EM-based Channel Estimation for Coded Multi-Carrier Transmissions

Yang Liu*†, Loïc Brunel†, and Joseph J. Boutros‡

***Abstract*—Expectation-maximization (EM) based iterative algorithms are investigated in order to estimate the impulse response of a frequency-selective multipath channel in a coded OFDM system. Two ways of choosing the EM complete data are compared: a complete data built from observations and transmitted symbols (CL-EM) and a complete data chosen by decomposing noise and observation components (NCD-EM). Both CL-EM and NCD-EM algorithms are derived for a coded OFDM system. The rate of convergence of both EM algorithms is theoretically determined. It is found that the rate of convergence of CL-EM is independent from the number of channel taps at high signal-to-noise ratio (SNR), while that of NCD-EM varies with the number of taps. It is shown that CL-EM converges in a few iterations. Furthermore, considering the complexity per iteration, CL-EM has a lower complexity than its counterpart. We also establish a Cramér-Rao bound (CRB) for coded OFDM transmission. Simulation results show that CL-EM has a good performance-complexity trade-off and it achieves the CRB.**

***Index Terms*—Orthogonal Frequency-Division Multiplexing, Expectation Maximization, Channel Estimation, Rate of Convergence, Cramér-Rao Bound.**

I. INTRODUCTION

Coded orthogonal frequency-division multiplexing (OFDM) has been chosen as the air interface for recent cellular and wireless local area network (WLAN) systems by virtue of its good performance, flexibility, and low implementation complexity.

A number of channel estimation methods have been proposed for OFDM. When pilot symbols are available on some sub-carriers, initial estimates are easily obtained and can be improved through frequency- and time-domain interpolation [1] or according to the minimum mean square error (MMSE) criterion [2]. Blind channel estimation methods, not relying on the presence of pilot symbols, have also been proposed. For instance, they take benefit from the cyclostationarity of the OFDM signal [3]. However, all these algorithms have their limitations: insufficient accuracy, low spectrum efficiency due to pilot overhead, or high sensitivity to Doppler in case of blind estimation.

As hardware capacity is continuously increasing, it becomes more feasible to implement iterative receivers allowing for improvement of physical layer functions. Among these functions, channel estimation especially benefits from data-aided methods requiring a feedback from the channel decoder. The iterative expectation-maximization (EM) estimation algorithm [4] is particularly well suited for OFDM systems with low pilot overhead. Instead of computing the maximum likelihood channel estimate from the observations only, it makes use of the so-called *complete data* κ, which are not observed directly but only through incomplete data. Since κ is a random variable, the log-likelihood can be averaged over κ knowing the incomplete data and a current channel estimate. A new channel estimate is then obtained by maximizing the average log-likelihood, which results in the EM iterative structure. The likelihood increases along EM iterations [4]. A classic way (CL-EM) to choose the complete data is $\kappa = (\mathbf{X}, \mathbf{Y})$, where $\mathbf{X}$ is the transmitted signal and $\mathbf{Y}$ is the observation [5]. Classic EM estimation for OFDM with space-frequency coding has been studied in [6]. In [7] and [8], for uncoded OFDM, complete data is obtained by decomposing the noise and observation components (NCD-EM). In [9], NCD-EM is also applied to a coded single-carrier system.

In this paper, we derive CL-EM and NCD-EM algorithms for coded OFDM systems and compare them from two aspects: complexity and performance. The complexity of the EM algorithm depends on two factors: the complexity in each iteration and the rate of convergence of the EM algorithm. In [4] and [10], the general rate of convergence of the EM algorithm is analyzed with mathematical derivation and graphical illustration respectively. Based on [4], we derive the rates of convergence of CL-EM and NCD-EM for a coded OFDM system. Together with the complexity in each iteration, the overall complexity of the EM algorithm can be obtained. Concerning performance, we compare CL-EM and NCD-EM for a coded OFDM system in terms of mean square error (MSE) and bit error rate (BER). In addition, we determine the Cramér-Rao bound for this coded OFDM system, as an ideal reference for MSE performance.

The rest of the paper is organized as follows. Section II describes the coded OFDM system and introduces the EM channel estimator. In section III, we develop the CL-EM and NCD-EM algorithms for the coded OFDM system. In section IV, we compare the two proposed EM algorithms in terms of complexity per iteration and rate of convergence, and discuss their overall complexities. Section V presents CRB for coded OFDM. Section VI includes performance results, and finally, some conclusions are drawn in section VII.

* Telecom ParisTech-École Nationale Supérieure des Télécommunications, 46 rue Barrault, 75013 Paris, France. Dr Yang Liu is now with Sequans Communications, Paris, France, yliu@sequans.com

†Mitsubishi Electric R&D Center Europe, 1 allée de Beaulieu, 35700 Rennes, France. l.brunel@fr.merce.mee.com

‡Texas A&M University at Qatar, PO Box 23874, Doha, Qatar. boutros@tamu.edu

This work has been presented in part at ISSSTA'2008, Bologna, Italy.

II. SYSTEM DESCRIPTION

We consider a coded OFDM signal transmitted over a single-input single-output (SISO) frequency-selective channel as shown in Fig. 1.

An information binary sequence $\mathbf{S}$ is encoded into a coded sequence $\mathbf{C}$. The encoded bits are then interleaved by a pseudo-random interleaver and modulated. After pilot insertion, the obtained sequence $\mathbf{X} = (X_0, \cdots, X_{N-1})^{\mathrm{T}}$ is processed by an inverse Fast Fourier transform (IFFT), which provides the time-domain sequence $\mathbf{x} = (x_0, \cdots, x_{N-1})^{\mathrm{T}} = \mathbf{U}^{\dagger}\mathbf{X}$, where $\mathbf{U}$ is the normalized $N \times N$ FFT matrix with the (m, n) entry equal to $\frac{1}{\sqrt{N}} e^{-\frac{j2\pi(m-1)(n-1)}{N}}$ $(m, n = 1, 2, \cdots, N)$ and $(\cdot)^{\dagger}$ stands for transpose-conjugate. After insertion of a cyclic prefix (CP) with length L_{CP}, the transmitted OFDM symbol is $\mathbf{x}^{'} = (x_{N-L_{\mathrm{CP}}}, \cdots, x_{N-1}, \mathbf{x}^{\mathrm{T}})^{\mathrm{T}}$. The received sequence is $\mathbf{y}^{'} = \left(y_0^{'}, \cdots, y_{N+L_{\mathrm{CP}}-1}\right)^{\mathrm{T}}$ with

$$y_k^{'} = \sum_{l=0}^{L-1} h_l x_{k-l}^{'} + n_k, \qquad L_{\mathrm{CP}} \leq k \leq N + L_{\mathrm{CP}} - 1, \quad (1)$$

where L is the number of taps in the channel, $L \leqslant L_{\mathrm{CP}}$, $\mathbf{h} = (h_0, \cdots, h_{L-1})^{\mathrm{T}}$ is the channel impulse response, and n_k is a complex Gaussian noise with zero mean and variance $2\sigma^2$. After CP removal, the received time domain sequence $\mathbf{y} = (y_0, \cdots, y_{N-1})^{\mathrm{T}}$ is processed by FFT. The received sequence in the frequency domain is $\mathbf{Y} = (Y_0, \cdots, Y_{N-1})^{\mathrm{T}} = \mathbf{U}\mathbf{y}$. We assume that the channel impulse response is constant over one OFDM symbol. Finally, the well-known OFDM discrete-time model is [11]

$$\mathbf{Y} = \mathrm{diag}\,(\mathbf{H})\,\mathbf{X} + \mathbf{N}, \quad (2)$$

where $\mathbf{N} = (N_0, \cdots, N_{N-1})^{\mathrm{T}} = \mathbf{U}\mathbf{n}$ has the same distribution as $\mathbf{n} = (n_0, \cdots, n_{N-1})^{\mathrm{T}}$, $\mathbf{H} = (H_0, \cdots, H_{N-1})^{\mathrm{T}}$ represents the channel frequency-response $\mathbf{H} = \mathbf{\Omega}\mathbf{h}$, where the $N \times L$ matrix $\mathbf{\Omega}$ is built from the L first columns of $\mathbf{U}$, and $\mathrm{diag}\,(\mathbf{H})$ represents a diagonal matrix with H_k as its (k, k) entry.

An initial channel estimate is obtained from pilots included in the sequence $\mathbf{Y}$. Soft information produced by the demodulator on coded bits is de-interleaved into sequence $\hat{\mathbf{C}}$, which is processed by the decoder. After decoding, *a posteriori* probabilities are fed back to the EM-based estimator which updates the channel estimate for next demodulation and decoding. Thus, the soft information and the channel estimate are improved along iterations.

III. EM-BASED CHANNEL ESTIMATION

A. Notations for the EM Algorithm

The EM algorithm provides a recursive solution to ML estimation [4] [12] and it performs a two-step procedure as shown in Fig. 2:

1) E-step: compute the auxiliary function $Q\left(\mathbf{h}|\mathbf{h}^{(i)}\right) = E_{\kappa}\left[\log p\left(\kappa\,|\mathbf{h}\right)|\mathbf{Y}, \mathbf{h}^{(i)}\right]$;
2) M-step: update the parameters $\mathbf{h}^{(i+1)} = \arg\max_{\mathbf{h}} Q\left(\mathbf{h}|\mathbf{h}^{(i)}\right)$,

where κ is called the complete data. It is possible to define the complete data in different ways leading to different types of EM algorithms. A classic definition is to choose $\kappa = (\mathbf{X}, \mathbf{Y})$. This is named classic EM (CL-EM). For superimposed signals, it has been shown in [13] that the complete data can be chosen by decomposing noise components into independent noise processes. This is named noise component decomposition EM (NCD-EM). NCD-EM has been utilized in some applications, such as uncoded OFDM with single and multiple transmit antennas (see [8] and [7] respectively) and coded single-carrier transmission on fading channel with uncorrelated and correlated paths (see [9] and [14] respectively).

