately from each column of $\tilde{\mathbf{R}}$ within a scalar ambiguity. Hence, an averaging step should be carried on to yield a better result. Since each estimates of $\mathbf{H}$ is different from the true one by an unknown factor $\sigma_s^2 H_{q-1}^*$, it is not possible to apply maximum ratio combining (MRC). Moreover, the direct averaging considered in [16] and [18] may not be a good choice, since some elements of the estimated channel vector may cancel with each other, e.g., when $H_{(q_1-1)} + H_{(q_2-1)} = 0$ for some q_1, $q_2 \in \{1,\ldots,M\}$. We now propose a new estimation algorithm that can jointly utilize all entries[1] of the matrix $\tilde{\mathbf{R}}$ and does not increase the computational complexity.

Joint Estimation Algorithm:

Step 1) From $\tilde{\mathbf{R}}(:,1)$ and (17), obtain a rough estimate of $\sigma_s^2|H_0|^2$ and then $\sigma_s|H_0|$.

Step 2) Use an arbitrary phase α and form an estimate of H_0 as $\hat{H}_0 = \sigma_s|H_0|e^{j\alpha}$.

Step 3) From $\tilde{\mathbf{R}}(2,1)$, obtain an estimate for H_1 as $\hat{H}_1 = \tilde{\mathbf{R}}(2,1)/\hat{H}_0^*$.

Step 4) From $\tilde{\mathbf{R}}(3,1:2)$, solve the overdetermined function using $\hat{H}_0$ and $\hat{H}_1$ and get an estimated value of H_2 as

$$\begin{aligned}\hat{H}_2 &= \begin{bmatrix}\hat{H}_0^* \\ \hat{H}_1^*\end{bmatrix}^\dagger \tilde{\mathbf{R}}(3,1:2)^T \\ &= \left(|\hat{H}_0|^2 + |\hat{H}_1|^2\right)^{-1} [\hat{H}_0 \hat{H}_1] \tilde{\mathbf{R}}(3,1:2)^T.\end{aligned} \tag{19}$$

[1]Except the diagonal entries.

$$\tilde{\mathbf{R}} = \begin{bmatrix} \sigma_s^2 P(1,1)|H_0|^2 + \sigma_n^2 & \sigma_s^2 H_0 H_1^* & \cdots & \sigma_s^2 H_0 H_{M-1}^* \\ \sigma_s^2 H_1 H_0^* & \sigma_s^2 P(2,2)|H_1|^2 + \sigma_n^2 & \ddots & \sigma_s^2 H_1 H_{M-1}^* \\ \vdots & \ddots & \ddots & \vdots \\ \sigma_s^2 H_{M-1} H_0^* & \sigma_s^2 H_{M-1} H_1^* & \cdots & \sigma_s^2 P(M,M)|H_{M-1}|^2 + \sigma_n^2 \end{bmatrix} \tag{13}$$

$$\mathbf{r}_q = [\tilde{\mathbf{R}}(1:q-1,q)^T,\ \tilde{\mathbf{R}}(q+1:M,q)^T]^T, \qquad q = 1,\ldots,M \tag{14}$$

$$\mathbf{F}_q = [\mathbf{F}(1:q-1,1:L+1)^T,\ \mathbf{F}(q+1:M,1:L+1)^T]^T \tag{15}$$

Step 5) Loop back to Step 4) and calculate $\hat{H}_{q-1}$ from $\tilde{\mathbf{R}}(q, 1:q-1)$ until all the elements of $\mathbf{H}$ are estimated.

Step 6) Denoising the whole estimate vector $\hat{\mathbf{H}} = [\hat{H}_0, \ldots, \hat{H}_{M-1}]$ and obtain the final estimation result as

$$\hat{\hat{\mathbf{H}}} = \mathbf{F}(:, 1:L+1)\mathbf{F}(:, 1:L+1)^H \hat{\mathbf{H}}. \tag{20}$$

Throughout the whole algorithm, we did not use the true value of H_0. Instead, the estimate of H_0 with an unknown scaling factor σ_s and an arbitrary phase α is considered. All the other parameters $H_{q-1}, q > 1$, are calculated subsequently. Hence, it is necessary to prove the validity of the proposed algorithm.

Proof: From $\sigma_s^2 H_0^* \mathbf{H}$, $\sigma_s|H_0|$ can be obtained by the previously stated denoising method. Let $\hat{H}_0 = \sigma_s|H_0|e^{j\alpha} = \sigma_s H_0 e^{j\beta}$ be the estimate of H_0. From the proposed algorithm, H_1 can be estimated by

$$\hat{H}_1 = \mathbf{R}(2,1)/\hat{H}_0^* = \sigma_s H_1 e^{j\beta}. \tag{21}$$

Suppose all the estimates $\hat{H}_{q-1}, (q-1) \le D$, have the form

$$\hat{H}_{q-1} = \sigma_s H_{q-1} e^{j\beta} \tag{22}$$

where D is some value greater than or equal to 1. Then, the estimate of $\hat{H}_{D+1}$ can be obtained according to

$$\begin{aligned}\hat{H}_{D+1} &= \begin{bmatrix} \hat{H}_0^* \\ \vdots \\ \hat{H}_D^* \end{bmatrix}^{\dagger} \tilde{\mathbf{R}}(D+2, 1:D+1)^T \\ &= \left(\sum_{q=1}^{D+1} |\hat{H}_{q-1}|^2\right)^{-1} [\hat{H}_0, \ldots, \hat{H}_D] \\ &\quad \times \tilde{\mathbf{R}}(D+2, 1:D+1)^T \\ &= \left(\sigma_s^{-2}\left(\sum_{q=1}^{D+1} |H_{q-1}|^2\right)^{-1}\right)\left(\sigma_s e^{j\beta}[H_0, \ldots, H_D]\right) \\ &\quad \times \begin{bmatrix} \sigma_s^2 H_0^* \\ \vdots \\ \sigma_s^2 H_D^* \end{bmatrix} H_{D+1} \\ &= \sigma_s H_{D+1} e^{j\beta}. \end{aligned} \tag{23}$$

Therefore, all $\hat{H}_{q-1}, 1 \le q \le M$, have the form $\sigma_s H_{q-1} e^{j\beta}$, and the final $\hat{\mathbf{H}}$ is the estimation of $\mathbf{H}$ within only a scalar ambiguity $\sigma_s e^{j\beta}$. ■

Remark 1: Since, except H_0 and H_1, all the other H_q, $q = 2, \ldots, M-1$ are obtained by jointly considering several entries of matrix $\tilde{\mathbf{R}}$, the final estimate of $\mathbf{H}$ is better than the one purely estimated from a single column.

