Joint Channel Estimation and Data Detection for OFDM Systems over Doubly Selective Channels

Lanlan He, Shaodan Ma, Yik-Chung Wu, Tung-Sang Ng
Department of Electrical and Electronic Engineering
The University of Hong Kong
Email:{llhe, sdma, ycwu, tsng}@eee.hku.hk

Abstract—In this paper, a joint channel estimation and data detection algorithm is proposed for OFDM systems under doubly selective channels (DSCs). After representing the DSC using Karhunen-Loève basis expansion model (K-L BEM), the proposed algorithm is developed based on the expectation-maximization (EM) algorithm. Basically, it is an iterative algorithm including two steps at each iteration. In the first step, the unknown coefficients in K-L BEM are first integrated out to obtain a function which only depends on data, and meanwhile, a maximum a posteriori (MAP) channel estimator is obtained. In the second step, data are directly detected by a novel approach based on the function obtained in the first step. Moreover, a Bayesian Cramer-Rao Lower Bound (BCRB) which is valid for any channel estimator is also derived to evaluate the performance of the proposed channel estimator. The effectiveness of the proposed algorithm is finally corroborated by simulation results.

I. INTRODUCTION

With high spectral efficiency and robustness against frequency selective fading, orthogonal frequency division multiplexing (OFDM) has been widely recognized as an efficient transmission technique for wireless communications and has been considered as a strong candidate for next generation mobile systems. Generally, in mobile systems high speed movement would cause Doppler spread and result in a multi-path time-varying channel, i.e., doubly selective channel (DSC). Due to the time variation of the channel, the number of channel parameters in one OFDM symbol significantly increases, which makes the channel estimation a challenge. Meanwhile, this doubly-selectivity in channel will destroy the orthogonality among subcarriers and induces intercarrier interference (ICI) in OFDM systems, which also complicates the data detection.

So far, channel estimation and data detection for OFDM systems over DSCs are usually tackled separately. In the literature [1]–[6], a number of pilot-aided channel estimation methods have been proposed by adopting various kinds of basis expansion models (BEMs). With the channel estimate, a minimum mean square error (MMSE) data detection method and an improved version with low-complexity have been proposed in [5] [6] respectively. In recent years, joint channel estimation and data detection methods have been developed for single carrier systems in [7] [8]. It is shown that those joint schemes can generally provide better performance than the separate approaches, since the data are also exploited for channel estimation. However, their extension to OFDM systems is by no means straightforward.

In this paper, a joint channel estimation and data detection algorithm is proposed for OFDM systems over DSCs. K-L

BEM channel model is utilized here to achieve minimum mean square modeling error and also reduce the complexity of the proposed algorithm. Basically, the proposed algorithm is an iterative EM algorithm including two steps (E-step and M-step) at each iteration. In E-step, the unknown BEM coefficients are integrated out to obtain a function which only depends on data. Meanwhile, a MAP channel estimator is also obtained. In M-step, the data are directly detected by a novel approach based on the function obtained in E-step. In order to evaluate the performance of the proposed algorithm, the BCRB which is valid for any channel estimator is then derived. Simulation results show that the proposed joint algorithm converges after a few iterations and it performs much better than those separate approaches in [4] [5]. Furthermore, the proposed channel estimator would touch the BCRB at high SNR and the performance of the proposed detector approaches that of the ideal detector with perfect channel state information after convergence.

Notation: Boldface uppercase and lowercase letters will be used for matrices and vectors. Superscripts H and T denotes Hermitian and transpose respectively. The symbol $\mathbf{I}_N$ denotes an $N \times N$ identity matrix, with $\mathbf{e}_l$ denoting the lth column of $\mathbf{I}_N$. diag$\{\mathbf{x}\}$ stands for the diagonal matrix with vector $\mathbf{x}$ on its diagonal. The $(m, n)^{th}$ entry of a matrix $\mathbf{X}$ is denoted by $[\mathbf{X}]_{m,n}$. The symbol $\otimes$ denotes the Kronecker product and $\odot$ denotes the Hadamard product. $\mathbb{E}\{\cdot\}$ denotes the expectation. Tr$\{\mathbf{X}\}$ and $|\mathbf{X}|$ are the trace and the determination of a square matrix $\mathbf{X}$ respectively. $\Re\{\cdot\}$ and $\Im\{\cdot\}$ are the real and imaginary parts respectively.

II. SYSTEM MODEL

In an OFDM system, the source data $\mathbf{x} = [x(0), x(1), \cdots, x(N - 1)]^T$ in frequency domain is modulated onto N parallel subcarriers to obtain the time domain signal $\mathbf{s} = \mathbf{F}^H \mathbf{x}$, where $\mathbf{F}$ is the FFT matrix with $[\mathbf{F}]_{m,n} = \frac{1}{\sqrt{N}} e^{-j2\pi mn/N}$. A cyclic prefix (CP) with length longer than the delay spread of the channel, is inserted at the beginning of each OFDM symbol to prevent intersymbol interference (ISI). The signal is then transmitted through a multi-path time-varying channel which has L independent taps with the average power of the lth tap denoted by σ_l^2. The auto-correlation of the lth tap follows the classical Jakes' model [3] given by $\mathbb{E}\{h_l(mT_s)h_l(nT_s)\} = \sigma_l^2 J_0(2\pi f_D(n - m)T_s)$, where $J_0(\cdot)$ represents the zero-order Bessel function of the first kind, f_D represents the Doppler spread normalized by the subcarrier spacing, and T_s is the sample interval. At the receiver side, assuming perfect synchronization is achieved,

