

# Frequency Synchronization for Multiuser MIMO-OFDM System Using Bayesian Approach

Jianwen Chen<sup>†</sup>, Jianwu Chen<sup>\*</sup>, Yik-Chung Wu<sup>\*</sup>

<sup>†</sup> IBM Research China, Bei Jing, China

<sup>\*</sup> Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong

Email: jianwenc@cn.ibm.com, {jwchen, ycwu}@eee.hku.hk

**Abstract**—This paper addresses the problem of frequency synchronization in multiuser multiple-input multiple-output orthogonal frequency-division multiplexing (MIMO-OFDM) systems. Different from existing work, a Bayesian approach is used in the parameter estimation problem. In this paper, the Bayes estimator for carrier frequency offset (CFO) estimation is proposed and the Bayesian Cramér-Rao bound (BCRB) is also derived in closed form. Direct implementation of the resultant estimation scheme with conventional methods is challenging since a high degree of mathematical sophistication is always required. To solve this problem, the Gibbs sampler is exploited with an efficient sample generation method. Simulation results illustrate the effectiveness of the proposed estimation scheme.

## I. INTRODUCTION

Multiple-input multiple-output (MIMO) system has attracted tremendous interest recently due to its ability to mitigate fading effects in wireless channels and to provide great capacity improvements [1]. The first advantage is achieved by increasing the diversity order. Besides the inherent diversity provided by multiple transmit antennas, efficient space-time coding strategies (e.g., space-time block coding) have been proposed in [2], [3]. To improve the spectral efficiency of MIMO system, several schemes denoted as layered space-time (V-BLAST) have been proposed in [4], [5].

In these schemes [2], [3], [4], [5], the perfect frequency synchronization is always assumed. However, such an assumption is too restrictive in practice due to mismatch of the local oscillator in transceivers and/or Doppler shifts caused by mobility. Since the performance of these schemes may seriously degrade due to the presence of frequency offsets, it is critical to compensate these offsets prior to performing detection. Thus in the literature, several works have been proposed on frequency synchronization in the MIMO context. In [6], the problem of joint carrier frequency offset (CFO) and channel estimation is investigated for a single user MIMO system. The obvious limit in the considered system of [6] is that only one CFO between the transmitter and receiver is assumed. In [7], a more general MIMO model with different CFOs for distinct transmit antennas is considered. However, in this scheme, the channel are assumed to be flat-fading. More recently, in [8], the optimal joint CFO and channel estimation scheme has been derived for a multiuser MIMO-OFDM. The channels in this paper are frequency selective fading channels, which are different from the case in [7]. In all these schemes, the synchronization parameters to be estimated are regarded

as deterministic unknown. However, in practice, it is more reasonable to model the CFO as a random process with known prior density (see remark 1 for details). It is valuable to exploit this quantitative information in the estimation procedure to improve the performance.

In this paper, we consider the frequency synchronization problem in a multiuser MIMO-OFDM system. Despite the fact that it may reduce the information rate, the use of training sequences to estimate frequency offsets remains an effective solution, and thus, we consider this situation. Different from the previously published papers (e.g., [6]-[8]), we assume that the channels are frequency selective fading and the prior information about the CFOs is available at receiver. To obtain a universal scheme using the available prior information, the Bayesian framework is exploited. The contributions of this paper are as follows. Firstly, exploiting the prior information of the parameters, the BCRB for the CFO estimation is derived in closed form. Secondly, by integrating out the nuisance parameters, the minimum mean squared error (MMSE) estimator is proposed. It is found that the direct implementation of this estimator with conventional methods always requires a high degree of mathematical sophistication. To overcome this problem and obtain a universal scheme, the Gibbs sampling algorithm is exploited.

The rest of this paper is organized as follows. In Section II, the considered multiuser MIMO-OFDM system model is presented. In Section III, the BCRB considering the prior information are derived. In Section IV, the Bayes estimator for the CFO estimation is derived from MMSE prospective and an efficient algorithm based on Gibbs sampling is proposed to reduce the complexity. Section V presents simulation results to validate the proposed estimation scheme. Concluding remarks are given in Section VI.

## II. SIGNAL MODEL

In the considered MIMO-OFDM system,  $K$  users transmit different data streams simultaneously using the same set of subcarriers to the base station, which is equipped with  $M_r$  antennas and is responsible for decoding the symbols for each user. For each user, only one transmit antenna is assumed. It is reasonable to assume that the receive antennas at base station share the same oscillator while all users are driven by different oscillators. Thus the data streams from different users will experience different CFOs. The data stream from

