CFO Estimation in OFDM Systems under Timing and Channel Length Uncertainties with Model Averaging

Kun Cai, Xiao Li, Jian Du, Yik-Chung Wu, and Feifei Gao

***Abstract*—In this letter, we investigate the problem of CFO estimation in OFDM systems when the timing offset and channel length are not exactly known. Instead of explicitly estimating the timing offset and channel length, we employ a multi-model approach, where the timing offset and channel length can take multiple values with certain probabilities. The effect of multi-model is directly incorporated into the CFO estimator. Results show that the proposed estimator outperforms the estimator selecting only the most probable model and the method taking the maximal model.**

***Index Terms*—Carrier frequency offset (CFO), Bayesian, multi-model, timing offset, channel length uncertainty, orthogonal frequency division multiplexing (OFDM).**

I. Introduction

DUE to its robustness against frequency selective fading channels, orthogonal frequency division multiplexing (OFDM) has been widely used in many communication systems such as wireless metropolitan area networks (WMAN), wireless local area networks (WLAN) and digital broadcasting (e.g., DAB and DVB) systems [1]. On the other hand, OFDM systems are highly sensitive to the carrier frequency offset (CFO) caused by the mismatch of the local oscillators in transceivers. Therefore, the topic of CFO estimation in OFDM systems attracts a lot of attention [2]-[4]. However, many of the existing CFO estimation schemes require perfect time synchronization and known channel length, but such assumptions are too restrictive in practice. In particular, the metrics for many existing timing synchronization algorithms have a plateau in the presence of frequency selective fading channel [3], [5], [6], making perfect time synchronization difficult to achieve. Furthermore, the channel length may be unknown in practice and may also vary with the propagation environment.

Recently, in [7], the ML joint CFO and channel estimatior with timing ambiguity is proposed. The tradeoff between jointly estimating timing offset and CFO versus using a channel model including the effect of timing offset is investigated. However, the channel length is assumed to be known in [7]. In this letter, we investigate the CFO estimation problem when both timing offset and channel length are not known. We model the channel as a mixture of possible models instead of explicitly selecting one single model. Based on the mixture models, we derive the optimal maximum a posterior (MAP) CFO estimator. However, the optimal MAP estimator may present some challenges for practical implementation. Therefore, we propose an empirical estimator, which also estimates the prior parameters from the received data with the aid of a training sequence. Furthermore, a model space reduction method is introduced, in order to exclude those models with low probabilities.

Notation : The operator $\text{diag}(\mathbf{x})$ denotes a diagonal matrix with the elements of $\mathbf{x}$ located on the main diagonal. Superscripts $(\cdot)^H$ and $(\cdot)^T$ denote the conjugate transpose and the transpose operators respectively. Notation $\mathbf{I}$ is the identity matrix and $\det(\mathbf{A})$ takes the determinant of matrix $\mathbf{A}$.

II. System Model

A packet-based OFDM system with N subcarriers is considered. For each data packet, it is preceded by some training blocks. Without loss of generality, we assume that there is only one OFDM training symbol in the preamble. At the transmitter, an OFDM symbol is generated by passing the data (or training) symbols $\mathbf{d} = [d(0), d(1), \ldots, d(N-1)]^T$ through an inverse fast Fourier transform (IFFT). A cyclic prefix (CP) of length L_{CP} is inserted ahead of the OFDM symbol to cope with the inter-symbol interference (ISI) caused by multipath channel. The discrete-time composite channel impulse response (encompassing the transmit/receive filters and the transmission medium) is denoted as $\mathbf{h} = [h_0, \ldots, h_{L-1}]^T$, and is static over one data packet. The CP length L_{CP} is assumed to be larger than the channel order L.

At the receiver, it is assumed that coarse timing synchronization has been achieved (e.g., by using correlation based timing synchronization scheme [3]), such that the FFT window starts within the ISI-free region. More specifically, defining the sample indexes of a perfectly synchronized OFDM symbol as $[-L_{\text{CP}}, ..., 0, ...N-1]$, the estimated starting position of the FFT window can be regarded in the range $[-(L_{\text{CP}} - L), 0]$. In the following, we denote the timing offset between the estimated starting position of the FFT window and the perfect timing point as θ_o. After CP removal, the received signal vector $\mathbf{x}$, which consists of N consecutive samples, is given by [7]

$$\mathbf{x} = \mathbf{\Gamma}(\omega_o)\mathbf{F}^H\mathbf{D}\mathbf{F}_{L_{\text{CP}}}\boldsymbol{\xi} + \mathbf{e} \triangleq \mathbf{H}(\omega_o)\boldsymbol{\xi} + \mathbf{e}, \tag{1}$$

where ω_o is the CFO between the transmitter and receiver; $\mathbf{\Gamma}(\omega_o) \triangleq \text{diag}(1, \ldots, e^{j(N-1)\omega_o})$ is the matrix modelling the CFO effect; $\mathbf{D} \triangleq \text{diag}(\mathbf{d})$; $\mathbf{F}$ is the IFFT matrix; $\mathbf{F}_{L_{\text{CP}}}$ denotes the first L_{CP} columns of $\mathbf{F}$; $\boldsymbol{\xi} \triangleq$

Manuscript received May 25, 2009; revised November 26, 2009; accepted January 15, 2010. The associate editor coordinating the review of this letter and approving it for publication was D. Dardari.