after discarding the CP, the received signal can be written as

$$\mathbf{y} = \mathbf{D}(\mathbf{s})\mathbf{h} + \mathbf{w} \tag{1}$$

where $\mathbf{y} = [y(0), y(1), \cdots, y(N-1)]^T$ is the received signal; $\mathbf{D}(\mathbf{s}) = [\text{diag}\{\boldsymbol{\Xi}_0\mathbf{s}\}, ..., \text{diag}\{\boldsymbol{\Xi}_{L-1}\mathbf{s}\}]$ with $\boldsymbol{\Xi}_l = [\mathbf{e}_{l+1}, \cdots, \mathbf{e}_N, \mathbf{e}_1, \cdots, \mathbf{e}_l]$; and $\mathbf{w} = [w(0), w(1), \cdots, w(N-1)]^T$ is the additive white Gaussian noise vector with zero-mean and covariance $\sigma_w^2 \mathbf{I}_N$. The channel vector $\mathbf{h}$ is defined as

$$\mathbf{h} = [\mathbf{h}_0^T, \cdots, \mathbf{h}_{L-1}^T]^T \tag{2}$$

with $\mathbf{h}_l = [h_l(0T_s), \cdots, h_l((N-1)T_s)]^T$ being the channel coefficient of the lth tap during the whole OFDM symbol. For notation simplicity, the sample interval T_s will be omitted in $\mathbf{h}_l$. Since the channel taps are independent of each other, the correlation matrix of $\mathbf{h}$ is given by

$$\mathbf{R}_h = \mathbf{R}_L \otimes \mathbf{J} \tag{3}$$

with $\mathbf{R}_L = \text{diag}\{\sigma_0^2, \cdots, \sigma_{L-1}^2\}$ and $[\mathbf{J}]_{m,n} = J_0(2\pi f_D(n-m)T_s)$.

III. PROPOSED EM-KL ALGORITHM

A. Channel model

Using basis expansion model (BEM) [1] [2] [3], the channel can be expressed as

$$\mathbf{h} \cong (\mathbf{I}_L \otimes \mathbf{W})\mathbf{h}_b \tag{4}$$

where $\mathbf{W}$ is a fixed $N \times K$ matrix with K being a design parameter balancing a tradeoff between the model accuracy and complexity, and $\mathbf{h}_b$ is the unknown BEM coefficient. In [1], the columns of $\mathbf{W}$ are chosen to be a set of complex exponentials with frequencies $(-\frac{K-1}{2N}, -\frac{K-3}{2N}, \cdots, 0, \cdots, \frac{K-3}{2N}, \frac{K-1}{2N})$ and K being an odd number, resulting a so-called complex-exponential BEM (CE-BEM). The CE-BEM attracts a lot of attention due to the orthogonality among columns of $\mathbf{W}$, however, it induces large modeling error when channel is varying rapidly. An improved version, which makes the set of complex exponentials more closely spaced in the frequency domain than those in CE-BEM, is proposed in [2], named generalized CE-BEM (GCE-BEM). Though GCE-BEM achieves higher model accuracy than CE-BEM, the over-sampling in the frequency domain destroys the orthogonality among columns of $\mathbf{W}$. Recently in [3], $\mathbf{W}$ is chosen to be the matrix which contains the K principal eigenvectors of $\mathbf{J}$, resulting in Karhunen-Loève BEM (K-L BEM). It is shown in [3] that K-L BEM achieves minimum mean square channel modeling error. More specifically, given eigen-decomposition of $\mathbf{J} = \mathbf{V}\boldsymbol{\Delta}\mathbf{V}^H$ where $\boldsymbol{\Delta}$ is a diagonal matrix composed of eigenvalues of $\mathbf{J}$ arranged in decreasing order and $\mathbf{V}$ is a unitary matrix formed by the corresponding eigenvectors, it follows that $\mathbf{W} = \mathbf{V}_{:,1:K}$ (i.e., the first K columns of $\mathbf{V}$). With K-L BEM, the elements of the BEM coefficient $\mathbf{h}_b$ are independent from each other and since $\mathbf{h}$ is Gaussian, it is easy to prove that $\mathbf{h}_b$ also follows Gaussian distribution with probability density function (pdf) given by

$$p(\mathbf{h}_b) = \frac{1}{(\pi)^{KL}|\mathbf{R}_{h_b}|} \exp(-\mathbf{h}_b^H \mathbf{R}_{h_b}^{-1} \mathbf{h}_b) \tag{5}$$

where $\mathbf{R}_{h_b} = \mathbf{R}_L \otimes \boldsymbol{\Delta}_K$ with $\boldsymbol{\Delta}_K$ denoting the $K \times K$ principal submatrix of $\boldsymbol{\Delta}$. Substituting (4) into (1), the received signal becomes

$$\mathbf{y} = \mathbf{B}(\mathbf{s})\mathbf{h}_b + \mathbf{w} \tag{6}$$

with $\mathbf{B}(\mathbf{s}) = \mathbf{D}(\mathbf{s})(\mathbf{I}_L \otimes \mathbf{W})$.