B. CL-EM Channel Estimation

Here, we define the complete data as in the classic method [5]: $\kappa = (\mathbf{X}, \mathbf{Y})$. $\mathbf{X}$ is called the missing data and $\mathbf{Y}$ the incomplete data. So, the E-step can be re-written as

$$Q\left(\mathbf{h}|\mathbf{h}^{(i)}\right) = E_{\mathbf{X}}\left[\log p\left(\mathbf{X},\, \mathbf{Y}\,|\mathbf{h}\right)|\mathbf{Y}, \mathbf{h}^{(i)}\right]. \quad (3)$$

If all values of $\mathbf{X}$ are equiprobable, the auxiliary function can be written as

$$\begin{aligned} Q\left(\mathbf{h}|\mathbf{h}^{(i)}\right) &= E_{\mathbf{X}}\left[\log p\left(\mathbf{Y}\,|\mathbf{X},\, \mathbf{h}\right)|\mathbf{Y}, \mathbf{h}^{(i)}\right] \\ &= \sum_{\mathbf{X}} \log p\left(\mathbf{Y}\,|\mathbf{X},\, \mathbf{h}\right) APP^{(i)}\left(\mathbf{X}\right), \end{aligned} \quad (4)$$

where $APP^{(i)}\left(\mathbf{X}\right) = \mathrm{P}\left(\mathbf{X}|\mathbf{Y}, \mathbf{h}^{(i)}\right)$ is the *a posteriori* probability of $\mathbf{X}$ in the ith iteration.

In [15], we derived an EM algorithm by considering an energy constraint, called EC-EM. However, we found that EC-EM does not perform so well because of a rough estimation of the energy constraint. Therefore, we focus here on the CL-EM without energy constraint. In order to estimate the channel impulse response $\mathbf{h}$, we re-write (2) as

$$\mathbf{Y} = \mathrm{diag}\,(\mathbf{X})\,\mathbf{\Omega}\mathbf{h} + \mathbf{N}. \quad (5)$$

With the Gaussian noise assumption, given $\mathbf{X}$ and $\mathbf{h}$, we have $\mathbf{Y} \sim \mathcal{CN}\left(\mathrm{diag}\,(\mathbf{X})\,\mathbf{\Omega}\mathbf{h}, 2\sigma^2\mathbf{I}_N\right)$, where $\mathbf{I}_N$ represents the unit matrix of size N. Then, the auxiliary function can be written as

$$\begin{aligned} Q\left(\mathbf{h}|\mathbf{h}^{(i)}\right) = &-\frac{1}{2\sigma^2}\sum_{\mathbf{X}} \left\|\mathbf{Y} - \mathrm{diag}\,(\mathbf{X})\,\mathbf{\Omega}\mathbf{h}\right\|^2 APP^{(i)}\left(\mathbf{X}\right) \\ &- \sum_{\mathbf{X}} N\,\log\,2\pi\sigma^2\, APP^{(i)}\left(\mathbf{X}\right). \end{aligned} \quad (6)$$

The new channel estimate $\mathbf{h}^{(i+1)}$ is the value of $\mathbf{h}$ satisfying $\frac{\partial}{\partial \mathbf{h}} Q\left(\mathbf{h}|\mathbf{h}^{(i)}\right) = 0$,

$$\mathbf{h}^{(i+1)} = \left(\mathbf{\Omega}^{\dagger}\mathbf{R}_{N\times N}^{(i)*}\mathbf{\Omega}\right)^{-1}\mathbf{\Omega}^{\dagger}\widetilde{\mathrm{diag}\,(\mathbf{X})}^{(i)\dagger}\mathbf{Y}, \quad (7)$$

where $\widetilde{\mathrm{diag}\,(\mathbf{X})}^{(i)}$ is the $N \times N$ diagonal matrix of soft

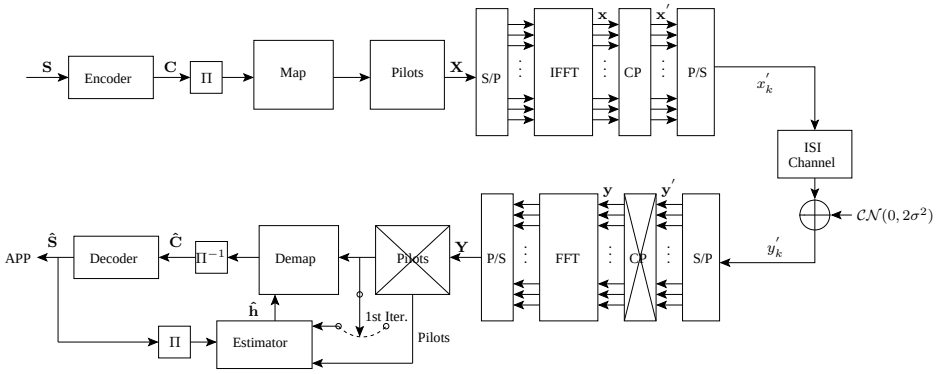


Fig. 1. Coded OFDM system model with the proposed EM based channel estimator. The initial estimate is obtained from pilot symbols only, whereas the estimates in following iterations are obtained from both pilot and data symbols.

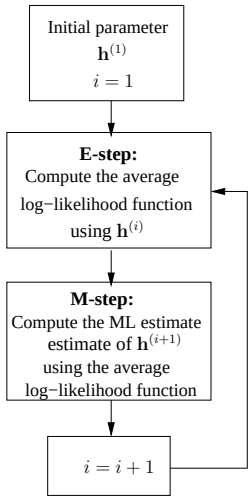


Fig. 2. Flowchart of the EM algorithm.

estimates of $\mathbf{X}$:

$$\widetilde{\text{diag}\,(\mathbf{X})}^{(i)} \triangleq \sum_{\mathbf{X}_j} APP^{(i)}\left(\mathbf{X}_j\right) \text{diag}\left(\mathbf{X}_j\right) = \text{diag}\left(\left[\sum_m APP^{(i)}\left(X_0 = \alpha_m\right)\alpha_m, \cdots, \sum_m APP^{(i)}\left(X_{N-1} = \alpha_m\right)\alpha_m\right]^{\mathrm{T}}\right), \quad (8)$$

where α_m $(m = 0, \cdots, M-1)$ represents the set of possible symbols in the mapping constellation, M represents the size of the mapping constellation and $APP^{(i)}\left(X_k = \alpha_m\right)$ is the *a posteriori* probability of $X_k = \alpha_m$ in the ith iteration; $\mathbf{R}^{(i)}_{N\times N}$ is a $N \times N$ matrix:

$$\mathbf{R}^{(i)}_{N\times N} = \text{diag}\left(\mathbf{X}\right)^{\widetilde{(i)\mathrm{T}}} \text{diag}\left(\mathbf{X}\right)^{(i)*} = \text{diag}\left(\left[\sum_m APP^{(i)}\left(X_0 = \alpha_m\right)\left|\alpha_m\right|^2, \cdots, \sum_m APP^{(i)}\left(X_{N-1} = \alpha_m\right)\left|\alpha_m\right|^2\right]^{\mathrm{T}}\right). \quad (9)$$

For a phase modulated system, all symbols have the same energy $\mathcal{E}$. Thus, (7) can be simplified into

$$\mathbf{h}^{(i+1)} = \frac{1}{\mathcal{E}N}\mathbf{\Omega}^{\dagger}\widetilde{\text{diag}\,(\mathbf{X})}^{(i)\dagger}\mathbf{Y}. \quad (10)$$

C. NCD-EM Channel Estimation

In this section, the NCD-EM algorithm is derived with the complete data chosen by decomposing the noise and observation components. NCD-EM channel estimation for an uncoded OFDM system has been introduced in [7] and the NCD-EM algorithm for a coded OFDM system is derived here.

In order to estimate the channel impulse response $\mathbf{h}$, we re-write (2) as:

$$\mathbf{Y} = \mathbf{A}\mathbf{h} + \mathbf{N} = \sum_{l=0}^{L-1}\mathbf{A}_l h_l + \mathbf{N}, \quad (11)$$

where $\mathbf{A} = \text{diag}\left(\mathbf{X}\right)\mathbf{\Omega}$, $\mathbf{A}_l$ is the lth column of matrix $\mathbf{A}$ and $[\mathbf{A}]_{k,l} = a_{k,l}$. From (11), we get

$$Y_k = \sum_{l=0}^{L-1} a_{k,l}h_l + N_k, \qquad 0 \le k \le N-1. \quad (12)$$

The noise and observed data are decomposed as in [7]:

$$\mathbf{N} = \sum_{l=0}^{L-1}\mathbf{N}_l, \quad \mathbf{Y} = \sum_{l=0}^{L-1}\left(\mathbf{A}_l h_l + \mathbf{N}_l\right) = \sum_{l=0}^{L-1}\mathbf{Z}_l, \quad (13)$$

where $\mathbf{N}_l = \left(N_{0,l}, \cdots, N_{N-1,l}\right)^{\mathrm{T}}$ with a variance $2\sigma_l^2 = 2\beta_l\sigma^2$, $\mathbf{Z}_l = \left(Z_{0,l}, \cdots, Z_{N-1,l}\right)^{\mathrm{T}}$ and $\mathbf{Z}_l = \mathbf{A}_l h_l + \mathbf{N}_l$. Here, the noise variance factors β_l's, which satisfy $\sum_{l=0}^{L-1}\beta_l = 1$, are important parameters which control the rate of convergence of the NCD-EM algorithm [13]. We choose the complete data as $\kappa = (\mathbf{Z}, \mathbf{A})$, where $\mathbf{Z} = (\mathbf{Z}_0, \cdots, \mathbf{Z}_{L-1})$, and the auxiliary function for the NCD-EM can be written as

$$Q\left(\mathbf{h}|\mathbf{h}^{(i)}\right) = E_\kappa\left[\log p\left(\kappa\,|\mathbf{h}\right)|\mathbf{Y}, \mathbf{h}^{(i)}\right] = \sum_{\mathbf{A}}\int_{\mathbf{Z}} \mathrm{P}\left(\mathbf{Z},\, \mathbf{A}\,|\mathbf{Y}, \mathbf{h}^{(i)}\right)\log p\left(\mathbf{Z},\, \mathbf{A}\,|\mathbf{h}\right) d\,\mathbf{Z}. \quad (14)$$

With equiprobable transmitted symbols, the auxiliary function becomes

$$Q\left(\mathbf{h}|\mathbf{h}^{(i)}\right) = \sum_{\mathbf{A}}\int_{\mathbf{Z}} \mathrm{P}\left(\mathbf{Z},\, \mathbf{A}\,|\mathbf{Y}, \mathbf{h}^{(i)}\right)\log p\left(\mathbf{Z}\,|\,\mathbf{A},\, \mathbf{h}\right) d\,\mathbf{Z}. \quad (15)$$

