EM-Based Channel Estimation Algorithms for OFDM

Xiaoqiang Ma

Department of Electrical Engineering, School of Engineering and Applied Science, Princeton University,
Princeton, NJ 08544-5263, USA
Email: xma@princeton.edu

Hisashi Kobayashi

Department of Electrical Engineering, School of Engineering and Applied Science, Princeton University,
Princeton, NJ 08544-5263, USA
Email: hisashi@princeton.edu

Stuart C. Schwartz

Department of Electrical Engineering, School of Engineering and Applied Science, Princeton University,
Princeton, NJ 08544-5263, USA
Email: stuart@princeton.edu

Estimating a channel that is subject to frequency-selective Rayleigh fading is a challenging problem in an orthogonal frequency division multiplexing (OFDM) system. We propose three EM-based algorithms to efficiently estimate the channel impulse response (CIR) or channel frequency response of such a system operating on a channel with multipath fading and additive white Gaussian noise (AWGN). These algorithms are capable of improving the channel estimate by making use of a modest number of pilot tones or using the channel estimate of the previous frame to obtain the initial estimate for the iterative procedure. Simulation results show that the bit error rate (BER) as well as the mean square error (MSE) of the channel can be significantly reduced by these algorithms. We present simulation results to compare these algorithms on the basis of their performance and rate of convergence. We also derive Cramer-Rao-like lower bounds for the unbiased channel estimate, which can be achieved via these EM-based algorithms. It is shown that the convergence rate of two of the algorithms is independent of the length of the multipath spread. One of them also converges most rapidly and has the smallest overall computational burden.

Keywords and phrases: OFDM, EM-algorithm, channel estimation, Cramer-Rao lower bound.

1. INTRODUCTION

Orthogonal frequency division multiplexing (OFDM) [1], a spectrally efficient form of frequency division multiplexing (FDM), divides its allocated channel spectrum into several parallel subchannels. OFDM is inherently robust against frequency-selective fading since each subchannel occupies a relatively narrowband, where the channel frequency characteristic is nearly flat. OFDM has an additional advantage of being computationally efficient because the fast Fourier transform (FFT) technique can be used to implement the modulation and demodulation functions [2]. Furthermore, the OFDM system can eliminate interframe interference (IFI[1]) through the use of a cyclic prefix (CP) that is longer than the order of the channel impulse re-

sponse (CIR). OFDM has already been used in European digital audio broadcasting (DAB), digital video broadcasting (DVB) systems, high performance radio local area network (HIPERLAN) and IEEE 802.11a wireless local area networks (WLAN). It has also been shown that OFDM is an effective way of increasing data rates and simplifying the equalization in wireless communications [3].

However, it is not possible to make reliable data decisions unless a good channel estimate is available for coherent demodulation. Although differential detection could be used to detect the transmitted signal in the absence of channel information, it would result in about a 3 dB loss in signal-to-noise ratio (SNR) compared with coherent detection. A number of channel estimation algorithms have been reported in the literature [4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18]. For some of these algorithms, however, the channel estimate is continuously updated by transmitting pilot symbols using specified time-frequency lattices. One class of

---

[1]In the literature, the term intersymbol interference (ISI) is used, but we believe IFI is more appropriate in this paper.
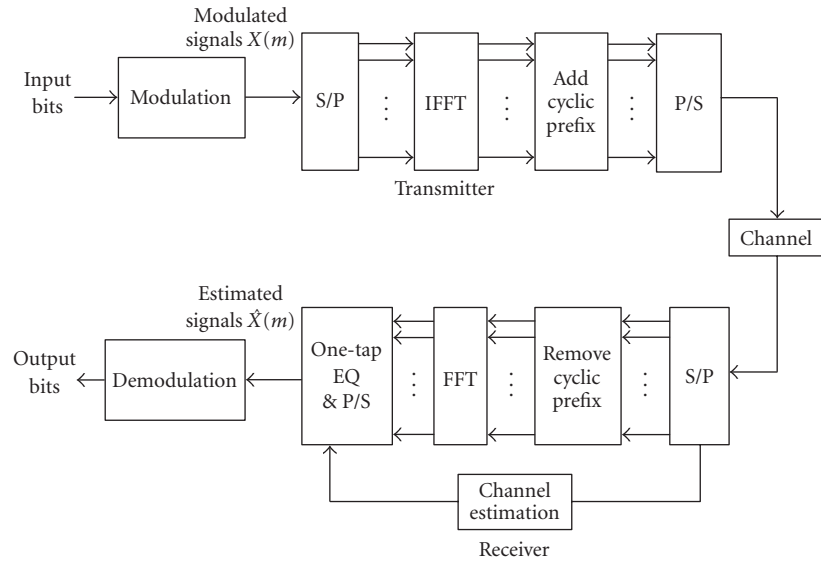
FIGURE 1: Baseband OFDM system model.

such pilot-assisted estimation algorithms adopt an interpolation technique with fixed parameters (two-dimensional (2D) [6, 7] or one-dimensional (1D) [5]) to estimate the channel frequency response by using the channel estimate obtained at the lattices assigned to the pilot tones. Linear, spline, and Gaussian filters have all been studied [5]. Another method in this category adopts a known channel frequency covariance matrix and uses a channel estimate at pilot tones to estimate the CIR in the sense of minimum mean square error (MMSE) [4, 8, 9, 11]. Shortcomings of these algorithms include (1) a large error floor that may be incurred by a mismatch between the estimated and real CIR, (2) difficulty in obtaining the channel frequency covariance matrix and the resultant error due to channel statistics mismatch, and (3) spectrum inefficiency due to the overhead (typically 20%) associated with use of pilot symbols. In addition, several kinds of blind channel estimation algorithms have been proposed in order to improve transmission efficiency. These algorithms are based on the statistical property of received signals (e.g., second-order statistics [12, 13, 14, 15]), the characteristic of virtual subcarriers [16], and the finite-alphabet property of transmitted signals [18]. However, each of these blind estimation algorithm has its limitation. For example, second-order statistics-based algorithms cannot be used in a high mobility environment (i.e., a large Doppler spread) since they require many blocks of data to carry out the estimation procedure. A finite-alphabet-based algorithm can be applied only to a constant modulus signal. In contrast, in this paper, we extend and enhance some existing pilot-based channel estimation algorithms by substantially reducing the number of pilot symbols using the expectation-maximization (EM) algorithm.

The EM algorithm [19, 20] is a technique for finding maximum likelihood (ML) estimates of system parameters in a broad range of problems where observed data are incomplete. The EM algorithm consists of two iterative steps:

the expectation (E) step and the maximization (M) step. The E-step is performed with respect to the unknown underlying parameters, using current estimates of the parameters, conditioned upon the incomplete observations. The M-step then provides new estimates of the parameters that maximize the expectation of the log-likelihood function defined over complete data, conditioned on the most recent observation and the last estimate. These two steps are iterated until the estimated values converge.

The main objective of this paper is to investigate the use of the EM algorithm for channel estimation of an OFDM system that is subject to slow time-varying frequency-selective fading. Three different algorithms have been developed and compared. In each of the algorithms, the initial channel estimate is obtained either from pilot symbols (that are inserted in the OFDM frame) or from the channel estimate of the previous OFDM frame (where there is no pilot symbol in the current OFDM frame).

The rest of the paper is organized as follows. In Section 2, we will describe the baseband OFDM system model and discuss some assumptions. In Section 3, the three different EM-based channel estimation algorithms are derived and fully discussed. The Cramer-Rao lower bound (CRLB) and modified CRLB (MCRB) are discussed in Section 4 for both constant and nonconstant modulus signals. Comprehensive simulation results and discussion are given in Section 5. Finally, we draw some conclusions in Section 6.

2. SYSTEM MODEL AND ASSUMPTIONS

The schematic diagram (Figure 1) is a baseband equivalent representation of an OFDM system. The input binary bits[2] are first fed into a serial-to-parallel (S/P) converter. Each

---

[2]We only consider uncoded OFDM systems.

data stream then modulates the corresponding subcarrier by MPSK or MQAM. The modulation scheme may vary from one subcarrier to another in order to achieve the maximum capacity or the minimum bit error rate (BER) for a given channel characteristic and total signal power constraint. In this paper, we assume, for simplicity, that only QPSK or 16 QAM is used in any of these subcarriers. We use M to denote the number of subcarriers in the OFDM system. The modulated data symbols, represented by complex variables $X(0), \ldots, X(M-1)$, are then transformed by the inverse fast Fourier transform (IFFT). The output symbols are denoted as $x(0), \ldots, x(M-1)$ and are given by

$$x(k) = \frac{1}{\sqrt{M}} \sum_{m=0}^{M-1} X(m)e^{j2\pi(km/M)}, \quad 0 \le k \le M-1. \quad (1)$$

In order to avoid IFI, CP symbols, which replicate the end part of the IFFT output symbols, are added in front of each frame, that is,

$$x(k) = x(M+k), \quad -N_{cp} \le k \le -1, \quad (2)$$

where N_{cp} denotes the length of the CP. The parallel data are converted back to a serial data stream, that is, $x(M - N_{cp}), \ldots, x(M-1), x(0), \ldots, x(M-1)$, and transmitted over the frequency-selective channel with additive white Gaussian noise (AWGN). The received data $y(-N_{cp}), \ldots, y(-1), y(0), \ldots, y(M-1)$ are converted back to $Y(0), \ldots, Y(M-1)$ after discarding the prefix symbols $y(-N_{cp}), \ldots, y(-1)$, and applying the FFT and demodulation to the remainder $y(0), \ldots, y(M-1)$.

The channel model we adopt in the present paper is a multipath slowly time-varying (unchanged in any one OFDM frame) fading channel, which can be described by

$$y(k) = \sum_{l=0}^{L-1} h_l x(k-l) + n(k), \quad 0 \le k \le M-1. \quad (3)$$

The CIR h_l's $(0 \le l \le L-1)$ are independent complex-valued Gaussian random variables (which represents a frequency-selective Rayleigh fading channel), and $n(k)$'s $(0 \le k \le M-1)$ are i.i.d. complex-valued Gaussian random variables with zero mean and variance σ^2 for both real and imaginary components. L is the length of the CIR.

If we add the CP in each OFDM data frame, with its length chosen to be longer than L, there will be no IFI between OFDM frames. Thus, we only need to consider one OFDM frame at a time in deriving the system model. After discarding the CP and performing an FFT at the receiver, we can obtain the received data frame in the frequency domain:

$$Y(m) = \frac{1}{\sqrt{M}} \sum_{k=0}^{M-1} y(k)e^{-j2\pi(km/M)}, \quad 0 \le m \le M-1. \quad (4)$$

Then using the CP condition (2), we obtain the following simple expression:

$$Y(m) = X(m)H(m) + N(m), \quad 0 \le m \le M-1, \quad (5)$$

where $H(m)$ is the frequency response of the channel at subcarrier m defined as follows:

$$H(m) = \sum_{l=0}^{L-1} h_l e^{-j2\pi(ml/M)}, \quad 0 \le m \le M-1, \quad (6)$$

and the set of the transformed noise variables $N(m)$, $0 \le m \le M-1$,

$$N(m) = \frac{1}{\sqrt{M}} \sum_{k=0}^{M-1} n(k)e^{-j2\pi(mk/M)}, \quad 0 \le m \le M-1, \quad (7)$$

are i.i.d. complex-valued Gaussian variables and have the same distribution as $n(k)$, that is, with mean zero and variance σ^2. In a regular OFDM system, the channel delay spread L is much smaller than the number of subcarriers. This leads to a high correlation between the channel frequency responses $H(m)$, $0 \le m \le M-1$, even when h_l, $0 \le l \le L-1$, are independent.

In this paper, we assume the CIR is constant in each OFDM frame and varies from frame to frame according to the fading rate. However, in the derivation below, we assume, for generality, that the channel is constant during D OFDM frames. Note that intercarrier interference (ICI) is also eliminated at the FFT output because of the orthogonality between the subcarriers under the assumption that the CP is longer than the channel delay spread. Furthermore, we assume the system has perfect timing and frequency synchronization.

Notation

We use the standard notation, that is, $(\cdot)^T$ denotes the transpose, $(\cdot)^*$ denotes the complex conjugate, $(\cdot)^H$ denotes the Hermitian, underscore letters stand for column vectors, and bold letters stand for matrices. We denote the pth estimates of the channel response in the frequency domain as $\underline{H}^{(p)}$ and in the time domain as $\underline{h}^{(p)}$, and transmitted signals as $\underline{X}^{(p)}$.

3. EM-BASED CHANNEL ESTIMATION ALGORITHMS

3.1. Introduction to the EM algorithm

The EM algorithm [19, 20] is an iterative method to find the ML estimates of parameters in the presence of unobserved data. The idea behind the algorithm is to augment the observed data with latent data, which can be either missing data or parameter values, so that the likelihood function conditioned on the data and the latent data has a form that is easy to manipulate. The algorithm can be broken down into two steps: the E-step and the M-step. We assume that the data $\mathcal{Z}$ ("complete" data) can be separated into two components, $\mathcal{Z} = (\mathcal{X}, \mathcal{Y})$, where $\mathcal{X}$ are the observed data ("incomplete" data) and $\mathcal{Y}$ are the missing data. We denote θ as the unknown parameter we try to estimate from $\mathcal{Y}$.