user  $i$  is first segmented into blocks of length  $N$  (denoted as  $\mathbf{d}_i = [d_i(0), \dots, d_i(N-1)]^T$ ) and then modulated onto different subcarriers by left multiplying an  $N$ -point inverse FFT matrix  $\mathbf{F}^H$ , where  $\mathbf{F}$  is the FFT matrix with  $\mathbf{F}(k, l) = \frac{1}{\sqrt{N}} \exp(-j2\pi kl/N)$ . After inserting a cyclic prefix (CP) of length  $L_{cp}$  into each block of the time domain signal (denoted as  $\mathbf{F}^H \mathbf{d}_i$ ), the augmented block is serially transmitted through the multipath channel. Let the channel impulse response (including all transmit/receive filtering effects) between the  $i$ th user and the  $j$ th receive antenna be denoted as  $\mathbf{h}_{i,j} = [h_{i,j}(0), \dots, h_{i,j}(L-1)]^T$ , where  $L$  is the upper bound on the length of the longest channel. For user  $i$ , the normalized CFO (between the oscillator at user  $i$  and that of the base station) is denoted as  $\varepsilon_i$ . At the base station, after timing synchronization and removal of CP, the signal from user  $i$  to the  $j$ th receive antenna is given by

$$\mathbf{x}_{ij} \triangleq [x_{ij}(0), \dots, x_{ij}(N-1)]^T = \mathbf{\Gamma}(\omega_i) \mathbf{A}_i \mathbf{h}_{i,j} \quad (1)$$

where

$$\mathbf{\Gamma}(\omega_i) \triangleq \text{diag}(1, \dots, \exp(j(N-1)\omega_i)) \quad (2)$$

$$\mathbf{A}_i = \mathbf{F}^H \mathbf{D}_i \mathbf{F}_L \quad (3)$$

$$\mathbf{D}_i \triangleq \text{diag}(\mathbf{d}_i) \quad (4)$$

$$\mathbf{F}_L = \mathbf{F}(:, 1:L) \quad (5)$$

$$\omega_i \triangleq 2\pi\varepsilon_i/N. \quad (6)$$

Since the received signal at the  $j$ th receive antenna of the base station is the sum of signals from all users and noise, the received signal at this antenna is given by

$$\mathbf{x}_j = \sum_{i=1}^K \mathbf{x}_{ij} + \mathbf{n}_j = \sum_{i=1}^K \mathbf{\Gamma}(\omega_i) \mathbf{A}_i \mathbf{h}_{i,j} + \mathbf{n}_j. \quad (7)$$

In above, the vector  $\mathbf{n}_j$  is complex white Gaussian noise with zero mean and covariance matrix  $C_{\mathbf{n}_j} = E\{\mathbf{n}_j \mathbf{n}_j^H\} = \sigma^2 \mathbf{I}_{N \times N}$ .

Denoting  $\mathbf{x} = [\mathbf{x}_1^T \ \mathbf{x}_2^T \ \dots \ \mathbf{x}_{M_r}^T]^T$ ,  $\boldsymbol{\omega} = [\omega_1 \ \omega_2 \ \dots \ \omega_K]$ ,  $\mathbf{n} = [\mathbf{n}_1^T \ \mathbf{n}_2^T \ \dots \ \mathbf{n}_{M_r}^T]^T$ ,  $\mathbf{h} = [\mathbf{h}_1^T \ \mathbf{h}_2^T \ \dots \ \mathbf{h}_{M_r}^T]^T$  with  $\mathbf{h}_j = [\mathbf{h}_{1,j}^T \ \mathbf{h}_{2,j}^T \ \dots \ \mathbf{h}_{K,j}^T]^T$ , the signal model from (7) can be rewritten as

$$\mathbf{x} = \mathbf{Q}(\boldsymbol{\omega}) \mathbf{h} + \mathbf{n} \quad (8)$$

where

$$\mathbf{Q}(\boldsymbol{\omega}) = \mathbf{I}_{M_r \times M_r} \otimes \mathbf{A} \mathbf{\Gamma}(\boldsymbol{\omega}) \quad (9)$$

$$\mathbf{A} \triangleq [\mathbf{A}_1 \ \mathbf{A}_2 \ \dots \ \mathbf{A}_K] \quad (10)$$

$$\mathbf{\Gamma}(\boldsymbol{\omega}) \triangleq \begin{bmatrix} \mathbf{\Gamma}(\omega_1) & 0 & \dots & 0 \\ 0 & \mathbf{\Gamma}(\omega_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{\Gamma}(\omega_K) \end{bmatrix}. \quad (11)$$

*Remark 1:* In practical communication systems, the short-term stability of oscillator at transceiver is usually evaluated by Allan Variance [9], which is defined as  $\sigma_f^2 = \frac{1}{2} \mathcal{E}\{(\Delta f)^2\}$ , with  $\Delta f$  as the difference between two consecutive measurements of frequency of the oscillator. Since in the signal model,

the CFO is assumed to be constant in one training block, which corresponds to one measurement for this block. Thus, the uncorrelated normalized CFO measurements in different training blocks can be modeled as a random process with zero mean and the variance  $\sigma_\omega^2 = \sigma_f^2/f_s^2$ , where  $f_s$  is sampling frequency. Since the Gaussian distribution is the most noninformative distribution having maximum entropy compared to other distributions having the same variance, it is further assumed that  $\omega_i \sim \mathcal{N}(0, \sigma_\omega^2)$ .