After some derivations given in Appendix A, we get (16) at the bottom of the next page, where $\mathbf{a}_k$ =

TABLE 1
NUMBER OF COMPLEX MULTIPLICATIONS OF CL-EM AND NCD-EM IN ONE ITERATION (QUADRATURE AMPLITUDE MODULATION)

	CL-EM	NCD-EM
$\left(\mathbf{\Omega}^\dagger \mathbf{R}_{N\times N}^{(i)*} \mathbf{\Omega}\right)^{-1} \mathbf{\Omega}^\dagger \widetilde{\mathbf{diag}\,(\mathbf{X})}^{(i)\dagger} \mathbf{Y}$	$N \times (L^2+2L+1) + \frac{4}{3}L^3 + L^2$	
$\mathbf{Tr}(\mathbf{R}_{N\times N}^{(i)})\mathbf{h}^{(i)}$		L
$\mathbf{\Omega}^\dagger \widetilde{\mathbf{diag}\,(\mathbf{X})}^{(i)\dagger} \mathbf{Y}$		$N \times (L+1)$
$\mathbf{\Omega}^\dagger \mathbf{R}_{N\times N}^{(i)} \mathbf{\Omega} \mathbf{h}^{(i)}$		$N \times (2L+1)$
Total	$N \times (L^2+2L+1) + \frac{4}{3}L^3 + L^2$	$N \times (3L+2) + L$

$(a_{k,0}, \cdots, a_{k,l}, \cdots, a_{k,L-1})$ is the kth row of the matrix $\mathbf{A}$, ς_l is the lth component of vector $\boldsymbol{\varsigma}$ and [9]

$$E\left\{Z_{k,l}|\mathbf{a}_k = \boldsymbol{\varsigma}, \mathbf{Y}, \mathbf{h}^{(i)}\right\} = h_l^{(i)}\varsigma_l + \beta_l\left(Y_k - \sum_{l=0}^{L-1} h_l^{(i)}\varsigma_l\right). \tag{17}$$

We choose $\beta_l = \dfrac{1}{L}$ [9] [14] in our simulations, which provides the best rate of convergence as it will be proved in the next section. From (11), $\mathbf{a}_k = X_k\mathbf{\Omega}_k$, where $\mathbf{\Omega}_k$ is the kth row of the matrix $\mathbf{\Omega}$. Hence, the *a posteriori* conditional probability in (16) can be written as $\mathrm{P}\left(\mathbf{a}_k = \boldsymbol{\varsigma}|\mathbf{Y}, \mathbf{h}^{(i)}\right) = \mathrm{P}\left(X_k = \alpha_m|\mathbf{Y}, \mathbf{h}^{(i)}\right)$. The *a posteriori* conditional probability is produced by the decoder. Taking the partial derivative of (16) with respect to each channel coefficient h_l and making the derivative equal to zero, the new channel coefficient estimates $h_l^{(i+1)}$ may be expressed as in (18). For phase modulated systems, (18) can be simplified into (19) at the bottom of the page.

IV. COMPLEXITY AND CONVERGENCE RATE COMPARISON

In this section, the complexity of CL-EM and NCD-EM are analyzed and compared.

A. Complexity per Iteration

In order to compare the complexity of CL-EM and NCD-EM in one iteration, we re-write (18) in a matrix form as in (20) at the bottom of the next page. For phase modulated system, (20) becomes:

$$\mathbf{h}^{(i+1)} = (1 - \frac{1}{L})\mathbf{h}^{(i)} + \frac{1}{\mathcal{E}NL}\mathbf{\Omega}^\dagger \widetilde{\mathrm{diag}\,(\mathbf{X})}^{(i)\dagger}\mathbf{Y}. \tag{21}$$

We check the number of complex multiplications in (7), (20), (10) and (21) as shown in Table 1 [16] and Table 2. For QAM modulation, CL-EM in most cases needs more computations than NCD-EM in one iteration, and the ratio of numbers of complex multiplications is ($N \gg L^3$ and $L \gg 1$)

$$\frac{N \times (L^2+2L+1) + \frac{4}{3}L^3 + L^2}{N \times (3L+2) + L} \approx \frac{L^2+2L+1}{3L+2} \approx \frac{L}{3}. \tag{22}$$

For phase modulated system, CL-EM needs L computations less than NCD-EM in one iteration.

TABLE 2
NUMBER OF COMPLEX MULTIPLICATIONS OF CL-EM AND NCD-EM IN ONE ITERATION (PHASE MODULATION)

	CL-EM	NCD-EM
$\frac{1}{\mathcal{E}N}\mathbf{\Omega}^\dagger \widetilde{\mathbf{diag}\,(\mathbf{X})}^{(i)\dagger}\mathbf{Y}$	$N \times (L+1) + L$	
$(1 - \frac{1}{L})\mathbf{h}^{(i)}$		L
$\frac{1}{\mathcal{E}NL}\mathbf{\Omega}^\dagger \widetilde{\mathbf{diag}\,(\mathbf{X})}^{(i)\dagger}\mathbf{Y}$		$N \times (L+1) + L$
Total	$N \times (L+1) + L$	$N \times (L+1) + 2L$

$$Q\left(\mathbf{h}|\mathbf{h}^{(i)}\right) = -\frac{1}{2\sigma^2}\sum_{l=0}^{L-1}|h_l|^2\sum_{k=0}^{N-1}\sum_{\boldsymbol{\varsigma}}|\varsigma_l|^2\mathrm{P}\left(\mathbf{a}_k = \boldsymbol{\varsigma}|\mathbf{Y}, \mathbf{h}^{(i)}\right) + \frac{1}{\sigma^2}\sum_{l=0}^{L-1}\sum_{k=0}^{N-1}\Re e\left\{h_l^*\sum_{\boldsymbol{\varsigma}}\varsigma_l^* E\left\{Z_{k,l}|\mathbf{a}_k = \boldsymbol{\varsigma}, \mathbf{Y}, \mathbf{h}^{(i)}\right\}\mathrm{P}\left(\mathbf{a}_k = \boldsymbol{\varsigma}|\mathbf{Y}, \mathbf{h}^{(i)}\right)\right\}. \tag{16}$$

$$h_l^{(i+1)} = \frac{\sum_{k=0}^{N-1}\sum_{\alpha_m}\alpha_m^* e^{j2\pi\frac{(l-1)(k-1)}{N}} E\left\{Z_{k,l}|X_k = \alpha_m, \mathbf{Y}, \mathbf{h}^{(i)}\right\}\mathrm{P}\left(X_k = \alpha_m|\mathbf{Y}, \mathbf{h}^{(i)}\right)}{\sum_{k=0}^{N-1}\sum_{\alpha_m}|\alpha_m|^2\mathrm{P}\left(X_k = \alpha_m|\mathbf{Y}, \mathbf{h}^{(i)}\right)}. \tag{18}$$

$$h_l^{(i+1)} = \frac{1}{\mathcal{E}N}\sum_{k=0}^{N-1}\sum_{\alpha_m}\alpha_m^* e^{j2\pi\frac{(l-1)(k-1)}{N}} E\left\{Z_{k,l}|X_k = \alpha_m, \mathbf{Y}, \mathbf{h}^{(i)}\right\}\mathrm{P}\left(X_k = \alpha_m|\mathbf{Y}, \mathbf{h}^{(i)}\right) \tag{19}$$

The reader should notice that the modulation size does not appear in Tables 1 and 2, because expectations based on *a posteriori* probability, which are common operations to both estimation methods, are not taken into account.

B. Analysis of the Convergence Rate

From [4] and [17], we know that the rate of convergence of the EM algorithm can be evaluated as:

$$\mathbf{DM}\left(\mathbf{h}^c\right) = \mathbf{D}^{11}H\left(\mathbf{h}^c|\mathbf{h}^c\right)\left[\mathbf{D}^{11}Q\left(\mathbf{h}^c|\mathbf{h}^c\right)\right]^{-1}, \quad (23)$$

where $\mathbf{h}^c$ represents the limit point of the converging sequence $\mathbf{h}^i$,

$$\mathbf{D}^{11}H\left(\mathbf{h}^c|\mathbf{h}^c\right) = E_\kappa\left\{\frac{\partial}{\partial \mathbf{h}^c}\log\, p\left(\kappa|\mathbf{Y},\mathbf{h}^c\right)\left[\frac{\partial}{\partial \mathbf{h}^c}\log\, p\left(\kappa|\mathbf{Y},\mathbf{h}^c\right)\right]^\dagger|\mathbf{Y},\mathbf{h}^c\right\}, \quad (24)$$

and

$$\mathbf{D}^{11}Q\left(\mathbf{h}^c|\mathbf{h}^c\right) = E_\kappa\left\{\frac{\partial}{\partial \mathbf{h}^c}\log\, p\left(\kappa|\mathbf{h}^c\right)\left[\frac{\partial}{\partial \mathbf{h}^c}\log\, p\left(\kappa|\mathbf{h}^c\right)\right]^\dagger|\mathbf{Y},\mathbf{h}^c\right\}. \quad (25)$$

The largest eigenvalue of $\mathbf{DM}\left(\mathbf{h}^c\right)$ gives the rate of convergence for the algorithm.

Recall the relationship between the Fisher information and the rate of convergence of the EM algorithm [4]:
a) $\mathbf{D}^{11}H\left(\mathbf{h}^c|\mathbf{h}^c\right)$ is the Fisher information in the unobserved part of κ about $\mathbf{h}^c$;
b) $\mathbf{D}^{11}Q\left(\mathbf{h}^c|\mathbf{h}^c\right)$ is the Fisher information in the complete data κ about $\mathbf{h}^c$, and $\mathbf{D}^{11}Q\left(\mathbf{h}^c|\mathbf{h}^c\right)$ can be written as:

$$\mathbf{D}^{11}Q\left(\mathbf{h}^c|\mathbf{h}^c\right) = \mathbf{D}^{11}H\left(\mathbf{h}^c|\mathbf{h}^c\right) + \mathbf{D}^2L\left(\mathbf{h}^c\right), \quad (26)$$

where $L\left(\mathbf{h}^c\right) = \log p(\mathbf{Y}|\mathbf{h^c})$; $\mathbf{D}^2L\left(\mathbf{h}^c\right) = E_{\mathbf{Y}}\left\{\frac{\partial}{\partial \mathbf{h}^c}L\left(\mathbf{h}^c\right)\left[\frac{\partial}{\partial \mathbf{h}^c}L\left(\mathbf{h}^c\right)\right]^\dagger\right\}$ is a measure of the information in the incomplete data $\mathbf{Y}$.