K. Cai, J. Du, and Y.-C. Wu are with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong (e-mail: {caikun, dujian, ycwu}@eee.hku.hk).

X. Li is with the Department of Electrical and Computer Engineering, University of California, Davis (e-mail: eceli@ucdavis.edu).

F. Gao is with the School of Engineering and Science, Jacobs University, Bremen, Germany, 28759 (e-mail: feifeigao@ieee.org).

The work was supported in part by the HKU Seed Funding Programme, Project No. 200711159056.

Digital Object Identifier 10.1109/TWC.2010.03.090762

$$P(\mathbf{x}|\omega,\mathcal{M}_{ij}) = \frac{\sigma^{2(j-i+1-N)}\det(\mathbf{\Lambda}_{ij}^{-1})}{\pi^N \det[\sigma^2\mathbf{\Lambda}_{ij}^{-1} + \mathbf{H}_{ij}^H(\omega)\mathbf{H}_{ij}(\omega)]}\exp\left(-\boldsymbol{\mu}_{ij}^H\mathbf{\Lambda}_{ij}^{-1}\boldsymbol{\mu}_{ij} - \sigma^{-2}\mathbf{x}^H\mathbf{x}\right)$$
$$\times \exp([\sigma^2\mathbf{\Lambda}_{ij}^{-1}\boldsymbol{\mu}_{ij} + \mathbf{H}_{ij}^H(\omega)\mathbf{x}]^H[\sigma^2\mathbf{\Lambda}_{ij}^{-1} + \mathbf{H}_{ij}^H(\omega)\mathbf{H}_{ij}(\omega)]^{-1}[\mathbf{\Lambda}_{ij}^{-1}\boldsymbol{\mu}_{ij} + \sigma^{-2}\mathbf{H}_{ij}^H(\omega)\mathbf{x}]). \quad (5)$$

$[\mathbf{0}_{\theta_o\times 1}^T \quad \mathbf{h}^T \quad \mathbf{0}_{(L_{\text{CP}}-\theta_o-L)\times 1}^T]^T$ is the composite channel including the timing offset θ_o; and vector $\mathbf{e} \triangleq [e_0, e_1, \cdots, e_{N-1}]^T$ denotes the complex white Gaussian noise with zero mean and covariance matrix $\mathbf{C}_e = E\{\mathbf{e}\mathbf{e}^H\} = \sigma^2\mathbf{I}$.

III. CFO Estimation Under Channel Model Uncertainty

In general, we do not know the channel length L and the timing offset θ_o. One way to estimate the CFO is to treat the whole $\boldsymbol{\xi}$ as unknown as suggested in [7]. It has been shown that this method performs asymptotically ($N \to \infty$) the same as if the timing and channel length are estimated together with CFO [7]. However, when the number of subcarriers is not very large, this method suffers from performance degradation, since unnecessary parameters are included in the estimator.

Since the true channel $\mathbf{h}$ can start and end at any position within $\boldsymbol{\xi}$ as long as the starting position occurs before the ending position, in the following, we model the channel as a mixture of possible models from a set Ω. Each model $\mathcal{M}_{ij} \in \Omega$ corresponds to the case $\mathbf{h}$ starts at the i^{th} position and ends at the j^{th} position in $\boldsymbol{\xi}$ ($0 \le i \le j \le L_{\text{CP}}-1$). More specifically

$$\mathcal{M}_{ij}: \quad \boldsymbol{\xi}_{ij} \triangleq [\xi_i \ \xi_{i+1} \ \cdots \ \xi_j]^T, \quad \xi_0 = \cdots = \xi_{i-1} = \xi_{j+1} = \cdots = \xi_{L_{\text{CP}}-1} = 0, \quad (2)$$

where $\boldsymbol{\xi}_{ij}$ is the subvector containing the i^{th} to the j^{th} elements of $\boldsymbol{\xi}$, and the symbol ξ_i denotes the i^{th} element in $\boldsymbol{\xi}$. It is assumed that the channel $\boldsymbol{\xi}_{ij}$ follows Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}_{ij}, \mathbf{\Lambda}_{ij})$ with the probability density function

$$P(\boldsymbol{\xi}_{ij}) = \frac{1}{\pi^{(j-i+1)}\det(\mathbf{\Lambda}_{ij})} \times \exp[-(\boldsymbol{\xi}_{ij}-\boldsymbol{\mu}_{ij})^H\mathbf{\Lambda}_{ij}^{-1}(\boldsymbol{\xi}_{ij}-\boldsymbol{\mu}_{ij})]. \quad (3)$$

