國 立 交 通 大 學

電 信 工 程 學 系

碩 士 論 文

雙選擇性通道下正交分頻多工系統中使用馬可夫鏈蒙地卡羅法之基於期望值最大化接收機設計

Design of an EM-based Receiver Using the Markov Chain Monte Carlo Method for OFDM Systems in Doubly Selective Channels

研 究 生：鍾曉顗

指導教授：黃家齊 博士

中 華 民 國 九 十 八 年 六 月

雙選擇性通道下正交分頻多工系統中使用馬可夫鏈蒙地卡羅法之基於期望值最大化接收機設計

Design of an EM-based Receiver Using Markov Chain Monte Carlo Method for OFDM Systems in Doubly Selective Channels

研 究 生：鍾曉顗　　Student：Siao-Yi Jhong

指導教授：黃家齊　　Advisor：Dr. Chia-Chi Huang

國 立 交 通 大 學

電 信 工 程 學 系 碩 士 班

碩 士 論 文

A Thesis

Submitted to Department of Communication Engineering
College of Electrical Engineering and Computer Engineering
National Chiao Tung University
in Partial Fulfillment of the Requirements
for the Degree of
Master of Science
in
Communication Engineering
June 2009

Hsinchu, Taiwan, Republic of China

中華民國 九十八 年 六 月

雙選擇性通道下正交分頻多工系統中使用馬可夫鏈蒙地卡羅法之基於期望值最大化接收機設計

學生：鍾曉顗　　指導教授：黃家齊 博士

國立交通大學電信工程學系 碩士班

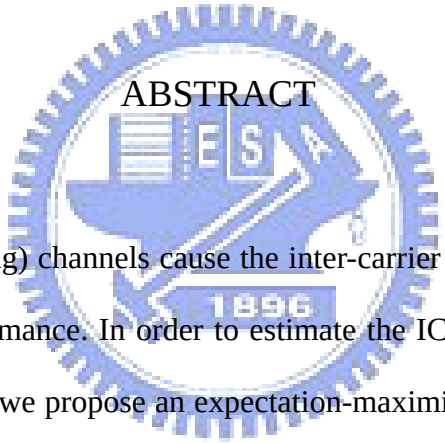
摘　　要

雙選擇性(衰減)通道造成載波間干擾(Inter-carrier Interference, ICI)的問題並降低系統效能，受到鄰近子載波之能量擴散而產生的載波間干擾效應，是本論文首要的估測目標，因而提出一種正交分頻多工(Orthogonal Frequency Division Multiplexing, OFDM)系統接收機，其為基於期望值最大化演算法(Expectation-Maximization, EM)設計而成。在頻域模式下進行系統分析，將 EM 演算法與馬可夫鏈蒙地卡羅法(Markov chain Monte Carlo, MCMC)結合，並以最大相似度(Maximum Likelihood, ML)作為判定準則，我們得到一套有系統地估測載波間干擾的方法，稱之為「EM 通道估測法」。此外，為了減低運算複雜度，以及充分利用時變通道所賦予的時間多樣性，ML-EM 接收機採用「分群式載波間干擾消除器」，針對通過此消除器之後的接收信號，估算適當的殘餘載波間干擾功率，以提高資料偵測的正確率。電腦模擬結果顯示，相較於以往一階等化器，我們提出的 ML-EM 接收機在錯誤率方面的表現，有著明顯的進步。

Design of an EM-based Receiver Using Markov Chain Monte Carlo Method for OFDM Systems in Doubly Selective Channels

Student：Siao-Yi Jhong　　Advisor：Dr. Chia-Chi Huang

Department of Communication Engineering
National Chiao Tung University



ABSTRACT

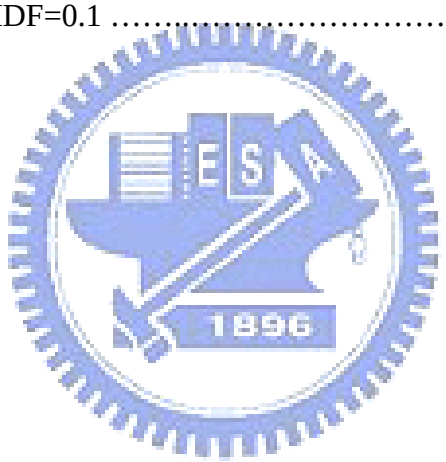
Doubly selective (fading) channels cause the inter-carrier interference (ICI) problem and thus degrade the system performance. In order to estimate the ICI effect made by the spreading energy of adjacent subcarriers, we propose an expectation-maximization (EM)-based receiver for orthogonal frequency division multiplexing (OFDM) systems. In this paper, we use the frequency domain model for system analysis and derive the EM channel estimation method by combining the Markov Chain Monte Carlo (MCMC) method with the EM algorithm according to the maximum-likelihood (ML) criterion. Besides, the proposed EM-based receiver is incorporated with the group-wise ICI cancellation method for reducing computational complexity and exploiting the inherent time diversity in time-variant channels. After the ICI cancellation, the residual ICI power is calculated for the ICI-reduced signals with the goal of making data detection more correctly. Results of computer simulation demonstrate that the ML-EM receiver performs much better than the conventional one-tap equalizer.

誌　謝

這一年來，投入研究的時間佔據了我大部分的生活，而研究過程的起伏帶來喜怒哀樂各種情緒，留下一段非常難忘的經驗。雖然現在的我，還是有很多事情弄不明白，但是跟一年之前的我相比，我想應該有了一點進步。由於許多人的幫忙與鼓勵，使我能夠順利完成畢業論文。首先，我要感謝黃家齊老師，與他談話有種如沐春風的愉悅，謝謝老師一直以來的關心與指導。另外，感謝實驗室學長姊的提攜，特別是古孟霖學長與陳文娟學姊，總是不厭其煩地教導、鼓勵我，為迷途的我指點前進的方向。接著，我想感謝實驗室同學，這四位男生有著截然不同的個性，卻都很平易近人，謝謝他們陪伴我度過研究所的時光。最後，感謝家人和朋友，包容我的任性、體諒我的慌張、接納我的執著，因為他們，使我更有勇氣。

Contents

List of Figures



Chapter 1 Introduction

Over the last few years, the mobile communication technology develops rapidly, and so do the wireless techniques. The wideband transmission became an inevitable trend because of the data rate demanded by users. The wireless network has the advantage of the mobility, the convenience and the high-coverage. Limits of time and place do not restrain the communication among people. However, the channel of the wireless communication is interfered with by severe noise, and the multipath effect is also a problem which should be overcome. The multipath propagation causes the frequency selective fading and the inter-symbol interference (ISI), and thus harms the quality of transmission and degrades the system performance. It would be significant to choose an appropriate system model according to the channel conditions and the requirements for transmission.