Therefore, the rate of convergence for the EM algorithm in (23) can be interpreted as:

$$\text{Rate of Convergence} = \frac{\text{Lost information due to unobserved data}}{\text{All information in complete data}}. \quad (27)$$

Thus, a smaller rate of convergence implies that less information is lost and the EM algorithm converges faster.

1) CL-EM Algorithm: For CL-EM, the complete data is $\kappa = (\mathbf{X},\mathbf{Y})$. From (24), we get

$$\mathbf{D}^{11}H\left(\mathbf{h}^c|\mathbf{h}^c\right) = E_{\mathbf{X}}\left\{\frac{\partial}{\partial \mathbf{h}^c}\log\, p\left(\mathbf{X},\mathbf{Y}|\mathbf{Y},\mathbf{h}^c\right)\left[\frac{\partial}{\partial \mathbf{h}^c}\log\, p\left(\mathbf{X},\mathbf{Y}|\mathbf{Y},\mathbf{h}^c\right)\right]^\dagger|\mathbf{Y},\mathbf{h}^c\right\}. \quad (28)$$

With the derivations in Appendix B assuming high SNR, we get

$$\mathbf{D}^{11}H\left(\mathbf{h}^c|\mathbf{h}^c\right) \approx 0. \quad (29)$$

From (25) and the complete data of CL-EM,

$$\mathbf{D}^{11}Q\left(\mathbf{h}^c|\mathbf{h}^c\right) = E_{\mathbf{X}}\left\{\frac{\partial}{\partial \mathbf{h}^c}\log\, p\left(\mathbf{X},\mathbf{Y}|\mathbf{h}^c\right)\left[\frac{\partial}{\partial \mathbf{h}^c}\log\, p\left(\mathbf{X},\mathbf{Y}|\mathbf{h}^c\right)\right]^\dagger\Big|\mathbf{Y},\mathbf{h}^c\right\}. \quad (30)$$

Considering the equiprobability of transmitted data, we rewrite (30) as (31), where $\mathbf{R}_{\mathrm{N}\times\mathrm{N}}$ is always greater than 0.

Therefore, at high SNR, CL-EM algorithm always leads to

$$\mathbf{DM}\left(\mathbf{h}^c\right) \approx 0. \quad (32)$$

With this rate of convergence close to zero, we can expect that CL-EM should be already very close to its fix point at the second iteration, i.e., after the first one based on pilots.

2) NCD-EM Algorithm: For NCD-EM, the complete data is $\kappa = (\mathbf{Z},\mathbf{A})$. From (24), we have

$$\mathbf{D}^{11}H\left(\mathbf{h}^c|\mathbf{h}^c\right) = E_{(\mathbf{Z},\mathbf{A})}\left\{\frac{\partial}{\partial \mathbf{h}^c}\log\, p\left(\mathbf{Z},\mathbf{A}|\mathbf{Y},\mathbf{h}^c\right)\left[\frac{\partial}{\partial \mathbf{h}^c}\log\, p\left(\mathbf{Z},\mathbf{A}|\mathbf{Y},\mathbf{h}^c\right)\right]^\dagger\Big|\mathbf{Y},\mathbf{h}^c\right\}, \quad (33)$$

With the derivations in Appendix C, where $\boldsymbol{\beta} = [\beta_0,\cdots,\beta_{L-1}]^{\mathrm{T}}$, $\mathcal{E}$ is the average energy of transmitted symbols and $\mathbf{I}_L$ represents the identity matrix of size L, we have, assuming high SNR,

$$\mathbf{D}^{11}H\left(\mathbf{h}^c|\mathbf{h}^c\right) \approx \frac{1}{2\sigma^2}N\mathcal{E}\left[\mathrm{diag}\left(\boldsymbol{\beta}\right)^{-1} - \mathbf{I}_L\right]. \quad (34)$$

From (25),

$$\mathbf{D}^{11}Q\left(\mathbf{h}^c|\mathbf{h}^c\right) = E_{(\mathbf{Z},\mathbf{A})}\left\{\frac{\partial}{\partial \mathbf{h}^c}\log\, p\left(\mathbf{Z},\mathbf{A}|\mathbf{h}^c\right)\left[\frac{\partial}{\partial \mathbf{h}^c}\log\, p\left(\mathbf{Z},\mathbf{A}|\mathbf{h}^c\right)\right]^\dagger|\mathbf{Y},\mathbf{h}^c\right\}. \quad (35)$$

$$\mathbf{h}^{(i+1)} = \frac{1}{\mathrm{Tr}(\mathbf{R}^{(i)}_{N\times N})}\left[\mathrm{Tr}(\mathbf{R}^{(i)}_{N\times N})\mathbf{h}^{(i)} + \frac{1}{L}\boldsymbol{\Omega}^\dagger\widetilde{\mathrm{diag}\left(\mathbf{X}\right)}^{(i)\dagger}\mathbf{Y} - \frac{1}{L}\boldsymbol{\Omega}^\dagger\mathbf{R}^{(i)}_{N\times N}\boldsymbol{\Omega}\mathbf{h}^{(i)}\right] \quad (20)$$

$$\frac{1}{\left(2\sigma^2\right)^2}E_{\mathbf{X}}\left\{\boldsymbol{\Omega}^{\mathrm{T}}\mathrm{diag}\left(\mathbf{X}\right)^{\mathrm{T}}\left[\mathbf{Y}^* - \mathrm{diag}\left(\mathbf{X}\right)^*\boldsymbol{\Omega}^*\mathbf{h}^{c*}\right]\left[\mathbf{Y}^* - \mathrm{diag}\left(\mathbf{X}\right)^*\boldsymbol{\Omega}^*\mathbf{h}^{c*}\right]^\dagger\mathrm{diag}\left(\mathbf{X}\right)^*\boldsymbol{\Omega}^*\Big|\mathbf{Y},\mathbf{h}^c\right\}$$
$$= \frac{1}{2\sigma^2}\sum_{\mathbf{X}_j}APP^{(i)}\left(\mathbf{X}_j\right)\boldsymbol{\Omega}^{\mathrm{T}}\mathrm{diag}\left(\mathbf{X}_j\right)^{\mathrm{T}}\mathrm{diag}\left(\mathbf{X}_j\right)^*\boldsymbol{\Omega}^* = \frac{1}{2\sigma^2}\boldsymbol{\Omega}^{\mathrm{T}}\mathbf{R}_{\mathrm{N}\times\mathrm{N}}\boldsymbol{\Omega}^* \quad (31)$$

Using $\log p(\mathbf{Z},\mathbf{A}|\mathbf{h}^c) = \log p(\mathbf{Z}|\mathbf{A},\mathbf{h}^c) + \log p(\mathbf{A})$ and (C-4), we get

$$\mathbf{D}^{11}Q(\mathbf{h}^c|\mathbf{h}^c) = \frac{1}{2\sigma^2}N\mathcal{E}\mathrm{diag}(\boldsymbol{\beta})^{-1}. \quad (36)$$

Substituting (34) and (36) into (23), we obtain $\mathbf{DM}(\mathbf{h}^c) = \mathbf{I}_L - \mathrm{diag}(\boldsymbol{\beta})$. The eigenvalues of $\mathbf{DM}(\mathbf{h}^c)$ are $1-\beta_l$, $0 \le l \le L-1$, which explains why $\beta_l = \frac{1}{L}$ is the best choice as suggested by simulation results in [9] and [14]: the largest eigenvalue is minimized and the best NCD-EM convergence is achieved. With this choice, $\mathbf{DM}(\mathbf{h}^c)$ can be written as $\mathbf{DM}(\mathbf{h}^c) = \mathrm{diag}\left(\left[1-\frac{1}{L},\cdots,1-\frac{1}{L}\right]^{\mathrm{T}}\right)$ and we see the relationship between the rate of convergence and the number of taps in the channel: the more taps there are in the channel, the slower the NCD-EM converges.

C. Overall complexity

Based on sections IV-A and IV-B, we can analyze the overall complexity of the CL-EM and the NCD-EM. Since the best rate of convergence of NCD-EM is $1-\frac{1}{L}$, we assume that the number of iterations for convergence is L. Thus, the total number of complex multiplications for convergence is $L^2 + N(3L^2+2L)$. However, the numerical results in Fig. 6 and Fig. 7 will show that the actual number of iterations of NCD-EM is larger than L. On the other hand, the rate of convergence of CL-EM is approximately 0. Therefore, the total number of complex multiplications for convergence is approximately equal to that of a single iteration.

First, we check the difference $\mathrm{D}(L,N)$ between the number of complex multiplications in NCD-EM denoted by $N_{\mathrm{NCD_{overall}}}$ and the number of complex multiplications in CL-EM denoted by $N_{\mathrm{CL_{overall}}}$. With QAM modulations:

$$\mathrm{D_{QAM}}(L,N) = N_{\mathrm{NCD_{overall}}} - N_{\mathrm{CL_{overall}}} = N\left(2L^2-1\right) - \frac{4}{3}L^3. \quad (37)$$

Since, for $N > \frac{4}{3}$, $\mathrm{D_{QAM}}(1,N) > 0$ and, for $1 \le L < N$, $\frac{\partial}{\partial L}\mathrm{D_{QAM}}(L,N) > 0$, $\mathrm{D_{QAM}}(L,N)$ is always greater than zero with $N > \frac{4}{3}$ and $1 \le L < N$ which is always true in practice. Especially, when $N \gg L^3$ and $L \gg 1$, together with (22), the ratio of overall complexities can be expressed as

$$\frac{N_{\mathrm{CL_{overall}}}}{N_{\mathrm{NCD_{overall}}}} \approx \frac{L}{3\times L} \approx \frac{1}{3}. \quad (38)$$

Since the assumption of L iterations underestimates the number of iterations of NCD-EM, we can expect a smaller ratio in practice. Therefore, with a QAM modulation, even though CL-EM needs more computations than NCD-EM in one iteration, CL-EM has a lower overall complexity thanks to the almost zero rate of convergence.