Remark 2: Let us take a look at the computational complexity of our proposed algorithm. The complexity of Step 1) can be approximated as $\mathrm{O}(M^2 L) + O(M \log M)$ due to the multiplication and the FFT operations in (17). The complexity through Steps 2)–5) can be approximated as $\mathrm{O}(M^2)$. The final step has the complexity $2\mathrm{O}(M \log M)$. Therefore, the total complexity of the proposed algorithm can be approximated as $\mathrm{O}(M^2 L)$. It can be easily shown that the complexity of the algorithm in [15] and [16] is also $\mathrm{O}(M^2 L)$ due to the denoising process. Then, the proposed algorithm requires similar complexity compared to [15] and [16]. Moreover, the popular subspace-based blind channel estimation algorithm [10] is well known to have the computational complexity around $\mathrm{O}(M^3)$ due to the eigenvalue decomposition (EVD) on the $M \times M$ covariance matrix $\mathbf{R}$. Since $L \ll M$ in most applications, the proposed method is much more computationally efficient than the subspace-based method.[2]

Remark 3: We can also start the algorithm from an arbitrary H_{q-1}, $q > 1$ at the very beginning. In this case, $\hat{H}_{q-1}$ should be obtained from the qth column in Step 1). However, different starting points may not cause a very big difference for the final estimation, because no matter where it is started, the joint estimation procedure always takes the full information into consideration. This declaration will be numerically proven in the simulation part.

C. *Criteria for the Design of Precoders*

Since both $\mathbf{W}$ and $\mathbf{P}$ are predefined parameters, we will only focus on the design of the matrix $\mathbf{P}$, while the matrix $\mathbf{W}$ can be readily obtained from $\mathbf{P}^{1/2}$.

1) Distortion Constraint: In reality, the number of snapshots received within the channel coherent time is not infinite. Therefore, the received signal covariance matrix is replaced by the sample covariance matrix. As a result, each entry of $\mathbf{R}$ contains the distortion due to the effect of both the noise and the lack of the number of snapshots. If one entry of $\mathbf{P}$ is much smaller than the other entries, the distortion in the corresponding entry of $\mathbf{R}$ will be greatly enlarged after the elimination of the effect of $\mathbf{P}$. Hence, the entry of $\mathbf{R}$ with possibly large distortion should be discarded, or the corresponding entry of $\mathbf{P}$ should be assigned a relatively large value. However, no prior information of the distortion can be obtained due to all unknown factors. A reasonable way is to assign equal value to all the nondiagonal entries and at the same time as large as possible.

2) Power Constraint: To keep the transmission power unchanged, matrix $\mathbf{P}$ should be designed such that

$$\text{Power} = \mathrm{E}\left\{\mathbf{s}(k)^H \mathbf{W}^H \mathbf{W} \mathbf{s}(k)\right\} = \sigma_s^2 \mathrm{tr}(\mathbf{P}) = M\sigma_s^2 \tag{24}$$

or equivalently

$$\mathrm{tr}(\mathbf{P}) = M. \tag{25}$$

Since the joint estimation algorithm equally considers all the entries of $\mathbf{R}$, there is no reason why the value of one diagonal

[2]Although the traditional subspace method cannot be applied directly for OFDM systems, the complexity to eigendecompose an $M \times M$ covariance matrix can be taken as a reference.

entry should dominate another. Consequently, we consider each diagonal element to be 1. Combined with the distortion constraint, the (m,q)th entry of $\mathbf{P}$ takes on the form

$$\mathbf{P}(m,q)=\begin{cases}1, & m=q\\ p\neq 0, & m\neq q\end{cases}\qquad m,\ q=1,\ldots,M \tag{26}$$

where p is some nonzero value.

3) Symbol Error Constraint: Due to its quadrature form, $\mathbf{P}$ should be a positive semidefinite matrix. It can be proven that one eigenvalue of $\mathbf{P}$ in (26) is $(M-1)p+1$, and all the other eigenvalues are $1-p$. Hence, for channel estimation, the range of p is $-(1/M-1)\leq p\leq 1$ and $p\neq 0$. However, $\mathbf{P}$ should be designed as a positive definite matrix such that the transmitted symbols can be restored after the channel is estimated. Hence, the range of p for data detection is $-(1/M-1)<p<1$ and $p\neq 0$. Suppose at the receiver the minimum mean square error (MMSE) equalizer is applied for detecting the symbols [17], [18], i.e.,

$$\hat{\mathbf{s}}(k)=\mathbf{G}\mathbf{y}(k) \tag{27}$$

where

$$\begin{aligned}\mathbf{G}&=\mathrm{E}\left\{\mathbf{s}(k)\mathbf{y}(k)^H\right\}\mathrm{E}\left\{\mathbf{y}(k)\mathbf{y}(k)^H\right\}^{-1}\\ &=\mathbf{W}^H\tilde{\mathbf{H}}^H\left(\tilde{\mathbf{H}}\mathbf{P}\tilde{\mathbf{H}}^H+\frac{\sigma_n^2}{\sigma_s^2}\mathbf{I}\right)^{-1}.\end{aligned} \tag{28}$$

The MMSE of the estimation can be obtained as

$$\begin{aligned}\mathrm{MMSE}&=\mathrm{E}\left\{\|\mathbf{G}\mathbf{y}(k)-\mathbf{s}(k)\|^2\right\}\\ &=\sigma_s^2\mathrm{tr}\left(\mathbf{I}-\left(\tilde{\mathbf{H}}\mathbf{P}\tilde{\mathbf{H}}^H+\frac{\sigma_n^2}{\sigma_s^2}\mathbf{I}\right)^{-1}\tilde{\mathbf{H}}\mathbf{P}\tilde{\mathbf{H}}^H\right)\\ &=\sigma_n^2\mathrm{tr}\left(\left(\tilde{\mathbf{H}}\mathbf{P}\tilde{\mathbf{H}}^H+\frac{\sigma_n^2}{\sigma_s^2}\mathbf{I}\right)^{-1}\right).\end{aligned} \tag{29}$$

The optimal value of $\mathbf{P}$ to minimize (29) can be proven as $\lambda(\tilde{\mathbf{H}}^H\tilde{\mathbf{H}})^{-1}$, where the scalar λ is designed to meet the power constraint. However, since channel is unknown, $\mathbf{P}$ should not be designed according to the value of $\tilde{\mathbf{H}}$. Instead, the following lemma is used.

Lemma 1: For any positive definite $M\times M$ matrix $\mathbf{A}$, the following inequality holds [20]:

$$\mathrm{tr}(\mathbf{A}^{-1})\geq\sum_{m=1}^{M}(\mathbf{A}(m,m))^{-1} \tag{30}$$

and the equality holds if $\mathbf{A}$ is diagonal. Since $\tilde{\mathbf{H}}$ is a diagonal matrix, the minimum value of MMSE can only be achieved when $\mathbf{P}$ is also a diagonal matrix. With the fact that $\tilde{\mathbf{H}}$ is unknown, a reasonable choice of $\mathbf{P}$ is to be an identity matrix. Note that a similar result is also derived in [17] based on the point of view of maximizing the received signal-to-noise ratio (SNR). In this case, p is expected to be 0, which makes the channel estimation impossible. To reduce the symbol detection error, one may try to decrease the value of $|p|$ as small as possible. However, this approach will enlarge the distortion effect as stated in the distortion constraint part.