Assuming that each model $\mathcal{M}_{ij}$ has an associated probability $P(\mathcal{M}_{ij})$, the CFO posterior density can be calculated by the Bayes rule as [8]

$$P(\omega|\mathbf{x}) = \sum_{i=0}^{L_{\text{CP}}-1}\sum_{j=i}^{L_{\text{CP}}-1} P(\mathcal{M}_{ij})\frac{P(\mathbf{x}|\omega,\mathcal{M}_{ij})P(\omega)}{P(\mathbf{x})}, \quad (4)$$

where $P(\omega)$ is the CFO prior, whose distribution could be obtained under some scenarios, e.g., in cooperative communication systems [9]. If there is no CFO prior information, we can set $P(\omega) = $ constant [8]. In (4), $P(\mathbf{x}|\omega,\mathcal{M}_{ij}) = \int P(\mathbf{x}|\omega,\boldsymbol{\xi}_{ij},\mathcal{M}_{ij})P(\boldsymbol{\xi}_{ij})d\boldsymbol{\xi}_{ij}$ is the averaged likelihood function of $\mathbf{x}$ over channel realization $\boldsymbol{\xi}_{ij}$. Since the vector $\mathbf{e}$ is Gaussian distributed and $\boldsymbol{\xi}_{ij}$ is also a Gaussian random vector drawn from $\mathcal{N}(\boldsymbol{\mu}_{ij}, \mathbf{\Lambda}_{ij})$, the likelihood function can be evaluated as

$$\begin{aligned} P(\mathbf{x}|\omega,\mathcal{M}_{ij}) &= \int \frac{1}{(\pi\sigma^2)^N}\exp\left(-\frac{\left\|\mathbf{x}-\mathbf{H}_{ij}(\omega)\boldsymbol{\xi}_{ij}\right\|^2}{\sigma^2}\right)P(\boldsymbol{\xi}_{ij})d\boldsymbol{\xi}_{ij} \\ &= \frac{1}{\pi^{(N+j-i+1)}\sigma^{2N}}\int \exp\left(-\frac{\left\|\mathbf{x}-\mathbf{H}_{ij}(\omega)\boldsymbol{\xi}_{ij}\right\|^2}{\sigma^2}\right) \\ &\quad \times \frac{1}{\det(\mathbf{\Lambda}_{ij})}\exp[-(\boldsymbol{\xi}_{ij}-\boldsymbol{\mu}_{ij})^H\mathbf{\Lambda}_{ij}^{-1}(\boldsymbol{\xi}_{ij}-\boldsymbol{\mu}_{ij})]d\boldsymbol{\xi}_{ij}, \end{aligned}$$

where $\mathbf{H}_{ij}(\omega)$ is the sub-matrix of $\mathbf{H}(\omega)$ from the i^{th} to the j^{th} column. Assuming that $\{\boldsymbol{\mu}_{ij}, \mathbf{\Lambda}_{ij}, \sigma^2\}$ and $P(\mathcal{M}_{ij})$ are known, and following the method in [10], after some straightforward but tedious derivations, the exact expression of $P(\mathbf{x}|\omega,\mathcal{M}_{ij})$ is shown in (5) at the top of this page. After putting (5) into (4), the MAP CFO estimator is obtained as

$$\hat{\omega} = \arg\max_{\omega}\{P(\omega|\mathbf{x})\}. \quad (6)$$

IV. Empirical MAP CFO Estimator

While the MAP Estimator in (6) is a rigorous solution to the CFO estimation problem under channel model uncertainty, it presents some challenges for practical implementation. Firstly, the number of terms in (4) can be enormous (more specifically, $(L_{\text{CP}}+1)L_{\text{CP}}/2$ in this case), rendering exhaustive summation impractical. This problem will be addressed in Section V. Secondly, in general, one does not know the prior probabilities $P(\mathcal{M}_{ij})$, nor the quantities $\{\boldsymbol{\mu}_{ij}, \mathbf{\Lambda}_{ij}, \sigma^2\}$. For the unknown quantities $\{\boldsymbol{\mu}_{ij}, \mathbf{\Lambda}_{ij}, \sigma^2\}$, one could treat them as nuisance parameters and integrate them out as

$$P(\mathbf{x}|\omega,\mathcal{M}_{ij}) = \int P(\mathbf{x}|\omega,\boldsymbol{\mu}_{ij},\mathbf{\Lambda}_{ij},\sigma^2,\mathcal{M}_{ij}) \times P(\boldsymbol{\mu}_{ij},\mathbf{\Lambda}_{ij},\sigma^2)d\{\boldsymbol{\mu}_{ij},\mathbf{\Lambda}_{ij},\sigma^2\}, \quad (7)$$