Remark 1: An equivalent model for $\mathbf{y}$ is given by

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{w} \tag{7}$$

where

$$\mathbf{H} = \begin{bmatrix} \tilde{h}_0(0) & \mathbf{0} & \tilde{h}_{L-1}(0) & \cdots & \tilde{h}_1(0) \\ \tilde{h}_1(1) & \tilde{h}_0(1) & \mathbf{0} & \tilde{h}_{L-1}(1) & \cdots \\ \cdots & \cdots & \cdots & \cdots & \\ & \mathbf{0} & \tilde{h}_{L-1}(N-1) & \cdots & \tilde{h}_0(N-1) \end{bmatrix} \tag{8}$$

with $\tilde{h}_l(n)$ being the $(l \times N + n)^{th}$ element of $(\mathbf{I}_L \otimes \mathbf{W})\mathbf{h}_b$. This expression will be used in the derivation of our proposed algorithm.

B. Joint channel estimation and data detection

Because the noise is Gaussian distributed, based on (6), the likelihood function for $\mathbf{h}_b$ and $\mathbf{s}$ is given by

$$p(\mathbf{y}|\mathbf{h}_b, \mathbf{s}) = \frac{1}{(\pi\sigma_w^2)^N} \exp(-\frac{1}{\sigma_w^2}\|\mathbf{y} - \mathbf{B}(\mathbf{s})\mathbf{h}_b\|^2). \tag{9}$$

The direct channel estimation and data detection require exhaustive search over a very high-dimensional space, which is prohibitively complex. In the following, based on EM algorithm, we propose an iterative joint channel estimation and data detection algorithm. For the derivation, the received signal $\mathbf{y}$ is referred as incomplete data and we take $(\mathbf{y}, \mathbf{h}_b)$ as the complete data. The EM algorithm alternates between E-step and M-step until convergence, and these two steps at the ith iteration follows [9]:
1) E-step: Compute $Q(\mathbf{s}|\hat{\mathbf{s}}^{i-1}) = \mathbb{E}\{\log(p(\mathbf{y}|\mathbf{h}_b, \mathbf{s})p(\mathbf{h}_b))|\mathbf{y}, \hat{\mathbf{s}}^{i-1}\}$;
2) M-step: Solve $\hat{\mathbf{s}}^i = \arg\max Q(\mathbf{s}|\hat{\mathbf{s}}^{i-1})$.
The expectation in E-step is with respect to the conditional pdf $p(\mathbf{h}_b|\mathbf{y}, \hat{\mathbf{s}}^{i-1})$. The symbol $\hat{\mathbf{s}}^{i-1}$ is the $(i-1)^{th}$ estimate of data. The derivations of E-step and M-step are detailed as follows.

E-step:
The expectation in E-step with respect to the conditional pdf $p(\mathbf{h}_b|\mathbf{y}, \hat{\mathbf{s}}^{i-1})$ can be specified as

$$Q(\mathbf{s}|\hat{\mathbf{s}}^{i-1}) = \int \log(p(\mathbf{y}|\mathbf{h}_b, \mathbf{s})p(\mathbf{h}_b))p(\mathbf{h}_b|\mathbf{y}, \hat{\mathbf{s}}^{i-1})d\mathbf{h}_b. \tag{10}$$

Based on (5) and (9),

$$\log[p(\mathbf{y}|\mathbf{h}_b, \mathbf{s})p(\mathbf{h}_b)] = -\frac{1}{\sigma_w^2}(\mathbf{y}^H\mathbf{y} - 2\Re\{\mathbf{y}^H\mathbf{B}(\mathbf{s})\mathbf{h}_b\} \\ + \mathbf{h}_b^H\mathbf{B}^H(\mathbf{s})\mathbf{B}(\mathbf{s})\mathbf{h}_b) - \mathbf{h}_b^H\mathbf{R}_{h_b}^{-1}\mathbf{h}_b + c_1 \tag{11}$$

with c_1 being a constant. On the other hand, using Bayes' rules, the conditional pdf $p(\mathbf{h}_b|\mathbf{y}, \hat{\mathbf{s}}^{i-1})$ can be expressed as

$$p(\mathbf{h}_b|\mathbf{y}, \hat{\mathbf{s}}^{i-1}) = \frac{p(\mathbf{y}|\mathbf{h}_b, \hat{\mathbf{s}}^{i-1})p(\mathbf{h}_b)}{p(\mathbf{y}|\hat{\mathbf{s}}^{i-1})}. \tag{12}$$