From the above discussion, it is seen that there exists a compromise between channel estimation error and data detection error (see also [17]). Therefore, different considerations should be made under different situations.

IV. Stochastic CRB

In this section, we derive the CRB for the proposed channel estimation algorithm. Since the source covariance is a multiple of an identity matrix, and the channel matrix $\bar{\mathbf{H}}$ is squared, the deterministic CRB is no longer applicable. Instead, we consider the stochastic CRB [21], [22].

Here, we do not use the covariance of $\mathbf{y}(k)$, since only $L+1$ elements of vector $\mathbf{H}$ are linearly independent. Instead, the signal covariance matrix of $\mathbf{x}(k)$ is considered and is parameterized by the $(2L+4)\times 1$ vector $\boldsymbol{\theta}=[\mathbf{h}_R^T,\mathbf{h}_I^T,\sigma_s^2,\sigma_n^2]^T$. The covariance matrix $\mathbf{R}_x$ is given by

$$\mathbf{R}_x=\mathrm{E}\left\{\mathbf{x}(k)\mathbf{x}(k)^H\right\}=\sigma_s^2\bar{\mathbf{H}}\mathbf{F}^H\mathbf{P}\mathbf{F}\bar{\mathbf{H}}^H+\sigma_n^2\mathbf{I}. \tag{31}$$

Define $M\times M$ matrices $\boldsymbol{\mathcal{H}}_l$ with the (m,q)th entry

$$\boldsymbol{\mathcal{H}}_l(m,q)=\begin{cases}1, & \text{for } ((m-q)\mod M)=l-1\\ 0, & \text{otherwise}\end{cases},\qquad l=1,\ldots,L+1 \tag{32}$$

and let

$$\mathbf{A}_l=\sigma_s^2\mathbf{R}_x^{-1/2}\boldsymbol{\mathcal{H}}_l\mathbf{F}^H\mathbf{P}\mathbf{F}\mathbf{H}^H\mathbf{R}_x^{-1/2} \tag{33}$$

$$\mathbf{G}_1(:,l)=\mathrm{vec}\left(\mathbf{A}_l+\mathbf{A}_l^H\right) \tag{34}$$

$$\mathbf{G}_2(:,l)=\mathrm{vec}\left(j\mathbf{A}_l-j\mathbf{A}_l^H\right) \tag{35}$$

$$\mathbf{v}=\mathrm{vec}\left(\mathbf{R}_x^{-1/2}\mathbf{H}\mathbf{F}^H\mathbf{P}\mathbf{F}\mathbf{H}^H\mathbf{R}_x^{-1/2}\right) \tag{36}$$

$$\mathbf{u}=\mathrm{vec}\left(\mathbf{R}_x^{-1}\right) \tag{37}$$

$$\mathbf{G}=[\mathbf{G}_1\ \mathbf{G}_2] \tag{38}$$

$$\boldsymbol{\Delta}=[\mathbf{v}\ \mathbf{u}]. \tag{39}$$

Suppose N data blocks are received, the stochastic CRB of channel vector $[\mathbf{h}_R^T\ \mathbf{h}_I^T]^T$ is given by

$$\mathrm{CRB}=\frac{1}{N}\left(\mathbf{G}^H\mathbf{P}_{\boldsymbol{\Delta}}^{\perp}\mathbf{G}\right)^{\dagger} \tag{40}$$

where

$$\mathbf{P}_{\boldsymbol{\Delta}}^{\perp}=\mathbf{I}-\boldsymbol{\Delta}(\boldsymbol{\Delta}^H\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}^H \tag{41}$$

is the projection matrix onto the orthogonal complement space of $\boldsymbol{\Delta}$.

Proof of (40): See the Appendix.

V. Simulation Results

In this section, we examine the performance of the proposed estimator under various scenarios. The channel model with the exponential power delay profile [23]

$$E\left\{|h_l|^2\right\} = \exp(-l/10), \qquad l = 0, \ldots, L \tag{42}$$

is used. The phase of each channel ray is uniformly distributed over $[0, 2\pi)$. The parameters are set according to the IEEE 802.11a standards [6], which calls for a block length of $M = 64$. A three-tap ($L = 2$) channel is selected following [15] and [16]. Unless otherwise mentioned, the signal constellation is quaternary phase-shift keying (QPSK), and the number of OFDM blocks is $N = 150$. All the results are averaged over $N_i = 100$ Monte Carlo runs. For simulation purposes, we do not use pilot symbols to remove the ambiguity. Instead, the scalar ambiguity is resolved by [24]

$$\alpha = \min_{\alpha} \|\mathbf{H} - \alpha\hat{\mathbf{H}}\|^2 = \min_{\alpha} \|\mathbf{h} - \alpha\hat{\mathbf{h}}\|^2. \tag{43}$$

Then, the normalized estimation mean square error (NMSE) is defined as

$$\begin{aligned}\text{NMSE} &= \frac{1}{N_i}\sum_{i=1}^{N_i}\frac{\|\alpha_i\hat{\mathbf{H}}_i - \mathbf{H}\|^2}{\|\mathbf{H}\|^2} \\ &= \frac{1}{N_i}\sum_{I=1}^{N_i}\frac{\|\alpha_i\hat{\mathbf{h}}_i - \mathbf{h}\|^2}{\|\mathbf{h}\|^2}\end{aligned} \tag{44}$$

where the subscript i refers to the ith simulation run.

A. Example 1

In the first example, we compare the performance of the proposed joint estimator with that of the single-column estimator under both the proposed precoder and the specific precoder in [15]. The predefined single column is taken as $M/4 = 16$, as assumed in [15]. For a relatively fair comparison, the value of p in the proposed precoder is obtained in a way that the MMSE (29) of both the precoders are the same, from which p is obtained as 0.54. In Fig. 2, the NMSEs of channel estimations versus SNR of these two different algorithms are displayed. The training-based channel estimation NMSE and the CRB (40) are also shown. The training is carried on following the method in [16], where two OFDM blocks are used as training sequences [6]. It is seen that both blind methods outperform the training-based method at a relatively lower SNR region, e.g., SNR < 18 dB. This phenomenon is consistent with that in [16]. Then, we see that the joint estimator is slightly better than the single-column estimator at the low SNR region but is comparable at the high SNR region when both apply the precoders in [15]. This is due to the fact that the precoder in [15] is specifically designed for the single-column method so the performance difference at high SNR is not apparent. Nevertheless, it is clearly seen that the proposed joint estimator under the proposed precoder can yield much better performance than the single-column method. The SNR gain at NMSE $= 0.4 \cdot 10^{-3}$ is about 5 dB, whereas the SNR gain for the higher SNR region can be claimed as infinity, although the vertical distance between two performance lines may look small. That is, no matter how we increase the SNR, the performance of the single-column estimator cannot achieve that of the joint estimator when SNR is higher than some threshold.