For phase modulated systems, since CL-EM needs less computations in one iteration and converges faster than NCD-EM, it is obvious that CL-EM has a lower overall complexity than NCD-EM. With phase modulated systems, the ratio of overall complexities can be expressed as ($N \gg L^2$)

$$\frac{N_{\mathrm{CL_{overall}}}}{N_{\mathrm{NCD_{overall}}}} = \frac{N(L+1)+L}{N(L^2+L)+2L^2} \approx \frac{1}{L}. \quad (39)$$

V. CRAMER-RAO BOUND FOR CODED OFDM

The Cramér-Rao Bound (CRB) provides a lower bound for MSE to evaluate how good an unbiased estimator can be [18] [17]. Besides, modified CRB (MCRB) is a looser bound assuming perfect knowledge of the transmitted signal [19] [20]. Its computation is less complex. The CRB for the uncoded OFDM system and the MCRB for the OFDM system have been given in [7]. Here, we derive the CRB for a coded OFDM system. For the vector parameter $\mathbf{h}$ [17]:

$$CRB(h_l) = I_{ll}^{-1}(\mathbf{h}) \ , \quad l = 0,\cdots,L-1 \quad (40)$$

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

$$I(\mathbf{h}) = E_{\mathbf{Y}}\left\{\frac{\partial}{\partial \mathbf{h}}\log p(\mathbf{Y}|\mathbf{h})\left(\frac{\partial}{\partial \mathbf{h}}\log p(\mathbf{Y}|\mathbf{h})\right)^{\dagger}\right\}. \quad (41)$$

Making use of (B-8), we have

$$\frac{\partial}{\partial \mathbf{h}}\log p(\mathbf{Y}|\mathbf{h}) = -\frac{1}{2\sigma^2}\boldsymbol{\Omega}^T\widetilde{\mathrm{diag}(\mathbf{X})}^{(i)\mathrm{T}}\mathbf{Y}^* + \frac{1}{2\sigma^2}\boldsymbol{\Omega}^T\mathbf{R}_{N\times N}^{(i)}\boldsymbol{\Omega}^*\mathbf{h}^*. \quad (42)$$

Substituting (42) into (41), we obtain the Fisher information matrix. The CRB for a single parameter is $I_{ii}^{-1}(\mathbf{h})$ and the overall CRB for all parameters is

$$CRB(\mathbf{h}) = \sum_{l=0}^{L-1} CRB(h_l) = \mathrm{Tr}\left(I^{-1}(\mathbf{h})\right). \quad (43)$$

In our simulations, the Fisher information is obtained from the average over a large number of OFDM symbols and the final CRB is obtained by substituting the Fisher information into (43).

VI. NUMERICAL RESULTS

In order to demonstrate the validity and the effectiveness of the proposed EM-based channel estimation algorithms, the coded OFDM system introduced in section II has been simulated. A 6-tap rectangular Rayleigh ISI channel has been considered. The entire channel bandwidth is divided into 128 sub-carriers and the cyclic prefix length is 16 samples. A 1/2-rate $(133,171)_8$ nonsystematic nonrecursive convolutional code and a pseudo-random interleaver of size 480 are used in simulations. The modulation scheme is 16-QAM. Pilot symbols are uniformly inserted into every OFDM symbol.

A. Cramér-Rao Bound

Figure 3 shows CRBs for OFDM systems without coding and with two different coding schemes: the first code is the rate 1/2 code mentioned above with minimum distance 10; the second code is a rate 1/5 convolutional code with minimum distance 25. Improving the code performance at a given signal-to-noise ratio (E_s/N_0) results in a lower CRB. For high

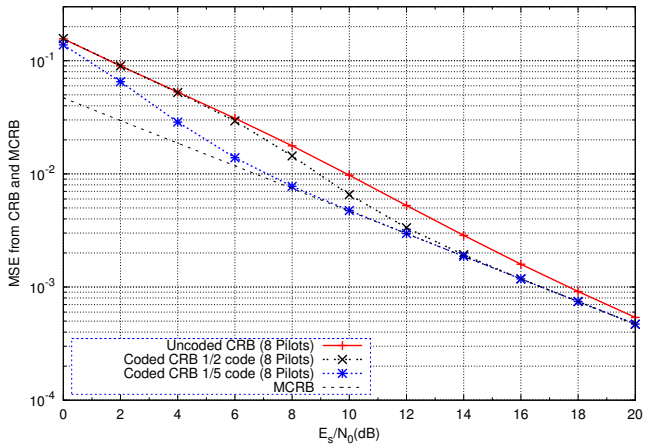


Fig. 3. Comparison of Cramér-Rao bounds and modified Cramér-Rao bound. Modulation is 16-QAM. The channel has 6 taps.

E_s/N_0, all CRBs converge to the MCRB as APP information from the decoder becomes perfect; for low E_s/N_0, CRBs for the coded systems converge to the CRB for the uncoded system as decoding does not bring improvement anymore.

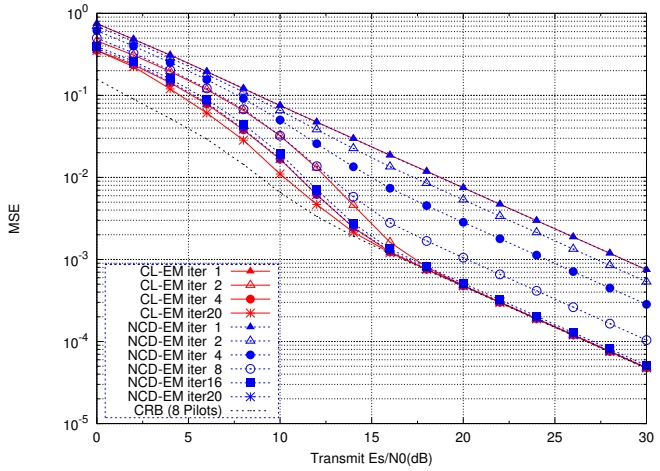


Fig. 4. Mean square error (MSE) performances for CL-EM and NCD-EM in coded OFDM system. The number of pilot symbols is 8. Modulation is 16-QAM. The channel has 6 taps.

B. CL-EM versus NCD-EM

We also compare CL-EM and NCD-EM. From Fig. 4, we see that NCD-EM has slower convergence than CL-EM, it does not achieve CRB before the 18-th iteration. Similar behavior is observed in Fig. 5 for BER performance, whereas CL-EM approaches performance with perfect channel state information (CSI) after 4 iterations only.

C. Rate of Convergence and Complexity

In order to show the convergence of CL-EM and NCD-EM more clearly, we draw the ratio of estimate differences in Fig. 6 and Fig. 7, where

$$\text{ratio of estimate differences} = \frac{\|\mathbf{h}^{(i+1)} - \mathbf{h}\|}{\|\mathbf{h}^{(i)} - \mathbf{h}\|}. \qquad (44)$$

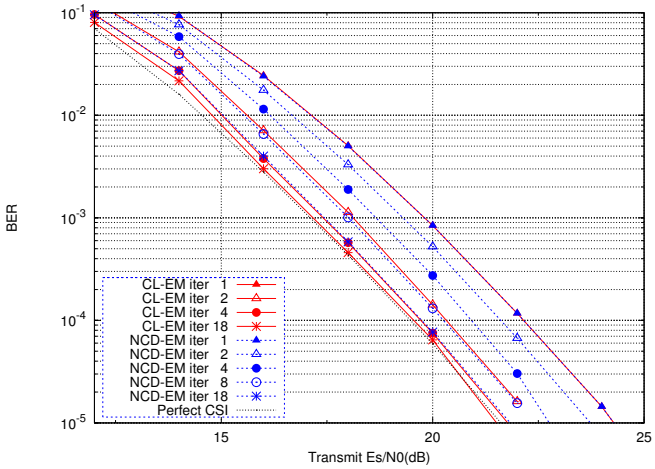


Fig. 5. Bit error rate (BER) performances for CL-EM and NCD-EM in coded OFDM system. The number of pilot symbols is 8. Modulation is 16-QAM. The channel has 6 taps.

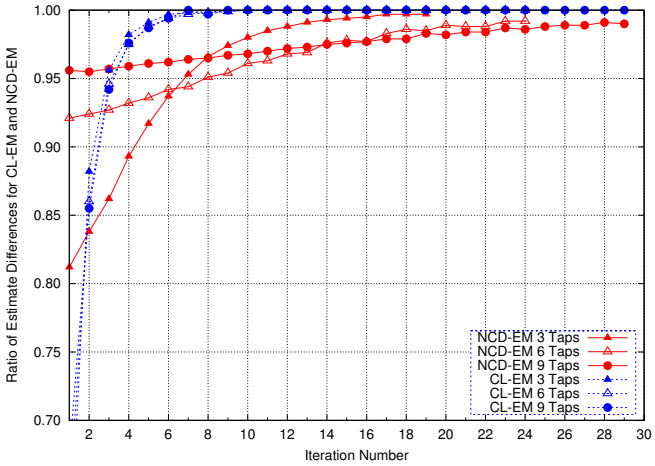


Fig. 6. Ratio of estimate differences for CL-EM and NCD-EM in coded OFDM at E_s/N_0=10dB. The number of pilot symbols is 16. Modulation is 16-QAM.

A ratio of estimate differences equal to 1 means that EM algorithm has converged.

In Fig. 6, at moderate SNR ($E_s/N_0 = 10$dB), we observe that CL-EM needs a dozen of iterations to converge; however, it is at least twice faster than NCD-EM under similar conditions. Furthermore, with different numbers of taps in the channel, CL-EM always converges with almost the same number of iterations. However, NCD-EM convergence speed depends on the number of channel taps. This result is consistent with the theoretical analysis.

In Fig. 7, at high SNR ($E_s/N_0 = 20$dB), we can see that CL-EM always converges from the second iteration (the first iteration includes a pilot-based estimation). As expected from (32), the rate of convergence of CL-EM is almost zero for high SNR.