There are many techniques invented for raising the utility rate and mitigating the influence of the multipath effect, and orthogonal frequency division multiplexing (OFDM) is one of the most famous schemes. In OFDM, subcarrier frequencies are chosen to be orthogonal to each other; namely, the crosstalk between the sub-channels is eliminated. The orthogonality also provides high spectral efficiency since almost the whole available frequency band can be utilized. The duration of each symbol is long enough to put in a guard interval to eliminate the ISI, and the cyclic prefix (CP) used as the guard interval consists of a copy of the end of the OFDM symbol. Besides, OFDM is equipped for coping with

attenuation of high frequencies in a long copper wire and narrowband interference. The effect of frequency selective fading can be considered as flat over an OFDM sub-channel if its band is sufficiently narrow. This makes the equalizer simpler at the receiver compared with conventional single-carrier modulation.

OFDM requires accurate frequency synchronization between the receiver and the transmitter. The subcarriers are no longer orthogonal if there is frequency deviation, inducing the inter-carrier interference (ICI). Frequency offsets are typically caused by mismatched transmitter and receiver oscillators, or by Doppler shift due to movement, and this effect worsens as speed increases or as the length of a symbol gets longer. In communication systems, the transmission often proceeds in the high- mobility condition, but the time-variant channels damage the orthogonality, cause the ICI effect and then lower the system performance. As a result, the ICI suppression is a significant issue for research in mobile communication, and it is also the main study in this paper.

There have been many techniques suggested for the ICI suppression; for example, minimum mean square error (MMSE) [3], minimum mean square error with successive detection (MMSE-SD) [3], polynomial cancellation coding (PCC) [4] and self-cancellation coding [5]. The method in [3] is efficient but with high computational complexity. The schemes in [4] and [5] provide good bit error rate (BER) performance at the expense of sacrificing bandwidth efficiency. In [6] and [7], the piece-wise linear model is proposed to

approximate the channel variation, helping the analysis of properties of the channel. It is explained that energy of a sub-carrier leaks to the adjacent sub-carriers owing to Doppler shift in [8] and [9]. The expectation-maximization (EM) algorithm can be utilized to solve the maximum-likelihood (ML) estimation problem in an iterative manner. Recently, some EM-based methods have been proposed for channel estimation and data detection in OFDM systems. But the wireless channel is assumed to be quasi-static, i.e., channel gain remains constant over the duration of one OFDM symbol.

In this paper, we propose an EM-based receiver for OFDM systems in doubly selective fading channels. By assuming channel varies in a linear fashion, we analyze the ICI effect in frequency domain and derive a channel estimation method based on the EM algorithm in [10] and [11] under the ML criterion. A Gibbs sampler (a Markov chain Monte Carlo procedure) is used for the calculation of Bayesian estimates and also for data detection in the EM algorithm. Moreover, we combine the EM-based receiver with a group-wise ICI cancellation scheme for the sake of reducing the computational complexity. The ML-EM receiver is implemented to iterate between a group-wise ICI canceller and an EM detector (including the Gibbs samplers inside). The MMSE estimator is employed in the receiver for the initial setting by exploiting the pilot tones in each frame, and the accuracy of initialization can be refined successfully through the mechanism of decision feedback.

The rest of this paper is organized as follows. In Chapter 2, the system under study is

described and the ICI effect is analyzed in frequency domain. And the basic idea of the Gibbs sampling method is briefly introduced in Chapter 3. In Chapter 4, a channel estimation method is derived from the EM algorithm combined with a Gibbs sampler using the ML criterion, and accordingly, we propose an ML-EM receiver for OFDM systems. And the ML-EM receiver is further united with the group-wise method in chapter 4.The problem of computing the ICI power (or the variance of ICI) for the Gibbs sampling is treated in Chapter 5. Results of computer simulation are presented and discussed in Chapter 6. In the final part of the paper, Chapter 7, we draw some conclusions for the study.

The following are some interpretations of notations used in the paper. Boldface capital letters denote matrices, whereas boldface lowercase letters denote column vectors. The superscripts $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$ stand for the transpose and the Hermitian transpose of a matrix, respectively. The column vector $\mathbf{x}$ can be explicitly expressed by $\langle x_1, x_2, \ldots, x_{|\mathbf{x}|} \rangle$ or $\langle x_i : i \in \{1, \ldots, |\mathbf{x}|\} \rangle$, where $|\mathbf{x}|$ is the dimension of the vector $\mathbf{x}$. The notation $\{\ldots\}$ represents a set, e.g. a set $\mathbf{x} = \{x_1, \ldots, x_{|\mathbf{x}|}\}$, and the cardinality of the set $\mathbf{x}$ is denoted by $|\mathbf{x}|$. The set can also be expressed in a compact form: $\{x_i : i \in \{1, \ldots, |\mathbf{x}|\}\}$.

Chapter 2 OFDM System Model

2.1 Frame Format

According to the frame format in IEEE 802.16e standard [12], a frame consists of the pilot preamble in the first symbol followed by many OFDM symbols carried by numerous subcarriers. The number of subcarriers depends on the size of the Fast Fourier Transform (FFT), and there are three types of subcarriers shown in Fig. 2.2. First, the data subcarriers are used to transmit data symbols. And the pilot subcarriers are used as virtual subcarriers to help the channel estimation. In an OFDM symbol, a sequence of values is inserted to be the pilot signals. Moreover, the null subcarriers can be the DC subcarriers or the guard band which alleviates the aliasing problem at the receiver.

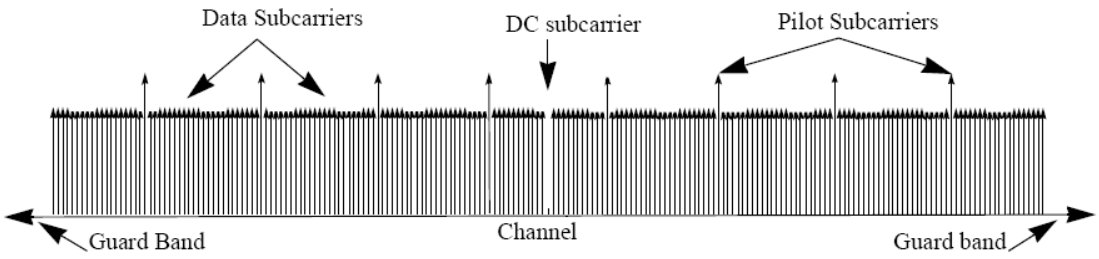


Fig. 2.1 The allocation of subcarriers in OFDM systems.