Putting (5) and (9) into (12) (with $\mathbf{s}$ in (9) replaced by $\hat{\mathbf{s}}^{i-1}$) and after some manipulations, we have

$$p(\mathbf{h}_b|\mathbf{y}, \hat{\mathbf{s}}^{i-1})$$
$$= \frac{1}{(\pi)^{KL}|\sigma_w^2 \mathbf{G}_i^{-1}|} \exp\{-(\mathbf{h}_b - \hat{\mathbf{h}}_b^i)^H \frac{1}{\sigma_w^2} \mathbf{G}_i(\mathbf{h}_b - \hat{\mathbf{h}}_b^i)\}$$
$$\times \underbrace{\frac{|\sigma_w^2 \mathbf{G}_i^{-1}|}{(\pi\sigma_w^2)^N |\mathbf{R}_{h_b}|} \exp\{-\frac{1}{\sigma_w^2}(\mathbf{y}^H\mathbf{y} - (\hat{\mathbf{h}}_b^i)^H \mathbf{G}_i \hat{\mathbf{h}}_b^i)\} \frac{1}{p(\mathbf{y}|\hat{\mathbf{s}}^{i-1})}}_{\triangleq c_2}$$

$$\tag{13}$$

where

$$\hat{\mathbf{h}}_b^i = \mathbf{G}_i^{-1} \mathbf{B}^H(\hat{\mathbf{s}}^{i-1})\mathbf{y} \tag{14}$$

with

$$\mathbf{G}_i = \mathbf{B}^H(\hat{\mathbf{s}}^{i-1})\mathbf{B}(\hat{\mathbf{s}}^{i-1}) + \sigma_w^2 \mathbf{R}_{h_b}^{-1}. \tag{15}$$

Notice that c_2 is a constant independent of $\mathbf{s}$ and $\mathbf{h}_b$. Applying (11) and (13) to (10), we have

$$Q(\mathbf{s}|\hat{\mathbf{s}}^{i-1}) = \int [-\frac{1}{\sigma_w^2}(\mathbf{y}^H\mathbf{y} - 2\Re\{\mathbf{y}^H\mathbf{B}(\mathbf{s})\mathbf{h}_b\}$$
$$+ \mathbf{h}_b^H\mathbf{B}^H(\mathbf{s})\mathbf{B}(\mathbf{s})\mathbf{h}_b) - \mathbf{h}_b^H\mathbf{R}_{h_b}^{-1}\mathbf{h}_b + c_1] \frac{c_2}{(\pi)^{KL}|\sigma_w^2 \mathbf{G}_i^{-1}|}$$
$$\times \exp\{-(\mathbf{h}_b - \hat{\mathbf{h}}_b^i)^H \frac{1}{\sigma_w^2} \mathbf{G}_i(\mathbf{h}_b - \hat{\mathbf{h}}_b^i)\}d\mathbf{h}_b$$

$$\tag{16}$$

After integrating out $\mathbf{h}_b$, we obtain [9]

$$Q(\mathbf{s}|\hat{\mathbf{s}}^{i-1}) = -\frac{c_2}{\sigma_w^2}(\mathbf{y}^H\mathbf{y} - 2\Re\{\mathbf{y}^H\mathbf{B}(\mathbf{s})\hat{\mathbf{h}}_b^i\}$$
$$+ \mathrm{Tr}\{\mathbf{B}^H(\mathbf{s})\mathbf{B}(\mathbf{s})[\sigma_w^2 \mathbf{G}_i^{-1} + \hat{\mathbf{h}}_b^i(\hat{\mathbf{h}}_b^i)^H]\}) \tag{17}$$
$$- c_2\mathrm{Tr}\{\mathbf{R}_{h_b}^{-1}[\sigma_w^2 \mathbf{G}_i^{-1} + \hat{\mathbf{h}}_b^i(\hat{\mathbf{h}}_b^i)^H]\} + c_2 c_1$$

Notice that, $\hat{\mathbf{h}}_b^i$ in (14) is the maximum a posteriori (MAP) estimator for BEM coefficients at the ith iteration. Accordingly, the estimate of $\mathbf{h}$ at the ith iteration can be updated as $\hat{\mathbf{h}}^i = (\mathbf{I}_L \otimes \mathbf{W})\hat{\mathbf{h}}_b^i$.

M-step:
In M-step, we maximize the $Q(\mathbf{s}|\hat{\mathbf{s}}^{i-1})$ in (17) with respect to the data $\mathbf{s}$. After dropping those terms and scalars independent of $\mathbf{s}$, the equivalent $Q(\mathbf{s}|\hat{\mathbf{s}}^{i-1})$ function is

$$Q(\mathbf{s}|\hat{\mathbf{s}}^{i-1}) = 2\Re\{\mathbf{y}^H\mathbf{B}(\mathbf{s})\hat{\mathbf{h}}_b^i\}$$
$$- \mathrm{Tr}\{\mathbf{B}^H(\mathbf{s})\mathbf{B}(\mathbf{s})[\sigma_w^2 \mathbf{G}_i^{-1} + \hat{\mathbf{h}}_b^i(\hat{\mathbf{h}}_b^i)^H]\} \tag{18}$$

Notice that $\mathbf{B}(\mathbf{s})$ depends on the data $\mathbf{s}$ in a nonlinear way, direct maximization of (18) with respect to $\mathbf{s}$ is difficult. In the following we derive an alternative expression for $Q(\mathbf{s}|\hat{\mathbf{s}}^{i-1})$ from which a solution for $\mathbf{s}$ can be directly obtained.