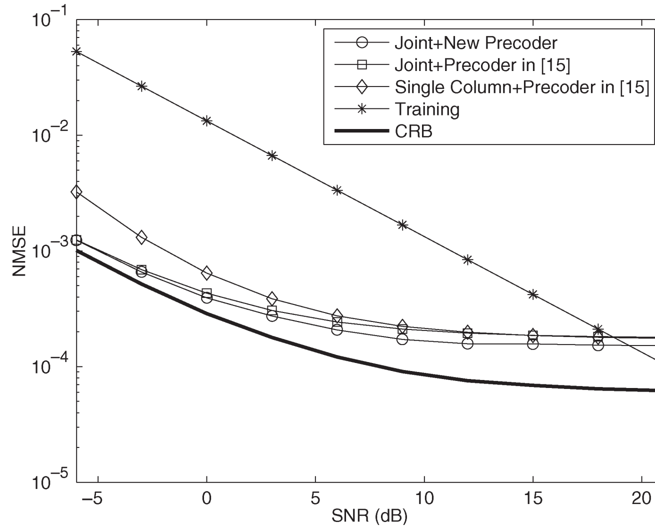


Fig. 2. Comparison between a joint estimator and a single-column estimator under different precoders.

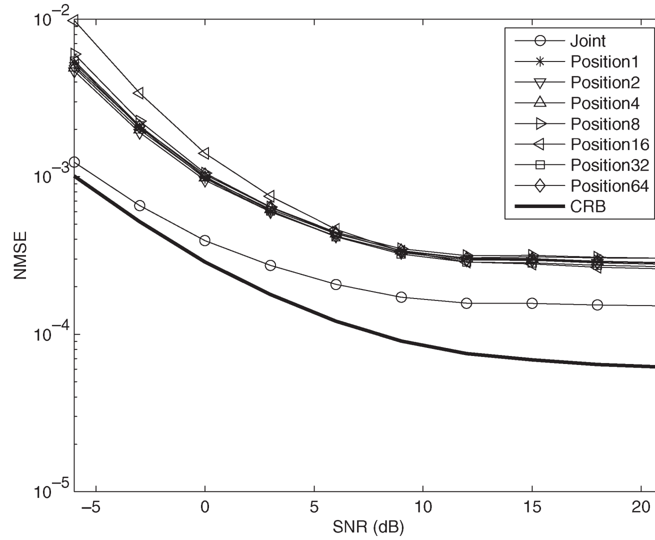


Fig. 3. NMSE versus SNR for both estimators under the proposed precoders.

B. Example 2

In this example, we fix $p = 0.54$ and compare the proposed joint estimator and the single-column estimator for columns 1, 2, 4, 8, 16, 32, and 64 under the proposed precoder. The NMSEs of channel estimation versus SNR of these eight different results are displayed in Fig. 3. The stochastic CRB for the proposed precoder is displayed as well. In Fig. 3, we see that the performance of the joint estimator is better than all the single-column estimators. We can also see that both the NMSE curves and the CRB curve meet a threshold after the SNR reaches a certain level. This is because our proposed algorithm is based on the assumption that the source covariance is a multiple of the identity matrix. Since only a finite number

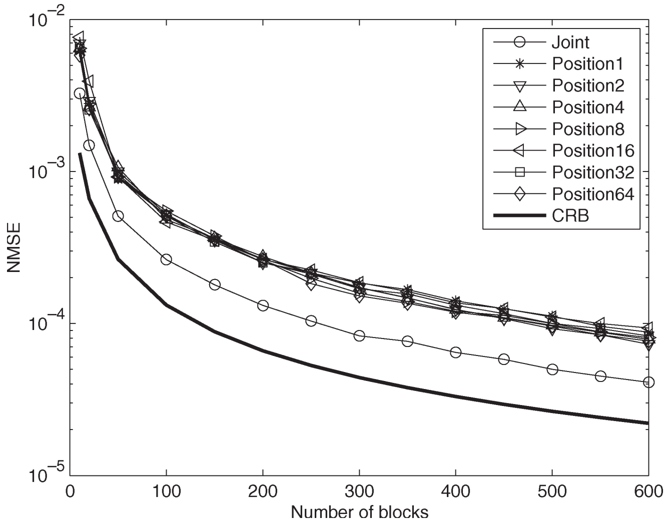


Fig. 4. NMSE versus number of OFDM blocks for both estimators under the proposed precoders.

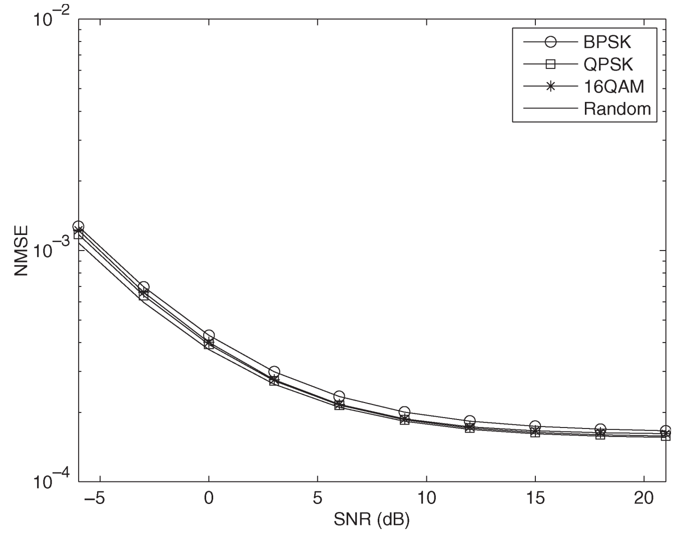


Fig. 5. NMSE versus SNR for the joint estimator under different constellations.

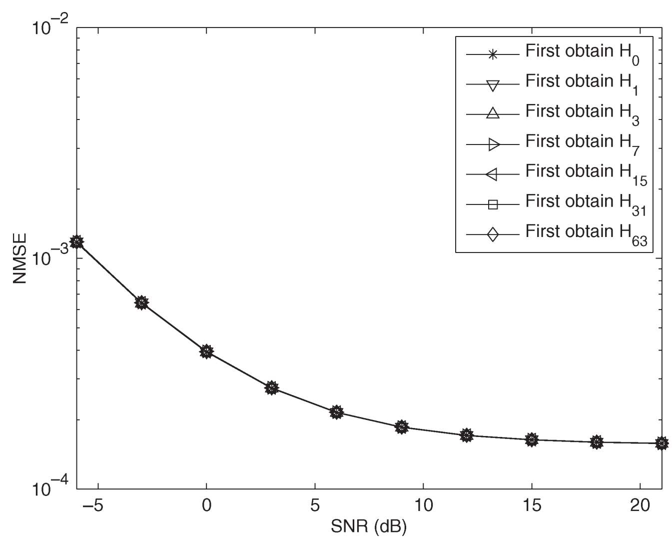


Fig. 6. NMSE versus SNR for the joint estimator under a starting point.

of snapshots are available in the simulation, the performance accuracy could not be further increased after the SNR reaches a certain level. Further improvement could be achieved by increasing the number of snapshots.