The E-step finds $Q(\theta|\theta^{(p)})$, the expected value of the log-likelihood of θ, $\log f(\mathcal{Z}|\theta)$, where the expectation is taken with respect to $\mathcal{Y}$ conditioned on $\mathcal{X}$ and the latest estimate

of θ, $\theta^{(p)}$:

$$Q(\theta \mid \theta^{(p)}) = E\{\log f(\mathcal{Z}|\theta) \mid \mathcal{X}, \theta^{(p)}\}. \qquad (8)$$

The M-step then finds $\theta^{(p+1)}$, the value of θ that maximizes $Q(\theta|\theta^{(p)})$ over all possible values of θ:

$$\theta^{(p+1)} = \arg\max_\theta Q(\theta \mid \theta^{(p)}). \qquad (9)$$

This procedure is repeated until the sequence $\theta^{(0)}, \theta^{(1)}, \theta^{(2)}, \ldots$ converges. The EM algorithm is constructed in such a way that the sequence of $\theta^{(p)}$'s converges to the ML estimate of θ.

Applications of the EM algorithm to estimation problems in communications systems have appeared a lot in the literature. Channel estimation [21] and signal detection [22, 23] are two typical applications of the EM algorithm. Georghiades and Han [22] provide a general study of data sequence estimation in the presence of random parameters. Zeger and Kobayashi [23] give a simplified algorithm to detect continuous phase modulated signals in fading channels. In the remainder of this section, we propose three different EM-based channel estimation and signal detection algorithms by defining different "complete" and "incomplete" data sets for these algorithms.

3.2. Algorithm 1: estimating the channel frequency response

OFDM divides its allocated channel spectrum into several parallel subchannels that are only subjected to frequency flat fading. Thus, we only need to estimate the individual $H(m)$, $0 \le m \le M-1$, separately, which will result in a considerable reduction in computational complexity. To simplify the expressions, we omit the subcarrier index m, and simply write Y, X, and H instead of $Y(m)$, $X(m)$, and $H(m)$.

We assume that the frequency-domain signal X of a given subcarrier represents a QPSK or 16 QAM signal with constellation size $C(= 4$ or 16, respectively). We denote the symbols in the signal constellation by $\{X_i, \ 1 \le i \le C\}$.

Due to the Gaussian noise assumption, the probability density function (pdf) of Y given X and H is given by

$$f(Y|X,H) = \frac{1}{2\pi\sigma^2} \exp\left\{ -\frac{1}{2\sigma^2}|Y - HX|^2 \right\}. \qquad (10)$$

By assuming that all C symbols are equally likely and averaging the conditional pdf of (10) over the variable X, we obtain the pdf of Y given H as follows:

$$f(Y|H) = \sum_{i=1}^{C} \frac{1}{2\pi\sigma^2 C} \exp\left\{ -\frac{1}{2\sigma^2}|Y - HX_i|^2 \right\}. \qquad (11)$$

Suppose the channel is static over the period of D OFDM frames. Different values of D can be applied in different applications depending on how rapidly the channel changes. We define the received signal vector $\underline{Y} = [Y^1, \ldots, Y^D]$ and the transmitted signal vector $\underline{X} = [X^1, \ldots, X^D]$ for a specific

subcarrier over D frames. Then we call $\underline{Y}$ and $(\underline{Y}, \underline{X})$ "incomplete" and "complete" data, respectively, following the terminology of the EM algorithm. Assuming that additive Gaussian noise is independent from frame to frame for each subcarrier, we can write the conditional pdf of the incomplete data as follows:

$$f(\underline{Y}|H,\underline{X}) = \prod_{d=1}^{D} f(Y^d \mid H, X^d). \qquad (12)$$

Thus, the log-likelihood function of the incomplete data is

$$\log f(\underline{Y}|H,\underline{X}) = \sum_{d=1}^{D} \log f(Y^d \mid H, X^d), \qquad (13)$$

and the log-likelihood function of the complete data is given by

$$\log f(\underline{Y},\underline{X}|H) = \sum_{d=1}^{D} \left\{ \log \frac{1}{C} f(Y^d \mid H, X^d) \right\}. \qquad (14)$$

In the conventional ML estimation, we try to find an estimate of H that maximizes $f(\underline{Y}|H)$. But since $\log f(\underline{Y}|H)$, (11), is not easy to manipulate (summation of several exponential functions), we resort to the EM algorithm, which increases the likelihood at each step. Each iterative process $p = 0, 1, 2, \ldots$ in the EM algorithm for estimating H from $\underline{Y}$ consists of the following two steps:

E-step:

$$Q(H \mid H^{(p)}) = E_{\underline{X}}\{\log f(\underline{Y},\underline{X}|H) \mid \underline{Y}, H^{(p)}\}; \qquad (15)$$

M-step:

$$\tilde{H}^{(p+1)} = \arg\max_H Q(H \mid H^{(p)}), \qquad (16)$$

where (see Appendix A)

$$Q(H \mid H^{(p)}) = \sum_{i=1}^{C} \sum_{d=1}^{D} \log\left\{ \frac{1}{C} f(Y^d \mid H, X_i) \right\} \frac{f(Y^d \mid H^{(p)}, X_i)}{C f(Y^d \mid H^{(p)})}. \qquad (17)$$

$\tilde{H}^{(p+1)}$ is the tentative estimate of H directly from (16). The final $(p+1)$st estimate of H, that is, $H^{(p+1)}$, will be obtained through additional manipulation on $\tilde{H}^{(p+1)}$. The conditional pdfs $f(Y^d|H^{(p)}, X_i)$ and $f(Y^d|H^{(p)})$ can be calculated from (10) and (11), where X_i is the ith signal in the constellation.

The value of H that maximize (17) is found as (see Appendix B) follows:

$$\tilde{H}^{(p+1)} = \left[\sum_{i=1}^{C} \sum_{d=1}^{D} |X_i|^2 \frac{f(Y^d \mid H^{(p)}, X_i)}{f(Y^d \mid H^{(p)})} \right]^{-1}$$
$$\times \left[\sum_{i=1}^{C} \sum_{d=1}^{D} Y^d X_i^* \frac{f(Y^d \mid H^{(p)}, X_i)}{f(Y^d \mid H^{(p)})} \right]. \qquad (18)$$

It should be pointed out that the above maximization problem is actually a weighted least square (LS) problem.
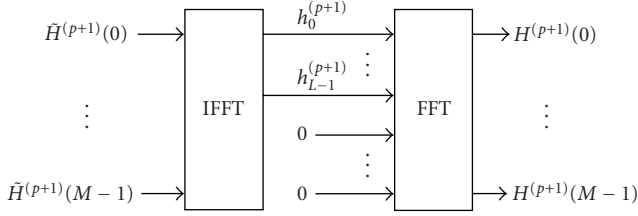
FIGURE 2: Lowpass filter structure.

In this paper, we assume that L, the delay spread in the CIR, is known. In practice, however, L is another unknown parameter. In such a case, we need to perform channel-order detection and parameter estimation. Alternatively, we may use some upper bound for L, which may be easier to obtain than trying to estimate the exact value of L. However, use of an upper bound of L would degrade the estimation performance. One obvious upper bound of L can be the length of the CP since its length is chosen to be longer than L.

The channel estimate of the form (18) obtained for the M subcarriers, which we denote $\tilde{H}^{(p+1)}(m), 0 \leq m \leq M - 1$, can be refined by taking advantage of the structure of OFDM systems and the fact that L is much smaller than M, the number of subcarriers. We will proceed as follows:

$$\underline{h}^{(p+1)} = \frac{1}{M} \mathbf{W}_L^H \underline{\tilde{H}}^{(p+1)}, \tag{19}$$

where we use the notation defined in Section 3.3 for mathematical simplification and $\mathbf{W}_L$ is an $M \times L$ matrix:

$$\mathbf{W}_L = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & e^{-j2\pi \frac{1}{M}} & \cdots & e^{-j2\pi \frac{L-1}{M}} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-j2\pi \frac{M-1}{M}} & \cdots & e^{-j2\pi \frac{(L-1)(M-1)}{M}} \end{bmatrix}_{M \times L} . \tag{20}$$

Finally, we can obtain the $(p+1)$st estimate of the channel frequency response as follows:

$$\underline{H}^{(p+1)} = \mathbf{W}_L \underline{h}^{(p+1)}. \tag{21}$$

The above procedure can be simply realized by applying the IFFT followed by the FFT, as schematically shown in Figure 2. The values $h_l^{(p+1)}, L \leq l \leq M - 1$, obtained by the IFFT must be set to zero before performing the FFT. The reason is to eliminate the estimation noise from paths that do not actually exist.

The iterative procedure should be terminated as soon as the difference between $\underline{H}^{(p+1)}$ and $\underline{H}^{(p)}$ is sufficiently small, since at this point, $\underline{H}^{(p)}$ has presumably converged to the estimate we are seeking. Once the frequency-domain channel response $\hat{H}$ is found, the ML estimate of the transmitted signal

can be obtained by solving

$$\hat{X}(m) = \arg \min_{X \in C} | Y(m) - \hat{H}(m)X(m)|^2, \quad 0 \leq m \leq M - 1, \tag{22}$$

which leads to the final estimates of the transmitted signals as follows:

$$\hat{X}(m) = \text{Quantization} \left\{ \frac{Y(m)}{\hat{H}(m)} \right\}, \quad 0 \leq m \leq M - 1. \tag{23}$$

For a constant modulus signal, for example, a PSK modulation signal $|X(m)|^2 = A$ for all m, where A is a positive constant. Thus, we can simplify (18) as follows:

$$\tilde{H}^{(p+1)} = (CDA)^{-1} \times \left[\sum_{i=1}^{C} \sum_{d=1}^{D} Y^d X_i^* \frac{f(Y^d | H^{(p)}, X_i)}{f(Y^d | H^{(p)})} \right]. \tag{24}$$

Notice that only the noise variance σ^2 is used to calculate $f(Y^d | H^{(p)}, X_i)$ in this algorithm. Any other statistical information about the channel is not necessary. Moreover, in Section 5, we will show that the accuracy of σ^2 will not affect the performance very much. Thus, this algorithm is fairly robust to the noise variance.

3.3. Algorithm 2: estimating the transmitted signals

In this algorithm, we try to improve the performance of the detection accuracy of the transmitted signal $X^d(m), 0 \leq m \leq M - 1, 1 \leq d \leq D$, as well as the CIR from the observation $Y^d(m), 0 \leq m \leq M - 1, 1 \leq d \leq D$, using the EM algorithm. To simplify the expressions, we use $\underline{H}, \underline{h}, \underline{X}, \underline{Y}, \underline{N}$ to denote the vectors of frequency-domain CIR, time-domain CIR, modulated input data, output data, and white Gaussian noise, respectively, where $\underline{h} = [h_0, \ldots, h_{L-1}]^T$, $\underline{X}^d = [X^d(0), \ldots, X^d(M-1)]^T$, $\underline{Y}^d = [Y^d(0), \ldots, Y^d(M-1)]^T$, $\underline{N}^d = [N^d(0), \ldots, N^d(M-1)]^T$, and $\underline{H} = \mathbf{W}_L \underline{h}$. We also use the notation $\mathbf{X}^d = \text{diag}(\underline{X}^d)$, which denotes an $M \times M$ matrix with $X(m)$ as its (m, m) entry and zeros elsewhere. The system model can be expressed in the vector form for the dth OFDM frame as follows:

$$\underline{Y}^d = \mathbf{X}^d \mathbf{W}_L \underline{h} + \underline{N}^d. \tag{25}$$

We still assume that the channel is static over the period of D frames for generality. To process the channel estimation algorithm using observed data in all D frames, we define some variables: $\underline{X} = [(\underline{X}^1)^T, \ldots, (\underline{X}^D)^T]^T$, $\underline{Y} = [(\underline{Y}^1)^T, \ldots, (\underline{Y}^D)^T]^T$, $\underline{N} = [(\underline{N}^1)^T, \ldots, (\underline{N}^D)^T]^T$, $\mathbf{X} = \text{diag}(\underline{X})$, $\mathbf{Y} = \text{diag}(\underline{Y})$, and $\mathbf{W}_{LD} = [\mathbf{W}_L, \ldots, \mathbf{W}_L]^T$ with D copies of $\mathbf{W}_L$. With this notation, the system model can be modified as follows:

$$\underline{Y} = \mathbf{X} \mathbf{W}_{LD} \underline{h} + \underline{N}. \tag{26}$$

The "incomplete" and "complete" data are defined as $(\underline{Y})$ and $(\underline{Y}, \underline{h})$, respectively. Each iterative process $p = 0, 1, 2, \ldots$ in the EM algorithm for estimating $\underline{X}$ from $\underline{Y}$ consists of the following two steps:

E-step:

$$Q(\underline{X}|\underline{X}^{(p)}) = E_{\underline{h}}\{\log f(\underline{Y},\underline{h}|\underline{X}) \mid \underline{Y},\underline{X}^{(p)}\}, \qquad (27)$$

M-step:

$$\tilde{\underline{X}}^{(p+1)} = \arg\max_{\underline{X}} Q(\underline{X}|\underline{X}^{(p)}). \qquad (28)$$

In the E-step at the $(p+1)$st iteration, we compute the expected value of $\log f(\underline{Y},\underline{h}|\underline{X})$, given $\underline{Y}$ and $\underline{X}^{(p)}$, the estimates obtained in the pth iteration. The M-step of the $(p+1)$st iteration determines the transmitted signal $\underline{X}^{(p+1)}$ that maximizes $Q(\underline{X}|\underline{X}^{(p)})$ given $\underline{X}^{(p)}$.