Since $\mathbf{G}_i$ is a Hermitian matrix, based on eigen-decomposition, we have

$$\mathbf{G}_i^{-1} = \sum_{m=1}^{KL} \beta_{m,i}\mathbf{u}_{m,i}\mathbf{u}_{m,i}^H \tag{19}$$

where $\beta_{m,i}$ is the mth eigenvalue and $\mathbf{u}_{m,i}$ is the corresponding eigenvector of $\mathbf{G}_i^{-1}$. Putting (19) into (18), the function

$Q(\mathbf{s}|\hat{\mathbf{s}}^{i-1})$ is equivalent to

$$Q(\mathbf{s}|\hat{\mathbf{s}}^{i-1}) = 2\Re\{\mathbf{y}^H\mathbf{B}(\mathbf{s})\hat{\mathbf{h}}_b^i\} - (\hat{\mathbf{h}}_b^i)^H\mathbf{B}^H(\mathbf{s})\mathbf{B}(\mathbf{s})\hat{\mathbf{h}}_b^i$$
$$- \sum_{m=1}^{KL} \sigma_w^2\beta_{m,i}\mathbf{u}_{m,i}^H\mathbf{B}^H(\mathbf{s})\mathbf{B}(\mathbf{s})\mathbf{u}_{m,i}. \tag{20}$$

From the result of *Remark 1*, notice that

$$\mathbf{B}(\mathbf{s})\hat{\mathbf{h}}_b^i = \hat{\mathbf{H}}_i\mathbf{s} \tag{21}$$

where $\hat{\mathbf{H}}_i$ has the same structure as $\mathbf{H}$ in (8) with $\tilde{h}_l(n)$ replaced by the $(l \times N + n)^{th}$ element of $\hat{\mathbf{h}}^i$. Similarly, we can write

$$\mathbf{B}(\mathbf{s})\mathbf{u}_{m,i} = \mathbf{U}_{m,i}\mathbf{s} \tag{22}$$

where $\mathbf{U}_{m,i}$ also has the same structure as $\mathbf{H}$ in (8) with $\tilde{h}_l(n)$ replaced by the $(l \times N + n)^{th}$ element of $(\mathbf{I}_L \otimes \mathbf{W})\mathbf{u}_{m,i}$. Putting (21) and (22) back into (20), the function to be maximized becomes

$$Q(\mathbf{s}|\hat{\mathbf{s}}^{i-1}) = 2\Re\{\mathbf{y}^H\hat{\mathbf{H}}_i\mathbf{s}\} - \mathbf{s}^H \underbrace{(\hat{\mathbf{H}}_i^H\hat{\mathbf{H}}_i + \sum_{m=1}^{KL}\sigma_w^2\beta_{m,i}\mathbf{U}_{m,i}^H\mathbf{U}_{m,i})}_{\triangleq \mathbf{P}_i}\mathbf{s}$$

$$\tag{23}$$

which is a quadratic function of $\mathbf{s}$. By setting the first derivative of $Q(\mathbf{s}|\hat{\mathbf{s}}^{i-1})$ in (23) to zero and with $\mathbf{s} = \mathbf{F}^H\mathbf{x}$, it follows that

$$\hat{\mathbf{x}}^i = \mathbf{F}\mathbf{P}_i^{-1}\hat{\mathbf{H}}_i^H\mathbf{y}. \tag{24}$$

After taking hard decision on $\hat{\mathbf{x}}^i$, $\hat{\mathbf{s}}^i$ is obtained by transforming the hard decision outputs of $\hat{\mathbf{x}}^i$ to the time domain, i.e., $\hat{\mathbf{s}}^i = \mathbf{F}^H(\text{demod}(\hat{\mathbf{x}}^i))$, where $\text{demod}(\hat{\mathbf{x}}^i)$ denotes making hard decision on $\hat{\mathbf{x}}^i$.

In summary, the proposed algorithm iterates between the following two equations:

$$\hat{\mathbf{h}}_b^i = (\mathbf{B}^H(\hat{\mathbf{s}}^{i-1})\mathbf{B}(\hat{\mathbf{s}}^{i-1}) + \sigma_w^2\mathbf{R}_{h_b}^{-1})^{-1}\mathbf{B}^H(\hat{\mathbf{s}}^{i-1})\mathbf{y} \tag{25}$$

and

$$\hat{\mathbf{s}}^i = \mathbf{F}^H(\text{demod}(\mathbf{F}(\hat{\mathbf{H}}_i^H\hat{\mathbf{H}}_i + \sum_{m=1}^{KL}\sigma_w^2\beta_{m,i}\mathbf{U}_{m,i}^H\mathbf{U}_{m,i})^{-1}\hat{\mathbf{H}}_i^H\mathbf{y})),$$

$$\tag{26}$$

which is referred as the EM-KL algorithm. In each iteration, the detected data in previous iteration is exploited to update the channel estimate. If the data detection becomes more accurate with the increase of iteration number, the channel estimation can be refined gradually. Meanwhile, the improved channel estimator could reduce the data detection error further in the next iteration. Therefore, with the proposed algorithm working in an iterative fashion, better performance could be expected (also shown in the simulation results) than that of those algorithms which solve these two problems separately [4] [5].