Consequently, we also examine the performance of these eight estimators under different numbers of OFDM blocks in Fig. 4. In this case, the SNR is fixed at 10 dB. It can be seen that the joint estimator always performs better for different numbers of OFDM blocks. An important observation is that the NMSEs of the proposed method is around 10^{-3}, even if the number of OFDM blocks is as small as 20. This is a comparable performance with that of the training-based method at $\mathrm{SNR} = 10$ dB, as can be seen in Fig. 2. Therefore, as long as the channel remains constant during 20 OFDM blocks, which is a relatively small number, the proposed algorithm can yield comparable performance as the training-based method, but it can avoid sending training sequence repeatedly every 20 blocks. Compared with the traditional subspace-based method, the proposed algorithm also brings benefits. It is well known that for the traditional subspace-based method [10], the number of OFDM blocks should be greater than the block length $M = 64$ so that the signal subspace can be formed. However, our algorithm works well even for 20 OFDM blocks. The reason is that we derive the channel estimation from the covariance matrix directly without using EVD. Therefore, the proposed method is robust to the varying channel where less OFDM blocks can be obtained during the channel coherent time. Moreover, we found that all NMSE curves keep on decreasing as the number of snapshots increases. This is different from the traditional subspace-based blind channel estimation because the assumption in the proposed method requires $\mathbf{R}_s$ to be an identity matrix, which is critically affected by the number of OFDM blocks.

C. Example 3

We examine the impact of the symbol constellation on the channel estimation accuracy in this example. Three NMSEs curves of the joint estimator corresponding to binary phase-shift keying, QPSK, and 16-quadratic-amplitude modulation are shown in Fig. 5. One more curve where the transmitted symbols are taken as Gaussian random variables is also displayed. One can see that different constellations give ignorable performance difference. Hence, the signal constellation does not affect the channel estimation result significantly.

D. Example 4

In this example, we compare the performance of the joint estimator by taking different starting points H_{q-1}. The performance NMSEs versus SNR is shown in Fig. 6. We see that all the joint estimations give the same performance. Hence, we numerically prove the robustness of our proposed algorithm to the selection of the starting point.

E. Example 5

We then show the performance of the proposed algorithm under different values of p as 0.25, 0.5, 0.8, and 1. The

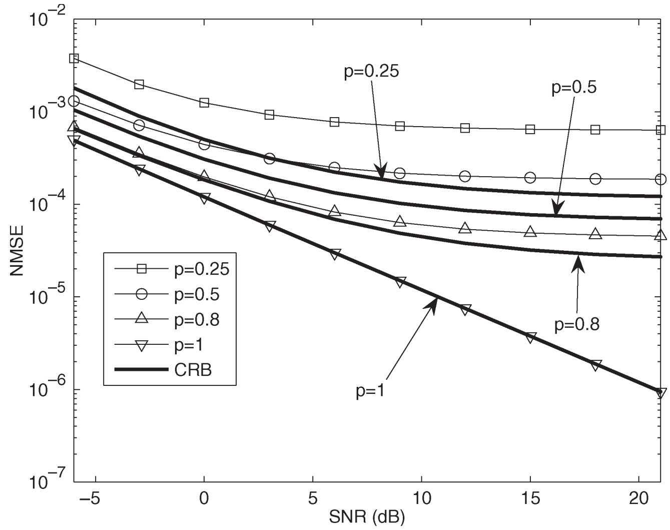


Fig. 7. NMSE versus SNR for the joint estimator under different values of p.

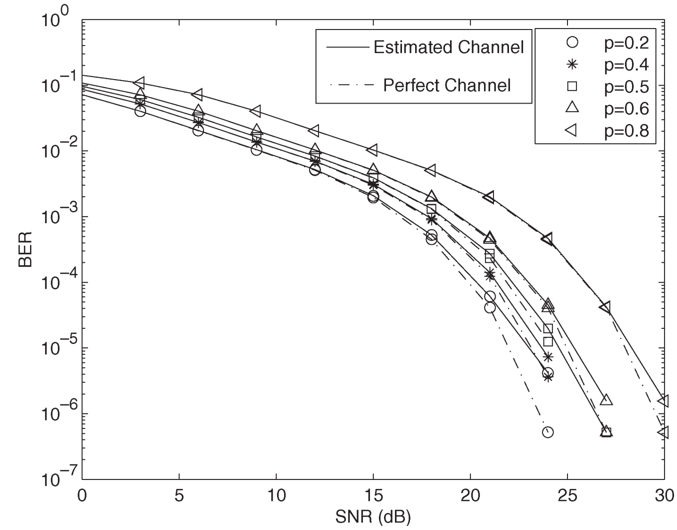


Fig. 8. BER versus SNR under different values of p.

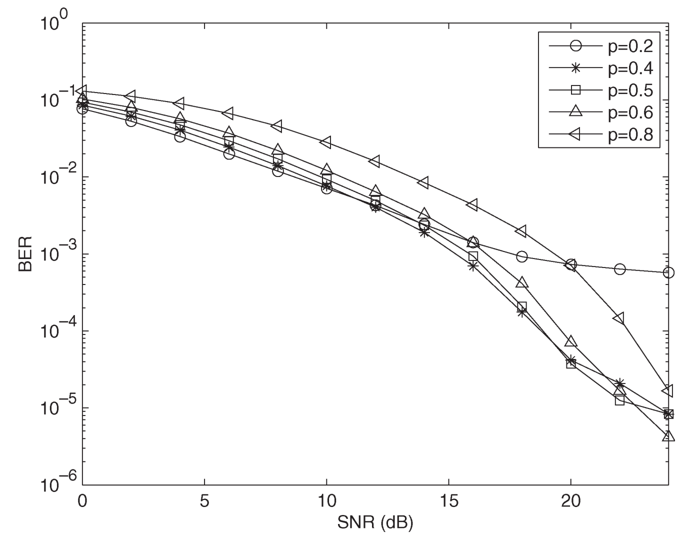


Fig. 9. BER versus SNR for $M = 8$ under different values of p.

corresponding NMSEs versus SNR is shown in Fig. 7. The four stochastic CRBs are also drawn for comparison. We can see that the value of p is critical to the performance of channel estimation. As analyzed in Section III, increasing p will cause a better channel estimation. At the same time, the gaps between the NMSEs and their corresponding CRBs become smaller. Specifically, when p reaches 1, the NMSE curve overlaps the CRB curve, indicating that the proposed algorithm gives as best channel estimation as it could. However, as claimed before, this value of p will caused a singular matrix $\mathbf{P}$ and cannot be used for signal detection.[3]