where $P(\mathbf{x}|\omega,\boldsymbol{\mu}_{ij},\mathbf{\Lambda}_{ij},\sigma^2,\mathcal{M}_{ij})$ has the same expression as (5) with $\{\boldsymbol{\mu}_{ij}, \mathbf{\Lambda}_{ij}, \sigma^2\}$ being treated as unknown parameters. Although this makes sense intuitively, we do not pursue this approach here because integrating (5) with respect to $\{\boldsymbol{\mu}_{ij}, \mathbf{\Lambda}_{ij}, \sigma^2\}$ is an intractable problem. Furthermore, studies in [8] show that the integration in (7) is asymptotically equivalent to substituting the estimate of the nuisance parameters as follows

$$P(\mathbf{x}|\omega,\mathcal{M}_{ij}) \approx P(\mathbf{x}|\omega,\hat{\boldsymbol{\mu}}_{ij},\hat{\mathbf{\Lambda}}_{ij},\hat{\sigma}^2,\mathcal{M}_{ij}), \quad (8)$$

where $\{\hat{\boldsymbol{\mu}}_{ij}, \hat{\mathbf{\Lambda}}_{ij}, \hat{\sigma}^2\}$ are the estimates of $\{\boldsymbol{\mu}_{ij}, \mathbf{\Lambda}_{ij}, \sigma^2\}$. Similarly, for unknown $P(\mathcal{M}_{ij})$ in (4), it can also be tackled by using its estimate $\hat{P}(\mathcal{M}_{ij}|\mathbf{x})$. This results in the so-called empirical Bayesian estimator [11]. In the following, we discuss how to obtain the estimates of $\{\boldsymbol{\mu}_{ij}, \mathbf{\Lambda}_{ij}, \sigma^2, P(\mathcal{M}_{ij})\}$.

For a fixed model $\mathcal{M}_{ij}$, the received data can be written as

$$\mathbf{x} = \mathbf{H}_{ij}(\omega_o)\boldsymbol{\xi}_{ij} + \mathbf{e}. \quad (9)$$

Based on the signal model in (9), the joint ML estimates of parameters $\{\boldsymbol{\xi}_{ij}, \omega_o\}$ are given by minimizing $\Upsilon(\mathbf{x}; \tilde{\boldsymbol{\xi}}_{ij}, \tilde{\omega}_{ij}) = \|\mathbf{x} - \mathbf{H}_{ij}(\tilde{\omega}_{ij})\tilde{\boldsymbol{\xi}}_{ij}\|^2$ [7], where $\tilde{\boldsymbol{\xi}}_{ij}, \tilde{\omega}_{ij}$ are trial values of $\boldsymbol{\xi}_{ij}$ and ω_o respectively. Due to the linearity of parameter $\boldsymbol{\xi}_{ij}$ in (9), the ML estimate for $\boldsymbol{\xi}_{ij}$ (when $\tilde{\omega}_{ij}$ is fixed) is $\hat{\boldsymbol{\xi}}_{ij} = [\mathbf{H}_{ij}^H(\tilde{\omega}_{ij})\mathbf{H}_{ij}(\tilde{\omega}_{ij})]^{-1}\mathbf{H}_{ij}^H(\tilde{\omega}_{ij})\mathbf{x}$. Substituting this result into the cost function $\Upsilon(\mathbf{x}; \hat{\boldsymbol{\xi}}_{ij}, \tilde{\omega}_{ij})$, the ML estimator for ω_o conditioned on $\mathcal{M}_{ij}$ is given by

$$\hat{\omega}_{ij} = \arg\max_{\tilde{\omega}}\{\mathbf{x}^H\mathbf{H}_{ij}(\tilde{\omega})[\mathbf{H}_{ij}^H(\tilde{\omega})\mathbf{H}_{ij}(\tilde{\omega})]^{-1}\mathbf{H}_{ij}^H(\tilde{\omega})\mathbf{x}\}, \tag{10}$$

and the ML estimator for $\boldsymbol{\xi}_{ij}$ is $\hat{\boldsymbol{\xi}}_{ij} = [\mathbf{H}_{ij}^H(\hat{\omega}_{ij})\mathbf{H}_{ij}(\hat{\omega}_{ij})]^{-1}\mathbf{H}_{ij}^H(\hat{\omega}_{ij})\mathbf{x}$. With the channel and CFO estimates under model $\mathcal{M}_{ij}$, an unbiased, consistent estimate of σ^2 can be obtained by taking [12]

$$\hat{\sigma}_{ij}^2 = \frac{1}{N-(j-i+1)}\left\|\mathbf{x} - \mathbf{H}_{ij}(\hat{\omega}_{ij})\hat{\boldsymbol{\xi}}_{ij}\right\|^2. \tag{11}$$