Remark 2: In case pilots are embedded to initialize the algorithm, we only need to estimate those data subcarriers. In this case, the transmitted time domain signal is given by

$$\mathbf{s} = \mathbf{F}^H\mathbf{x} = \mathbf{F}_p^H\mathbf{x}_p + \mathbf{F}_d^H\mathbf{x}_d \tag{27}$$

where $\mathbf{F}_p$ collects those rows of $\mathbf{F}$ corresponding to pilot subcarriers, $\mathbf{F}_d$ collects those rows of $\mathbf{F}$ corresponding to data subcarriers, $\mathbf{x}_p$ and $\mathbf{x}_d$ denote pilots and data respectively. Putting (27) into (23), through some straightforward manipulations, it is obtained that

$$Q(\mathbf{x}_d|\hat{\mathbf{x}}_d^{i-1}) = 2\Re\{(\mathbf{y}^H\hat{\mathbf{H}}_i - \mathbf{x}_p^H\mathbf{F}_p\mathbf{P}_i)\mathbf{F}_d^H\mathbf{x}_d\} - \mathbf{x}_d^H\mathbf{F}_d\mathbf{P}_i\mathbf{F}_d^H\mathbf{x}_d. \tag{28}$$

By setting the first derivative of $Q(\mathbf{x}_d|\hat{\mathbf{x}}_d^{i-1})$ to zero, it follows

$$\hat{\mathbf{x}}_d^i = (\mathbf{F}_d\mathbf{P}_i\mathbf{F}_d^H)^{-1}\mathbf{F}_d(\hat{\mathbf{H}}_i^H\mathbf{y} - \mathbf{P}_i\mathbf{F}_p^H\mathbf{x}_p). \tag{29}$$

Accordingly, $\hat{\mathbf{s}}^i$ is updated using (27) with $\mathbf{x}_d$ replaced by its hard decision outputs demod($\hat{\mathbf{x}}_d^i$).

C. Initialization

A good initialization is essential to EM algorithm, and therefore, the problem now becomes how to obtain the initial estimates $\hat{\mathbf{h}}_b^0$ and $\hat{\mathbf{s}}^0$. Over slow fading channels, the channel estimates obtained in a given symbol could be directly used to initialize the iterative algorithm in the next symbol. However, over DSCs, since the channel is varying rapidly, the channel estimates from the previous symbol might be outdated. To deal with the fast varying channel, assuming the BEM coefficients over each symbol obey a simple first-order auto-regressive (AR) model, prediction of the BEM coefficients could be carried out to initialize the channel estimates for the next symbol [10]. However, in general, there is no guideline for selecting the parameters of the AR model, which significantly affects the performance of prediction. In this paper, we consider another simple solution based on inserted scattered pilots in each symbol. The initial estimate of $\mathbf{h}_b$ is obtained as [4]

$$\hat{\mathbf{h}}_b^0 = \mathbf{R}_{h_b}\mathbf{S}_p^H(\mathbf{S}_p\mathbf{R}_{h_b}\mathbf{S}_p^H + (\mathbf{W}\boldsymbol{\Delta}_K\mathbf{W}^H) \odot (\mathbf{F}_d^H\boldsymbol{\Lambda}_d\mathbf{F}_d) + \sigma_w^2\mathbf{I}_N)^{-1}\mathbf{y} \tag{30}$$

where $\mathbf{S}_p = [\mathrm{diag}\{\boldsymbol{\Xi}_0\mathbf{F}_p^H\mathbf{x}_p\}, ..., \mathrm{diag}\{\boldsymbol{\Xi}_{L-1}\mathbf{F}_p^H\mathbf{x}_p\}](\mathbf{I}_L \otimes \mathbf{W})$ and $\boldsymbol{\Lambda}_d$ is a diagonal matrix whose elements depend on the average power of $\mathbf{x}_d$. After channel estimation, based on MMSE criterion and following the derivation in [5], the data estimator is then given by

$$\hat{\mathbf{x}}_d^0 = (\mathbf{F}_d\hat{\mathbf{H}}_0^H\hat{\mathbf{H}}_0\mathbf{F}_d^H + \sigma_w^2\boldsymbol{\Lambda}_d^{-1})^{-1}\mathbf{F}_d\hat{\mathbf{H}}_0^H(\mathbf{y} - \hat{\mathbf{H}}_0\mathbf{F}_p^H\mathbf{x}_p). \tag{31}$$

Then $\hat{\mathbf{s}}^0$ is given by (27) with $\mathbf{x}_d$ replaced by demod($\hat{\mathbf{x}}_d^0$). The data estimator in (31) is a generalization of that proposed in [5] with $\mathbf{x}_p = 0$. Notice that in the initialization, channel estimation and data detection are actually tackled separately.

IV. BAYESIAN CRAMER-RAO LOWER BOUNDS

To provide a performance benchmark for the proposed EM-KL algorithm, in this section we derive the Carmer-Rao lower bound for the channel estimator by assuming that data on all subcarriers are training, since the performance of the proposed EM-KL algorithm is expected to approach the all pilot case.

Notice that

$$p(\mathbf{y}|\mathbf{h}) = \frac{1}{(\pi\sigma_w^2)^N}\exp(-\frac{1}{\sigma_w^2}\|\mathbf{y} - \mathbf{D}(\mathbf{s})\mathbf{h}\|^2), \tag{32}$$

$$p(\mathbf{h}) = \frac{1}{(\pi)^{NL}|\mathbf{R}_h|}\exp(-\mathbf{h}^H\mathbf{R}_h^{-1}\mathbf{h}). \tag{33}$$