After some calculations (see Appendix C), we obtain the solution of (28):

$$\begin{aligned}\tilde{\underline{X}}^{(p+1)} &= \arg\max_{\underline{X}} Q(\underline{X}|\underline{X}^{(p)}) \\ &= \mathbf{C}_D^{-1}\left((\underline{h}^{(p)})^H \mathbf{W}_{LD}^H \underline{Y}\right)^T,\end{aligned} \qquad (29)$$

where

$$\mathbf{C}_D = \mathrm{diag}(\mathbf{C},\dots,\mathbf{C})_{MD\times MD}, \qquad (30)$$

$$\mathbf{C} = \mathrm{diag}\left(C_0,\dots,C_{M-1}\right), \qquad (31)$$

$$C_m = \sum_{k=0}^{L-1}\sum_{n=0}^{L-1} e^{j2\pi((k-n)m/M)}\left(\mathbf{\Sigma}^{(p)}(k,n) + h_k^{(p)*} h_n^{(p)}\right), \qquad (32)$$

$$\underline{h}^{(p)} = \mathbf{\Sigma}^{(p)}\left(\frac{\mathbf{W}_{LD}^H (\mathbf{X}^{(p)})^H \underline{Y}}{\sigma^2} + \mathbf{\Sigma}^{-1} E\{\underline{h}\}\right), \qquad (33)$$

$$\mathbf{\Sigma}^{(p)} = \left(\frac{\mathbf{W}_{LD}^H (\mathbf{X}^{(p)})^H \mathbf{X}^{(p)} \mathbf{W}_{LD}}{\sigma^2} + \mathbf{\Sigma}^{-1}\right)^{-1}. \qquad (34)$$

$\underline{h}^{(p)}$ and $\mathbf{\Sigma}^{(p)}$ are called the estimated posterior mean and posterior covariance matrix at the pth iteration. Therefore, in each iteration, the updated estimation of CIR $\underline{h}^{(p)}$ is obtained automatically as a by-product. After quantizing $\tilde{\underline{X}}^{(p+1)}$, we obtain the $(p+1)$st estimate

$$\underline{X}^{(p+1)} = \mathrm{Quantization}\{\tilde{\underline{X}}^{(p+1)}\}. \qquad (35)$$

The limitation of this algorithm is that the mean $E\{\underline{h}\}$ and the covariance matrix $\mathbf{\Sigma}$ of time-domain CIR are also assumed to be known. In a practical situation, these channel statistics may not be known. Fortunately, as we examine (33) and (34), we find that when σ^2 is small (i.e., SNR is high), the contribution of $\mathbf{\Sigma}^{-1}$ and $\mathbf{\Sigma}^{-1}E\{\underline{h}\}$ is so small that we can eliminate them and still expect similar performance. Furthermore, for an MPSK modulated signal, that is, $|X(m)|^2 = A$ for all m, the signal estimation can be performed by using only the phase information. Thus, we can simplify (35) to

$$\underline{X}^{(p+1)} = \mathrm{Quantization}\left\{\left(\underline{Y}^H \mathbf{X}^{(p)} \mathbf{W}_{LD} \mathbf{W}_{LD}^H \underline{Y}\right)^T\right\}. \qquad (36)$$

Consequently, only multiplication and addition operations are required. Furthermore, $\mathbf{W}_{LD}\mathbf{W}_{LD}^H$ can be calculated and stored ahead of time. Thus, the computational complexity is considerably reduced for the high SNR case.

A closer examination of (36) reveals that the simplified Algorithm 2 is a combination of ML channel estimation assuming $\underline{X}^{(p)} = \underline{X}$ and ML signal detection assuming $\underline{h}^{(p)} = \underline{h}$. This has been proposed in [17] in a different context. To conclude, Algorithm 2 is the extension of the iterative ML channel estimation algorithm when we take advantage of the channel statistics. The corresponding simplified algorithm is the same as the iterative ML channel estimation algorithm.

3.4. Algorithm 3: estimating the channel impulse response

In this section, we try to estimate the time-domain channel response by applying the parameter estimation algorithm proposed by Feder and Weinstein [24] for the general estimation problem based on the EM algorithm. We still assume that the channel is static over the period of D frames for generality. The system model used here is the same as the previous algorithm stated in (26). We define $\mathbf{A} = \mathbf{X}\mathbf{W}_{LD}$ which is a $MD \times L$ matrix, and rewrite the system model as follows:

$$\underline{Y} = \mathbf{A}\underline{h} + \underline{N} = \sum_{i=0}^{L-1}\mathbf{A}_i h_i + \underline{N}, \qquad (37)$$

where $\mathbf{A}_i$ is the ith column of the matrix $\mathbf{A}$. Note from (37) that each element of $\underline{Y}$, $Y(m)$, consists of L superimposed signals and AWGN which can be represented by

$$Y(m) = \sum_{i=0}^{L-1} a_i(m) h_i + N(m), \quad 0 \le m \le MD - 1. \qquad (38)$$

Following [24], a natural choice for the "complete" data $\underline{Z}_m$ is defined by decomposing the observed data $Y(m)$ into L components, that is, $\underline{Z}_m = [Z_0(m),\dots,Z_{L-1}(m)]^T$, where

$$Z_i(m) = a_i(m) h_i + N_i(m), \quad 0 \le m \le MD - 1. \qquad (39)$$

Here, $a_i(m)$ is the (m,i)th entry of the matrix $\mathbf{A}$ and $N_i(m)$, $0 \le i \le L-1$, are obtained by arbitrarily decomposing the total noise $N(m)$ into L components such that

$$\sum_{i=0}^{L-1} N_i(m) = N(m). \qquad (40)$$

Thus, the relation between the "complete" data $\underline{Z}_m$ and "incomplete" data $Y(m)$ is given by

$$Y(m) = \sum_{i=0}^{L-1} Z_i(m). \qquad (41)$$

It is convenient to choose the $N_i(m)$ to be statistically independent Gaussian random variables with zero mean and variance σ_i^2, where

$$\sigma^2 = \sum_{i=0}^{L-1} \sigma_i^2. \qquad (42)$$

The EM-based algorithm is used here to obtain an estimation of $\underline{h}$ that maximizes $f(\underline{Y}|\underline{h})$. The "incomplete" and

"complete" data for mth element of $\underline{Y}$, as stated before, are $(Y(m))$ and $(\underline{Z}_m)$, respectively. We then group all $\underline{Z}_m$ for all D OFDM frames and all M subcarriers into a new vector $\underline{Z} = [\underline{Z}_0^T, \ldots, \underline{Z}_{MD-1}^T]^T$. Each iterative process $p = 0, 1, 2, \ldots$ in the EM algorithm for estimating $\underline{h}$ from $\underline{Y}$ consists of the following two steps:

(i) E-step:

$$Q(\underline{Z} \mid \underline{h}^{(p)}) = E_{\underline{Z}} \{ \log f(\underline{Z} \mid \underline{h}) \mid \underline{Y}, \underline{h}^{(p)} \}, \qquad (43)$$

(ii) M-step:

$$\underline{h}^{(p+1)} = \arg \max_{\underline{h}} Q(\underline{Z} \mid \underline{h}^{(p)}). \qquad (44)$$

In the E-step at the $(p+1)$st iteration, we compute the expected log-likelihood function $\log f(\underline{Z} \mid \underline{h})$, given $\underline{Y}$ and $\underline{h}^{(p)}$, the estimates obtained in the pth iteration. The M-step of the $(p + 1)$st iteration determines the transmitted CIR $\underline{h}^{(p+1)}$ that maximizes $Q(\underline{Z} \mid \underline{h}^{(p)})$.

After some calculation (see Appendix D), we obtain the solution of (44):

$$h_i^{(p+1)} = \frac{1}{MD} \sum_{m=0}^{MD-1} \frac{\hat{Z}_i^{(p+1)}(m)}{a_i(m)}, \quad 0 \le i \le L - 1, \qquad (45)$$

where

$$\hat{Z}_i^{(p)}(m) = Z_i^{(p)}(m) + \beta_i \left(Y(m) - \sum_{j=0}^{L-1} Z_j^{(p)}(m) \right), \qquad (46)$$

$$\sum_{i=0}^{L-1} \beta_i = 1, \quad \beta_i \ge 0, \qquad (47)$$

$$Z_i^{(p)}(m) = a_i(m) h_i^{(p)}. \qquad (48)$$

Observe that β_i, the ith decomposition factor, can be arbitrarily selected with the constraint (47) due to the arbitrary selection of the independent noise components $N_i(m)$. Different sets of β_i will give different system performance and we will discuss the selection of β_i with simulation results in the next section.

Note that the elements of $\mathbf{A} = \mathbf{X} \mathbf{W}_{LD}$ are dependent on the transmitted signals $\underline{X}$. However, we do not know all these transmitted signals in the OFDM frames except for some pilot symbols. Thus, in order to proceed, we adopt the pth estimates $\underline{X}^{(p)}$ instead of the actual values (which are unknown) to calculate the matrix $\mathbf{A}$. In this case, the elements of $\underline{X}^{(p)}$ are given by

$$X^{(p)}(m) = \text{Quantization} \left(\frac{Y(m)}{\mathbf{W}_m \underline{h}^{(p)}} \right), \quad 0 \le m \le MD - 1, \qquad (49)$$

where $\mathbf{W}_m$ is the $(m + 1)$st row of matrix $\mathbf{W}_{LD}$.

Notice that we do not need any information about the channel in this algorithm except the choice of the set β_i. However, we can always choose $\beta_i = 1/L$ which will give near optimum performance as demonstrated in the simulation results. Thus, this algorithm is also very robust.

3.5. Initialization

As is known from the general convergence property of the EM algorithm, there is no guarantee that the iterative steps converge to the global maximum. For a likelihood function with multiple local maxima, the convergence point may be one of these local maxima, depending on the initial estimates $\underline{H}^{(0)}$, $\underline{X}_0$, and $\underline{h}^{(0)}$. Therefore, we propose to use pilot symbols distributed at certain locations in the OFDM time-frequency lattices to find appropriate initial values of $\underline{H}^{(0)}$, $\underline{X}_0$, and $\underline{h}^{(0)}$ if there are pilot symbols inserted in the current OFDM frame. On the other hand, if there is no pilot symbol, we just set the initial channel estimates of the current OFDM frame as the final channel estimates of the previous OFDM frame assuming the channel is changing slowly. This is more likely to lead us to the true maximum point, as can be observed in the numerical results. Another benefit of this selection of the initial estimates of the CIR is that we do not need to do time-domain filtering or interpolation. Thus, we can considerably reduce the detection latency since we can carry out channel estimation and signal detection procedures as soon as we have received signals for each OFDM frame.

For those OFDM frames with pilot symbols, we define the pilot position set $S = \{s_1, \ldots, s_{|S|}\}$. The corresponding FFT matrix only with those rows belonging to S is denoted as $\mathbf{W}_S$. Thus, we use the simple LS algorithm to obtain the channel frequency response [8] at each pilot position by

$$\tilde{H}^{(0)}(s_i) = \frac{Y(s_i)}{X(s_i)}, \quad 0 \le i \le |S|. \qquad (50)$$

Then, we apply the IFFT on $\tilde{H}^{(0)}(s_i), \ldots, \tilde{H}^{(0)}(s_{|S|})$ and obtain the initial CIR by

$$\underline{h}^{(0)} = \frac{1}{M} \mathbf{W}_S^H \tilde{\underline{H}}^{(0)}, \qquad (51)$$

where $\tilde{\underline{H}}^{(0)} = [\tilde{H}^{(0)}(s_1), \ldots, \tilde{H}^{(0)}(s_{|S|})]^T$. Next, we apply the FFT on $\underline{h}^{(0)}$ and obtain the initial estimates of the channel frequency response for all subcarriers as $\underline{H}^{(0)} = \mathbf{W}_L \underline{h}^{(0)}$. Finally, the initial estimates of the transmitted signals are obtained from

$$X^{(0)}(m) = \text{Quantization} \left\{ \frac{Y(m)}{H^{(0)}(m)} \right\}, \quad 0 \le m \le M - 1. \qquad (52)$$

4. CRAMER-RAO LOWER BOUND

The CRLB is an important criterion to evaluate how good any unbiased estimator can be since it provides the MMSE bound among all unbiased estimators. In this section, we will derive the CRLB for the CIR in OFDM systems. In Section 5, we will show the performance of the three proposed EM-based channel estimation algorithms and compare it to the CRLB. We note that in [25], Morelli and Mengali discuss the CRLB for channel estimators in OFDM, but they only treat PSK modulation in their discussion. We will discuss below the modified and averaged CRLB for the CIR with nonconstant modulus modulation.