The theoretical analysis of convergence and complexity accomplished in section IV-C assumes high SNR. Good results are also measured at low SNR as can be seen in Fig. 6. For 9 taps, the ratio of number of iterations between CL-EM and

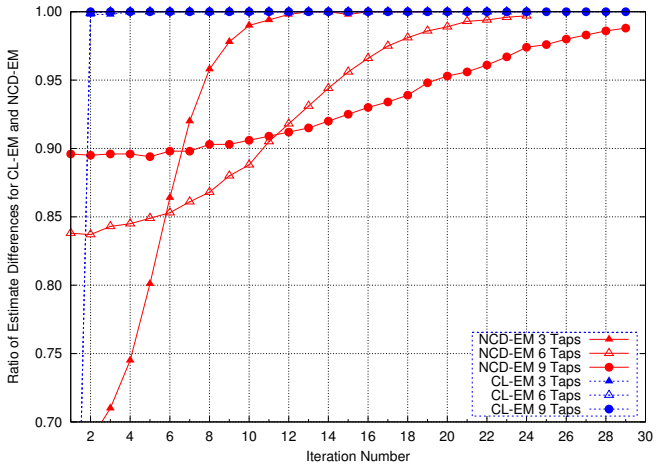


Fig. 7. Ratio of estimate differences for CL-EM and NCD-EM in coded OFDM at E_s/N_0=20dB. The number of pilot symbols is 16. Modulation is 16-QAM.

NCD-EM is about 20% and the complexity ratio is about 69%. These ratios are 25% and 61% respectively for a channel with 6 taps.

According to section IV-C, the complexity of CL-EM should be about 41% of NCD-EM. This ratio is obtained with the assumption that NCD-EM converges after L iterations which is underestimated. When considering MSE performance, in Fig. 4, NCD-EM achieves CRB after $3L$ iterations. Thus, the ratio of overall complexities between CL-EM and NCD-EM is actually 14%. Concerning BER performance, as shown in Fig. 5, NCD-EM needs 18 iterations to achieve the same performance as CL-EM with only 4 iterations and the ratio of overall complexities is approximately 55%. However, in the iterative OFDM receiver, the soft demapping and decoding procedures are also included in every iteration. If we consider these additional processes, the iterative receiver with CL-EM channel estimator will be much less complex than that with NCD-EM channel estimator. Therefore, CL-EM algorithm has a better performance-complexity trade-off.

VII. Conclusions

We have derived EM based iterative algorithms for the estimation of the channel impulse response in a coded OFDM system and compared the CL-EM algorithm using the classic complete data with the NCD-EM algorithm where the complete data is obtained via decomposing noise and observation components. The CRB for a coded OFDM system has also been established as a lower bound for MSE performance. Theoretical analysis shows that the convergence rate of CL-EM is almost zero at high SNR, while that of NCD-EM depends on the number of channel taps and the choice of noise variance factors. Taking into account the complexity per iteration, CL-EM is less complex than NCD-EM for all linear modulations at the same error rate performance. Thus, CL-EM is an excellent OFDM channel estimator with a reasonable complexity.

VIII. Acknowledgement

This work has been performed in the framework of the CELTIC project CP5-026 WINNER+.

Appendix A
Derivation of (16)

Using (13), we get

$$p\left(\mathbf{Z}|\mathbf{A},\mathbf{h}\right)=\frac{1}{\left(2\pi\sigma^2\right)^N}\exp\left\{-\frac{1}{2\sigma^2}\sum_{l=0}^{L-1}\left\|\mathbf{Z}_l-\mathbf{A}_l h_l\right\|^2\right\} \tag{A-1}$$

and

$$\log p\left(\mathbf{Z}|\mathbf{A},\mathbf{h}\right)=-\frac{1}{2\sigma^2}\sum_{l=0}^{L-1}\left\|\mathbf{Z}_l-\mathbf{A}_l h_l\right\|^2-\gamma. \tag{A-2}$$

Using (15) and (A-2) and neglecting the constant number γ which will not impact the subsequent partial derivative, we get the auxiliary function of NCD-EM:

$$Q\left(\mathbf{h}|\mathbf{h}^{(i)}\right)= \sum_{\mathbf{A}}\int_{\mathbf{Z}}-\frac{1}{2\sigma^2}\sum_{l=0}^{L-1}\left\|\mathbf{Z}_l-\mathbf{A}_l h_l\right\|^2 \mathrm{P}\left(\mathbf{Z},\,\mathbf{A}\,|\mathbf{Y},\mathbf{h}^{(i)}\right)d\,\mathbf{Z}. \tag{A-3}$$

In (A-3),

$$\sum_{l=0}^{L-1}\left\|\mathbf{Z}_l-\mathbf{A}_l h_l\right\|^2=\sum_{l=0}^{L-1}\sum_{k=0}^{N-1}\left|Z_{k,l}-a_{k,l}h_l\right|^2 =\sum_{l=0}^{L-1}\sum_{k=0}^{N-1}\left|Z_{k,l}\right|^2-2\Re e\left\{Z_{k,l}a_{k,l}^{*}h_l^{*}\right\}+\left|a_{k,l}h_l\right|^2. \tag{A-4}$$

Substituting (A-4) into (A-3), we get

$$Q\left(\mathbf{h}|\mathbf{h}^{(i)}\right)=\mathcal{A}+\mathcal{B}+\mathcal{C}, \tag{A-5}$$

where

$$\mathcal{A}=-\frac{1}{2\sigma^2}\sum_{\mathbf{A}}\int_{\mathbf{Z}}\sum_{l=0}^{L-1}\sum_{k=0}^{N-1}\left|Z_{k,l}\right|^2\mathrm{P}\left(\mathbf{Z},\,\mathbf{A}\,|\mathbf{Y},\mathbf{h}^{(i)}\right)d\,\mathbf{Z}, \tag{A-6a}$$

$$\mathcal{B}=\frac{1}{2\sigma^2}\sum_{\mathbf{A}}\int_{\mathbf{Z}}\sum_{l=0}^{L-1}\sum_{k=0}^{N-1}2\Re e\left\{Z_{k,l}a_{k,l}^{*}h_l^{*}\right\} \mathrm{P}\left(\mathbf{Z},\,\mathbf{A}\,|\mathbf{Y},\mathbf{h}^{(i)}\right)d\,\mathbf{Z}, \tag{A-6b}$$

$$\mathcal{C}=-\frac{1}{2\sigma^2}\sum_{\mathbf{A}}\int_{\mathbf{Z}}\sum_{l=0}^{L-1}\sum_{k=0}^{N-1}\left|a_{k,l}h_l\right|^2\mathrm{P}\left(\mathbf{Z},\,\mathbf{A}\,|\mathbf{Y},\mathbf{h}^{(i)}\right)d\,\mathbf{Z}. \tag{A-6c}$$

Since the channel parameter $\mathbf{h}$ is not contained in (A-6a), $\frac{\partial\mathcal{A}}{\partial\mathbf{h}}=0$; the second item (A-6b) can be transformed to (A-7), where

$$E\left\{Z_{k,l}|\mathbf{A},\mathbf{Y},\mathbf{h}^{(i)}\right\}=\int_{\mathbf{Z}}Z_{k,l}\,\mathrm{P}\left(\mathbf{Z}\,|\mathbf{A},\mathbf{Y},\mathbf{h}^{(i)}\right)d\,\mathbf{Z}. \tag{A-8}$$

From (12) and (13), $Z_{k,l}$ and $a_{k,l}$ only depend on the kth row of $\mathbf{A}$, denoted as $\mathbf{a}_k = (a_{k,0}, \cdots, a_{k,l}, \cdots, a_{k,L-1})$. Thus, (A-7) can be written as

$$\mathcal{B} = \frac{1}{2\sigma^2} \sum_{l=0}^{L-1} \sum_{k=0}^{N-1} 2\Re e \left\{ h_l^* \sum_{\boldsymbol{\varsigma}} E\left\{ Z_{k,l} | \mathbf{a}_k = \boldsymbol{\varsigma}, \mathbf{Y}, \mathbf{h}^{(i)} \right\} \varsigma_l^* \right. \\ \left. \mathrm{P}\left(\mathbf{a}_k = \boldsymbol{\varsigma} \,|\mathbf{Y}, \mathbf{h}^{(i)} \right) \right\}, \quad \text{(A-9)}$$

where the vector $\boldsymbol{\varsigma}$ represents a possible value of vector $\mathbf{a}_k$ and ς_l is the lth component of $\boldsymbol{\varsigma}$. With (12), (13) and (A-8), we get (17). Furthermore, with the same notations, the item (A-5) can be written as (A-10). Substituting (A-9) and (A-10) into (A-6), we get the auxiliary function for NCD-EM in (16).

Appendix B
Derivation of (29)

In (28),

$$\log p(\mathbf{X}, \mathbf{Y} | \mathbf{Y}, \mathbf{h}^c) = \log p(\mathbf{X} | \mathbf{Y}, \mathbf{h}^c) \\ = \log p(\mathbf{X}, \mathbf{Y} | \mathbf{h}^c) - \log p(\mathbf{Y} | \mathbf{h}^c). \quad \text{(B-1)}$$

In (B-1), considering the independence between the equiprobable transmitted data and the channel parameters, we have

$$\log p(\mathbf{X}, \mathbf{Y} | \mathbf{Y}, \mathbf{h}^c) = \log \frac{1}{M} p(\mathbf{Y} | \mathbf{X}, \mathbf{h}^c) - \log p(\mathbf{Y} | \mathbf{h}^c). \quad \text{(B-2)}$$

Substituting (B-2) into (28), we obtain (B-3). For the first item in (B-3), the expectation can be written as

$$E_{\mathbf{X}} \left\{ \left\| \frac{\partial}{\partial \mathbf{h}^c} \log p(\mathbf{Y} | \mathbf{X}, \mathbf{h}^c) \right\|^2 \middle| \mathbf{Y}, \mathbf{h}^c \right\} = \\ \sum_{\mathbf{X}_j} \mathrm{P}(\mathbf{X}_j | \mathbf{Y}, \mathbf{h}^c) \left\| \frac{\partial}{\partial \mathbf{h}^c} \log p(\mathbf{Y} | \mathbf{X}_j, \mathbf{h}^c) \right\|^2, \quad \text{(B-4)}$$

where j enumerates all possible symbol sequences $\mathbf{X}_j$ of length N. For coded transmission, the log-likelihood function $\log p(\mathbf{Y} | \mathbf{h}^c)$ can be written as [21]:

$$\log p(\mathbf{Y} | \mathbf{h}^c) = \log \sum_{\mathbf{X}_j} \mathrm{P}(\mathbf{X}_j) \, p(\mathbf{Y} | \mathbf{X}_j, \mathbf{h}^c). \quad \text{(B-5)}$$

Differentiating (B-5) yields

$$\frac{\partial}{\partial \mathbf{h}^c} \log p(\mathbf{Y} | \mathbf{h}^c) = \\ \sum_{\mathbf{X}_j} \frac{\mathrm{P}(\mathbf{X}_j) \, p(\mathbf{Y} | \mathbf{X}_j, \mathbf{h}^c)}{p(\mathbf{Y}, \mathbf{h}^c)} \frac{\partial}{\partial \mathbf{h}^c} \log p(\mathbf{Y} | \mathbf{X}_j, \mathbf{h}^c). \quad \text{(B-6)}$$