It is known that the ML estimate $\hat{\boldsymbol{\xi}}_{ij}$ is asymptotically Gaussian distributed with mean $\boldsymbol{\xi}_{ij}$ and covariance matrix equals to the Cramer Rao Lower Bound (CRLB) [13]. That is, $\hat{\boldsymbol{\xi}}_{ij} \overset{a}{\sim} \mathcal{N}(\boldsymbol{\xi}_{ij}, \text{CRLB})$. On the other hand, given $\hat{\boldsymbol{\xi}}_{ij}$, it can be easily shown that [14] $\boldsymbol{\xi}_{ij} \overset{a}{\sim} \mathcal{N}(\hat{\boldsymbol{\xi}}_{ij}, \text{CRLB})$. With the CRLB derived for $\boldsymbol{\xi}_{ij}$ in the presence of unknown ω in [15], we can take

$$\hat{\boldsymbol{\mu}}_{ij} = \hat{\boldsymbol{\xi}}_{ij} = [\mathbf{H}_{ij}^H(\hat{\omega}_{ij})\mathbf{H}_{ij}(\hat{\omega}_{ij})]^{-1}\mathbf{H}_{ij}^H(\hat{\omega}_{ij})\mathbf{x} \tag{12}$$

$$\hat{\boldsymbol{\Lambda}}_{ij} = \frac{\hat{\sigma}_{ij}^2}{2}(2[\mathbf{H}_{ij}^H(\hat{\omega}_{ij})\mathbf{H}_{ij}(\hat{\omega}_{ij})]^{-1} + \gamma^{-1}\boldsymbol{\beta}\boldsymbol{\beta}^H), \tag{13}$$

where

$$\begin{aligned}
\gamma &= \hat{\boldsymbol{\xi}}_{ij}^H\mathbf{H}_{ij}^H(\hat{\omega}_{ij})\mathbf{T}\boldsymbol{\Psi}(\hat{\omega}_{ij})\mathbf{T}\mathbf{H}_{ij}(\hat{\omega}_{ij})\hat{\boldsymbol{\xi}}_{ij},\\
\boldsymbol{\beta} &= [\mathbf{H}_{ij}^H(\hat{\omega}_{ij})\mathbf{H}_{ij}(\hat{\omega}_{ij})]^{-1}\mathbf{H}_{ij}^H(\hat{\omega}_{ij})\mathbf{T}\mathbf{H}_{ij}(\hat{\omega}_{ij})\hat{\boldsymbol{\xi}}_{ij},\\
\mathbf{T} &= \text{diag}(0, \cdots, N-1),\\
\boldsymbol{\Psi}(\hat{\omega}_{ij}) &= \mathbf{I} - \mathbf{H}_{ij}(\hat{\omega}_{ij})[\mathbf{H}_{ij}^H(\hat{\omega}_{ij})\mathbf{H}_{ij}(\hat{\omega}_{ij})]^{-1}\mathbf{H}_{ij}^H(\hat{\omega}_{ij}).
\end{aligned}$$

Note that the estimated values of ω, $\boldsymbol{\xi}_{ij}$ and σ_{ij}^2 are used instead to find an approximate covariance because the CRLB depends on the true value of ω, $\boldsymbol{\xi}_{ij}$ and σ_{ij}^2. Therefore, putting (11), (12) and (13) into (8), we can obtain an approximation to $P(\mathbf{x}|\omega, \mathcal{M}_{ij})$.

Finally, to estimate the posterior probabilities $P(\mathcal{M}_{ij})$ for each possible $\mathcal{M}_{ij}$, we could follow the derivation from [16] and obtain

$$\hat{P}(\mathcal{M}_{ij}|\mathbf{x}) = \frac{\exp(-\frac{1}{2}\text{BIC}_{ij})}{\sum_{i=0}^{L_{\text{CP}}-1}\sum_{j=i}^{L_{\text{CP}}-1}\exp(-\frac{1}{2}\text{BIC}_{ij})}, \tag{14}$$

where

$$\begin{aligned}
\text{BIC}_{ij} \triangleq &-2\ln[P(\mathbf{x}|\hat{\omega}_{ij}, \hat{\boldsymbol{\xi}}_{ij}, \hat{\boldsymbol{\Lambda}}_{ij}, \hat{\sigma}_{ij}^2, \mathcal{M}_{ij})]\\
&+ 2(j-i+1)\ln(2N)
\end{aligned} \tag{15}$$

is the Bayesian information criterion (BIC) for model $\mathcal{M}_{ij}$.

V. Reduction of Model Space

The number of possible models in (4) often renders the exhaustive summation impractical. In general, the models to be averaged can be reduced to a subset of models Φ that are strongly supported by the data [17]. For example, in [17], Madigan and Raftery argued that if a model predicts the data far less well than the model which provides the best predictions, then it should no longer be considered. More specifically, defining the most likely model as $\mathcal{M}_{\hat{m}\hat{n}}$ with

$$\{\hat{m}, \hat{n}\} = \underset{0\le m\le n\le L_{\text{CP}}-1}{\arg\max}\ \hat{P}(\mathcal{M}_{mn}|\mathbf{x}), \tag{16}$$