Denote the $\boldsymbol{\theta} = [\Re\{\mathbf{h}\}^T\Im\{\mathbf{h}\}^T]^T$, the Bayesian CRB (BCRB) for channel estimator can be expressed as [11]

$$BCRB_h = [\mathbf{J}_H]^{-1} = [\mathbf{J}_F + \mathbf{J}_P]^{-1} \tag{34}$$

where $\mathbf{J}_F$ is given by [11]

$$\begin{aligned}\mathbf{J}_F &= \mathbb{E}_{\mathbf{h}}\{\mathbb{E}_{\mathbf{y}|\mathbf{h}}\{\frac{\partial^2\log p(\mathbf{y}|\mathbf{h})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\}\} \\ &= \frac{2}{\sigma_w^2}\begin{bmatrix}\Re\{\mathbf{D}^H(\mathbf{s})\mathbf{D}(\mathbf{s})\} & -\Im\{\mathbf{D}^H(\mathbf{s})\mathbf{D}(\mathbf{s})\} \\ \Im\{\mathbf{D}^H(\mathbf{s})\mathbf{D}(\mathbf{s})\} & \Re\{\mathbf{D}^H(\mathbf{s})\mathbf{D}(\mathbf{s})\}\end{bmatrix}\end{aligned} \tag{35}$$

with the second equality obtained easily as an extension of the derivation in [12] and $\mathbf{J}_P$ is the prior information matrix defined in [11] as

$$\mathbf{J}_P = -\mathbb{E}_{\mathbf{h}}\{\frac{\partial^2\log p(\mathbf{h})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\} = \begin{bmatrix}2\mathbf{R}_h^{-1} & 0 \\ 0 & 2\mathbf{R}_h^{-1}\end{bmatrix} \tag{36}$$

with the second equality due to (33). Accordingly,

$$\mathbf{J}_H = \frac{2}{\sigma_w^2}\begin{bmatrix}\Re\{\mathbf{D}^H(\mathbf{s})\mathbf{D}(\mathbf{s})\} + \sigma_w^2\mathbf{R}_h^{-1} & -\Im\{\mathbf{D}^H(\mathbf{s})\mathbf{D}(\mathbf{s})\} \\ \Im\{\mathbf{D}^H(\mathbf{s})\mathbf{D}(\mathbf{s})\} & \Re\{\mathbf{D}^H(\mathbf{s})\mathbf{D}(\mathbf{s})\} + \sigma_w^2\mathbf{R}_h^{-1}\end{bmatrix}. \tag{37}$$

After inverting $\mathbf{J}_H$ and combing together the real part and imaginary part of $\mathbf{h}$, we can finally obtain the bound as

$$BCRB_h = \sigma_w^2(\mathbf{D}^H(\mathbf{s})\mathbf{D}(\mathbf{s}) + \sigma_w^2\mathbf{R}_h^{-1})^{-1}. \tag{38}$$

The BCRB has been derived without using channel model and therefore is applicable to evaluate performance of any channel estimator.

V. SIMULATION RESULTS

We now present the simulation results to illustrate the effectiveness of the proposed EM-KL algorithm for a practical OFDM system. Each OFDM symbol has 128 subcarriers (N=128) and the length of CP is 8; carrier frequency is f_c=2GHz; the sample interval $T_s = 2\mu s$; the speed of vehicle is $v = 105km/hour$ which results the normalized Doppler spread to be $f_D = NT_s\frac{vf_c}{c} = 0.048$ with c being the speed of light. The channel has three taps ($L = 3$) with an exponential power delay profile. Each tap is Rayleigh distributed and is assumed to experience the same f_D, and the time correlation of each tap follows the Jakes' model. The parameter K is set to 4. The pilot structure in [4] is adopted for the initialization. Four pilot clusters are used with each cluster occupying three pilots, where only one non-zero pilot is transmitted in the middle of the cluster, which means roughly 90.6% of the subcarriers are used for transmitting data symbols. The non-zero pilots are chosen according to complex Gaussian distribution and the data are chosen from QPSK constellation. The average power of the non-zero pilots and the data is normalized to unity without loss of generality. The MSE of channel estimation at the ith iteration is defined as

$$MSE_h = \sum_{n=0}^{N-1}\sum_{l=0}^{L-1}\|\hat{h}_l^i(n) - h_l(n)\|^2.$$

1000 Monte carlo trials are averaged to give each point in the figures.

In Fig. 1, the MSE of channel estimation under different signal-to-noise ratio (SNR) versus the number of iterations
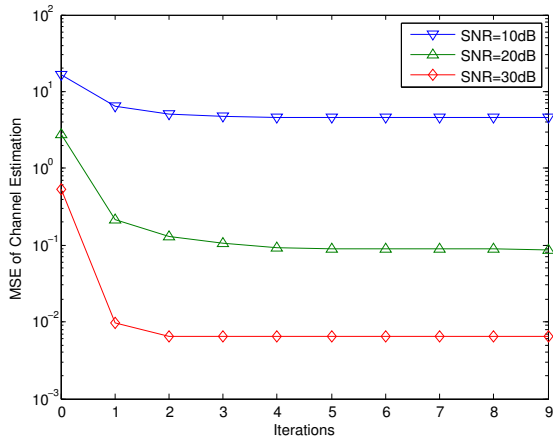
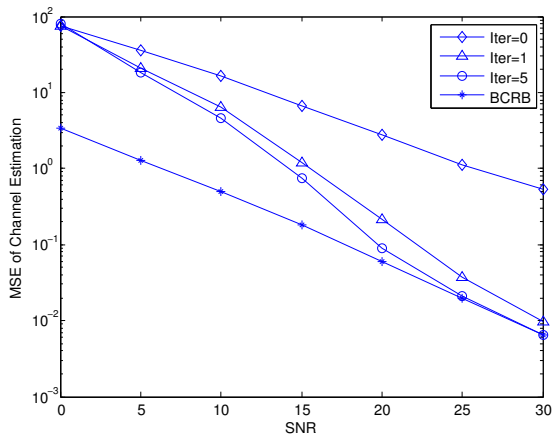Fig. 1.   MSE of channel estimation versus number of iterations