*Remark 2:* The considered signal model in this paper can be straightforwardly extended to other systems, such as OFDMA and OFDM/SDMA. The only difference lies in how the matrices  $\mathbf{A}_i, i = 1, \dots, K$  are constructed.

### III. BAYESIAN CRAMÉR-RAO BOUND

Different from the previous work using ML approach, in this paper, the channels are assumed to be random processes. In particular, the Rayleigh fading channel is considered. Thus, the channel vector  $\mathbf{h}_{i,j}$  is assumed to be a complex Gaussian vector with zero mean and covariance matrix  $\boldsymbol{\eta}_{i,j} \triangleq \text{diag}(\boldsymbol{\lambda})_{i,j}$ . The vector  $\boldsymbol{\lambda}_{i,j} = [\lambda_{i,j}(0), \dots, \lambda_{i,j}(L-1)]$  is the delay power profile.

From the stochastic received signal, we have [10]

$$\mathbf{x} \sim \mathcal{N}(0, \mathbf{R}_\mathbf{x}) \quad (12)$$

where

$$\mathbf{R}_\mathbf{x} = \mathbf{Q}(\boldsymbol{\omega}) \boldsymbol{\eta} \mathbf{Q}^H(\boldsymbol{\omega}) + \sigma^2 \mathbf{I}_{M_r N} \quad (13)$$

$$\boldsymbol{\eta} = \text{diag}(\boldsymbol{\lambda}_{1,1}, \dots, \boldsymbol{\lambda}_{K,1}, \dots, \boldsymbol{\lambda}_{1,M_r}, \dots, \boldsymbol{\lambda}_{K,M_r}). \quad (14)$$

For the unknown parameters  $\boldsymbol{\omega}$ , the covariance matrix of any unbiased estimator  $\hat{\boldsymbol{\omega}}$  cannot be lower than the CRB, which is the inverse of the Fisher information matrix (FIM) [11]

$$\mathcal{E}\{(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega})(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega})^T\} \geq \text{CRB}(\boldsymbol{\omega}) = \mathbf{J}_{\boldsymbol{\omega}}^{-1}. \quad (15)$$

Since the unknown parameters in  $\boldsymbol{\omega}$  are random, the FIM is expressed in the following form [12]

$$\mathbf{J}_{\boldsymbol{\omega}, \boldsymbol{\omega}} = \mathbf{J}_D + \mathbf{J}_P \quad (16)$$

where the  $K \times K$  matrices  $\mathbf{J}_D$  and  $\mathbf{J}_P$  (representing information from the data and from the known prior distributions respectively) are given by [13]

$$\mathbf{J}_D(k, l) = \text{tr}\left\{ \mathbf{R}_\mathbf{x}^{-1} \frac{\partial \mathbf{R}_\mathbf{x}}{\partial \omega_k} \mathbf{R}_\mathbf{x}^{-1} \frac{\partial \mathbf{R}_\mathbf{x}}{\partial \omega_l} \right\} \quad (17)$$

$$\mathbf{J}_P(k, l) = -\mathcal{E}\left[ \frac{\partial^2 \ln \rho(\boldsymbol{\omega})}{\partial \omega_k \partial \omega_l} \right] \quad (18)$$

where  $f(\boldsymbol{\omega})$  is the prior joint density function for the vector  $\boldsymbol{\omega}$ . Based on (13), we have

$$\frac{\partial \mathbf{R}_\mathbf{x}}{\partial \omega_k} = \mathcal{J}[\boldsymbol{\beta}_k - \boldsymbol{\beta}_k^H] \quad (19)$$

where

$$\boldsymbol{\beta}_k = [\mathbf{I}_{M_r} \otimes \mathbf{M} \mathbf{\Gamma}(\omega_k) \mathbf{A}_k] \mathbf{A}_k [\mathbf{I}_{M_r} \otimes \mathbf{A}_k^H \mathbf{\Gamma}^H(\omega_k)] \quad (20)$$

$$\boldsymbol{\Lambda}_k = \text{diag}(\boldsymbol{\eta}_{k,1}, \dots, \boldsymbol{\eta}_{k,M_r}). \quad (21)$$