the models not belonging to

$$\Phi = \{\mathcal{M}_{ij}\ :\ \frac{\hat{P}(\mathcal{M}_{\hat{m}\hat{n}}|\mathbf{x})}{\hat{P}(\mathcal{M}_{ij}|\mathbf{x})} \le \mathcal{C}\}, \tag{17}$$

should be excluded from (4), where $\mathcal{C}$ is a design parameter. However, this method involves a subjective selection of parameter $\mathcal{C}$, while there is no general guideline for determining its optimal value. Here, we focus on the most likely model $\mathcal{M}_{\hat{m}\hat{n}}$ and then average the estimator over an expanded class of models "near" $\mathcal{M}_{\hat{m}\hat{n}}$ as [18]

$$\begin{aligned}
\Phi = \{\mathcal{M}_{ij}\ :\ &i \in \{\hat{m}-1, \hat{m}, \hat{m}+1\},\\
&j \in \{\hat{n}-1, \hat{n}, \hat{n}+1\}, i \le j\}
\end{aligned} \tag{18}$$

where in this case, only the adjacent models are examined.

It is noticed that (16) by nature is a model selection problem. As we see in (16), a two dimensional search over (m, n) is needed to locate the most likely model. The computational complexity could be quite high if L_{CP} is large. In the following, we propose a low complexity model selection algorithm by exploiting the information provided by the local behavior of the BIC rule.

Based on (15), we compare the BIC values of two adjacent models

$$\text{BIC}_{(i,j)} - \text{BIC}_{(i,j-1)} = \lambda_{(i,j)} + 2\ln(2N), \tag{19}$$

where

$$\begin{aligned}
\lambda_{(i,j)} =&2\ln P(\mathbf{x}|\hat{\omega}_{(i,j-1)}, \hat{\boldsymbol{\xi}}_{(i,j-1)}, \hat{\boldsymbol{\Lambda}}_{(i,j-1)}, \hat{\sigma}_{(i,j-1)}^2, \mathcal{M}_{(i,j-1)})\\
&- 2\ln P(\mathbf{x}|\hat{\omega}_{(i,j)}, \hat{\boldsymbol{\xi}}_{(i,j)}, \hat{\boldsymbol{\Lambda}}_{(i,j)}, \hat{\sigma}_{(i,j)}^2, \mathcal{M}_{(i,j)})
\end{aligned}$$

is the difference of the "goodness-of-fit" to the data between two adjacent models. In order to avoid confusions, here we use $\text{BIC}_{(i,j)}$ to denote BIC_{ij}, and the same notation is also applied to $\hat{\omega}_{(i,j)}$, $\hat{\boldsymbol{\xi}}_{(i,j)}$, $\hat{\boldsymbol{\Lambda}}_{(i,j)}$ and $\hat{\sigma}_{(i,j)}^2$. If both $\mathcal{M}_{(i,j)}$ and $\mathcal{M}_{(i,j-1)}$ fit the data very well, $\lambda_{(i,j)}$ is approximately zero, and $\text{BIC}_{(i,j)} - \text{BIC}_{(i,j-1)}$ will be close to $2\ln(2N)$. On the other hand, if $\mathcal{M}_{(i,j)}$ fits the data much better than $\mathcal{M}_{(i,j-1)}$, $\lambda_{(i,j)}$ is a negative number with large magnitude, and $\text{BIC}_{(i,j)} - \text{BIC}_{(i,j-1)}$ will be a number very different from $2\ln(2N)$.

Therefore, by comparing the difference in (19) to a threshold, we can decide whether the additional element ξ_j (at the end of $\boldsymbol{\xi}_{(i,j)}$) is different from zero. If the threshold is taken as $2\ln(2N)$ [19], the decision criterion can be mathematically stated as

$$\text{BIC}_{(i,j)} - \text{BIC}_{(i,j-1)} \underset{\xi_j\neq 0}{\overset{\xi_j=0}{\gtrless}} 0. \tag{20}$$

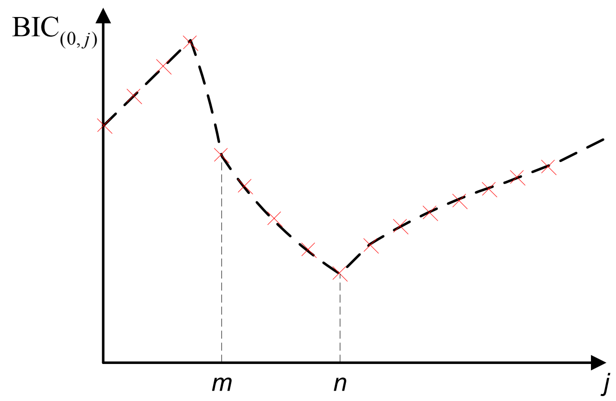

Fig. 1. A typical variation of $\text{BIC}_{(0,j)}$ for increasing value of $j \in \{0, ...L_{\text{CP}} - 1\}$.