The CRLB for the channel estimation is given by (see Appendix E)

$$\text{CRLB}(\underline{h}) = \text{trace}\left(I^{-1}(\underline{h})\right), \tag{53}$$

where

$$I(\underline{h}) = \frac{1}{2\sigma^2}\mathbf{W}_L^H \sum_{d=1}^{D} (\mathbf{X}^d)^H \mathbf{X}^d \mathbf{W}_L. \tag{54}$$

Clearly, the CRLB changes from a set of D frames to another due to the different sets of transmitted signals. We define the average CRLB [26] denoted $\overline{\text{CRLB}}(\underline{h})$ as follows:

$$\overline{\text{CRLB}}(\underline{h}) = E\{\text{CRLB}(\underline{h})\}, \tag{55}$$

where the expectation is carried out with respect to the transmitted data $\underline{X}$ in D frames.

Another CRLB is called the modified CRLB [27], denoted by MCRB. It is defined as

$$\begin{aligned}
\text{MCRB}(\underline{h}) &= \sum_{i=0}^{L-1} \frac{1}{E\{I(\theta)_{ii}\}} \\
&= \frac{2L\sigma^2}{ME\left\{\sum_{d=1}^{d=D}|X^d|^2\right\}} \\
&= \frac{2L\sigma^2}{MD}\frac{1}{E\{|X^d|^2\}}.
\end{aligned} \tag{56}$$

We note that we use M to denote the number of subcarriers in this paper. It also could be the number of effective subcarriers which exclude the null subcarriers as the guard frequency band. Of course, in the presence of null subcarriers, we have to make some modifications on $\mathbf{W}_L$ by deleting those rows corresponding to the null subcarriers.

It is easy to show that $\overline{\text{CRLB}}(\underline{h}) \geq \text{MCRB}(\underline{h})$ by simply applying the Cauchy-Schwarz inequality. This is equivalent to saying that the $\overline{\text{CRLB}}(\underline{h})$ is always tighter than the MCRB($\underline{h}$) [27]. We will discuss the specific CRLB for constant and nonconstant modulus signals in the following.

4.1. CRLB for constant modulus signals

For constant modulus signals, $|X^d(m)|^2 = A$ for all d's and m's (for instance, PSK modulated signals). Thus, we can simplify (53) as follows:

$$\text{CRLB}(\underline{h}) = \frac{2L\sigma^2}{MDA}. \tag{57}$$

It is obvious that the above CRLB is inversely proportional to the number of observed OFDM frames D, number of subcarriers M, and SNR $A/2\sigma^2$. Note that CRLBs of different frames for OFDM channel estimation are constant and do not depend on the channel response $\underline{H}$ or $\underline{h}$. Consequently, this CRLB can be applied to any multipath fading channel. Another important observation is that

$$\text{CRLB}(\underline{h}) = \overline{\text{CRLB}}(\underline{h}) = \text{MCRB}(\underline{h}) \tag{58}$$

in the case of constant modulus signals.

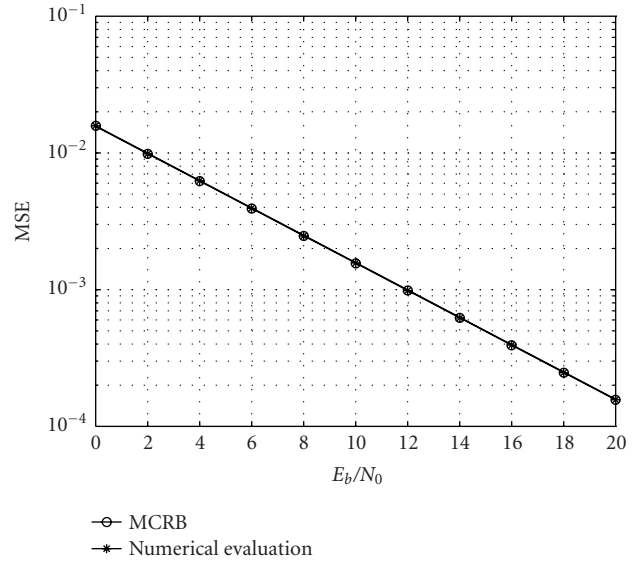

Figure 3: Analytical and numerical evaluation of MCRB($\underline{h}$) with 16 QAM modulated signals for each subcarrier.

4.2. CRLB for nonconstant modulus signals

For nonconstant modulus signals, $|X^d(m)|^2$ is no longer constant (e.g., 16 QAM modulated signals). Thus, the CRLB in this case changes from D frames to another D frames. In addition, it is not straightforward to obtain an explicit expression for the $\overline{\text{CRLB}}(\underline{h})$ because $I(\underline{h})$ can no longer be easily inverted. However, the MCRB($\underline{h}$) can be computed assuming the transmitted signals are independent. This results in

$$E\left\{(\underline{X}^d)^H \underline{X}^d\right\} = AI_{M\times M}. \tag{59}$$

Thus, the MCRB($\underline{h}$) can be calculated as follows:

$$\text{MCRB}(\underline{h}) = \frac{2L\sigma^2}{MDA}, \tag{60}$$

which is the same as the constant modulus CRLB in the case of the same average signal energy A. Figure 3 shows the theoretical curve of MCRB($\underline{h}$) and the numerically evaluated curve of 16 QAM signals. These two curves agree and justify the use of MCRB($\underline{h}$) as a performance measure for unbiased channel estimation algorithms in OFDM systems, both for constant modulus and nonconstant modulus signals.

5. SIMULATION AND DISCUSSION

We constructed an OFDM simulation model, which is similar to the specifications of 802.11a, to demonstrate the validity and effectiveness of the EM-based channel estimation and signal detection algorithms. The entire channel bandwidth is 800 kHz, and is divided into 64 subcarriers (or tones). To make the tones orthogonal to each other, the symbol duration is chosen as 80 microseconds. An additional 20 microseconds CP ($N_{\text{cp}} = 16$) is used to provide protection from IFI and ICI due to channel delay spread. Thus, the total

OFDM frame length is $T_s = 100$ microseconds and sub-channel symbol rate is 10 kbaud. The modulation scheme used in the system is QPSK. One OFDM frame out of 8 OFDM frames ($N_t = 8$) has pilot symbols and 8 pilot symbols ($N_f = 8$) are inserted into such a frame with equal space, where N_t and N_f denote the pilot spacing along the frequency and time domains, respectively. Thus, the overhead caused by pilot symbols is only 1/64. The simulated system can transmit uncoded data at 1.28 Mbps. The maximum Doppler frequency f_d is chosen to be 100 Hz, which implies $f_d T_s = 0.01$. The CIRs used in the simulations are given by

$$h_1(n) = 0.8\alpha_0\delta(n) + 0.6\alpha_1\delta(n-1),$$

$$h_2(n) = \frac{1}{C_2}\sum_{k=0}^{4} e^{-k}\alpha_k\delta(n-k),$$

$$h_3(n) = \frac{1}{C_3}\sum_{k=0}^{7} e^{-k/2}\alpha_k\delta(n-k),$$

(61)

where $C_2 = \sqrt{\sum_{k=0}^{4} e^{-2k}}$ and $C_3 = \sqrt{\sum_{k=0}^{7} e^{-k}}$ are the normalization constants and α_k, $0 \le k \le 7$, are independent complex-valued Gaussian random variables with unit variance, which vary in time according to the Doppler frequency. The amplitude of α_k are Rayleigh distributed. This is a conventional exponential decay multipath channel model. We set the stopping criterion as $\|\underline{h}^{(p+1)} - \underline{h}^{(p)}\|^2 \le 10^{-3}$.

5.1. Simulation results of Algorithm 1

The channel model we use to test the performance of Algorithm 1 is $h_3(n)$. Since we normalize the average channel power, the BER performance of different channel models should be the same. However, the MSE is proportional to the channel length L as shown in (57). For those OFDM frames containing pilot symbols, the initial estimate of CIR is obtained by using these 8 equally spaced pilot symbols. For those OFDM frames without pilot symbols, the initial estimate of CIR comes from the channel estimate of the previous OFDM frame.

From Figures 4 and 5, we observe that the EM-based Algorithm 1 reduces the BER and MSE simultaneously. Furthermore, the BER can achieve performance close to the known channel case and the MSE can almost achieve the CRLB in the high SNR region. For example, the MSE is very close to the CRLB when $E_b/N_0 > 14$ dB, which is a very favorable result since we only sacrifice 1/64 spectral efficiency ignoring the effect caused by the CP. One drawback of the algorithm is that the BER cannot be improved from the initial estimate when SNR is low. It is clear from Figure 6 that the algorithm needs more iterations in the low SNR region than in the high SNR region for the iterative procedure to converge. Indeed, for low SNR case, the BER may increase after a few iterations, while the MSE still decreases from the initial value. That is because the EM algorithm is used to obtain the true values of the CIR and better estimates of CIR (less MSE) do not necessarily lead to lower BER. Therefore, this algorithm is practical only when SNR is large. We see that the number of necessary iterations decreases rapidly as E_b/N_0
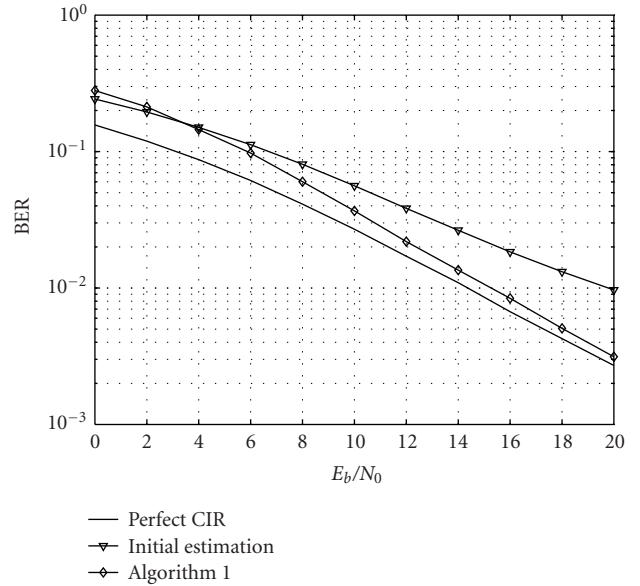

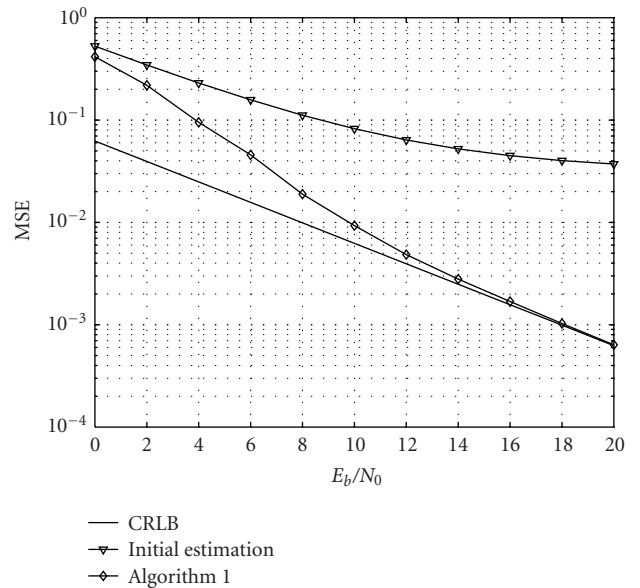FIGURE 4: BER versus E_b/N_0 in the 8-path channel model using Algorithm 1.

FIGURE 5: MSE versus E_b/N_0 in the 8-path channel model using Algorithm 1.

increases. When $E_b/N_0 = 20$ dB, for instance, only three or four iterations are needed to achieve the convergence in the 8-path channel. It turns out that the number of iterations does not depend on the channel delay spread L, which is not illustrated here.

For this algorithm, we need to know the Gaussian noise variance σ^2 in order to compute $f_i(Y^d|H^{(p)}(m))$ in each iteration. In practice, the noise variance is not known directly at the receiver and the error in the noise variance estimate will degrade channel estimation accuracy. We performed such a
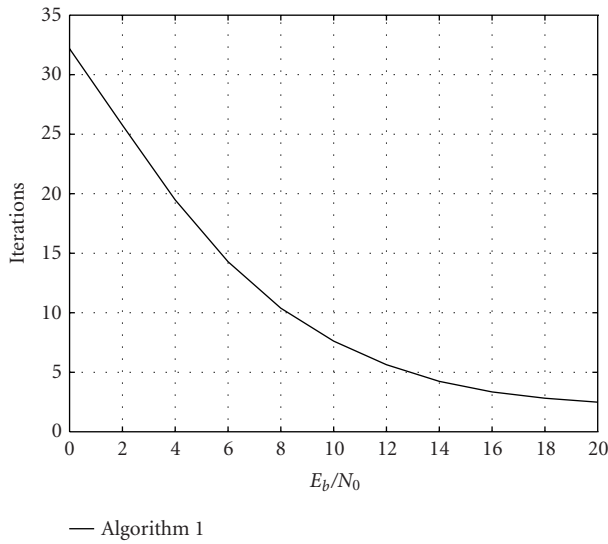
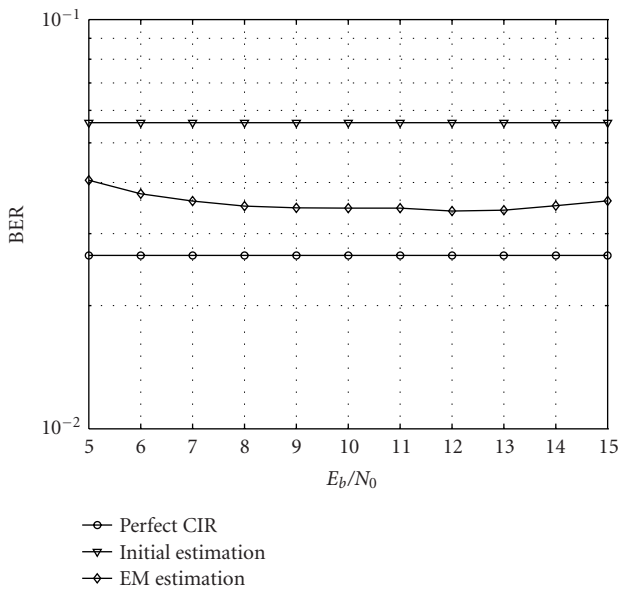FIGURE 6: The number of iterations versus E_b/N_0 in the 8-path channel model using Algorithm 1.