Substituting (19)–(21) into (17), the elements in  $\mathbf{J}_D$  are given by

$$\mathbf{J}_D(k, l) = -\text{tr}\left\{\mathbf{R}^{-1}[\bar{\boldsymbol{\beta}}_k - \bar{\boldsymbol{\beta}}_k^H]\mathbf{R}^{-1}[\bar{\boldsymbol{\beta}}_l - \bar{\boldsymbol{\beta}}_l^H]\right\} \quad (22)$$

where

$$\mathbf{R} = \mathbf{Q}(\boldsymbol{\omega})\bar{\boldsymbol{\eta}}\mathbf{Q}^H(\boldsymbol{\omega}) + \mathbf{I}_{M_r N} \quad (23)$$

$$\bar{\boldsymbol{\eta}} = \frac{1}{\sigma^2} = \text{diag}\left(\frac{1}{\sigma^2}\boldsymbol{\lambda}_{1,1}, \dots, \frac{1}{\sigma^2}\boldsymbol{\lambda}_{K,1}, \dots, \frac{1}{\sigma^2}\boldsymbol{\lambda}_{1,M_r}, \dots, \frac{1}{\sigma^2}\boldsymbol{\lambda}_{K,M_r}\right) \quad (24)$$

$$\begin{aligned} \bar{\boldsymbol{\beta}}_k &= \frac{\boldsymbol{\beta}_k}{\sigma^2} \\ &= [\mathbf{I}_{M_r} \otimes \mathbf{M}\boldsymbol{\Gamma}(\omega_k)\mathbf{A}_k]\bar{\boldsymbol{\Lambda}}_k[\mathbf{I}_{M_r} \otimes \mathbf{A}_k^H\boldsymbol{\Gamma}^H(\omega_k)] \end{aligned} \quad (25)$$

$$\bar{\boldsymbol{\Lambda}}_k = \text{diag}\left(\frac{1}{\sigma^2}\boldsymbol{\lambda}_{k,1}, \dots, \frac{1}{\sigma^2}\boldsymbol{\lambda}_{k,M_r}\right). \quad (26)$$

Since the CFOs for different users are independent, the joint distribution function  $\rho(\boldsymbol{\omega})$  is given by

$$\rho(\boldsymbol{\omega}) = \rho(\omega_1) \cdots \rho(\omega_K) \quad (27)$$

where

$$\rho(\omega_i) = \frac{1}{\sqrt{2\pi\sigma_{\omega_i}^2}} \exp\left(-\frac{\omega_i^2}{2\sigma_{\omega_i}^2}\right). \quad (28)$$

Thus, it is obvious that only the diagonal elements in  $\mathbf{J}_P$  will be non-zero and are given by

$$\mathbf{J}_P(k, k) = \frac{1}{\sigma_{\omega_i}^2}. \quad (29)$$

The exact BCRB of the estimation can be calculated through substituting (22) and (29) into (15).

#### IV. CFO ESTIMATION UNDER BAYESIAN FRAMEWORK

##### A. Posterior Density for CFOs

In the signal model in (8), the likelihood distribution of the observation vector  $\mathbf{x}$  is denoted as  $P(\mathbf{x}|\boldsymbol{\omega}, \mathbf{h}, \sigma)$ , which depends on the value of the three unknown parameters ( $\boldsymbol{\omega}$ ,  $\mathbf{h}$  and the noise variance  $\sigma^2$ ). The researcher's subjective knowledge about the three unknown parameters is comprised in the prior distribution  $P(\boldsymbol{\omega}, \mathbf{h}, \sigma)$ . Bayesian theory then tell us:

$$\begin{aligned} P(\mathbf{x}|\boldsymbol{\omega}, \mathbf{h}, \sigma)P(\boldsymbol{\omega}, \mathbf{h}, \sigma) &= P(\mathbf{x}, \boldsymbol{\omega}, \mathbf{h}, \sigma) \\ &= P(\boldsymbol{\omega}, \mathbf{h}, \sigma|\mathbf{x})P(\mathbf{x}) \end{aligned} \quad (30)$$

Thus, given the observation  $\mathbf{x}$ , the posterior density for  $(\boldsymbol{\omega}, \mathbf{h}, \sigma)$  is

$$\begin{aligned} P(\boldsymbol{\omega}, \mathbf{h}, \sigma|\mathbf{x}) &= \frac{P(\mathbf{x}|\boldsymbol{\omega}, \mathbf{h}, \sigma)P(\boldsymbol{\omega}, \mathbf{h}, \sigma)}{P(\mathbf{x})} \\ &\propto P(\mathbf{x}|\boldsymbol{\omega}, \mathbf{h}, \sigma)P(\boldsymbol{\omega})P(\mathbf{h})P(\sigma) \end{aligned} \quad (31)$$

where the marginal density of  $P(\mathbf{x})$  can be regarded as normalizing constant as it is independent of  $(\boldsymbol{\omega}, \mathbf{h}, \sigma)$ . In the synchronization procedure, the channel statistics are usually unavailable. Thus in this ignorant case, we assign uniform

priors to the  $\mathbf{h}$ . According to the arguments in remark 1, for the uncorrelated CFOs of all users, we have

$$P(\boldsymbol{\omega}) = \frac{1}{(\sqrt{2\pi})^K \sqrt{|\boldsymbol{\phi}|}} \exp(-\boldsymbol{\omega}^T \boldsymbol{\phi}^{-1} \boldsymbol{\omega}) \quad (32)$$

where  $\boldsymbol{\phi} = \text{diag}(\sigma_{\omega_1}^2 \dots \sigma_{\omega_K}^2)$ .