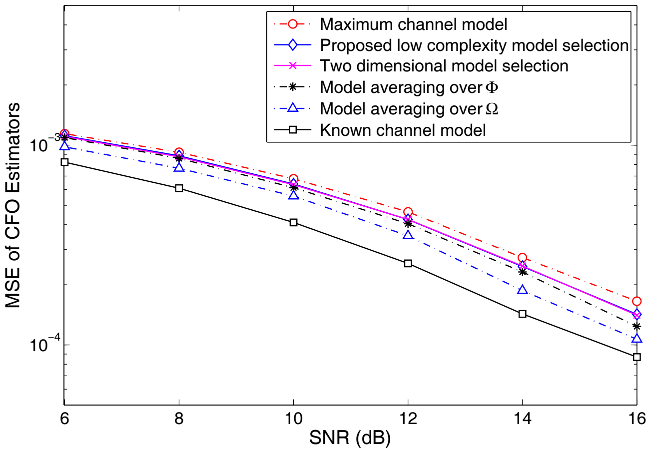

Fig. 2. MSE performance of different estimators at low and medium SNR.

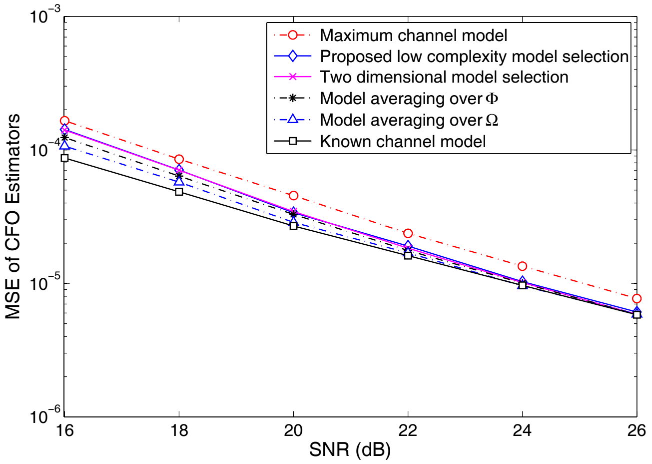

Fig. 3. MSE performance of different estimators at high SNR.

Similarly, we can obtain

$$\text{BIC}_{(i,j)} - \text{BIC}_{(i+1,j)} \underset{\xi_i \neq 0}{\overset{\xi_i = 0}{\gtrless}} 0, \tag{21}$$

for determining whether the additional element at the front of $\boldsymbol{\xi}_{(i,j)}$ is different from zero. Summarizing the discussion above, the two dimensional search in (16) can be reduced to one dimensional search by locating the first and last propagation channel path with significant energy. More specifically, the first propagation channel path $\xi_{\hat{m}}$ and the last propagation channel path $\xi_{\hat{n}}$ can be located as

$$\hat{m} = \min\{j\} \text{ subject to } \text{BIC}_{(i,j)} - \text{BIC}_{(i,j-1)} < 0, \tag{22}$$

$$\hat{n} = \max\{j\} \text{ subject to } \text{BIC}_{(i,j)} - \text{BIC}_{(i,j-1)} < 0, \tag{23}$$

where i should be smaller than the position of the first propagation path. Without loss of generality, i is set to be 0. A typical variation of $\text{BIC}_{(0,j)}$, as j increases, is shown in Fig. 1. As it is shown in the figure, m and n are the starting and ending order whose BIC value would decrease from the previous order respectively.

VI. Simulation Results and Discussions

In this section, simulation results are presented to illustrate the MSE performance of different empirical CFO MAP estimators. In particular, the following six estimators are compared: 1) Model averaged over full model space Ω; 2) Model selection by 2-D search in (16); 3) Model selection using local BIC information in (22) and (23); 4) Model averaged over the reduced set Φ in (18); 5) ML CFO estimator using maximum channel model [7]; and 6) CFO estimator with known timing offset and channel length.

In all simulations, $N = 64, L_{\text{CP}} = 16$, which is consistent with the WLAN standard [20]. The OFDM training symbol is constructed by transmitting a Chu-sequence in frequency domain [7]. A multipath Rayleigh fading channel of length $L = 4$ is considered. Exponential power delay profile (normalized to unit power) is used for the channel. The timing offset is set as $\theta_o = 4$ before the ideal timing. Without loss of generality, the CFO ω_o is generated as a random variable uniformly distributed in $2\pi \times [-0.5, 0.5]/N$ and it is assumed that there is no prior information for CFO. All results are averaged over 5000 Monte Carlo runs.

Fig. 2 shows the performance of the above estimators in low to medium signal-to-noise ratios (SNR). It can be seen that the algorithm with two dimensional model selection has the same performance as the proposed low complexity method using local BIC information in (22) and (23), and they both perform better than the maximal model method. Furthermore, it is obvious that the algorithms with model averaging have noticeable performance improvements over the algorithms based on model selection. It is also found that averaging algorithm over the full model space Ω in general has a better performance than that over the reduced set Φ in (18), forming a performance-complexity tradeoff.