FIGURE 7: The effect of noise variance error on the system performance. The exact E_b/N_0 is 10 dB.
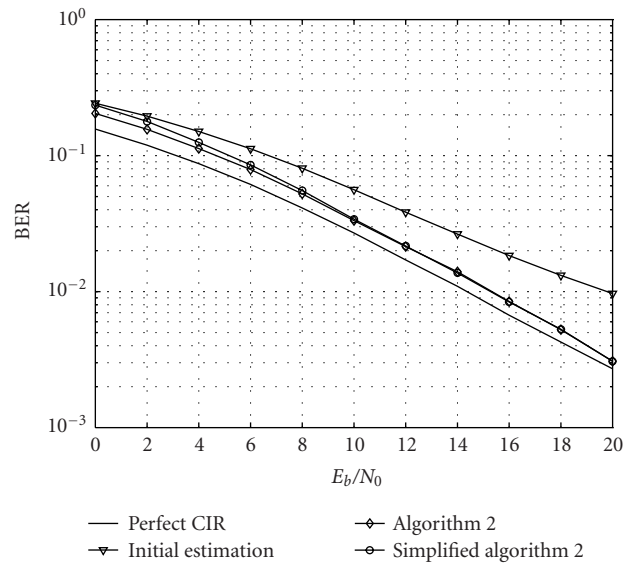

FIGURE 8: BER versus E_b/N_0 in the 8-path channel model using Algorithm 2.
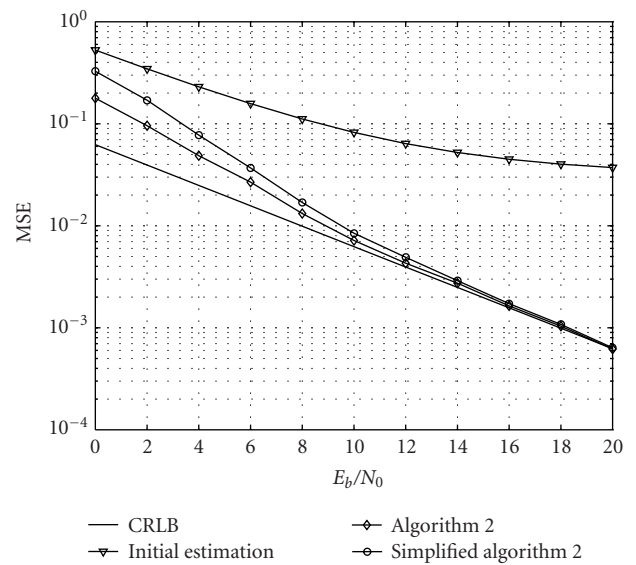
FIGURE 9: MSE versus E_b/N_0 in the 8-path channel model using Algorithm 2.

simulation to illustrate this effect. This is shown in Figure 7. The exact E_b/N_0 of the system is 10 dB and the horizontal axis is the E_b/N_0 we adopted in the EM-based algorithm. From this figure, it is seen that the effect of noise variance error is relatively small on the system performance (BER) when the noise variance error is within -2 dB and 3 dB. Therefore, we can use the following method to estimate the Gaussian noise variance on the fly with only negligible effect on the system performance:

$$\hat{\sigma}^2 = \frac{1}{M} \sum_{m=0}^{M} \left\| Y(m) - \hat{H}(m)\hat{X}(m) \right\|^2. \tag{62}$$

The E_b/N_0 computed by the above equation is about 11 dB by using the initial estimates of the CIR and transmitted signals. In this way, the performance degradation caused by using the estimated noise variance would be relatively small.

5.2. Simulation results of Algorithm 2

The channel model we use to test the performance of Algorithm 2 is still $h_3(n)$. Figure 8 shows the BER performance of EM-based Algorithm 2 and Figure 9 displays the corresponding MSE. The initial channel estimates are obtained using the same method stated above. In the EM-based Algorithm 2, we use the estimate of the previous OFDM frame as the
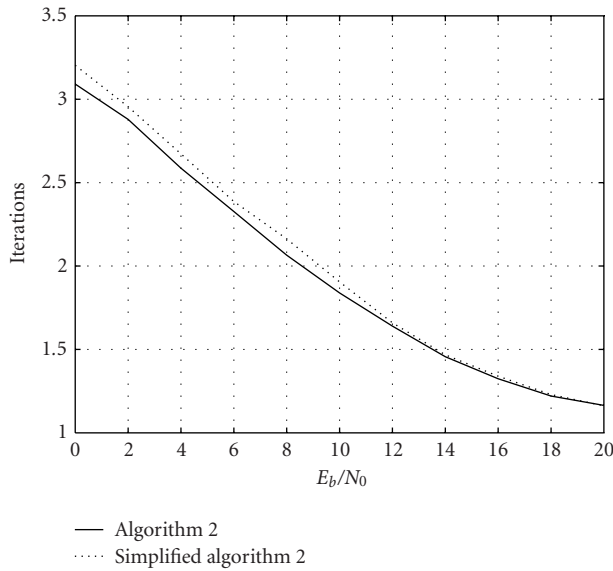
FIGURE 10: The number of iterations versus E_b/N_0 in the 8-path channel model using Algorithm 2.

FIGURE 11: BER versus E_b/N_0 in the 8-path channel model using Algorithm 3.

FIGURE 12: MSE versus E_b/N_0 in the 8-path channel model using Algorithm 3.

initial value for the current OFDM frame if there are no pilot symbols in the current frame. From these two figures, we can see that the EM-based Algorithm 2 can also achieve almost as good performance as the ideal case in terms of BER, where the channel characteristics are completely known and $f_d Ts = 0.01$, that is, the channel does not change very fast. Furthermore, the MSE of the EM-based channel estimation 2 converges to the CRLB when SNR becomes large. Another interesting result obtained from our simulation is that the performance degradation is quite small when we use the simplified Algorithm 2 that does not use the channel statistics. Degradation occurs only when the E_b/N_0 is less than 10 dB. Thus, this algorithm is well suited in practical situations.

In Figure 10, we plot the number of iterations required for the estimates $\underline{X}_p$ to converge versus E_b/N_0 at the receiver input. We see that the numbers of necessary iterations for both the simplified and nonsimplified algorithms are relatively small for a broad range of SNR. And the simplified algorithm causes only a very small increase in the number of iterations required to converge. This demonstrates that the algorithm can achieve a substantial performance improvement with only a modest increase in the computational complexity (details are in Section 5.5). This is due to the additional computation for the iterations. Furthermore, it turns out that the number of iterations does not depend on the channel length L, which is the same as Algorithm 1. Here, however, there are fewer iterations compared to Algorithm 1.

5.3. Simulation results of Algorithm 3

The channel model we use to test the performance BER and MSE of Algorithm 3 is also $h_3(n)$. However, we tested all three channel models to study the effect of the number of iterations needed to converge. Figure 11 shows the BER performance of Algorithm 3 and Figure 12 displays the corresponding MSE.
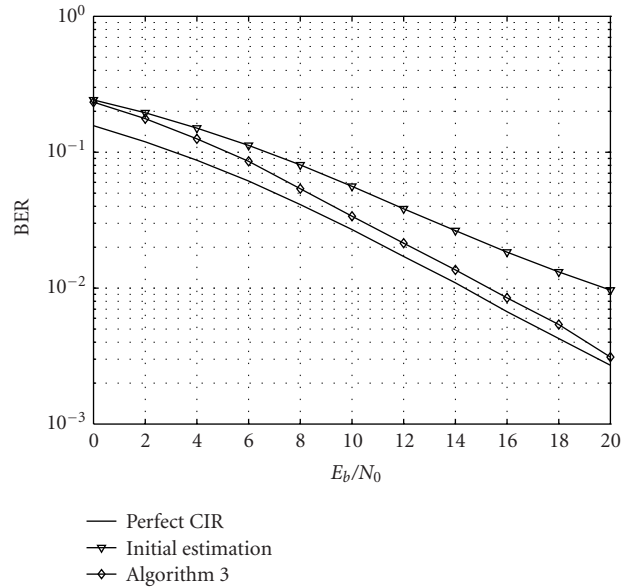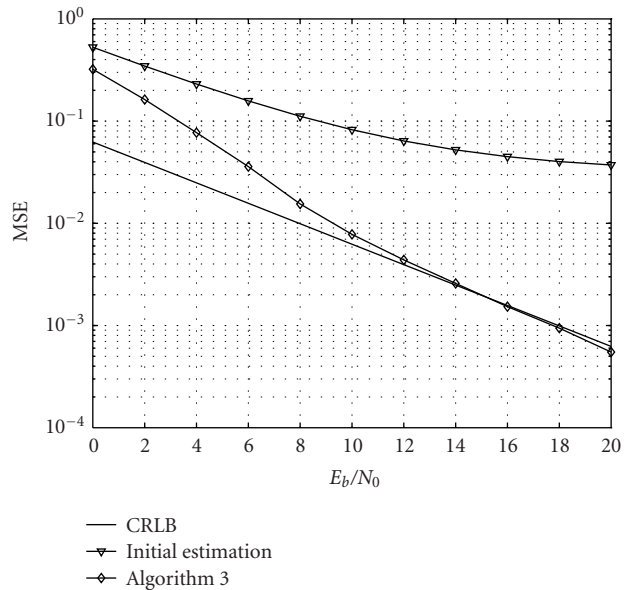
Similar conclusions can be drawn about the performance of BER and MSE as with Algorithms 1 and 2.

Figure 13 shows the relationship between the number of iterations needed for convergence versus the channel delay spread L for different E_b/N_0. From this figure, we observe that the number of iterations decreases when E_b/N_0 increases; the number of iterations increases when the channel delay spread increases under the same E_b/N_0. Furthermore, we see that the number of iterations is almost proportional to the channel delay spread L; for example, when L doubles, the number of
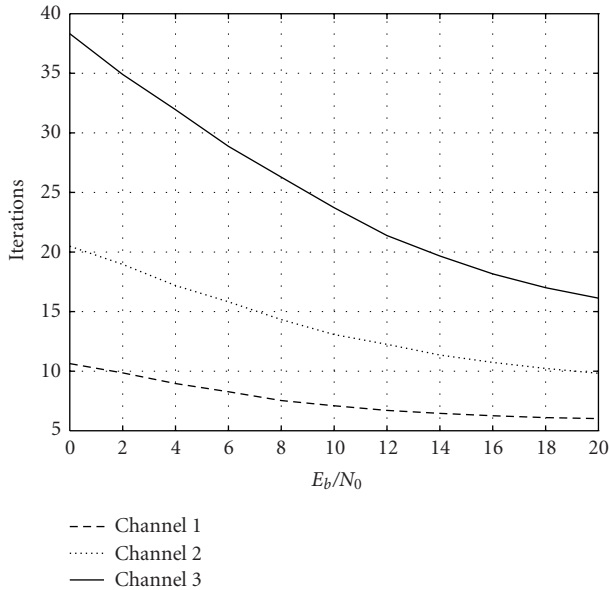
FIGURE 13: The number of iterations versus E_b/N_0 in the 8-path channel model using Algorithm 3.
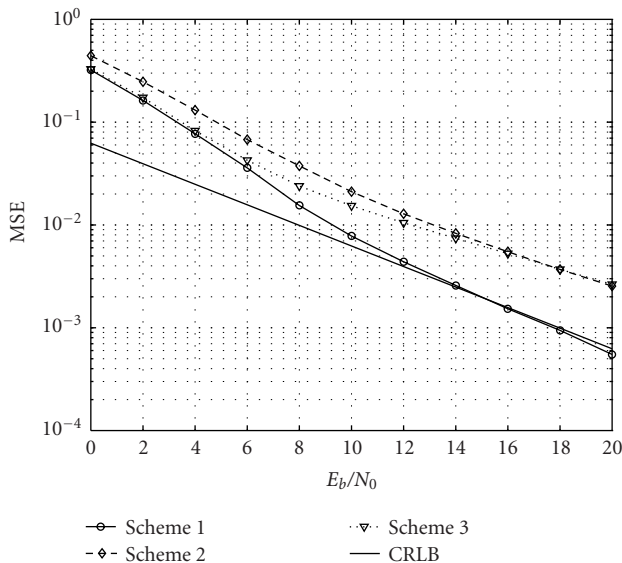

FIGURE 14: MSE versus E_b/N_0 in the 8-path model using Algorithm 3. Three different schemes of the set β_i are compared.

iterations approximately doubles. Therefore, this algorithm is more suitable for the case of small channel delay spread.