Since our main interest here is to estimate the synchronization parameters  $\boldsymbol{\omega}$ , the unknown channel vector  $\mathbf{h}$  and the noise variance  $\sigma^2$  are considered as nuisance parameters.

Proposition 1: Under the non-informative prior for the noise variance  $P(\sigma) \propto 1/\sigma$  and the above prior distributions for  $\mathbf{h}$  and  $\boldsymbol{\omega}$ , the marginal density of  $\boldsymbol{\omega}$  is given by

$$\begin{aligned} P(\boldsymbol{\omega}|\mathbf{x}) &\propto \int P(\mathbf{x}|\boldsymbol{\omega}, \mathbf{h}, \sigma)P(\boldsymbol{\omega}, \mathbf{h}, \sigma)d\mathbf{h}d\sigma \\ &= P(\mathbf{x}|\boldsymbol{\omega}, \mathbf{h}, \sigma)P(\boldsymbol{\omega})P(\mathbf{h})P(\sigma)d\mathbf{h}d\sigma \\ &\propto \frac{\exp(-\boldsymbol{\omega}^T \boldsymbol{\phi}^{-1} \boldsymbol{\omega})}{|\mathbf{Q}^H \mathbf{Q}|^{(M_r N - M_r K L)}} \end{aligned} \quad (33)$$

where  $\mathbf{Q}$  represents  $\mathbf{Q}(\boldsymbol{\omega})$  for expression simplicity and  $\mathbf{Q}^\perp = (\mathbf{I}_{M_r N} - \mathbf{Q}(\mathbf{Q}^H \mathbf{Q})^{-1} \mathbf{Q}^H)$ .

Proof: see Appendix A.

Under Bayesian framework, in order to obtain the estimate of  $\omega_k$  and its statistic characteristics (e.g. mean and variance), the marginal density  $P(\omega_k|\mathbf{x})$  is indispensable. It is obvious that  $P(\omega_k|\mathbf{x})$  can be obtained as

$$\begin{aligned} P(\omega_k|\mathbf{x}) &= \int \cdots \int P(\boldsymbol{\omega}|\mathbf{x})d\omega_1 \cdots d\omega_{k-1}d\omega_{k+1} \cdots d\omega_K \\ & \quad k = 1, \dots, K. \end{aligned} \quad (34)$$

With the above marginal densities, the minimum mean-squared error (MMSE) estimator can be derived as:

$$\hat{\omega}_{k-mmse} = \int P(\omega_k|\mathbf{x})\omega_k d\omega_k \quad k = 1, \dots, K. \quad (35)$$

In the implementation of the above estimator, the calculation of the multidimensional integration in (34) is required. It is obvious that the direct mathematical integration of (34) is impossible. In the literature, there are several approaches to handle the multidimensional integration. In [14], [15], the Laplace's method has been proposed to approximate the integrals in calculating marginal posterior distributions. In this scheme, the function to be integrated is expanded at a certain point and then some approximations are made to derive the integration result. However, the performance of this scheme is highly depended on the choice of the optimal expanding point, which is usually difficult to obtain. Monte Carlo method, which is numerical quadrature using pseudorandom numbers, has been proposed to approximate evaluation of definite integrals [16]. The traditional version of this algorithm suffers a low efficiency since the evaluation points are uniformly distributed over the integration region. To improve the efficiency, some techniques such as importance sampling and stratified sampling have been proposed in [17],

[18], [19]. However, the implementation of these algorithms requires a high degree of mathematical sophistication (i.e., the choice of importance function). For the implementation of above Bayes estimator, universal and efficient solutions are demanded.

### B. Estimation Via Gibbs Sampling

Markov Chain Monte Carlo (MCMC) methods have been regraded as a major breakthrough for the implementation of Bayesian analysis since any multidimensional integration can be performed by using MCMC. In MCMC methods, a Markov chain is constructed, whose equilibrium distribution is the joint posterior. Thus, after running the Markov chain for a certain “burn-in” period (after that the Markov chain reached convergence), one obtains samples from the limiting distribution. Then the resulting samples are used to approximate the required high dimensional integrations.

The Gibbs sampler is a special form of MCMC method which provides an alternative method for obtaining  $P(\omega_k|\mathbf{x})$ . Rather than compute or approximate  $P(\omega_k|\mathbf{x})$  directly, the Gibbs sampler allows us effectively to generate samples  $\omega_k^1, \dots, \omega_k^m \sim P(\omega_k|\mathbf{x})$  without requiring  $P(\omega_k|\mathbf{x})$ .