Finally, we examine the bit error rate (BER) performance after the channel is estimated by the proposed algorithm. The MMSE receiver is applied [18], and the uncoded transmission is adopted. In Fig. 8, the BERs versus SNR are plotted for p equal to 0.2, 0.4, 0.5, 0.6, and 0.8, respectively. The BER curves for the perfect channel knowledge under different p's are displayed as well. In Fig. 8, it is first noted that the proposed algorithm can provide a comparable BER performance as when the channel is perfectly known. Meanwhile, we see that $p = 0.2$ gives the best performance, whereas $p = 0.8$ gives the worst performance, although the latter can provide the best channel estimation. The best choice of p on symbol detection is affected by system parameters, e.g., the block length, the channel length, and the SNR. We then consider an OFDM system with $M = 8$, while still using the same channel. The BER performance for this new consideration is shown in Fig. 9. It is seen that the BER performance for $p = 0.2$ is the best for the lower SNR region but meets an error floor at the higher SNR region. Moreover, $p = 0.5$ performs the best for the relatively higher SNR, whereas $p = 0.6$ performs the best for the high SNR region.

VI. Conclusion

In this work, we developed a new blind channel estimation technique for OFDM transmissions based on the second-order statistical analysis. By considering a generalized linear precoding, a joint estimation method is proposed in which the full information of signal covariance matrix is utilized. A proper design criterion of the precoding matrix is provided, and the proposed method is shown to be more effective than the existing estimator where the channel information is extracted from a single column of the signal covariance matrix. At the same time, the complexity of our proposed estimator does not increase much compared to the previous work. In addition, the stochastic CRB is derived in a closed form. Simulation results clearly show the effectiveness and the improvement of our proposed algorithm in reducing the estimation error. We finally make the conclusion that the nonredundant precoding method basically provides two tradeoffs.

1) *Tradeoff Between the SNR and the Number of OFDM Blocks*: This can be understood as follows. When channel remains constant for a relatively longer time, we can always use longer observations to obtain an accurate channel estimation for a lower SNR transmission.

[3] $p = 1$ is used here, but only to show the extreme case on channel estimation.

2) *Tradeoff Between the Accuracy of the Channel Estimation and the Spectral Efficiency*: When channel varies relatively fast, where the channel coherent time spreads tens of OFDM blocks, we can apply the proposed algorithm to achieve a relatively accurate channel estimation while avoiding frequently sending training sequences.[4]

APPENDIX
PROOF OF THE STOCHASTIC CRB

For a circular complex zero-mean Gaussian random variable $\mathbf{x} = \mathbf{x}_R + j\mathbf{x}_I$ with covariance matrices $\mathbf{R}_x = \mathrm{E}\{\mathbf{x}\mathbf{x}^H\}$ parameterized by a real vector $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_K]$, the Fisher information matrix (FIM) of this vector $\boldsymbol{\theta}$ is given by [21], [22]

$$\mathrm{FIM}(m,q) = N\,\mathrm{tr}\left(\frac{d\mathbf{R}_x}{d\theta_m}\mathbf{R}_x^{-1}\frac{d\mathbf{R}_x}{d\theta_q}\mathbf{R}_x^{-1}\right), \quad m,q = 1,\ldots,K, \tag{45}$$

where N is the number of available snapshots. For the proposed method, the signal covariance matrix $\mathbf{R}_x$ can be parameterized by $\boldsymbol{\theta} = [\mathbf{h}_R^T, \mathbf{h}_I^T, \sigma_s^2, \sigma_n^2]^T$ and is given by

$$\mathbf{R}_x = \mathrm{E}\left\{\mathbf{x}(k)\mathbf{x}(k)^H\right\} = \sigma_s^2\bar{\mathbf{H}}\mathbf{F}^H\mathbf{P}\mathbf{F}\bar{\mathbf{H}}^H + \sigma_n^2\mathbf{I}. \tag{46}$$

The following properties are very useful for the derivation of the CRB:

$$\mathrm{tr}(\mathbf{X}\mathbf{Y}) = \mathrm{vec}(\mathbf{X}^H)^H\mathrm{vec}(\mathbf{Y}) \tag{47}$$

$$\mathrm{vec}(\mathbf{X}\mathbf{Y}\mathbf{Z}) = (\mathbf{Z}^T\otimes\mathbf{X})\mathrm{vec}(\mathbf{Y}) \tag{48}$$

$$(\mathbf{X}\otimes\mathbf{Y})(\mathbf{Z}\otimes\mathbf{W}) = (\mathbf{X}\mathbf{Z})\otimes(\mathbf{Y}\mathbf{W}) \tag{49}$$

which hold for all matrices $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$, and $\mathbf{W}$. Using these properties, we can rewrite (45) as

$$\frac{1}{N}\mathrm{FIM}(m,q) = \mathrm{vec}\left(\frac{d\mathbf{R}_x}{d\theta_m}\right)^H\left(\mathbf{R}_x^{-T}\otimes\mathbf{R}_x^{-1}\right)\mathrm{vec}\left(\frac{d\mathbf{R}_x}{d\theta_q}\right) \tag{50}$$

or equivalently

$$\frac{1}{N}\mathrm{FIM} = \left(\frac{d\mathbf{r}_x}{d\boldsymbol{\theta}^T}\right)^H\left(\mathbf{R}_x^{-T}\otimes\mathbf{R}_x^{-1}\right)\left(\frac{d\mathbf{r}_x}{d\boldsymbol{\theta}^T}\right) \tag{51}$$

where

$$\begin{aligned}\mathbf{r}_x &= \mathrm{vec}(\mathbf{R}_x)\\ &= \sigma_s^2(\bar{\mathbf{H}}^*\otimes\bar{\mathbf{H}})\mathrm{vec}(\mathbf{F}^H\mathbf{P}\mathbf{F}) + \sigma_n^2\mathrm{vec}(\mathbf{I}).\end{aligned} \tag{52}$$

Equation (51) can be partitioned as

$$\frac{1}{N}\mathrm{FIM} = \begin{bmatrix}\mathbf{G}^H\\ \boldsymbol{\Delta}^H\end{bmatrix}[\mathbf{G}\ \boldsymbol{\Delta}] \tag{53}$$

[4]This property is not held by the traditional subspace-based blind channel estimation method, as has been indicated in the simulation part.

where

$$\begin{aligned}[\mathbf{G}|\boldsymbol{\Delta}] &= \left(\mathbf{R}_x^{-T/2}\otimes\mathbf{R}_x^{-1/2}\right)\frac{d\mathbf{r}_x}{d\boldsymbol{\theta}^T}\\ &= \left(\mathbf{R}_x^{-T/2}\otimes\mathbf{R}_x^{-1/2}\right)\left[\frac{d\mathbf{r}_x}{d\mathbf{h}_R^T}\frac{d\mathbf{r}_x}{d\mathbf{h}_I^T}\middle|\frac{d\mathbf{r}_x}{d\sigma_s^2}\frac{d\mathbf{r}_x}{d\sigma_n^2}\right].\end{aligned} \tag{54}$$

From [25], we know that for blind channel estimation, the FIM is singular such that its inverse does not exist. Then, some constraints should be utilized to make FIM a nonsingular matrix. Instead of taking any specific constraint, we will use the minimal constrained CRB defined as [25]

$$\mathrm{CRB} = \mathrm{FIM}^{\dagger}. \tag{55}$$

This is a particularly constrained CRB that yields the lowest value for $\mathrm{tr}\{\mathrm{CRB}\}$ among all sets of minimal numbers of independent constraints.