Figure 14 shows the MSEs using different sets of β_i over the 8-path channel. Scheme 1 corresponds to $\beta_i = 1/8$. Scheme 2 corresponds to $\beta_i = E\{h_i^2\} = e^{-i}/C_3$. Scheme 3 corresponds to β_i equal to the energy of ith path obtained in each iteration. It changes from iteration to iteration. From this figure, we find that Scheme 1 has the best MSE performance and the other two schemes have larger MSE, especially in the large E_b/N_0 region. Furthermore, Scheme 1 is simpler than the other two since it only needs to know L, while Scheme 2 has to know the average energy of each path of the

channel and Scheme 3 has additional computation in each iteration. Based on these preliminary simulations, it appears that Scheme 1 has the best performance without additional knowledge of the channel and additional computation.

5.4. Performance in different time-varying channels

Since the three algorithms show similar performance patterns, we choose Algorithm 2 to display the channel estimation performance in various time-varying channels with $f_d T = 0.01, 0.03$, and 0.05, respectively. Some interesting observations can be noted from Figures 15 and 16. As $f_d T$ increases, that is, the channel varies faster, the performance of our algorithms degrades. However, the degradation is not so significant, especially when $f_d T \leq 0.03$. The pilot pattern definitely affects the channel estimation and system performance [10]. However, our algorithms show similar performance with different number of pilot symbols in the frequency domain as long as $N_f \geq M/L$. Most performance degradation comes from the time-varying nature of the channel. For those fast time-varying channels (e.g., $f_d T = 0.05$), reducing the pilot spacing along the time domain will improve the system performance as well as channel estimation accuracy. But, for slowly and moderately time-varying channels (e.g., $f_d T = 0.01$ and 0.03), it does not help much as our algorithms have already achieved near-optimal performance. Thus, our algorithms are able to achieve near-optimal performance by using very few pilot symbols both for slowly and fast time-varying channels.

5.5. Comparison with existing methods

In order to compare the performance with existing methods, we choose the method with MMSE in the frequency domain [8] and linear interpolation (LI) in the time domain [7] (MMSE + LI) as an example, since MMSE estimation in the frequency domain demonstrates a good and robust performance. The reason to use LI in the time domain is to make the demodulation latency and necessary memory requirement as small as possible. For example, if $N_t = 8$, the largest demodulation latency is 7 OFDM frames' period and 9 OFDM frames' data must be stored in the memory. Comparing Figure 17 with Figure 15 and Figure 18 with Figure 16, it is obvious that MMSE + LI performs much worse under the same channel model and pilot pattern, especially in the high SNR region. Our algorithms are CRLB achievable, whereas MMSE + LI is not. Furthermore, there is an error floor for MSE in fast time-varying channels using MMSE + LI. In order to remove it, more pilot symbols in the time domain must be inserted. This will reduce the system spectrum efficiency. Another advantage of our algorithms is that they do not have demodulation latency except the processing time because our algorithms are based on the received signals of the current OFDM frame. MMSE + LI or other pilot-symbols-assisted channel estimation methods have some extent of demodulation latency as long as they apply some kinds of time-domain interpolation or filtering.

One major disadvantage of our algorithms is the large computational complexity due to the iterative nature. We
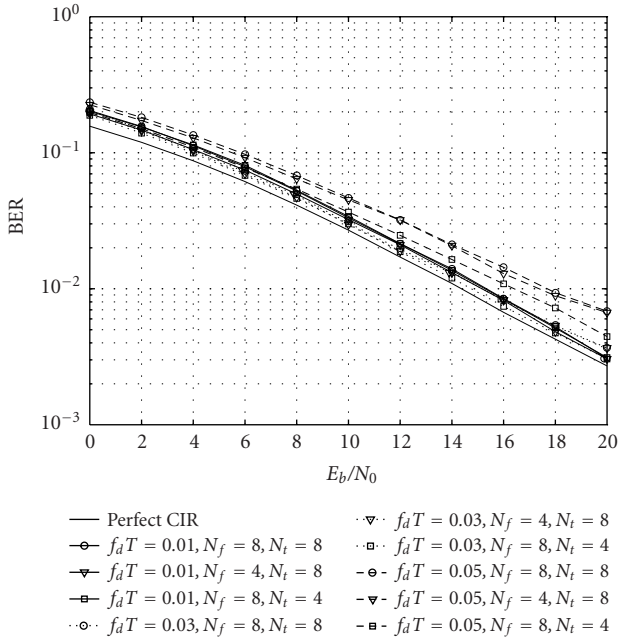
FIGURE 15: BRE through time-varying channels with different $f_d T$ of Algorithm 2.
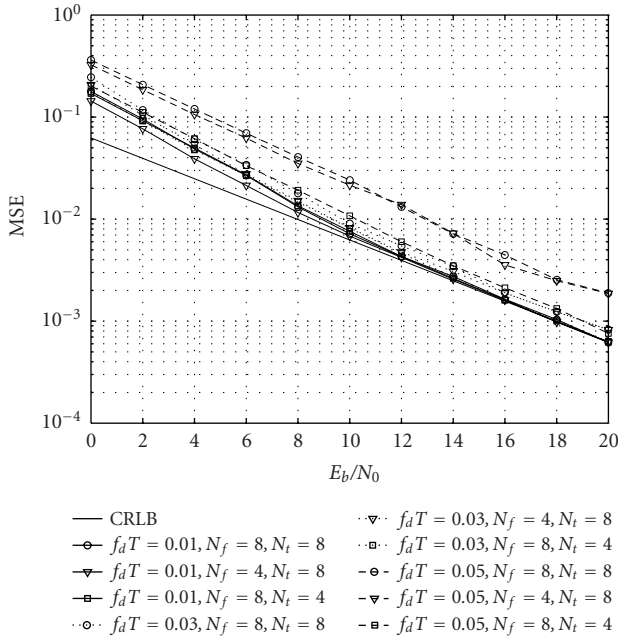

FIGURE 17: BER through time-varying channels with different $f_d T$ of MMSE + LI.

FIGURE 16: MSE of channel estimates through time-varying channels with different $f_d T$ of Algorithm 2.
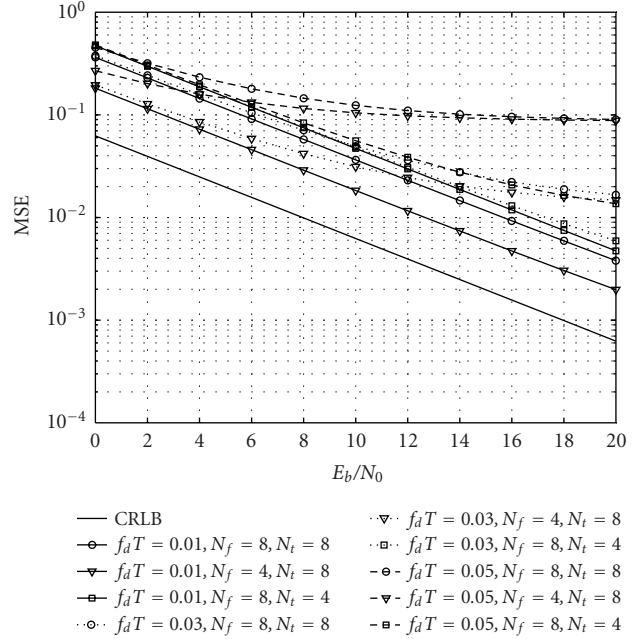
FIGURE 18: MSE of channel estimates through time-varying channels with different $f_d T$ of MMSE + LI.

select $N_t = 8$, $N_f = 8$, and $L = 8$ as the corresponding system parameters. MMSE + LI needs $M/8 + 2$ complex-variable multiplications per subcarrier. Algorithms 1, 2, and 3 need $17 + 2\log_2 M$, $10 + \log_2 M$, and $26 + \log_2 M$ complex-variable multiplications per subcarrier per iteration. Obviously, the complexity of our algorithms is larger than the
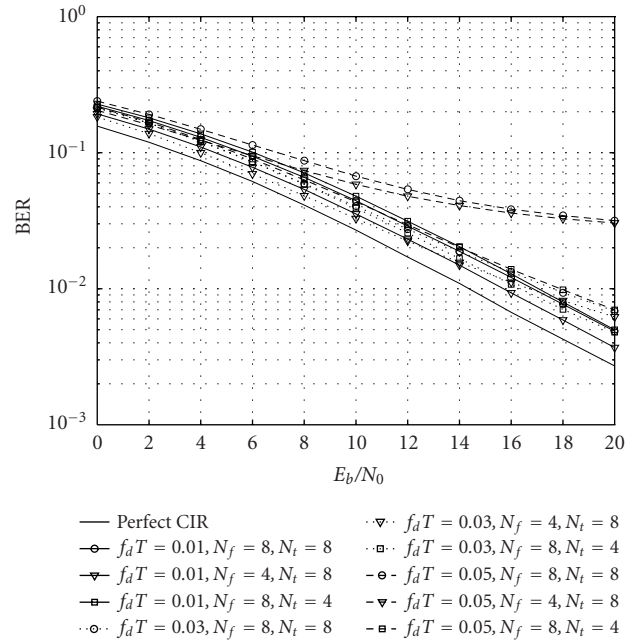
simple MMSE + LI method shown in Figure 19, especially when SNR is small. Among the three EM-based algorithms, Algorithm 2 has the least complexity and Algorithm 3 has the most. This suggests that our algorithms are suitable for a high SNR environment with comparable complexity as existing methods, while achieving much better performance.

FIGURE 19: Complexity comparison measured by complex-variable multiplications per subcarrier.

6. CONCLUSION

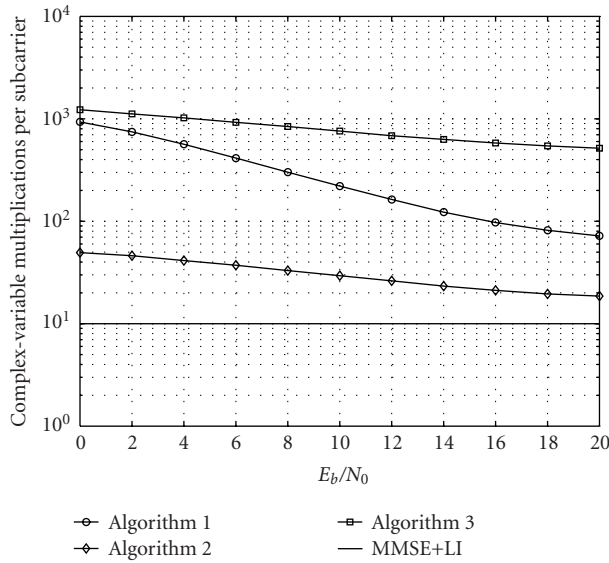In this paper, we proposed three EM-based iterative algorithms to efficiently estimate the CIR and demodulate the transmitted signals in an OFDM system. By defining different "complete" data sets for the EM algorithm, we are led to the three algorithms: (1) estimating the frequency response of the channel; (2) estimating the transmitted signals; and (3) estimating the CIR. Making use of a modest number of pilot symbols or the channel estimate of the previous OFDM frame to obtain the initial estimate, these algorithms can achieve near-optimal estimates after a few iterations. We also derived the CRLB and MCRB for the channel estimate for both constant and nonconstant modulus signals. Advantages and disadvantages of each algorithm have been discussed and illustrated by means of simulation. The simulations reveal that the first two (estimating the frequency response and demodulating the transmitted signals directly) converge at a rate independent of the multipath spread. Bypassing the channel estimate by demodulating the transmitted signals directly (Algorithm 2) is the fastest converging procedure and, thus, has the smallest overall computational burden. Algorithm 2 has the least complexity among the three and is comparable with MMSE + LI. However, the performance is much better than that of MMSE + LI, especially in the high SNR region. All three algorithms are able to work in a fast fading environment. These algorithms can easily be extended to estimate multiple-input multiple-output (MIMO) channels in MIMO-OFDM systems. Our results indicate that the performance is acceptable only when E_b/N_0 is larger than 8 dB. In the small SNR region, these algorithms need to be utilized with channel coding schemes to further improve performance.

APPENDICES

A. DERIVATION OF (17)

$$
\begin{aligned}
&Q\big(H \,\big|\, H^{(p)}\big) \\
&= E\big\{ \log f(\underline{Y}, \underline{X} | H) \,\big|\, \underline{Y}, H^{(p)} \big\} \\
&= \sum_{i=1}^{C} \sum_{d=1}^{D} \log \left\{ \frac{1}{C} f(Y^d | H, X_i) \right\} f(X_i | Y^d, H^{(p)}) \\
&= \sum_{i=1}^{C} \sum_{d=1}^{D} \log \left\{ \frac{1}{C} f(Y^d | H, X_i) \right\} \frac{f(Y^d | H^{(p)}, X_i) f(X_i | H^{(p)})}{f(Y^d | H^{(p)})} \\
&= \sum_{i=1}^{C} \sum_{d=1}^{D} \log \left\{ \frac{1}{C} f(Y^d | H, X_i) \right\} \frac{f(Y^d | H^{(p)}, X_i)}{C f(Y^d | H^{(p)})},
\end{aligned}
$$
(A.1)

where we assume X_i, $1 \le i \le C$, and $H^{(p)}$ are independent random variables.