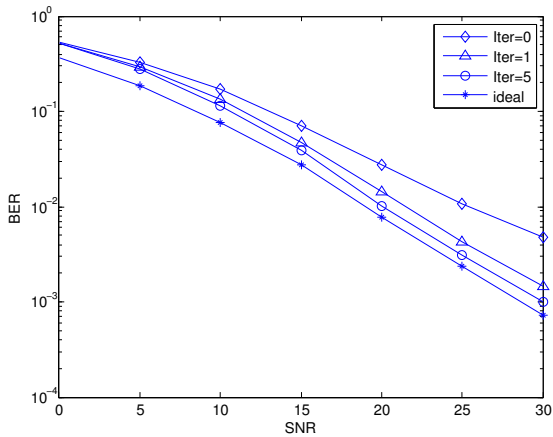Fig. 2.   MSE of channel estimation versus different SNR

Fig. 3.   BER versus different SNR

is depicted. It can be seen that the channel estimation performance improves significantly in the first iteration. Then the decrease of MSE slows down, and finally the proposed algorithm converges after about five iterations.

Fig. 2 shows MSE of channel estimation as a function of SNR. The channel estimate for initialization [4] is labeled as 'Iter=0'. As we can see, due to the low density of pilots (only

twelve pilots inserted), the initial channel estimator performs poorly. However, with more iterations, the MSE decreases. In particular, for high SNR, the proposed algorithm can touch the derived BCRB. For low SNR, though our estimator cannot touch the bound, compared to the estimator used for initialization, obvious improvement exists.

Fig. 3 shows the BER performance of the proposed algorithm as a function of SNR. The data detector for initialization in (31) and the ideal case which assumes perfect channel state information are labeled as 'Iter=0' and 'ideal' respectively. It is clear that with even one iteration, the proposed data detector performs much better than the separate approach ('Iter=0'). Furthermore, after five iterations, the BER performance is very close to the ideal case for the considered range of SNR.

VI. CONCLUSIONS

An EM-KL algorithm was derived to jointly estimate channel and detect data in OFDM systems over doubly selective channels. KL-BEM was used to reduce the complexity of the derived estimator. Moreover, a novel approach was proposed to directly detect data. The BCRB of channel estimation was also derived. The BCRB does not depends on the KL-BEM and is valid for any channel estimator. Extensive simulations showed that the proposed algorithm converged in a few iterations and the MSE of channel estimation touches the BCRB at high SNR. Further, the BER performance approaches the ideal case after convergence.

REFERENCES

[1] G. Giannakis and C. Tepedelenlioglu, "Basis expansion models and diversity techniques for blind identification and equalization of time-varying channels," IEEE *Proceedings*, vol. 86, no. 10, pp. 1969-1986, Oct. 1998.

[2] G. Leus, "On the estimation of rapidly time-varying channels,"in *Proc. EUSIPCO*, pp. 357-360, Sept. 2004.

[3] K. Teo and S. Ohno, "Optimal MMSE finite parameter model for doubly-selective channels," in *Proc. IEEE Globecom*, pp. 3501-3507, Dec. 2005.

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

[5] Y. Choi, J. Voltz and F. Cassara, "On channel estimation and detection for multicarrier signals in fast and selective rayleigh fading channels," IEEE *Trans. Commun.*, vol.45, no.12, pp. 1375-1387, Aug. 2001.

[6] X. Cai, and G. Giannakis, "Bounding performance and suppressing inter-carrier interference in wireless mobile OFDM," IEEE *Trans. Commun.*, vol.51, no.12, pp. 2047-2056, Dec. 2003.

[7] G. Kutz and D. Raphaeli, "Maximum likelihood semi-blind equalization of doubly selective channels," in *Proc. IEEE ISCCSP*, pp. 1580 - 1584, Mar. 2008.

[8] S. He and J. Tugnait, "On doubly selective channel estimation using superimposed training and discrete prolate spheroidal sequences," IEEE *Trans. Signal Processing*, vol. 56, no. 7, pp. 3214-3228, July 2008.

[9] J. Choi, *Adaptive and iterative signal processing in communications,* U.K.: Cambridge, 2006.

[10] S. He and J. Tugnait, "Decision-directed tracking of doubly-selective channels using exponential basis models," in *Proc. IEEE ICC*, pp. 5098 - 5102, May 2008.

[11] H. Tree and K. Bell, *Bayesian bounds for parameter estimation and nonlinear filtering/tracking,* Wiley-IEEE Press, 2007.

[12] P. Stoica and O. Besson, "Training sequence design for frequency offset and frequency-selective channel estimation," IEEE *Trans. Commun.*, vol. 51, no. 11, pp. 1910-1917, Nov. 2003.