2.2 Doubly Selective Channels

Nowadays, wireless applications are expected to operate at high level of mobility and at

high capacities that result in doubly selective channels. Doubly selective fading means frequency selective fading induced by the multipath propagation and time selective fading caused by the time-varying channel.

The multipath propagation is a phenomenon that the transmitted signals arrive at different times at the receiver through more than one path. The time difference between the arrival moment of the first multipath component and the last one is called delay spread. The coherence bandwidth measures the separation in frequency which two signals experience uncorrelated fading. If the coherence bandwidth of the channel is smaller than the bandwidth of the signal, different frequency components of the signal suffer from decorrelated fading.

The channel varies with time due to Doppler shift resulted from rapid traversing of the transmitter or the receiver. The coherence time is a measure of the minimum time required for the magnitude change of the channel to become decorrelated from its previous value. When the coherence time of the channel is small relative to the symbol duration, the amplitude and phase of the signal change imposed by the channel varies considerably, causing a fast fading channel.

That is, the doubly selective channel is also called the frequency-selective fast-fading channel. The equivalent impulse response can be expressed as follows:

$$h(t,\tau)=\sum_{l=1}^{L}\alpha_l(t)\delta\left(\tau-\tau^{(l)}\right) \tag{2.1}$$

where $\alpha_l(t)$ and $\tau^{(l)}$ are the complex fading gain and time delay of l th path respectively

and $\delta(\cdot)$ is the Kronecker delta function. The complex fading gain is a function of t which denotes the time index, and the value L stands for the number of paths. Fig. 2.2 below is intended to show the equivalent impulse response.

The variation of the fading gain, $\alpha_l(t)$, which depends upon Doppler shift is proportional to the carrier frequency and the speed of a motor vehicle. The maximum Doppler frequency (MDF) is defined as

$$f_{\mathrm{D}} = \frac{f_c v}{\mathrm{c}} \tag{2.2}$$

where v is the velocity of the source relative to the receiver, f_c is the carrier frequency and c is the speed of light (e.g. 3×10^8 m/s for light travelling in a vacuum). In OFDM systems, the normalized MDF $f_{\mathrm{D}}T_s$ (in which T_s is the sampling period) is used to indicate the range of variation in the channel. By keeping T_s and f_c constant, it can be observed from (2.2) that f_{D} becomes larger when the car is driven faster.

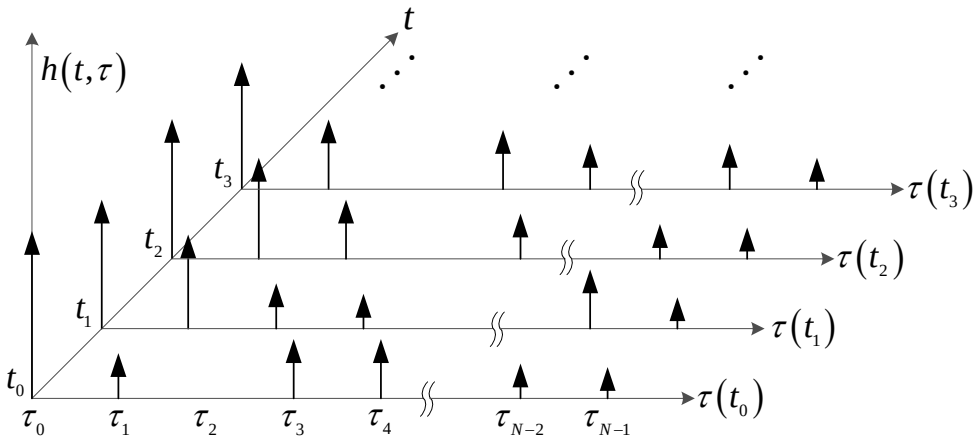


Fig. 2.2 An illustration of the equivalent impulse response.

2.3 Transmitted Signals and Received Signals

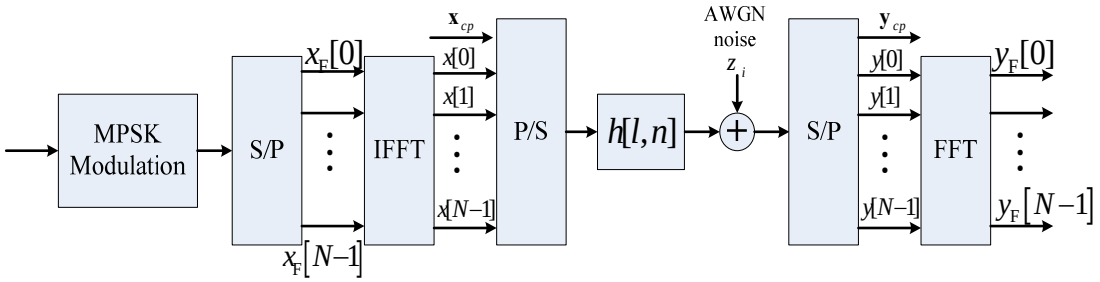


Fig. 2.3 OFDM Systems.

Fig.2.3 shows an OFDM system. The information source bits are mapped into MPSK data symbols and are converted into N parallel data streams through a serial to parallel (S/P) block. Then the data streams are modulated onto N subcarriers by an N-point inverse Fast Fourier Transform (IFFT) unit to produce samples in the time domain, and the transmitted signal is expressed as

$$x[n] = \frac{1}{N}\sum_{m=0}^{N-1} X[m] e^{j\frac{2\pi mn}{N}} \tag{2.3}$$

where $X[m]$ represents the data in frequency domain at the m th subcarrier. For the purpose of eliminating ISI due to multipath channels, a CP is added at the head of each data symbol, being presented as

$$x_c[n] = x[N+n], \qquad n = -N_G, \ldots, -1 \tag{2.4}$$

where N_G is the length of the guard interval. We assume that the maximum delay spread of the channel is always smaller than N_G to make sure that there is no ISI after removing the guard interval.