B. DERIVATION OF (18)

$$
\begin{aligned}
&\arg \max_{H} Q\big(H \,\big|\, H^{(p)}\big) \\
&= \arg \max_{H} \sum_{i=1}^{C} \sum_{d=1}^{D} \log \{ f(Y^d | H, X_i) \} \frac{f(Y^d | H^{(p)}, X_i)}{f(Y^d | H^{(p)})} \\
&= \arg \min_{H} \sum_{i=1}^{C} \sum_{d=1}^{D} | Y^d - X_i H |^2 \frac{f(Y^d | H^{(p)}, X_i)}{f(Y^d | H^{(p)})}.
\end{aligned}
$$
(B.1)

Then differentiating the last expression with respect to H, and setting it to zero, we have

$$
\begin{aligned}
\tilde{H}^{(p+1)} &= \left[\sum_{i=1}^{C} \sum_{d=1}^{D} |X_i|^2 \frac{f(Y^d | H^{(p)}, X_i)}{f(Y^d | H^{(p)})} \right]^{-1} \\
&\quad \times \left[\sum_{i=1}^{C} \sum_{d=1}^{D} Y^d X_i^* \frac{f(Y^d | H^{(p)}, X_i)}{f(Y^d | H^{(p)})} \right].
\end{aligned}
$$
(B.2)

C. DERIVATION OF ALGORITHM 2

Equation (27) can be rewritten as follows:

$$
Q(\underline{X} | \underline{X}^{(p)}) = \int \big[\log f(\underline{Y}, \underline{h} | \underline{X}) \big] f(\underline{h} | \underline{Y}, \underline{X}^{(p)}) d\underline{h}, \quad \text{(C.1)}
$$

where the log-likelihood function can be expressed as follows:

$$
\log f(\underline{Y}, \underline{h} | \underline{X}) = \log f(\underline{Y} | \underline{h}, \underline{X}) + \log f(\underline{h} | \underline{X}). \quad \text{(C.2)}
$$

The conditional pdf $f(\underline{h} | \underline{Y}, \underline{X}^{(p)})$ is used in (C.1) to take the conditional expectation over the unknown parameters $\underline{h}$. We assume that $\underline{h}$ and $\underline{X}$ are independent of each other. This is a reasonable assumption since the CIR does not depend on the transmitted signal in general. Thus, for the purpose

of maximization in (28), the Q function of (C.1) can be replaced by

$$Q(\underline{X}\,|\,\underline{X}^{(p)}) \propto \int [\log f(\underline{Y}\,|\,\underline{h},\underline{X})] f(\underline{h}\,|\,\underline{Y},\underline{X}^{(p)}) d\underline{h}. \quad (C.3)$$

The conditional pdf $f(\underline{h}\,|\,\underline{Y},\underline{X}^{(p)})$ can be calculated by

$$
\begin{aligned}
f(\underline{h}\,|\,\underline{Y},\underline{X}^{(p)}) &= \frac{f(\underline{Y}\,|\,\underline{h},\underline{X}^{(p)}) f(\underline{h}\,|\,\underline{X}^{(p)})}{f(\underline{Y}\,|\,\underline{X}_p)} \\
&= \frac{f(\underline{Y}\,|\,\underline{h},\underline{X}^{(p)}) f(\underline{h})}{f(\underline{Y}\,|\,\underline{X}^{(p)})},
\end{aligned}
\quad (C.4)
$$

where we use the assumption that $\underline{h}$ and $\underline{X}^{(p)}$ are independent of each other. Thus, (C.3) can be further reduced to

$$Q(\underline{X}\,|\,\underline{X}^{(p)}) \propto \int [\log f(\underline{Y}\,|\,\underline{h},\underline{X})] f(\underline{Y}\,|\,\underline{h},\underline{X}^{(p)}) f(\underline{h}) d\underline{h}, \quad (C.5)$$

since $f(\underline{Y}\,|\,\underline{X}^{(p)})$ does not depend on $\underline{X}$. Hence, it can be discarded in the last expression.

We now compute the above $Q(\underline{X}\,|\,\underline{X}^{(p)})$ for a fading channel with AWGN. The conditional pdfs $f(\underline{Y}\,|\,\underline{h},\underline{X})$ and $f(\underline{Y}\,|\,\underline{h},\underline{X}^{(p)})$ take the forms

$$f(\underline{Y}\,|\,\underline{h},\underline{X}) = (2\pi\sigma^2)^{-MD} \exp\left\{ -\frac{||\underline{Y} - \mathbf{X}\mathbf{W}_{LD}\underline{h}||^2}{2\sigma^2} \right\},$$

$$f(\underline{Y}\,|\,\underline{h},\underline{X}^{(p)}) = (2\pi\sigma^2)^{-MD} \exp\left\{ -\frac{||\underline{Y} - \mathbf{X}^{(p)}\mathbf{W}_{LD}\underline{h}||^2}{2\sigma^2} \right\},$$
$$(C.6)$$

where σ^2 is the variance of both real and imaginary components of complex-valued Gaussian white noise. The pdf $f(\underline{h})$ is given by

$$f(\underline{h}) = \frac{1}{(2\pi)^L |\det \boldsymbol{\Sigma}|} \exp\left\{ -\frac{1}{2}(\underline{h} - E\{\underline{h}\})^H \boldsymbol{\Sigma}^{-1}(\underline{h} - E\{\underline{h}\}) \right\}, \quad (C.7)$$

where $E\{\underline{h}\}$ and $\boldsymbol{\Sigma}$ are the mean and covariance matrix of the complex-valued CIR vector $\underline{h}$. By omitting the constant term and the scaling factor, (C.3) can be expressed as follows:

$$Q(\underline{X}\,|\,\underline{X}^{(p)}) \propto -\int ||\underline{Y} - \mathbf{X}\mathbf{W}_{LD}\underline{h}||^2 f(\underline{h}\,|\,\underline{Y},\underline{X}^{(p)}) d\underline{h}, \quad (C.8)$$

and $f(\underline{h}\,|\,\underline{Y},\underline{X}^{(p)})$ can be represented as follows:

$$
\begin{aligned}
&f(\underline{h}\,|\,\underline{Y},\underline{X}^{(p)}) \\
&= K_2 \exp\left\{ -\frac{1}{2}(\underline{h} - \underline{h}^{(p)})^H (\boldsymbol{\Sigma}^{(p)})^{-1}(\underline{h} - \underline{h}^{(p)}) \right\},
\end{aligned}
\quad (C.9)
$$

where K_1 and K_2 are constants. $\underline{h}^{(p)}$ and $\boldsymbol{\Sigma}^{(p)}$ are called the estimated posterior mean and posterior covariance matrix at

the pth iteration and are given by

$$
\begin{aligned}
\underline{h}^{(p)} &= \boldsymbol{\Sigma}^{(p)} \left(\mathbf{W}_{LD}^H (\mathbf{X}^{(p)})^H \underline{Y}/\sigma^2 + \boldsymbol{\Sigma}^{-1} E\{\underline{h}\} \right), \\
\boldsymbol{\Sigma}^{(p)} &= \left(\mathbf{W}_{LD}^H (\mathbf{X}^{(p)})^H \mathbf{X}^{(p)} \mathbf{W}_{LD}/\sigma^2 + \boldsymbol{\Sigma}^{-1} \right)^{-1}.
\end{aligned}
\quad (C.10)
$$

Maximizing (C.8) is equivalent to minimizing the distance

$$\arg\max_{\underline{X}} Q(\underline{X}\,|\,\underline{X}^{(p)}) = \arg\min_{\underline{X}} E\left\{ ||\underline{Y} - \mathbf{X}\mathbf{W}_{LD}\underline{h}||^2 \,\big|\, \underline{Y}, \underline{X}^{(p)} \right\}. \quad (C.11)$$

This minimization can be further simplified as follows:

$$\max_{\underline{X}} E\{\underline{h}^H \underline{F} + \underline{F}^H \underline{h} - \underline{h}^H \mathbf{G}\underline{h}\,|\,\underline{Y},\underline{X}^{(p)}\}, \quad (C.12)$$

where

$$\underline{F} = \mathbf{W}_{LD}^H \mathbf{X}^H \underline{Y}, \quad (C.13)$$

$$\mathbf{G} = \mathbf{W}_{LD}^H \mathbf{X}^H \mathbf{X}\mathbf{W}_{LD}. \quad (C.14)$$

Since the distribution of random vector $\underline{Y}$ given $\underline{h}$ and $\underline{X}^{(p)}$ is Gaussian with mean $\underline{h}^{(p)}$ and covariance matrix $\boldsymbol{\Sigma}^{(p)}$, it is easy to compute

$$E\{\underline{h}^H \underline{F} + \underline{F}^H \underline{h}\,|\,\underline{Y},\underline{X}^{(p)}\} = (\underline{h}^{(p)})^H \underline{F} + \underline{F}^H \underline{h}^{(p)}. \quad (C.15)$$

Moreover, all entries of matrix $\mathbf{G}$ are given in terms of the signal energies, that is, $|X(0)|^2, \ldots, |X(M-1)|^2$. Thus, we can compute the third part of (C.12) as follows:

$$E\{\underline{h}^H \mathbf{G}\underline{h}\,|\,\underline{Y},\underline{X}^{(p)}\} = \sum_{d=1}^{D} \sum_{m=0}^{M-1} C_m |X^d(m)|^2, \quad (C.16)$$

where C_m, $0 \le m \le M-1$, are real values dependent on $\underline{h}^{(p)}$ and $\boldsymbol{\Sigma}^{(p)}$ and can be obtained by the following equation:

$$C_m = \sum_{k=0}^{L-1} \sum_{n=0}^{L-1} e^{j2\pi((k-n)m/M)} \left(\boldsymbol{\Sigma}^{(p)}(k,n) + h_k^{(p)*} h_n^{(p)} \right). \quad (C.17)$$

In order to calculate $Q(\underline{X}\,|\,\underline{X}^{(p)})$ completely with respect to $\underline{X}$, we write (C.13) as follows:

$$\underline{F} = \mathbf{W}_{LD}^H \mathbf{Y}\underline{X}^*. \quad (C.18)$$

Thus, maximizing $Q(\underline{X}\,|\,\underline{X}_p)$ is the same as

$$
\begin{aligned}
\max_{\underline{X}} \bigg\{ &(\underline{h}^{(p)})^H \mathbf{W}_{LD}^H \mathbf{Y}\underline{X}^* + \underline{X}^T \mathbf{Y}^H \mathbf{W}_{LD} \underline{h}^{(p)} \\
&- \sum_{d=1}^{D} \sum_{m=0}^{M-1} C_m |X^d(m)|^2 \bigg\}.
\end{aligned}
\quad (C.19)
$$

Equation (C.19) can be solved as follows:

$$\tilde{\underline{X}}^{(p+1)} = \arg\max_{\underline{X}} Q(\underline{X}\,|\,\underline{X}^{(p)}) = \mathbf{C}_D^{-1}\left((\underline{h}^{(p)})^H \mathbf{W}_{LD}^H \mathbf{Y} \right)^T, \quad (C.20)$$

where $\mathbf{C}_D = \mathrm{diag}(\mathbf{C},\ldots,\mathbf{C})_{MD \times MD}$ and $\mathbf{C} = \mathrm{diag}(C_0^2,\ldots, C_{M-1}^2)$. After quantizing $\underline{\tilde{X}}^{(p+1)}$, we obtain the $(p+1)$st estimate

$$\underline{X}^{(p+1)} = \text{Quantization}\{\underline{\tilde{X}}^{(p+1)}\}. \qquad \text{(C.21)}$$

D.   DERIVATION OF ALGORITHM 3

Equation (43) can be rewritten as follows:

$$
\begin{aligned}
Q(\underline{Z}\,&|\,\underline{h}^{(p)}) \\
&= K_1 - E\left\{ \sum_{m=0}^{MD-1} \sum_{i=0}^{L-1} \frac{1}{2\sigma_i^2} ||Z_i(m) - a_i(m)h_i||^2 \,|\, \underline{Y},\underline{h}^{(p)} \right\} \\
&= K_2 - \sum_{m=0}^{MD-1} \sum_{i=0}^{L-1} \frac{1}{2\sigma_i^2} ||\hat{Z}_i^{(p)}(m) - a_i(m)h_i||^2,
\end{aligned}
$$
$$\text{(D.1)}$$

where K_1 and K_2 are some constants independent of $\underline{h}$, and

$$\hat{Z}_i^{(p)}(m) = E\{Z_i(m)\,|\,Y(m),\underline{h}^{(p)}\}. \qquad \text{(D.2)}$$

Here we use the fact that $Z_i(m)$ only depends on $Y(m)$. Since $Z_i(m)$ and $Y(m)$ are jointly Gaussian distributed and satisfy (41), then it is easy to obtain

$$\hat{Z}_i^{(p)}(m) = Z_i^{(p)}(m) + \rho_i \frac{\sigma_i}{\sigma}\left(Y(m) - \sum_{j=0}^{L-1} Z_j^{(p)}(m) \right), \quad \text{(D.3)}$$

where

$$Z_i^{(p)}(m) = a_i(m)h_i^{(p)}, \qquad \text{(D.4)}$$

and ρ_i is the cross-correlation between $Z_i(m)$ and $Y(m)$ given $\underline{h}^{(p)}$, which is

$$\rho_i = \frac{E\{N_i(m)N(m)\}}{\sqrt{\mathrm{Var}(N_i(m))\,\mathrm{Var}(N(m))}} = \frac{\sigma_i}{\sigma}. \qquad \text{(D.5)}$$