Lemma 2 [26]: Suppose the FIM for $\boldsymbol{\theta} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2]^T$ is

$$\mathrm{FIM} = \begin{bmatrix}\mathbf{J}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1} & \mathbf{J}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_2}\\ \mathbf{J}_{\boldsymbol{\theta}_2\boldsymbol{\theta}_1} & \mathbf{J}_{\boldsymbol{\theta}_2\boldsymbol{\theta}_2}\end{bmatrix} \tag{56}$$

and assume that FIM is singular but $\mathbf{J}_{\boldsymbol{\theta}_2\boldsymbol{\theta}_2}$ is nonsingular. Then, the minimal constrained CRB for $\boldsymbol{\theta}_1$ separately is

$$\mathrm{CRB}_{\boldsymbol{\theta}_1} = [\mathbf{J}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1} - \mathbf{J}_{\boldsymbol{\theta}_1\boldsymbol{\theta}_2}\mathbf{J}_{\boldsymbol{\theta}_2\boldsymbol{\theta}_2}^{-1}\mathbf{J}_{\boldsymbol{\theta}_2\boldsymbol{\theta}_1}]^{\dagger}. \tag{57}$$

In our case, $\boldsymbol{\theta}_1 = [\mathbf{h}_R^T, \mathbf{h}_I^T]^T$, $\boldsymbol{\theta}_2 = [\sigma_s^2, \sigma_n^2]$, and the minimal CRB for $\boldsymbol{\theta}_1$ can be written as

$$\begin{aligned}\mathrm{CRB}_{\boldsymbol{\theta}_1} &= \frac{1}{N}\left[\mathbf{G}^H\mathbf{G} - \mathbf{G}^H\boldsymbol{\Delta}(\boldsymbol{\Delta}^H\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}^H\mathbf{G}\right]^{\dagger}\\ &= \frac{1}{N}\left(\mathbf{G}^H\mathbf{P}_{\boldsymbol{\Delta}}^{\perp}\mathbf{G}\right)^{\dagger}.\end{aligned} \tag{58}$$

We need to evaluate the derivatives of $\mathbf{r}_x$ with respect to $\boldsymbol{\theta}$. Let us first make the following partition:

$$\begin{aligned}[\mathbf{G}\boldsymbol{\Delta}] &= [\mathbf{G}_1|\mathbf{G}_2|\mathbf{v}|\mathbf{u}]\\ &= \left(\mathbf{R}_x^{-T/2}\otimes\mathbf{R}_x^{-1/2}\right)\left[\frac{d\mathbf{r}_x}{d\mathbf{h}_R^T}\middle|\frac{d\mathbf{r}_x}{d\mathbf{h}_I^T}\middle|\frac{d\mathbf{r}_x}{d\sigma_s^2}\middle|\frac{d\mathbf{r}_x}{d\sigma_n^2}\right].\end{aligned} \tag{59}$$

After some algebraic manipulations, the following results can be obtained:

$$\begin{aligned}\mathbf{G}_1(:,l) &= \left(\mathbf{R}_x^{-T/2}\otimes\mathbf{R}_x^{-1/2}\right)\mathrm{vec}\left(\frac{d\mathbf{R}_x}{d\mathbf{h}_{R,l}}\right)\\ &= \mathrm{vec}\left(\mathbf{R}_x^{-1/2}\frac{d\mathbf{R}_x}{d\mathbf{h}_{R,l}}\mathbf{R}_x^{-1/2}\right)\\ &= \mathrm{vec}\left(\mathbf{A}_l + \mathbf{A}_l^H\right)\end{aligned} \tag{60}$$

$$\begin{aligned}\mathbf{G}_2(:,l) &= \left(\mathbf{R}_x^{-T/2}\otimes\mathbf{R}_x^{-1/2}\right)\mathrm{vec}\left(\frac{d\mathbf{R}_x}{d\mathbf{h}_{I,l}}\right)\\ &= \mathrm{vec}\left(\mathbf{R}_x^{-1/2}\frac{d\mathbf{R}_x}{d\mathbf{h}_{I,l}}\mathbf{R}_x^{-1/2}\right)\\ &= \mathrm{vec}\left(j\mathbf{A}_l - j\mathbf{A}_l^H\right)\end{aligned} \tag{61}$$

$$\begin{aligned}\mathbf{v} &= \left(\mathbf{R}_x^{-T/2} \otimes \mathbf{R}_x^{-1/2}\right) \mathrm{vec}\left(\frac{d\mathbf{R}_x}{d\sigma_s^2}\right)\\ &= \mathrm{vec}\left(\mathbf{R}_x^{-1/2} \frac{d\mathbf{R}_x}{d\sigma_s^2} \mathbf{R}_x^{-1/2}\right)\\ &= \mathrm{vec}\left(\mathbf{R}_x^{-1/2} \mathbf{H}\mathbf{F}^H \mathbf{P}\mathbf{F}\mathbf{H}^H \mathbf{R}_x^{-1/2}\right) \end{aligned} \tag{62}$$

$$\begin{aligned}\mathbf{u} &= \left(\mathbf{R}_x^{-T/2} \otimes \mathbf{R}_x^{-1/2}\right) \mathrm{vec}\left(\frac{d\mathbf{R}_x}{d\sigma_n^2}\right)\\ &= \mathrm{vec}\left(\mathbf{R}_x^{-1/2} \frac{d\mathbf{R}_x}{d\sigma_n^2} \mathbf{R}_x^{-1/2}\right)\\ &= \mathrm{vec}\left(\mathbf{R}_x^{-1}\right) \end{aligned} \tag{63}$$

where $\mathbf{h}_{R,l}$ and $\mathbf{h}_{I,l}$ are the lth elements of $\mathbf{h}_R$ and $\mathbf{h}_I$, respectively. Matrix $\mathbf{A}_l$ is given in (33).

REFERENCES

[1] J. A. C. Bingham, "Multicarrier modulation for data transmission: An idea whose time has come," *IEEE Commun. Mag.*, vol. 28, no. 5, pp. 5–14, May 1990.

[2] W. Y. Zou and Y. Wu, "COFDM: An overview," *IEEE Trans. Broadcast.*, vol. 41, no. 1, pp. 1–8, Mar. 1995.