It is assumed that both timing and carrier frequency synchronization are perfect. The

received OFDM symbols in time domain can be expressed as the circular convolution of the transmitted symbols and the channel impulse response; hence, received signals are given by

$$\begin{aligned} y[n] &= x[n] \otimes h[l,n] + z[n] \\ &= \sum_{l=1}^{L} h[l,n] x[((n-l))_N] + z[n] \end{aligned} \tag{2.5}$$

where $h[l,n]$ represents the channel impulse response of the l th channel tap at the n th sample time, l is the delay time of the l th path relative to the first path, n is the discrete time index, $z[n]$ is a sample of additive white Gaussian noise (AWGN) with zero-mean and variance σ_z^2, and $((\bullet))_N$ means a cyclic shift in the base of N.

After removing the guard interval and taking the Fast Fourier Transform (FFT), the received signals in frequency domain can be described by

$$\begin{aligned} y_{\mathrm{F}}[k] &= \sum_{m=0}^{N-1} H[k,m] x_{\mathrm{F}}[m] + z_{\mathrm{F}}[k] \\ &= H[k,k] x_{\mathrm{F}}[k] + \underbrace{\sum_{m=0, m\neq k}^{N-1} H[k,m] x_{\mathrm{F}}[m]}_{\text{ICI term}} + z_{\mathrm{F}}[k] \end{aligned} \tag{2.6}$$

in which $x_{\mathrm{F}}[k]$ is the transmitted signals in frequency domain and $H[k,k]$ is the frequency response of the average gain of the channel. And $H[k,m]$, which represents the leakage factor of ICI from the m th subcarrier to the k th subcarrier, is provided in the following:

$$H[k,m] = \sum_{l=0}^{L-1} \beta[k,m,l] e^{-j\frac{2\pi ml}{N}} \tag{2.7}$$

where $\beta[k,m,l]$ is the frequency response of the time-varying channel and can be given by

$$\beta[k,m,l] = \frac{1}{N} \sum_{n=0}^{N-1} h[l,n] e^{-j\frac{2\pi n((k-m))_N}{N}} . \tag{2.8}$$

Accordingly, $H[k,m]$ is regarded as the summation of $\alpha[k,m,l]$ multiplied by the linear phase resulted from the time delay of the l th tap.

If the channel is time-invariant in the OFDM symbol duration; namely, it is in the slow fading mode, $H[k,m]$ turns out to be zero when k and m are different values, and thus $H[k,m]$ is equivalent to $H[k,k]$. That is, there is no ICI among subcarriers.

2.4 The ICI Model

In [6] and [7], with the normalized MDF up to 0.1, a first-order polynomial function is adequate to model the time variation of each channel tap in an OFDM symbol and it is defined by

$$h[l,n] = a[l,1]n + a[l,0] \tag{2.9}$$

where $a[l,p]$ is the coefficient of p th monomial in the function of the l th path. By substituting (2.9) to (2.8), $\beta[k,m,l]$ can be presented as follows:

(i) $m = k$: $\beta[k,m,l] = \frac{N-1}{2}a[l,1] + a[l,0]$ (2.10)

(ii) $m \neq k$: $\beta[k,m,l] = \Phi[k,m]a[l,1]$ (2.11)

where $\Phi[k,m] = \frac{1}{N}\sum_{n=0}^{N-1} n e^{-j\frac{2\pi n(k-m)}{N}} = -\frac{1}{2} + j\frac{1}{2\tan\left(\frac{\pi((k-m))_N}{N}\right)}$. (2.12)

And the Malcaurin series can be used to replace the tangent function in (2.12). A parameter λ is in the interval between $-\frac{\pi}{2}$ and $\frac{\pi}{2}$, and then $\tan\lambda$ can approximate to a polynomial

in one variable expressed as follows:

$$\tan\lambda = \lambda + \frac{1}{3}\lambda^3 + \frac{2}{15}\lambda^5 + \ldots\ldots \tag{2.13}$$

in which the high-order terms, i.e. λ^3, λ^5, may be neglected as the value of λ is smaller than one, and thus (2.13) can be written as

$$\tan\lambda \approx \lambda . \tag{2.14}$$

The values of m and k are within the range from 0 to $N-1$, so $\left((k-m)\right)_N$ is between 0 and $N-1$, and then $\frac{\pi\left((k-m)\right)_N}{N}$ in (2.12) is in the interval of $\left[\frac{\pi}{N}, \frac{\pi(N-1)}{N}\right]$ or $(0,\pi)$.

$\tan\left(\frac{\pi\left((k-m)\right)_N}{N}\right)$ is simplified by the Malcaurin series, and so is $\Phi[k,m]$ which can be classified in the following cases:

(i) $\frac{\pi((k-m))_N}{N} = \frac{\pi}{2} \Rightarrow \tan\left(\frac{\pi((k-m))_N}{N}\right) = \tan\left(\frac{\pi}{2}\right)$, but $\tan\left(\frac{\pi}{2}\right)$ is infinite, and therefore $\Phi[k,m] = -\frac{1}{2}$.

(ii) $0 < \frac{\pi((k-m))_N}{N} < \frac{\pi}{2}$, so $\Phi[k,m] = -\frac{1}{2} + j\frac{N}{2\pi\left((k-m)\right)_N}$.

(iii) Subtracting π from $\frac{\pi}{2} < \frac{\pi((k-m))_N}{N} < \pi$, we can get a new inequality of $-\frac{\pi}{2} < \frac{\pi\left[\left((k-m)\right)_N - N\right]}{N} < 0$ which conforms to the range required by the Malcaurin series, and thus $\Phi[k,m] = -\frac{1}{2} + j\frac{N}{2\pi\left[\left((k-m)\right)_N - N\right]}$.

When k is different from m, $\Phi[k,m]$ can be formed in accordance with the results of (i), (ii) and (iii), and given by

$$\Phi[k,m] \approx \begin{cases} -\frac{1}{2} + j\frac{N}{2\pi((k-m))_N} & \text{for } 0 < ((k-m))_N < \frac{N}{2} \\ -\frac{1}{2} & \text{for } ((k-m))_N = \frac{N}{2} \\ -\frac{1}{2} + j\frac{N}{2\pi[((k-m))_N - N]} & \text{for } \frac{N}{2} < ((k-m))_N < N \end{cases} . \quad (2.15)$$

As a result, (2.12) can be replaced by the approximation in (2.15).