Substituting (D.5) into (D.3), we obtain

$$\hat{Z}_i^{(p)}(m) = Z_i^{(p)}(m) + \beta_i\left(Y(m) - \sum_{j=0}^{L-1} Z_j^{(p)}(m) \right), \qquad \text{(D.6)}$$

where $\beta_i = \sigma_i^2/\sigma^2$ are real-valued scalars satisfying

$$\sum_{i=0}^{L-1} \beta_i = 1, \quad \beta_i \geq 0. \qquad \text{(D.7)}$$

After obtaining $\hat{Z}_i^p(m)$, $0 \leq i \leq L-1$, for each data index m, we need to maximize $Q(\underline{Z}|\underline{h}^{(p)})$, which is the same as

$$\underline{h}^{(p+1)} = \arg\min_{\underline{h}} \sum_{m=0}^{MD-1} \sum_{i=0}^{L-1} ||\hat{Z}_i^{(p)}(m) - a_i(m)h_i||^2. \quad \text{(D.8)}$$

The above minimization problem can be separated into L simple minimization problems by just exchanging the order of the two summations. These L minimization problems are

$$h_i^{(p+1)} = \arg\min_{h_i} \sum_{m=0}^{MD-1} ||\hat{Z}_i^{(p)}(m) - a_i(m)h_i||^2, \quad 0 \leq i \leq L-1,$$
$$\text{(D.9)}$$

which can be solved by

$$h_i^{(p+1)} = \frac{1}{MD} \sum_{m=0}^{MD-1} \frac{\hat{Z}_i^{(p)}(m)}{a_i(m)}, \quad 0 \leq i \leq L-1. \qquad \text{(D.10)}$$

E.   DERIVATION OF CRLB

In [25], the authors derive the CRLB by separating the complex vector $\underline{h}$ into real and imaginary parts. We will simplify the derivation by applying derivatives with respect to the complex vector itself. Recall the OFDM system model can be represented as in (26), where we use the same notation as in Section 3.3. We again assume for generality that the channel is static over the period of D frames. We will derive a CRLB by using all the data from the D frames. The parameter vector here is obviously $\theta = \underline{h}$. The CRLB gives a lower bound [28] for the variance of an unbiased estimate:

$$\text{CRLB}\,(h_i) = I^{-1}(\underline{h})_{ii}, \quad 0 \leq i \leq L-1, \qquad \text{(E.1)}$$

where $I(\underline{h})$ is the Fisher information matrix

$$I(\underline{h}) = -E\left\{ \frac{\partial}{\partial \underline{h}} \log f(\underline{Y}|\underline{h}) \left(\frac{\partial}{\partial \underline{h}} \log f(\underline{Y}|\underline{h}) \right)^H \right\}. \qquad \text{(E.2)}$$

Following (26), we have the conditional pdf of $\underline{Y}$ given $\underline{h}$:

$$f(\underline{Y}|\underline{h}) = \frac{1}{(2\pi\sigma^2)^{MD}} \exp\left\{ -\frac{1}{2\sigma^2} ||\underline{Y} - \mathbf{X}\mathbf{W}_{LD}\underline{h}||^2 \right\}, \quad \text{(E.3)}$$

where we assume the data matrix $\mathbf{X}$ is known. Therefore, the pdf is not conditioned on $\mathbf{X}$. After differentiating the logarithm of (E.3) with respect to $\underline{h}$, we obtain

$$\frac{\partial}{\partial \underline{h}} \log f(\underline{Y}|\underline{h}) = \frac{1}{2\sigma^2}\left[(\underline{Y}^H - \underline{h}^H \mathbf{W}_{LD}^H \mathbf{X}^H)\mathbf{X}\mathbf{W}_{LD} \right]^T. \quad \text{(E.4)}$$

Then the Fisher information matrix is obtained as follows:

$$
\begin{aligned}
I(\underline{h}) &= E\left\{ \frac{\partial}{\partial \underline{h}} \log f(\underline{Y}|\underline{h}) \left(\frac{\partial}{\partial \underline{h}} \log f(\underline{Y}|\underline{h}) \right)^H \right\} \\
&= E\left\{ \frac{1}{2\sigma^2}\left[(\underline{Y}^H - \underline{h}^H \mathbf{W}_{LD}^H \mathbf{X}^H)\mathbf{X}\mathbf{W}_{LD} \right]^T \right. \\
&\qquad \left. \times \frac{1}{2\sigma^2}\left[(\underline{Y}^H - \underline{h}^H \mathbf{W}_{LD}^H \mathbf{X}^H)\mathbf{X}\mathbf{W}_{LD} \right]^* \right\} \\
&= \frac{1}{2\sigma^2} \mathbf{W}_{LD}^H \mathbf{X}^H \mathbf{X}\mathbf{W}_{LD} \\
&= \frac{1}{2\sigma^2} \mathbf{W}_L^H \sum_{d=1}^{D} (\mathbf{X}^d)^H \mathbf{X}^d \mathbf{W}_L.
\end{aligned}
$$
$$\text{(E.5)}$$

The CRLB for the CIR of the ith tap is $I^{-1}(\underline{h})_{ii}$. We define the CRLB for the overall CIR as follows:

$$\text{CRLB}(\underline{h}) = \sum_{i=0}^{L-1} \text{CRLB}\,(h_i) = \text{trace}\,(I^{-1}(\underline{h})). \qquad \text{(E.6)}$$

ACKNOWLEDGMENT

REFERENCES

[1] L. J. Cimini Jr., "Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing," *IEEE Trans. Communications*, vol. 33, no. 7, pp. 665–675, 1985.

[2] S. B. Weinstein and P. M. Ebert, "Data transmission by frequency-division multiplexing using the discrete Fourier transform," *IEEE Trans. Communications*, vol. 19, no. 5, pp. 628–634, 1971.

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

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

[5] J. K. Moon and S. I. Choi, "Performance of channel estimation methods for OFDM systems in a multipath fading channels," *IEEE Transactions on Consumer Electronics*, vol. 46, no. 1, pp. 161–170, 2000.

[6] P. Hoeher, S. Kaiser, and P. Robertson, "Two-dimensional pilot-symbol-aided channel estimation by Wiener filtering," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '97)*, vol. 3, pp. 1845–1848, Munich, Germany, April 1997.

[7] F. Said and A. H. Aghvami, "Linear two dimensional pilot assisted channel estimation for OFDM systems," in *Proc. 6th IEE Conference on Telecommunications*, pp. 32–36, Edinburgh, UK, March–April 1998.

[8] J.-J. van de Beek, O. Edfors, M. Sandell, S. K. Wilson, and P. O. Borjesson, "On channel estimation in OFDM systems," in *Proc. 45th IEEE Vehicular Technology Conference (VTC '95)*, vol. 2, pp. 815–819, Chicago, Ill, USA, July 1995.

[9] O. Edfors, M. Sandell, J.-J. van de Beek, S. K. Wilson, and P. O. Borjesson, "OFDM channel estimation by singular value decomposition," *IEEE Trans. Communications*, vol. 46, no. 7, pp. 931–939, 1998.

[10] R. Nilsson, O. Edfors, M. Sandell, and P. O. Borjesson, "An analysis of two-dimensional pilot-symbol assisted modulation for OFDM," in *Proc. IEEE International Conference on Personal Wireless Communications (ICPWC '97)*, pp. 71–74, Mumbai, India, December 1997.

[11] B. Yang, K. B. Letaief, R. S. Cheng, and Z. Cao, "Channel estimation for OFDM transmission in multipath fading channels based on parametric channel modeling," *IEEE Trans. Communications*, vol. 49, no. 3, pp. 467–479, 2001.

[12] R. W. Heath Jr. and G. B. Giannakis, "Exploiting input cyclostationarity for blind channel identification in OFDM systems," *IEEE Trans. Signal Processing*, vol. 47, no. 3, pp. 848–856, 1999.

[13] B. Muquet and M. de Courville, "Blind and semi-blind channel identification methods using second order statistics for OFDM systems," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '99)*, vol. 5, pp. 2745–2748, Phoenix, Ariz, USA, March 1999.

[14] X. Cai and A. N. Akansu, "A subspace method for blind channel identification in OFDM systems," in *Proc. IEEE International Conference on Communications (ICC '00)*, vol. 2, pp. 929–933, New Orleans, La, USA, June 2000.

[15] X. Zhuang, Z. Ding, and A. L. Swindlehurtst, "A statistical subspace method for blind channel identification in OFDM communications," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '00)*, vol. 5, pp. 2493–2496, Istanbul, Turkey, June 2000.

[16] C. Li and S. Roy, "Subspace based blind channel estimation for OFDM by exploiting virtual carrier," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM '01)*, vol. 1, pp. 295–299, San Antonio, Tex, USA, November 2001.

[17] P. Chen and H. Kobayashi, "Maximum likelihood channel estimation and signal detection for OFDM systems," in *Proc. IEEE International Conference on Communications (ICC '02)*, vol. 3, pp. 1640–1645, New York, NY, USA, April–May 2002.

[18] S. Zhou and G. B. Giannakis, "Finite-alphabet based channel estimation for OFDM and related multicarrier systems," *IEEE Trans. Communications*, vol. 49, no. 8, pp. 1402–1414, 2001.

[19] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society (B)*, vol. 39, no. 1, pp. 1–38, 1977.

[20] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Processing Magazine*, vol. 13, no. 6, pp. 47–60, 1996.

[21] H. Zamiri-Jafarian and S. Pasupathy, "EM-based recursive estimation of channel parameters," *IEEE Trans. Communications*, vol. 47, no. 9, pp. 1297–1302, 1999.

[22] C. N. Georghiades and J. C. Han, "Sequence estimation in the presence of random parameters via the EM algorithm," *IEEE Trans. Communications*, vol. 45, no. 3, pp. 300–308, 1997.

[23] L. M. Zeger and H. Kobayashi, "A simplified EM algorithm for detection of CPM signals in a fading multipath channel," *Wireless Networks*, vol. 8, no. 6, pp. 649–658, 2002.

[24] M. Feder and E. Weinstein, "Parameter estimation of superimposed signals using the EM algorithm," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 36, no. 4, pp. 477–489, 1988.

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

[26] R. W. Miller and C. B. Chang, "A modified Cramér-Rao bound and its applications (Corresp.)," *IEEE Transactions on Information Theory*, vol. 24, no. 3, pp. 398–400, 1978.

[27] A. N. D'Andrea, U. Mengali, and R. Reggiannini, "The modified Cramer-Rao bound and its application to synchronization problems," *IEEE Trans. Communications*, vol. 42, no. 234, pp. 1391–1399, 1994.

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

Xiaoqiang Ma received the B.E. and M.S. degrees in electrical engineering from Tsinghua University, Beijing, China, in 1996 and 1999, respectively, and is now pursuing the Ph.D. degree in electrical engineering from Princeton University. He was with Mitsubishi Lab, Murryhill, NJ, during the summer of 2000 and with Microsoft Research, Redmond, Wash, during the summer of 2003. His research interests are in the general areas of digital communication theory and signal processing applications in wireless communications. He focuses more on the multicarrier related system design and performance analysis, including equalization, channel estimation, synchronization, and peak-to-average power ratio reduction.

Hisashi Kobayashi is the Sherman Fairchild University Professor of Electrical Engineering and Computer Science at Princeton University, NJ, since 1986, when he joined the Princeton faculty as the Dean of the School of Engineering and Applied Science (1986–1991). He worked for the IBM Research Center in Yorktown Heights for fifteen years (1967–1982), and then served as the Founding Director of the IBM Tokyo Research Laboratory (1982–1986). He was a radar designer at Toshiba, Japan (1963-1965). He received his Ph.D. from Princeton University in 1967, and his B.E. and M.E. degrees from the University of Tokyo in 1961 and 1963, respectively. His research experiences include radar systems, high speed data transmission, seismic signal processing, coding for high density digital recording, image compression algorithms, performance modeling and analysis of computers and communication systems. His current research activities include network security, OFDM and UWB communication systems, and network performance modeling. He has been a Fellow (1977) and a Life Fellow (2003) of IEEE, and received the Humboldt Prize from Germany (1979). He was elected a Member of the Engineering Academy of Japan (1992) and a Fellow of IEICE, Japan (2004).

Stuart C. Schwartz received the B.S. and M.S. degrees from MIT in 1961 and the Ph.D. degree from the University of Michigan in 1966. At MIT, he was associated with the Naval Supersonic Laboratory and the Instrumentation Laboratory (now the Draper Laboratories). During the year 1961–1962, he was at the Jet Propulsion Laboratory in Pasadena, California, working on problems in orbit estimation and telemetry. During the academic year 1980–1981, he was a member of the technical staff at the Radio Research Laboratory, Bell Telephone Laboratories, Crawford Hill, NJ, working in the area of mobile telephony. He is currently a Professor of electrical engineering at Princeton University. He was Chair of the department during the period 1985–1994, and served as Associate Dean for the School of Engineering during the period from July 1977 till June 1980. During the academic year 1972–1973, he was a John S. Guggenheim Fellow and Visiting Associate Professor at the Department of Electrical Engineering, Technion, Haifa, Israel. He has also held visiting academic appointments at Dartmouth, University of California, Berkeley, and the Image Sciences Laboratory, ETH, Zurich. His principal research interests are in the application of probability and stochastic processes to problems in statistical communication theory and signal processing.