With arbitrary starting values  $\omega_1^0, \dots, \omega_K^0$ , the Gibbs sampler proceeds through an iterative scheme. In a cycle, samples are generated from each of the full conditional posterior distributions:

$$\begin{aligned} &\omega_1^1 \text{ from } P(\omega_1|\omega_2^0, \dots, \omega_K^0, \mathbf{x}) \\ &\vdots \\ &\omega_k^1 \text{ from } P(\omega_k|\omega_1^1, \dots, \omega_{k-1}^1, \omega_{k+1}^0, \dots, \omega_K^0, \mathbf{x}) \\ &\vdots \\ &\omega_K^1 \text{ from } P(\omega_K|\omega_1^1, \dots, \omega_{K-1}^1, \mathbf{x}). \end{aligned} \quad (36)$$

After  $m$  such iterations, we have  $\omega^m = \omega_1^m, \dots, \omega_K^m$ . Denoting the burn-in period as  $t$  iterations, (which is usually obtained by some convergence detection algorithms), points  $\omega_k^t, \dots, \omega_k^m$  will be dependent samples approximately from the marginal densities  $P(\omega_k|\mathbf{x})$   $k = 1, \dots, K$ . Usually, the full conditional distributions required in (36) are derived from the joint distributions as [20]

$$\begin{aligned} &P(\omega_k|\omega_1, \dots, \omega_{k-1}, \omega_{k+1}, \dots, \omega_K, \mathbf{x}) \\ &= \frac{P(\omega_k, \omega_1, \dots, \omega_{k-1}, \omega_{k+1}, \dots, \omega_K, \mathbf{x})}{P(\omega_1, \dots, \omega_{k-1}, \omega_{k+1}, \dots, \omega_K, \mathbf{x})} \\ &\propto P(\omega_k, \omega_1, \dots, \omega_{k-1}, \omega_{k+1}, \dots, \omega_K, \mathbf{x}) \\ &\propto P(\omega_k, \omega_1, \dots, \omega_{k-1}, \omega_{k+1}, \dots, \omega_K|\mathbf{x}) \\ &\quad k = 1, \dots, K. \end{aligned} \quad (37)$$

Here, the proportionalities follow because  $P(\omega_k|\omega_1, \dots, \omega_{k-1}, \omega_{k+1}, \dots, \omega_K, \mathbf{x})$  is a distribution for  $\omega_k$ , and the denominator of the equation above and  $P(\mathbf{x})$  do not depend on  $\omega_k$ .

Based on the generated samples from Gibbs sampling, we can obtain a close approximation to  $P(\omega_k|\mathbf{x})$  as [22]

$$\begin{aligned} \hat{P}(\omega_k|\mathbf{x}) &= \frac{1}{m-t} \cdot \\ &\sum_{i=t+1}^m P(\omega_k|\omega_1^i, \dots, \omega_{k-1}^i, \omega_{k+1}^i, \dots, \omega_K^i, \mathbf{x}) \\ &\quad k = 1, \dots, K \end{aligned} \quad (38)$$

Substituting the above result into (35), the MMSE estimator can be approximated by the generated samples as

$$\hat{\omega}_{k-mmse} = \frac{1}{m-t} \sum_{i=t+1}^m \omega_k^i \quad k = 1, \dots, K. \quad (39)$$

Due to the fact that the frequency offsets ( $\omega_k$ ,  $k = 1, \dots, K$ ) has the properties of a circular random variable, the above estimate for  $\omega_k$  can be rewritten as

$$\hat{\omega}_{k-mmse} = \frac{1}{2\pi} \angle \sum_{i=t+1}^m \exp(j2\pi\omega_k^i) \quad k = 1, \dots, K. \quad (40)$$

where  $\angle$  denotes the operation of finding the angle of the complex number.

### C. Sampling From Full Conditional Distributions

As illustrated in the above section, applying Gibbs sampling into the problem at hand needs sampling from full conditional distributions. However, in this case, the difficulty lies in the fact that full conditionals change from iteration to iteration as the conditioning parameters change. That means each full conditional is used only once and then disposed of. Thus it is essential that sampling from full conditional distributions should be highly computationally efficient. Considering the efficiency and implementation difficulty of the random number generation methods in literature ([23],[24] and [25]), the Metropolized version of adaptive rejection sampling (ARMS) [25] method is proposed in this paper.