By making use of (2.11) and (2.15), the equation in (2.6) is rewritten as

$$y_F[k] = H[k,k]x_F[k] + \sum_{\substack{m=0 \\ m\neq k}}^{N-1} \Phi[k,m]w[m]x_F[m] + z_F[k] \quad (2.16)$$

where $w[m] = \sum_{l=0}^{L-1} a[l,1]e^{-j\frac{2\pi ml}{N}}$ is defined as the channel variable in frequency domain and $\Phi[k,m]$ represents a fixed-valued coefficient of the ICI term. In order to provide a more compact representation, we rewrite (2.16) in a matrix form which is given by

$$\mathbf{y} = \mathbf{Hx} + \mathbf{z} = (\mathbf{M} + \mathbf{\Phi W})\mathbf{x} + \mathbf{z} = \mathbf{Mx} + \mathbf{\Phi X w} + \mathbf{z} \quad (2.17)$$

where the (k,m)th entry of $\mathbf{H}$ is $H[k,m]$, $\mathbf{y} = [y_F[0],\ldots,y_F[N-1]]^T$, $\mathbf{X} = diag\{[x_F[0],\ldots,x_F[N-1]]^T\}$, $\mathbf{x} = [x_F[0],\ldots,x_F[N-1]]^T$, $\mathbf{z} = [z_F[0],\ldots,z_F[N-1]]^T$, $\mathbf{M} = diag\{[H[0,0],\ldots,H[N-1,N-1]]^T\}$, $\mathbf{W} = diag\{[w[0],\ldots,w[N-1]]^T\}$, $\mathbf{w} = [w[0],\ldots,w[N-1]]^T$ and $\mathbf{\Phi}$ is an $N \times N$ matrix expressed in (2.18) below in which $\Phi[k,m]$ represents the (k,m)th entry. Furthermore, $\mathbf{w} = \mathbf{Fa}$ where $\mathbf{a} = [a[0,1],\ldots,a[L-1,1]]^T$ is a vector composed of slopes of channel paths and $\mathbf{F}$ is a DFT matrix of size $N \times L$ with the (m,l)th entry provided by $F[m,l] = \exp\left\{-j\frac{2\pi ml}{N}\right\}$.

$$\mathbf{\Phi} \approx \frac{1}{2}\begin{bmatrix} 0 & \left(-1+\frac{N}{j\pi}\right) & \left(-1+\frac{N}{j2\pi}\right) & \cdots & \left(-1-\frac{N}{j\pi}\right) \\ \left(-1-\frac{N}{j\pi}\right) & 0 & \left(-1+\frac{N}{j\pi}\right) & \cdots & \left(-1-\frac{N}{j2\pi}\right) \\ \left(-1-\frac{N}{j2\pi}\right) & \left(-1-\frac{N}{j\pi}\right) & 0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \left(-1+\frac{N}{j\pi}\right) \\ \left(-1+\frac{N}{j\pi}\right) & \left(-1+\frac{N}{j2\pi}\right) & \cdots & \left(-1-\frac{N}{j\pi}\right) & 0 \end{bmatrix} \quad (2.18)$$

Chapter 3 Gibbs Sampling

Named after the physicist J. W. Gibbs, Gibbs sampling is an algorithm invented by Stuart Geman and Donald Geman to generate a sequence of samples from the joint probability distribution of two or more random variables. The purpose of making such a sequence is to approximate the joint distribution, or to compute an integral such as an expected value. And it is a special case of the Metropolis-Hastings algorithm and also an example of a Markov chain Monte Carlo algorithm.

Now we would like to interpret the concept of Gibbs sampling through the following case [18] [19]. Let $\boldsymbol{\theta} = [\theta_1 \cdots\cdots \theta_{|\boldsymbol{\theta}|}]^T$ be a vector of unknown parameters and $\mathbf{y}$ be the observed data. Suppose that we are interested in the a posteriori marginal distribution of θ_j (where $1 \le j \le |\boldsymbol{\theta}|$) conditioned on the observation $\mathbf{y}$.

$$P(\theta_j \mid \mathbf{y}) = \int\int \cdots \int P(\boldsymbol{\theta} \mid \mathbf{y}) d\theta_1 d\theta_2 \cdots d\theta_{j-1} d\theta_{j+1} \cdots d\theta_{|\boldsymbol{\theta}|} \tag{3.1}$$

The calculation of integration in (2.19) may be infeasible if $|\boldsymbol{\theta}|$ is very large. Gibbs sampling is a Monte Carlo procedure for numerical evaluation of the multidimensional integrals. The basic idea is to generate random samples from the joint posterior distribution $P(\boldsymbol{\theta} \mid \mathbf{y})$ and to estimate the marginal distribution by these samples.

Given an initial vector $\boldsymbol{\theta}^{(0)} = [\theta_1^{(0)} \cdots\cdots \theta_{|\boldsymbol{\theta}|}^{(0)}]^T$, the algorithm iterates from $n = 1$ to $n = n_0 + N$ and is implemented as follows.

- Draw a sample $\theta_1^{(n)}$ from $P(\theta_1 \mid \theta_2^{(n-1)}, \theta_3^{(n-1)} \ldots, \theta_{|\boldsymbol{\theta}|}^{(n-1)}, \mathbf{y})$.

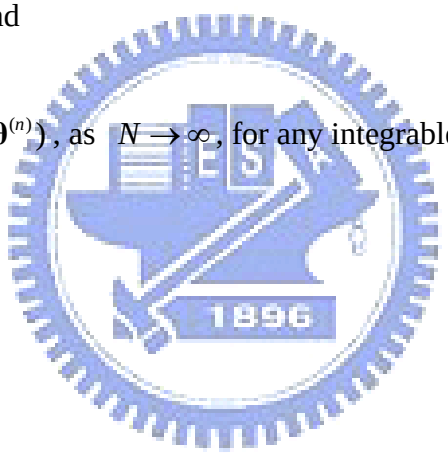
- Draw a sample $\theta_2^{(n)}$ from $P(\theta_2 \mid \theta_1^{(n-1)}, \theta_3^{(n-1)} \ldots, \theta_{|\boldsymbol{\theta}|}^{(n-1)}, \mathbf{y})$.

$\vdots$

- Draw a sample $\theta_{|\boldsymbol{\theta}|}^{(n)}$ from $P(\theta_{|\boldsymbol{\theta}|} \mid \theta_2^{(n-1)}, \theta_3^{(n-1)} \ldots, \theta_{|\boldsymbol{\theta}|-1}^{(n-1)}, \mathbf{y})$.