[3] *Radio Broadcasing System, Digital Audio Broadcasting (DAB) to Mobile, Portable, and Fixed Receiver*, Eur. Telecommun. Stand. Inst., Sophia-Antipolis, Valbonne, France, 1995–1997. ETS 300 401.

[4] *Digital Broadcasting System Television, Sound, and Data Services; Framing Structure, Channel Coding, and Modulation Digital Terrestrial Television*, Eur. Telecommun. Stand. Inst., Sophia-Antipolis, Valbonne, France, 1996. ETS 300 744.

[5] *Broadband Radio Access Networks (BRAN): High Performance Radio Local Area Networks (HIPERLAN), Type 2; Systems Overview*, Eur. Telecommun. Stand. Inst., Sophia-Antipolis, Valbonne, France, 1999. ETR 101 683 114.

[6] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: High Speed Physical Layer in the 5 GHZ Band*, 1999. IEEE802.11a.

[7] H. Sari, G. Karam, and I. Jeanclaude, "Transmission techniques for digital terrestrial TV broadcasting," *IEEE Commun. Mag.*, vol. 33, no. 2, pp. 100–109, Feb. 1995.

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

[9] L. Tong and S. Perreau, "Multichannel blind identification: From subspace to maximum likelihood methods," *Proc. IEEE*, vol. 86, no. 10, pp. 1951–1968, Oct. 1998.

[10] E. Moulines, P. Duhamel, J. F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Process.*, vol. 43, no. 2, pp. 516–525, Feb. 1995.

[11] Z. Wang and G. B. Giannakis, "Wireless multicarrier communications," *IEEE Signal Process. Mag.*, vol. 17, no. 3, pp. 29–48, May 2000.

[12] B. Muquet, Z. Wang, G. B. Giannakis, M. de Courville, and P. Duhamel, "Cyclic prefixing or zero padding for wireless multicarrier transmissions," *IEEE Trans. Commun.*, vol. 50, no. 12, pp. 2136–2148, Dec. 2002.

[13] R. Zhang, "Blind OFDM channel estimation through linear precoding: A subspace approach," in *Proc. 36th Asilomar Conf.*, Pacific Grove, CA, Nov. 2002, vol. 1, pp. 631–633.

[14] C. Li and S. Roy, "Subspace-based blind channel estimation for OFDM by exploiting virtual carriers," *IEEE Trans. Wireless Commun.*, vol. 2, no. 1, pp. 141–150, Jan. 2003.

[15] A. P. Petropulu and R. Zhang, "Blind channel estimation for OFDM systems," in *Proc. DSP/SPE*, Atlanta, GA, Oct. 2002, pp. 366–370.

[16] A. P. Petropulu, R. Zhang, and R. Lin, "Blind OFDM channel estimation through simple linear precoding," *IEEE Trans. Wireless Commun.*, vol. 3, no. 2, pp. 647–655, Mar. 2004.

[17] R. Lin and A. P. Petropulu, "Linear block precoding for blind channel estimation in OFDM systems," in *Proc. ISSPA*, Paris, France, Jul. 2003, vol. 2, pp. 395–398.

[18] R. Lin and A. P. Petropulu, "Blind channel estimation for OFDM systems based on non-redundant linear precoding," in *Proc. Stat. Process.*, St. Louis, MO, Oct. 2003, pp. 351–354.

[19] J.-J. van de Beek, O. Edfors, M. Sandell, S. K. Wilson, and P. O. Borjesson, "On channel estimation in OFDM system," in *Proc. VTC*, Chicago, IL, Jul. 1995, vol. 2, pp. 815–819.

[20] M. Biguesh and A. B. Gershman, "MIMO channel estimation: Optimal training and tradeoffs between estimation techniques," in *Proc. ICC*, Paris, France, Jun. 2004, vol. 5, pp. 2658–2662.

[21] P. Stoica, E. G. Larsson, and A. B. Gershman, "The stochastic CRB for array processing: A textbook derivation," *IEEE Signal Process. Lett.*, vol. 8, no. 5, pp. 148–150, May 2001.

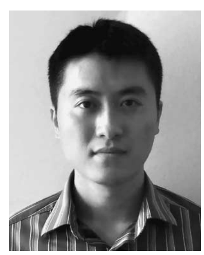
[22] J.-P. Delmas and H. Abeida, "Stochastic Cramér–Rao bound for noncircular signals with application to DOA estimation," *IEEE Trans. Signal Process.*, vol. 52, no. 11, pp. 3192–3199, Nov. 2004.

[23] M. Morelli and U. Mengali, "A comparison of pilot-aided channel estimation methods for OFDM systems," *IEEE Trans. Signal Process.*, vol. 49, no. 12, pp. 3065–3073, Dec. 2001.

[24] H. Gazzah, P. A. Regalia, J. P. Delmas, and K. Abed-Meraim, "A blind multichannel identification algorithm robust to order over estimation," *IEEE Trans. Signal Process.*, vol. 50, no. 6, pp. 1449–1458, Jun. 2002.

[25] E. de Carvalho, J. Cioffi, and D. Slock, "Cramér–Rao bounds for blind multichannel estimation," in *Proc. GLOBECOM*, San Francisco, CA, Nov. 2000, vol. 2, pp. 1036–1040.

[26] K. Wei and B. Champagne, "Generalized blind subspace channel estimation," in *Proc. VTC—Fall*, Orlando, FL, Oct. 2003, vol. 2, pp. 1209–1213.



Feifei Gao (S'05) received the B.Eng. degree in information engineering from Xi'an Jiaotong University, Xi'an, China, in 2002, the M.Sc. degree from McMaster University, Hamilton, ON, Canada, in 2004, and is currently working toward the Ph.D. degree in the Department of Electrical Engineering, National University of Singapore, Singapore.

His research interests are in communication theory, broadband wireless communications, signal processing for communications, MIMO systems, and array signal processing.

Mr. Gao received the President Scholarship from the National University of Singapore.

Arumugam Nallanathan (S'97–M'00–SM'05) received the B.Sc. degree (with honors) from the University of Peradeniya, Peradeniya, Sri Lanka, in 1991, the C.P.G.S. degree from Cambridge University, Cambridge, U.K., in 1994, and the Ph.D. degree from the University of Hong Kong, Hong Kong, in 2000, all in electrical engineering.

Since then, he has been an Assistant Professor with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore. His research interests include OFDM systems, ultra-wide-bandwidth (UWB) communication and localization, MIMO systems, and cooperative diversity techniques. In these areas, he has published over 100 journal and conference papers. He currently serves on the Editorial Board of John Wiley's *Wireless Communications and Mobile Computing* and the EURASIP *Journal of Wireless Communications and Networking* as an Associate Editor. He has also served as a Guest Editor for the EURASIP *Journal of Wireless Communications and Networking Special Issue on UWB Communication Systems—Technology and Applications*.

Dr. Nallanathan currently serves on the Editorial Board of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY as an Associate Editor. He served as a Technical Program Cochair and as a Technical Program Committee Member for more than 25 IEEE international conferences.