## V. SIMULATION RESULTS AND DISCUSSIONS

In this section, simulation results are presented to demonstrate the effectiveness of the proposed schemes. If there is no extra specification, the considered multiuser MIMO-OFDM system has the following parameters:  $N = 16$ ,  $L = 3$ ,  $K = 2, M_r = 2$ . The channel taps are modeled as independent and complex Gaussian random variables with zero mean and an exponential power delay profile. In the simulation, it is assumed that the users' signals arrive at the base station with equal power. The CFO for user  $k$  ( $\omega_k$ ,  $k = 1, \dots, K$ ) in each estimation is generated as a random variable with the distribution  $\mathcal{N}(0, \sigma_{\omega_k}^2)$ . Without of generality, we choose  $\sigma_{\omega_1}^2 = \dots = \sigma_{\omega_K}^2 = 0.001$  in the simulations. For each CFO estimation, the length of the Gibbs sequences is choose as 1000 ( $m = 1000$ ) and the turn-in period is set to be 200 iterations ( $t = 200$ ) (the exact value for these two parameters should be chosen carefully to improve the performance). The SNR used in the simulation is defined as the  $\text{SNR} = \frac{\|\mathbf{x}\|^2}{\sigma^2}$ .

Since there are multiple CFOs to be estimated in the scheme, the MSE performance is defined as

$$\text{MSE}_{CFO} = \|\hat{\omega} - \omega\|^2$$

All the simulation results of the proposed algorithm are averaged over 200 Monte Carlo runs.

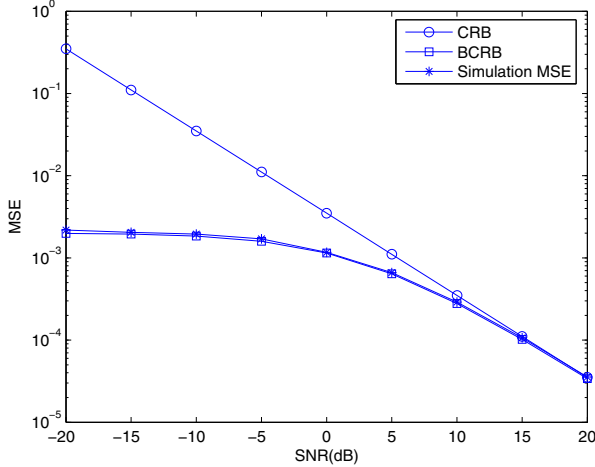


Fig. 1. The MSE performance for the proposed CFO estimator

In Fig. 1, the MSE performance of the proposed CFO are plotted versus SNR. The BCRB and CRB are also shown in the two figures as references. For the proposed CFO estimator, the MSE always coincides with the BCRB. In the low SNR range, there is an obvious gap between BCRB and CRB. With the increase of SNR, the BCRB is approaching CRB.

## VI. CONCLUSION

In this paper, the problem of CFO estimation for multiuser MIMO-OFDM system has been considered. With the prior information of the parameters, the BCRB for the CFO estimation has been derived in closed form. By integrating out the nuisance parameters, one MMSE estimator is proposed. To mitigate the mathematical sophistication in the derived estimator, Gibbs sampler with an efficient sample generation algorithm is suggested. Simulation results clearly verified that the proposed scheme is efficient.

## APPENDIX A

From the signal model (8) and the definitions in section II, the related density functions are given by

$$P(\mathbf{x}|\omega, \mathbf{h}, \sigma) = (\pi\sigma^2)^{-M_r N} \exp\left(-\frac{(\mathbf{x} - \mathbf{Qh})^H (\mathbf{x} - \mathbf{Qh})}{\sigma^2}\right). \quad (41)$$

$$P(\omega) \propto \exp(-\omega^T \phi^{-1} \omega) \quad (42)$$

$$P(\sigma) = 1/\sigma \quad (43)$$

$$P(\mathbf{h}) = \text{Const.} \quad (44)$$

Thus, the posterior density of  $\omega$  can be obtained as

$$\begin{aligned} P(\omega|\mathbf{x}) &\propto \int P(\mathbf{x}|\omega, \mathbf{h}, \sigma) P(\omega, \mathbf{h}, \sigma) d\mathbf{h} d\sigma \\ &\propto \exp(-\omega^T \phi^{-1} \omega) \cdot \int \sigma^{-2M_r N + 1} \exp\left(-\frac{(\mathbf{x} - \mathbf{Qh})^H (\mathbf{x} - \mathbf{Qh})}{\sigma^2}\right) d\mathbf{h} d\sigma. \end{aligned} \quad (45)$$

According to Gauss integrations [21], the first step is to integrate out the nuisance parameter  $\mathbf{h}$ :

$$\begin{aligned} \int \exp\left(-\frac{(\mathbf{x} - \mathbf{Qh})^H (\mathbf{x} - \mathbf{Qh})}{\sigma^2}\right) d\mathbf{h} &\propto |\mathbf{Q}^H \mathbf{Q}|^{-1} \sigma^{2M_r K L} \exp\left(-\frac{\mathbf{x}^H \mathbf{Q}^\perp \mathbf{x}}{\sigma^2}\right) \end{aligned} \quad (46)$$

where  $\mathbf{Q}^\perp$  is defined in (33). After incorporating the above result into (45),  $\sigma$  is integrated out based on the Gamma integration. The resultant marginal distribution is given by

$$P(\omega|\mathbf{x}) \propto \frac{\exp(-\omega^T \phi^{-1} \omega)}{|\mathbf{Q}^H \mathbf{Q}| (\mathbf{x}^H \mathbf{Q}^\perp \mathbf{x})^{(M_r N - M_r K L)}} \quad (47)$$