After going through the iterations of Gibbs sampling, $n_0 + N$ vectors will be obtained from the Gibbs sampler; however, only the last N vectors are regarded as samples that can be used. And the initial period of length n_0 is known as the "burn-in" period for the transient period required to converge to equilibrium. Consequently, the distribution of $\boldsymbol{\theta}^{(n)}$ converges to $P(\boldsymbol{\theta} \mid \mathbf{y})$ when $n \to \infty$ and

$\int f(\boldsymbol{\theta}) P(\boldsymbol{\theta} \mid \mathbf{y}) d\boldsymbol{\theta} \approx \frac{1}{N} \sum_{n=n_0+1}^{n_0+N} f(\boldsymbol{\theta}^{(n)})$, as $N \to \infty$, for any integrable function f.



Chapter 4 The ML-EM Receiver

4.1 Channel Estimation

At the receiver, the channel information is an unknown factor for detecting the transmitted data. Based on (2.17), the optimal ML channel estimation problem can be formulated as

$$\mathbf{w}_{\mathrm{ML}} = \arg\max_{\mathbf{w}} L(\mathbf{y}\,|\,\mathbf{w}) = \arg\max_{\mathbf{w}} \int L(\mathbf{y}\,|\,\mathbf{w},\mathbf{x}) P(\mathbf{x})\, d\mathbf{x} \tag{4.1}$$

where $L(\mathbf{y}\,|\,\mathbf{w})$ is a log-likelihood function given by taking logarithm of the corresponding probability density function (PDF) $P(\mathbf{y}\,|\,\mathbf{w})$; that is, $L(\mathbf{y}\,|\,\mathbf{w})$ is equivalent to $\ln P(\mathbf{y}\,|\,\mathbf{w})$. Nevertheless, if we directly calculate the integral in (2.20), the great complexity of multidimensional integrations is difficult to be solved. Thus the EM algorithm is used for avoiding direct calculation of those complicated integrals. And $L(\mathbf{y}\,|\,\mathbf{w})$ can be maximized by the way of iterating between the E-step and the M-step of the EM algorithm.

Applying Bayes' theorem, $P(\mathbf{y}\,|\,\mathbf{w})$ is able to be expressed as

$$P(\mathbf{y}\,|\,\mathbf{w}) = \frac{P(\mathbf{y},\mathbf{x}\,|\,\mathbf{w})}{P(\mathbf{x}\,|\,\mathbf{y},\mathbf{w})} \tag{4.2}$$

and then $L(\mathbf{y}\,|\,\mathbf{w})$ is obtained by taking logarithm of (4.2) as

$$L(\mathbf{y}\,|\,\mathbf{w}) = L(\mathbf{y},\mathbf{x}\,|\,\mathbf{w}) - L(\mathbf{x}\,|\,\mathbf{y},\mathbf{w}). \tag{4.3}$$

And the expected value of $L(\mathbf{y}\,|\,\mathbf{w})$ with respect to $P(\mathbf{x}\,|\,\mathbf{y},\hat{\mathbf{w}}_{(\mathrm{m\text{-}1})})$ is given by

$$L(\mathbf{y}\,|\,\mathbf{w}) = E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(\mathrm{m-1})}}\left[L(\mathbf{y},\mathbf{x}\,|\,\mathbf{w})\right] - E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(\mathrm{m-1})}}\left[L(\mathbf{x}\,|\,\mathbf{y},\mathbf{w})\right] \tag{4.4}$$

where $\hat{\mathbf{w}}_{(\mathrm{m\text{-}1})}$ represents the estimated channel information at the (m-1)th iteration. Because

of Jensen's inequality, $E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(\mathbf{m}-1)}}\left[L(\mathbf{x}\,|\,\mathbf{y},\mathbf{w})\right]$ becomes smaller when m is larger, and hence $E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(\mathbf{m}-1)}}\left[L(\mathbf{y},\mathbf{x}\,|\,\mathbf{w})\right]$ is thought of as the dominant term. Therefore, we can rewrite (4.4) as

$$L(\mathbf{y}\,|\,\mathbf{w}) \propto E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(\mathbf{m}-1)}}\left[L(\mathbf{y},\mathbf{x}\,|\,\mathbf{w})\right] \tag{4.5}$$

in which $L(\mathbf{y},\mathbf{x}\,|\,\mathbf{w})$ can be described by

$$L(\mathbf{y},\mathbf{x}\,|\,\mathbf{w}) = L(\mathbf{y}\,|\,\mathbf{x},\mathbf{w}) + L(\mathbf{x}\,|\,\mathbf{w}) = L(\mathbf{y}\,|\,\mathbf{x},\mathbf{w}) + L(\mathbf{x}) \tag{4.6}$$

based on Bayes' law. The second equality in (4.6) holds since the transmitted data $\mathbf{x}$ is independent of the channel information $\mathbf{w}$. Substituting (4.6) into (4.5), we find that (4.4) can also be written as

$$L(\mathbf{y}\,|\,\mathbf{w}) \propto E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(\mathbf{m}-1)}}\left[L(\mathbf{y}\,|\,\mathbf{x},\mathbf{w})\right] + E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(\mathbf{m}-1)}}\left[L(\mathbf{x})\right] \tag{4.7}$$

where $E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(\mathbf{m}-1)}}\left[L(\mathbf{x})\right]$ is a constant rather than a function of $\mathbf{w}$. Consequently, the formulation of (4.1) is replaced by

$$\begin{aligned}\mathbf{w}_{\mathrm{ML}} &= \arg\max_{\mathbf{w}} L(\mathbf{y}\,|\,\mathbf{w}) \\ &= \arg\max_{\mathbf{w}} \left\{ E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(\mathbf{m}-1)}}\left[L(\mathbf{y}\,|\,\mathbf{x},\mathbf{w})\right] + const \right\}.\end{aligned} \tag{4.8}$$

4.1.1 E-step and M-step

The E-step and the M-step associated with the optimization problem in (4.8) are described as follows:

E-step: $\Omega\left(\mathbf{w}\,|\,\mathbf{y},\hat{\mathbf{w}}_{(\mathrm{m}-1)}\right) = E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(\mathbf{m}-1)}}\left[L(\mathbf{y}\,|\,\mathbf{x},\mathbf{w})\right] + const$ (4.9)

M-step: $\hat{\mathbf{w}}_{(\mathrm{m})} = \arg\max_{\mathbf{w}} \Omega\left(\mathbf{w}\,|\,\mathbf{y},\hat{\mathbf{w}}_{(\mathrm{m}-1)}\right)$ (4.10)

where $\Omega(\mathbf{w}\,|\,\mathbf{y},\hat{\mathbf{w}}_{(\mathrm{m}-1)})$ is the expected complete log-likelihood (ECLL) function to be

maximized in the M-step of (4.10).