Fig. 3 shows the corresponding results at high SNR. It can be shown that the performance of algorithms with model averaging are better than those based on model selection and maximum possible model. But at high enough SNR, the performance of algorithms based on the model averaging and model selection all converge to that of the algorithm based on known channel model. It is because at high SNR, the model selection algorithm can correctly identify the timing offset and channel length. On the other hand, the algorithm based on the maximum possible model has an irreducible gap from the known channel model case.

Notice that in terms of complexity, model-averaging algorithm may not be a favorable choice. However, the work reported in this letter represents a systematic study and rigorous derivation of optimal CFO estimator when the timing offset and channel length are not known. The optimal estimator automatically weighs all the available information and achieves the best performance over the whole SNR range. The framework presented includes the model selection method as a special case. Furthermore, the method in [7] can also be viewed as a special case by putting all the weighting to one model only.

VII. Conclusions

In this letter, a Bayesian multi-model based CFO estimator was developed for OFDM systems, with timing offset and channel length being unknown. We also derived an empirical version of the estimator and proposed a low complexity model space reduction method. Results showed that model-averaging based estimator performs better than the estimator based on model selection and provides performance close to the estimator with known timing offset and channel length.

References

[1] A. R. S. Bahai, B. R. Saltzberg, and M. Ergen, *Multi-Carrier Digital Communications: Theory and Applications of OFDM*. New York: Springer, 2005.

[2] P. H. Moose, “A technique for orthogonal frequency division multiplexing frequency offset correction," *IEEE Trans. Commun.*, vol. 41, pp. 2908-2914, Nov. 1994.

[3] J. J. Van de Beek, M. Sandell, and P. O. Borjesson, “ML estimation of time and frequency offset in OFDM systems," *IEEE Trans. Commun.*, vol. 45, pp. 1800-1805, Nov. 1997.

[4] U. Tureli, H. Liu, and M. D. Zoltowski, “OFDM blind carrier offset estimation: ESPRIT," *IEEE Trans. Commun.*, vol. 48, no. 9, pp. 1459-1461, Nov. 2000.

[5] T. M. Schmidl and D. C. Cox, “Robust frequency and timing synchronization for OFDM," *IEEE Trans. Commun.*, vol. 45, no. 12, pp. 1613-1621, 1997.

[6] H. Minn, M. Zeng, and V. K. Bhargava, “On timing offset estimation for OFDM systems," *IEEE Commun. Lett.*, vol. 4, pp. 242-244, July 2000.

[7] J.-W. Chen, Y.-C. Wu, S.-D. Ma, and T.-S. Ng, “ML joint CFO and channel estimation in OFDM systems with timing ambiguity," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 2436-2440, July 2008.

[8] G. L. Bretthorst, *Bayesian Spectrum Analysis and Parameter Estimation, Lecture Notes in Statistics.* New York: Springer-Verlag, 1988.

[9] P. A. Parker, P. Mitran, D. W. Bliss, and V. Tarokh, “On bounds and algorithm for frequency synchronization for collaborative communication systems," *IEEE Trans. Signal Procss.*, vol. 56, pp. 3742-3752, Aug. 2008.

[10] K. Cai, X. Li, and Y.-C. Wu, “Bayesian CFO estimation in OFDM systems," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, 2009.

[11] B. P. Carlin and T. A. Louis, *Bayes and Empirical Bayes Methods for Data Analysis.* London: Chapman Hall, 1996.

[12] E. G. Larsson and Y. Selen,“Linear regression with a sparse parameter vector," *IEEE Trans. Signal Process.*, vol. 55, pp. 451-460, Feb. 2007.

[13] S. M. Kay, *Fundamentals of Statistical Signal Processing.* Upper Saddle River, NJ: Prentice-Hall, 1998.

[14] E. G. Larsson, Y. Selen, and P. Stoica, “Adaptive equalization for frequency-selective channels of unknown length," *IEEE Trans. Veh. Technol.*, vol. 54, pp. 568-579, Mar. 2005.

[15] 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.

[16] P. Stoica and Y. Selen, “Model order selection: a review of information criterion rules," *IEEE Trans. Signal Process.*, vol. 21, pp. 36-47, July 2004.

[17] D. Madigan and A. E. Raferty, “Model selection and accounting for model uncertainty in graphical models using Occam’s window," *J. Amer. Statist. Assoc.*, vol. 89, no. 428, pp. 1535-1546, Dec. 1994.

[18] D. Draper, “Assessment and propagation of model uncertainty," *J. Roy. Statist. Soc. Ser. B,* vol. 57, pp. 45-97, 1995.

[19] P. Stoica, Y. Selen, and J. Li, “On information criteria and the generalized likelihood ratio test of model order selection," *IEEE Signal Process. Lett.*, vol. 21, pp. 794-797, Oct. 2004.

[20] IEEE Std 802.11a, Supplement to Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specification: High-speed Physical Layer in the 5GHz Band, IEEE Std. 802.11a-1999, 1999