## REFERENCES

- [1] H. Sampath, S. Talwar, J. Tellado, V. Erceg, A. Paulraj, "A fourth-generation MIMO-OFDM broadband wireless system: design, performance, and field trial results," *IEEE Commun. Mag.*, vol. 40, pp. 143–149, Sept. 2002.
- [2] S. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1451–1458, Oct. 1998.
- [3] V. Tarokh, H. Jafarkhani, and A. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1456–1467, July 1999.
- [4] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs. Tech. J.*, vol. 1, pp. 41–59, 1996.
- [5] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel," in *Proc. URSI Int. Symp. Signals, Syst., Electron.*, pp. 295–300, Sep. 1998.
- [6] M. Ghogho and A. Swami, "Training design for multipath channel and frequency-offset estimation in MIMO systems," *IEEE Trans. on signal processing*, vol. 54, no. 10, pp. 3957–3965, Oct. 2006.
- [7] O. Besson and P. Stoica, "On parameter estimation of MIMO flat-fading channels with frequency offsets," *IEEE Trans. on signal processing*, vol. 51, no. 3, pp. 602–613, Mar. 2003.
- [8] Jianwu Chen, Yik-Chung Wu, Shaodan Ma and Tung-Sang Ng, "Joint CFO and channel estimation for multiuser MIMO-OFDM systems with optimal training sequences," *IEEE Trans. on Signal Process.*, vol. 56, no. 8, pp. 4008–4019, Aug. 2008.
- [9] D.B. Sullivan, D.W. Allan, D.A. Howe, and F.L. Walls, "Characterization of Clocks and Oscillators," NIST Tech Note 1337, 1990. (BIN: 868)
- [10] P. Stoica and A. Nehorai, "Performance Study of Conditional and Unconditional Direction-of-Arrival Estimation," *IEEE Tran. Acoustic. Speech. Signal Process.*, vol. 38, no. 10, pp. 1783–1795, Oct. 1990.
- [11] M. Ghogho, A. Swami, and T. Durrani, "Frequency estimation in the presence of Doppler spread: performance analysis," *IEEE Trans. Signal Process.*, vol. 49, no. 4, pp. 777–789, Apr. 2001.
- [12] H. L. V. Trees, "Detection, Estimation, and Modulation Theory: Part 1," Wiley Interscience, 1968.
- [13] S. M. Kay, "Fundamentals of Statistical Signal Processing," Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [14] A. Azevedo-Filho and R. D. Shachter, "Laplace's Method Approximations for Probabilistic Inference in Belief Networks with Continuous Variables," *Proceeding of Uncertainty in Artificial Intelligence (UAI)*, 1994.

- [15] R. Wong(1989), *Asymptotic Approximations of Integrals*. Academic Press, San Diego.
- [16] J. M. Hammersley, "Monte Carlo Methods for Solving Multivariable Problems." *Ann. New York Acad. Sci.* 86, 844-874, 1960.
- [17] G.P. Lepage, A New Algorithm for Adaptive Multidimensional Integration, *Journal of Computational Physics* 27, 192-203, (1978)
- [18] G.P. Lepage, VEGAS: An Adaptive Multi-dimensional Integration Program, Cornell preprint CLNS 80-447, March 1980
- [19] W. H. Press and G. R. Farrar, Recursive Stratified Sampling for Multi-dimensional Monte Carlo Integration, *Computers in Physics*, v4 (1990), pp190-195.
- [20] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, "Markov Chain Monte Carlo in Practice," Chapman and Hall, London, 1996.
- [21] Papoulis, A. "Probability, Random Variables, and Stochastic Processes," 2nd ed. New York: McGraw-Hill, pp. 147-148, 1984.
- [22] Alan E. Gelfand and Adrian F. M. Smith, "Sampling-Based Approaches to Calculating Marginal Densities" *Journal of the American Statistical Association*, vol. 85, no. 410, pp. 398-409, Jun., 1990.
- [23] J. E. Gentle, *Random number generation and monte carlo methods*, Second ed. Springer, New York, NY, 2003.
- [24] W. R. Gilks and P. Wild, "Adaptive Rejection Sampling for Gibbs Sampling," *Appl. Stat.*, vol. 41, no. 2, pp. 337-348, 1992.
- [25] W. R. Gilks, N. G. Best, and K. K. C. Chan, "Adaptive Rejection Sampling for Gibbs Sampling," *Appl. Stat.*, vol. 44, no. 4, pp. 455-472, 1995.