Dropping the constant in (4.9) would not affect the result of the M-step and the conditional PDF $P(\mathbf{y}\,|\,\mathbf{x},\mathbf{w})$ is observed to be a Gaussian distribution from (2.17), so the E-step can be presented as

$$\begin{aligned}\Omega(\mathbf{w}\,|\,\mathbf{y},\hat{\mathbf{w}}_{(\mathrm{m}-1)}) &= E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(\mathbf{m}-1)}}\left[L(\mathbf{y}\,|\,\mathbf{x},\mathbf{w})\right] \\ &= E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(\mathrm{m}-1)}}\left[\ln P(\mathbf{y}\,|\,\mathbf{x},\mathbf{w})\right] \\ &= \frac{-1}{\sigma^2}E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(\mathrm{m}-1)}}\left[\left\|\mathbf{y}-\mathbf{H}\mathbf{x}\right\|^2\right]\end{aligned} \tag{4.11}$$

where $\left\|\mathbf{y}-\mathbf{H}\mathbf{x}\right\|^2$ is given by

$$\left\|\mathbf{y}-\mathbf{H}\mathbf{x}\right\|^2 = \mathbf{y}^{\mathrm{H}}\mathbf{y} - \mathbf{y}^{\mathrm{H}}\mathbf{H}\mathbf{x} - \mathbf{x}^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\mathbf{y} + \mathbf{x}^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\mathbf{H}\mathbf{x}. \tag{4.12}$$

By using (4.12), it is straightforward to calculate

$$\begin{aligned}&E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(\mathrm{m}-1)}}\left[\left\|\mathbf{y}-\mathbf{H}\mathbf{x}\right\|^2\right] \\ &= E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(\mathrm{m}-1)}}\left[\mathbf{y}^{\mathrm{H}}\mathbf{y} - \mathbf{y}^{\mathrm{H}}\mathbf{H}\mathbf{x} - \mathbf{x}^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\mathbf{y} + \mathbf{x}^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\mathbf{H}\mathbf{x}\right] \\ &= E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(\mathrm{m}-1)}}\left[\mathbf{y}^{\mathrm{H}}\mathbf{y}\right] - E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(\mathrm{m}-1)}}\left[\mathbf{y}^{\mathrm{H}}\mathbf{H}\mathbf{x}\right] - E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(\mathrm{m}-1)}}\left[\mathbf{x}^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\mathbf{y}\right] + E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(\mathrm{m}-1)}}\left[\mathbf{x}^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\mathbf{H}\mathbf{x}\right] \\ &= \mathbf{y}^{\mathrm{H}}\mathbf{y} - \mathbf{y}^{\mathrm{H}}\mathbf{H}E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(\mathrm{m}-1)}}\left[\mathbf{x}\right] - E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(\mathrm{m}-1)}}\left[\mathbf{x}^{\mathrm{H}}\right]\mathbf{H}^{\mathrm{H}}\mathbf{y} + E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(\mathrm{m}-1)}}\left[\mathbf{x}^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\mathbf{H}\mathbf{x}\right] \\ &= \mathbf{y}^{\mathrm{H}}\mathbf{y} - \mathbf{y}^{\mathrm{H}}\mathbf{H}E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(\mathrm{m}-1)}}\left[\mathbf{x}\right] - E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(\mathrm{m}-1)}}\left[\mathbf{x}^{\mathrm{H}}\right]\mathbf{H}^{\mathrm{H}}\mathbf{y} + tr\left(E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(\mathrm{m}-1)}}\left[\mathbf{x}^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\mathbf{H}\mathbf{x}\right]\right) \\ &= \mathbf{y}^{\mathrm{H}}\mathbf{y} - \mathbf{y}^{\mathrm{H}}\mathbf{H}E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(\mathrm{m}-1)}}\left[\mathbf{x}\right] - E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(\mathrm{m}-1)}}\left[\mathbf{x}^{\mathrm{H}}\right]\mathbf{H}^{\mathrm{H}}\mathbf{y} + tr\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(\mathrm{m}-1)}}\left[\mathbf{x}\mathbf{x}^{\mathrm{H}}\right]\right)\end{aligned} \tag{4.13}$$

where we have $\mathbf{H} = \mathbf{M} + \mathbf{\Phi}\mathbf{W}$ from (2.17), both $\mathbf{y}$ and $\mathbf{H}$ contain nothing about the random variable $\mathbf{x}$, $E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(\mathrm{m}-1)}}\left[\mathbf{x}^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\mathbf{H}\mathbf{x}\right]$ is a value so that it is obviously equal to the sum of the diagonal elements given by $tr\left(E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(\mathrm{m}-1)}}\left[\mathbf{x}^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\mathbf{H}\mathbf{x}\right]\right)$, and the last equality is true because $tr\left(\mathbf{A}*\mathbf{B}\right)$ can be always converted into $tr\left(\mathbf{B}*\mathbf{A}\right)$ of which $\mathbf{A}$ and $\mathbf{B}$ are matrices.