Making use of Bayes' rule, we obtain

$$\frac{\mathrm{P}(\mathbf{X}_j) \, p(\mathbf{Y} | \mathbf{X}_j, \mathbf{h}^c)}{p(\mathbf{Y}, \mathbf{h})} = \mathrm{P}(\mathbf{X}_j | \mathbf{Y}, \mathbf{h}^c). \quad \text{(B-7)}$$

With (B-6), we get

$$\frac{\partial}{\partial \mathbf{h}^c} \log p(\mathbf{Y} | \mathbf{h}^c) = \sum_{\mathbf{X}_j} \mathrm{P}(\mathbf{X}_j | \mathbf{Y}, \mathbf{h}^c) \frac{\partial}{\partial \mathbf{h}^c} \log p(\mathbf{Y} | \mathbf{X}_j, \mathbf{h}^c). \quad \text{(B-8)}$$

Using (B-8) yields

$$E_{\mathbf{X}} \left\{ \left\| \frac{\partial}{\partial \mathbf{h}^c} \log p(\mathbf{Y} | \mathbf{h}^c) \right\|^2 \middle| \mathbf{Y}, \mathbf{h}^c \right\} = \\ \left\| \sum_{\mathbf{X}_j} \mathrm{P}(\mathbf{X}_j | \mathbf{Y}, \mathbf{h}^c) \frac{\partial}{\partial \mathbf{h}^c} \log p(\mathbf{Y} | \mathbf{X}_j, \mathbf{h}^c) \right\|^2 \quad \text{(B-9)}$$

and (B-10). With perfect APP (high SNR),

$$\sum_{\mathbf{X}_j} \mathrm{P}(\mathbf{X}_j | \mathbf{Y}, \mathbf{h}^c) \left\| \frac{\partial}{\partial \mathbf{h}^c} \log p(\mathbf{Y} | \mathbf{X}_j, \mathbf{h}^c) \right\|^2 \approx \\ \left\| \sum_{\mathbf{X}_j} \mathrm{P}(\mathbf{X}_j | \mathbf{Y}, \mathbf{h}^c) \frac{\partial}{\partial \mathbf{h}^c} \log p(\mathbf{Y} | \mathbf{X}_j, \mathbf{h}^c) \right\|^2. \quad \text{(B-11)}$$

Substituting (B-4), (B-9) and (B-10) into (B-3) and using (B-11), we get (29).

Appendix C
Derivation of (34)

In (33),

$$\log p(\mathbf{Z}, \mathbf{A} | \mathbf{Y}, \mathbf{h}^c) = \log p(\mathbf{Z} | \mathbf{A}, \mathbf{h}^c) - \log p(\mathbf{Y} | \mathbf{h}^c) \\ + \log p(\mathbf{Y} | \mathbf{Z}, \mathbf{A}, \mathbf{h}^c) + \log p(\mathbf{A}). \quad \text{(C-1)}$$

$$\mathcal{B} = \frac{1}{2\sigma^2} \sum_{l=0}^{L-1} \sum_{k=0}^{N-1} 2\Re e \left\{ h_l^* \sum_{\mathbf{A}} \int_{\mathbf{Z}} Z_{k,l} a_{k,l}^* \, \mathrm{P}\left(\mathbf{Z} \,|\mathbf{A}, \mathbf{Y}, \mathbf{h}^{(i)} \right) \mathrm{P}\left(\mathbf{A} \,|\mathbf{Y}, \mathbf{h}^{(i)} \right) d\mathbf{Z} \right\} \\ = \frac{1}{2\sigma^2} \sum_{l=0}^{L-1} \sum_{k=0}^{N-1} 2\Re e \left\{ h_l^* \sum_{\mathbf{A}} E\left\{ Z_{k,l} | \mathbf{A}, \mathbf{Y}, \mathbf{h}^{(i)} \right\} a_{k,l}^* \, \mathrm{P}\left(\mathbf{A} \,|\mathbf{Y}, \mathbf{h}^{(i)} \right) \right\} \quad \text{(A-7)}$$

$$\mathcal{C} = -\frac{1}{2\sigma^2} \sum_{l=0}^{L-1} \sum_{k=0}^{N-1} \sum_{\mathbf{A}} \int_{\mathbf{Z}} \mathrm{P}\left(\mathbf{Z} \,|\mathbf{A}, \mathbf{Y}, \mathbf{h}^{(i)} \right) |a_{k,l} h_l|^2 \mathrm{P}\left(\mathbf{A} \,|\mathbf{Y}, \mathbf{h}^{(i)} \right) d\mathbf{Z} \\ = -\frac{1}{2\sigma^2} \sum_{l=0}^{L-1} \sum_{k=0}^{N-1} \sum_{\mathbf{A}} |a_{k,l} h_l|^2 \mathrm{P}\left(\mathbf{A} \,|\mathbf{Y}, \mathbf{h}^{(i)} \right) = -\frac{1}{2\sigma^2} \sum_{l=0}^{L-1} |h_l|^2 \sum_{k=0}^{N-1} \sum_{\boldsymbol{\varsigma}} |\varsigma_l|^2 \mathrm{P}\left(\mathbf{a_k} = \boldsymbol{\varsigma} | \mathbf{Y}, \mathbf{h}^{(i)} \right) \quad \text{(A-10)}$$

With (13), $\mathbf{Y}$ is completely known from $\mathbf{Z}$, thus $\frac{\partial}{\partial \mathbf{h}^c} p\left(\mathbf{Y}|\mathbf{Z},\mathbf{A},\mathbf{h}^c\right) = 0$; since $p\left(\mathbf{A}\right)$ does not depend on $\mathbf{h}^c$, it will not impact the subsequent partial derivative. Therefore, we get (C-2). Using (12) and (13), we get

$$p\left(\mathbf{Z}|\mathbf{A},\mathbf{h}^c\right) \propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{l=0}^{L-1}\frac{1}{\beta_l}\left\|\mathbf{Z}_l - \mathbf{A}_l h_l^c\right\|^2\right\}. \quad \text{(C-3)}$$

Thus,

$$\begin{aligned}
&\frac{\partial}{\partial \mathbf{h}^c}\log p\left(\mathbf{Z}|\mathbf{A},\mathbf{h}^c\right) \\
&= \frac{1}{2\sigma^2}\left[\frac{1}{\beta_0}\left(\mathbf{Z}_0 - \mathbf{A}_0 h_0^c\right)^\dagger \mathbf{A}_0, \cdots,\right. \\
&\qquad \left.\frac{1}{\beta_{L-1}}\left(\mathbf{Z}_{L-1} - \mathbf{A}_{L-1} h_{L-1}^c\right)^\dagger \mathbf{A}_{L-1}\right]^{\mathrm{T}} \\
&= \frac{1}{2\sigma^2}\left[\frac{1}{\beta_0}\mathbf{N}_0^{c\,\dagger}\mathbf{A}_0, \cdots, \frac{1}{\beta_{L-1}}\mathbf{N}_{L-1}^{c\,\dagger}\mathbf{A}_{L-1}\right]^{\mathrm{T}}. \quad \text{(C-4)}
\end{aligned}$$

Using (B-8), we also have (C-5), where $\mathbf{A}_j = \mathbf{\Omega}\mathrm{diag}\left(\mathbf{X}_j\right)$. Substituting (C-4) and (C-5) into (C-2) and considering the perfect APP (high SNR), we get

$$E_{(\mathbf{Z},\mathbf{A})}\left\{\left\|\frac{\partial}{\partial \mathbf{h}^c}\log p\left(\mathbf{Z}|\mathbf{A},\mathbf{h}^c\right)\right\|^2 \middle| \mathbf{Y},\mathbf{h}^c\right\} = \frac{1}{2\sigma^2} N\mathcal{E}\mathrm{diag}\left(\boldsymbol{\beta}\right)^{-1}, \quad \text{(C-6)}$$

$$E_{(\mathbf{Z},\mathbf{A})}\left\{\left\|\frac{\partial}{\partial \mathbf{h}^c}\log p\left(\mathbf{Y}|\mathbf{h}^c\right)\right\|^2 \middle| \mathbf{Y},\mathbf{h}^c\right\} = \frac{1}{2\sigma^2} N\mathcal{E}\mathbf{I}_L, \quad \text{(C-7)}$$

and (C-8). Substituting (C-6), (C-7) and (C-8) into (C-2), we get (34).

REFERENCES

[1] J.K. Moon and S.I. Choi, "Performance of channel estimation methods for OFDM systems in a multipath fading channels," *IEEE Trans. Comsum. Electron.*, vol. 46, no. 1, pp. 161-170, Feb. 2000.

[2] J.-J. van de Beek, O. Edfors, and M. Sandell, "On channel estimation in OFDM systems," in *Proc. 45th IEEE Vehicular Technology Conference, 1995 (VTC'95)*, vol. 2, pp. 815-819, Sep. 1995.

$$\mathbf{D}^{11} H\left(\mathbf{h}^c|\mathbf{h}^c\right) = E_{\mathbf{X}}\left\{\left\|\frac{\partial}{\partial \mathbf{h}^c}\log p\left(\mathbf{Y}|\mathbf{X},\mathbf{h}^c\right)\right\|^2 \middle| \mathbf{Y},\mathbf{h}^c\right\} + E_{\mathbf{X}}\left\{\left\|\frac{\partial}{\partial \mathbf{h}^c}\log p\left(\mathbf{Y}|\mathbf{h}^c\right)\right\|^2 \middle| \mathbf{Y},\mathbf{h}^c\right\} - 2\Re e\left\{E_{\mathbf{X}}\left\{\frac{\partial}{\partial \mathbf{h}^c}\log p\left(\mathbf{Y}|\mathbf{X},\mathbf{h}^c\right)\left[\frac{\partial}{\partial \mathbf{h}^c}\log p\left(\mathbf{Y}|\mathbf{h}^c\right)\right]^\dagger \middle| \mathbf{Y},\mathbf{h}^c\right\}\right\}. \quad \text{(B-3)}$$

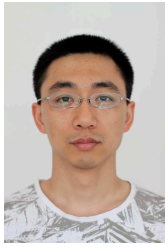
$$\begin{aligned}
&2\Re e\left\{E_{\mathbf{X}}\left\{\frac{\partial}{\partial \mathbf{h}^c}\log p\left(\mathbf{Y}|\mathbf{X},\mathbf{h}^c\right)\left[\frac{\partial}{\partial \mathbf{h}^c}\log p\left(\mathbf{Y}|\mathbf{h}^c\right)\right]^\dagger \middle| \mathbf{Y},\mathbf{h}^c\right\}\right\} \\
&= 2\Re e\left\{\left[\sum_{\mathbf{X}_j}\mathrm{P}\left(\mathbf{X}_j|\mathbf{Y},\mathbf{h}^c\right)\frac{\partial}{\partial \mathbf{h}^c}\log p\left(\mathbf{Y}|\mathbf{X}_j,\mathbf{h}^c\right)\right]^\dagger \sum_{\mathbf{X}_j}\mathrm{P}\left(\mathbf{X}_j|\mathbf{Y},\mathbf{h}^c\right)\frac{\partial}{\partial \mathbf{h}^c}\log p\left(\mathbf{Y}|\mathbf{X}_j,\mathbf{h}^c\right)\right\} \\
&= 2\left\|\sum_{\mathbf{X}_j}\mathrm{P}\left(\mathbf{X}_j|\mathbf{Y},\mathbf{h}^c\right)\frac{\partial}{\partial \mathbf{h}^c}\log p\left(\mathbf{Y}|\mathbf{X}_j,\mathbf{h}^c\right)\right\|^2. \quad \text{(B-10)}
\end{aligned}$$

$$\mathbf{D}^{11} H\left(\mathbf{h}^c|\mathbf{h}^c\right) = E_{(\mathbf{Z},\mathbf{A})}\left\{\left\|\frac{\partial}{\partial \mathbf{h}^c}\log p\left(\mathbf{Z}|\mathbf{A},\mathbf{h}^c\right)\right\|^2 \middle| \mathbf{Y},\mathbf{h}^c\right\} + E_{(\mathbf{Z},\mathbf{A})}\left\{\left\|\frac{\partial}{\partial \mathbf{h}^c}\log p\left(\mathbf{Y}|\mathbf{h}^c\right)\right\|^2 \middle| \mathbf{Y},\mathbf{h}^c\right\} - 2\Re e\left\{E_{(\mathbf{Z},\mathbf{A})}\left\{\frac{\partial}{\partial \mathbf{h}^c}\log p\left(\mathbf{Z}|\mathbf{A},\mathbf{h}^c\right)\left[\frac{\partial}{\partial \mathbf{h}^c}\log p\left(\mathbf{Y}|\mathbf{h}^c\right)\right]^\dagger \middle| \mathbf{Y},\mathbf{h}^c\right\}\right\} \quad \text{(C-2)}$$

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{h}^c}\log p\left(\mathbf{Y}|\mathbf{h}^c\right) &= \frac{1}{2\sigma^2}\sum_{\mathbf{X}_j}\mathrm{P}\left(\mathbf{X}_j|\mathbf{Y},\mathbf{h}^c\right)\mathbf{\Omega}^{\mathrm{T}}\mathrm{diag}\left(\mathbf{X}_j\right)^{\mathrm{T}}\left[\mathbf{Y}^* - \mathrm{diag}\left(\mathbf{X}_j\right)^*\mathbf{\Omega}^*\mathbf{h}^{c*}\right] \\
&= \frac{1}{2\sigma^2}\sum_{\mathbf{A}_j}\mathrm{P}\left(\mathbf{A}_j|\mathbf{Y},\mathbf{h}^c\right)\left(\mathbf{N}^{c\,\dagger}\mathbf{A}_j\right)^{\mathrm{T}} = \frac{1}{2\sigma^2}\sum_{\mathbf{A}_j}\mathrm{P}\left(\mathbf{A}_j|\mathbf{Y},\mathbf{h}^c\right)\left[\mathbf{N}^{c\,\dagger}\mathbf{A}_{j,0}, \cdots, \mathbf{N}^{c\,\dagger}\mathbf{A}_{j,L-1}\right]^{\mathrm{T}} \quad \text{(C-5)}
\end{aligned}$$

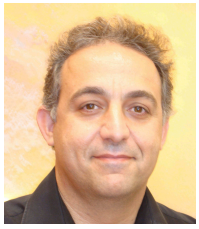
$$2\Re e\left\{E_{(\mathbf{Z},\mathbf{A})}\left\{\frac{\partial}{\partial \mathbf{h}^c}\log p\left(\mathbf{Z}|\mathbf{A},\mathbf{h}^c\right)\left[\frac{\partial}{\partial \mathbf{h}^c}\log p\left(\mathbf{Y}|\mathbf{h}^c\right)\right]^\dagger \middle| \mathbf{Y},\mathbf{h}^c\right\}\right\} \approx \frac{2}{2\sigma^2} N\mathcal{E}\mathbf{I}_L \quad \text{(C-8)}$$

[3] R.W.J. Health and G.B. Giannakis, "Exploiting input cyclostationarity for blind channel identification in OFDM systems," *IEEE Trans. Signal Process.*, vol. 47, no. 3, pp. 848-856, Mar. 1999.
[4] 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.
[5] G.K. Kaleh and R. Vallet, "Joint parameter estimation and symbol detection for linear or nonlinear unknown channels," *IEEE Trans. Commun.*, vol. 42, no. 7, pp. 2406-2413, Jul. 1994.
[6] H.A. Çirpan, E. Panayirci, and H. Doğan, "Iterative channel estimation approach for space-time/frequency coded ofdm systems with transmitter diversity," *Euro. Trans. Telecomms.*, pp. 235-248, 2004.
[7] X. Ma, H. Kobayashi, and S.C. Schwartz, "EM-based channel estimation algorithms for OFDM," *EURASIP Journal on Applied Signal Processing*, vol. 2004, no. 10, pp. 1460-1477, 2004.
[8] Y. Xie and C.N. Georghiades, "Two EM-type channel estimation algorithms for OFDM with transmitter diversity," *IEEE Trans. Commun.*, vol. 51, no. 1, pp. 106-115, Jan. 2003.
[9] B.S. Ünal, A.O. Berthet, and R. Visoz, "Iterative channel estimation and coded symbol detection for dispersive channels," in *Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, vol. 1, pp. 100-106, Sep./Oct. 2001.
[10] W. Navidi, "A graphical illustration of the EM algorithm," *The American Statistical Association*, vol. 51, no. 1, Feb. 1997.
[11] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge University Press, June 2005.
[12] T. Moon, "The expectation-maximization algorithm," *IEEE Signal Processing Magazine*, vol. 13, no. 6, pp. 47-60, Nov 1996.
[13] M. Feder and E. Weinstein, "Parameter estimation of superimposed signals using the EM algorithm," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 4, pp. 477-489, Apr. 1988.
[14] P. Spasojević and C.N. Georghiades, "Implicit diversity combining based on the EM algorithm for fading channels with correlated path components," in *Proc. IEEE Wireless Communications and Networking Conference, 1999 (WCNC 1999)*, vol. 1, pp. 15-19, Sep. 1999.
[15] Y. Liu, L. Brunel, and J.J. Boutros, "EM channel estimation for coded OFDM transmissions over frequency-selective channel," in *Proc. IEEE 10th International Symposium on Spread Spectrum Techniques and Applications, 2008 (ISSSTA08)*, pp. 544-549, Aug. 2008.
[16] G.H. Golub and C.F. Loan, *Matrix Computations*, 3rd edition, The Johns Hopkins University Press, 1996.
[17] S.M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Upper Saddle River, NJ: Prentice-Hall, Inc., 1993.
[18] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, New York: John Wiley and Sons, 1991.
[19] A.N. D'Andrea, U. Mengali, and R. Reggiannini, "The modified Cramer-Rao bound and its application to synchoronization problems," *IEEE Trans. Commun.*, vol. 42, no. 2/3/4, pp. 1391-1399, Feb./Mar./Apr. 1994.
[20] F. Gini, R. Reggiannini, and U. Mengali, "The modified Cramer-Rao bound in vector parameter estimation," *IEEE Trans. Commun.*, vol. 46, no. 1, pp. 52-60, Jan. 1998.
[21] N. Noels, H. Steendam, and M. Moeneclaey, "Carrier and clock recovery in (turbo-) coded systems: Cramer-Rao bound and synchronizer performance," *EURASIP Journal on Applied Signal Processing*, vol. 2005, no. 6, pp. 972-980, 2005.



Yang Liu received the B.E. and M.S. degrees in electrical engineering from Shanghai Jiao Tong University, Shanghai, China, in 2001 and 2004, respectively. From 2004 to 2006, he was with Communication Systems Group of Sharp, Hiroshima, Japan. He received the Ph.D. degree in electrical engineering from the Nationale Supérieure des Télécommunications (ENST, Telecom ParisTech), Paris, France, in 2009. Since December 2009, he has been an engineer at Sequans Communications in Paris, France. His research interests are multi-carrier transmissions, channel estimation, and multiple-antenna techniques.

Loïc Brunel received the M.S. degree in electrical engineering in 1996 and the Ph.D. degree in 1999, both from Ecole Nationale Supérieure des Télécommunications (ENST, Telecom ParisTech), Paris, France. Since January 2000, he has been a Research Engineer in the Communications Technology Division of the Mitsubishi Electric R&D Centre Europe in Rennes, France. His research interests are multi-carrier transmissions, channel estimation, channel coding and multiple-antenna techniques.



Joseph Jean Boutros received the M.S. degree in electrical engineering in 1992 and the Ph.D. degree in 1996, both from Ecole Nationale Supérieure des Télécommunications (ENST, Telecom ParisTech), Paris, France. From 1996 to 2006, he was with the Communications and Electronics Department at ENST as an Associate Professor. Also, Dr Boutros was a member of the research unit UMR-5141 of the French National Scientific Research Center (CNRS). In 2007, Dr Boutros joined Texas A&M University at Qatar as a Professor in the electrical engineering program. Dr Boutros has been a scientific consultant for Alcatel Space, Philips Research, and Motorola Semiconductors, and member of the Digital Signal Processing team at Juniper Networks Cable. His fields of interest are codes on graphs, iterative decoding, joint source-channel coding, cooperative communications, space-time coding, security and network coding for the physical layer, and lattice sphere packings.