With further calculation, we can obtain

$$
\begin{aligned}
&E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(m-1)}}\left[\left\|\mathbf{y}-\mathbf{H}\mathbf{x}\right\|^2\right] \\
&=\frac{-1}{\sigma^2}\left\{\begin{array}{l}
\mathbf{y}^{\mathrm{H}}\mathbf{y}-\mathbf{y}^{\mathrm{H}}\mathbf{M}E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(m-1)}}\left[\mathbf{x}\right]-\mathbf{y}^{\mathrm{H}}\mathbf{\Phi}\mathbf{W}E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(m-1)}}\left[\mathbf{x}\right]-E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(m-1)}}\left[\mathbf{x}^{\mathrm{H}}\right]\mathbf{M}^{\mathrm{H}}\mathbf{y}- \\
E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(m-1)}}\left[\mathbf{x}^{\mathrm{H}}\right]\mathbf{W}^{\mathrm{H}}\mathbf{\Phi}^{\mathrm{H}}\mathbf{y}+tr\left(\mathbf{M}^{\mathrm{H}}\mathbf{M}E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(m-1)}}\left[\mathbf{x}\mathbf{x}^{\mathrm{H}}\right]\right)+ \\
tr\left(E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(m-1)}}\left[\mathbf{x}\mathbf{x}^{\mathrm{H}}\right]\mathbf{M}^{\mathrm{H}}\mathbf{\Phi}\mathbf{W}\right)+tr\left(\mathbf{W}^{\mathrm{H}}\mathbf{\Phi}^{\mathrm{H}}\mathbf{M}E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(m-1)}}\left[\mathbf{x}\mathbf{x}^{\mathrm{H}}\right]\right)+ \\
tr\left(\mathbf{W}^{\mathrm{H}}\mathbf{\Phi}^{\mathrm{H}}\mathbf{\Phi}\mathbf{W}E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(m-1)}}\left[\mathbf{x}\mathbf{x}^{\mathrm{H}}\right]\right)
\end{array}\right\} \\
&=\frac{-1}{\sigma^2}\left\{\begin{array}{l}
\mathbf{y}^{\mathrm{H}}\mathbf{y}-\mathbf{y}^{\mathrm{H}}\mathbf{M}E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(m-1)}}\left[\mathbf{x}\right]-\mathbf{y}^{\mathrm{H}}\mathbf{\Phi}\mathbf{W}E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(m-1)}}\left[\mathbf{x}\right]-E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(m-1)}}\left[\mathbf{x}^{\mathrm{H}}\right]\mathbf{M}^{\mathrm{H}}\mathbf{y}- \\
E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(m-1)}}\left[\mathbf{x}^{\mathrm{H}}\right]\mathbf{W}^{\mathrm{H}}\mathbf{\Phi}^{\mathrm{H}}\mathbf{y}+tr\left(\mathbf{M}^{\mathrm{H}}\mathbf{M}E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(m-1)}}\left[\mathbf{x}\mathbf{x}^{\mathrm{H}}\right]\right)+ \\
tr\left(E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(m-1)}}\left[\mathbf{x}\mathbf{x}^{\mathrm{H}}\right]\mathbf{M}^{\mathrm{H}}\mathbf{\Phi}\mathbf{W}\right)+tr\left(\mathbf{W}^{\mathrm{H}}\mathbf{\Phi}^{\mathrm{H}}\mathbf{M}E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(m-1)}}\left[\mathbf{x}\mathbf{x}^{\mathrm{H}}\right]\right)+ \\
tr\left(\mathbf{W}E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(m-1)}}\left[\mathbf{x}\mathbf{x}^{\mathrm{H}}\right]\mathbf{W}^{\mathrm{H}}\mathbf{\Phi}^{\mathrm{H}}\mathbf{\Phi}\right)
\end{array}\right\} \\
&=\frac{-1}{\sigma^2}\left\{\begin{array}{l}
\mathbf{y}^{\mathrm{H}}\mathbf{y}-\mathbf{y}^{\mathrm{H}}\mathbf{M}E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(m-1)}}\left[\mathbf{x}\right]-\mathbf{y}^{\mathrm{H}}\mathbf{\Phi}E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(m-1)}}\left[\mathbf{X}\right]\mathbf{F}\mathbf{a}-E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(m-1)}}\left[\mathbf{x}^{\mathrm{H}}\right]\mathbf{M}^{\mathrm{H}}\mathbf{y}- \\
\mathbf{a}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(m-1)}}\left[\mathbf{X}^{\mathrm{H}}\right]\mathbf{\Phi}^{\mathrm{H}}\mathbf{y}+tr\left(\mathbf{M}^{\mathrm{H}}\mathbf{M}E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(m-1)}}\left[\mathbf{x}\mathbf{x}^{\mathrm{H}}\right]\right)+ \\
\left[Diag\left(E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(m-1)}}\left[\mathbf{x}\mathbf{x}^{\mathrm{H}}\right]\mathbf{M}^{\mathrm{H}}\mathbf{\Phi}\right)\right]^{\mathrm{T}}\mathbf{F}\mathbf{a}+\mathbf{a}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}Diag\left(\mathbf{\Phi}^{\mathrm{H}}\mathbf{M}E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(m-1)}}\left[\mathbf{x}\mathbf{x}^{\mathrm{H}}\right]\right)+ \\
tr\left(\left(\left(\mathbf{w}\mathbf{w}^{\mathrm{H}}\right)\odot E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(m-1)}}\left[\mathbf{x}\mathbf{x}^{\mathrm{H}}\right]\right)\mathbf{\Phi}^{\mathrm{H}}\mathbf{\Phi}\right)
\end{array}\right\} \\
&=\frac{-1}{\sigma^2}\left\{\begin{array}{l}
\mathbf{y}^{\mathrm{H}}\mathbf{y}-\mathbf{y}^{\mathrm{H}}\mathbf{M}E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(m-1)}}\left[\mathbf{x}\right]-\mathbf{y}^{\mathrm{H}}\mathbf{\Phi}E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(m-1)}}\left[\mathbf{X}\right]\mathbf{F}\mathbf{a}-E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(m-1)}}\left[\mathbf{x}^{\mathrm{H}}\right]\mathbf{M}^{\mathrm{H}}\mathbf{y}- \\
\mathbf{a}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(m-1)}}\left[\mathbf{X}^{\mathrm{H}}\right]\mathbf{\Phi}^{\mathrm{H}}\mathbf{y}+tr\left(\mathbf{M}^{\mathrm{H}}\mathbf{M}E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(m-1)}}\left[\mathbf{x}\mathbf{x}^{\mathrm{H}}\right]\right)+ \\
\left[Diag\left(E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(m-1)}}\left[\mathbf{x}\mathbf{x}^{\mathrm{H}}\right]\mathbf{M}^{\mathrm{H}}\mathbf{\Phi}\right)\right]^{\mathrm{T}}\mathbf{F}\mathbf{a}+\mathbf{a}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}Diag\left(\mathbf{\Phi}^{\mathrm{H}}\mathbf{M}E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(m-1)}}\left[\mathbf{x}\mathbf{x}^{\mathrm{H}}\right]\right)+ \\
tr\left(\mathbf{F}\mathbf{a}\mathbf{a}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\left(E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(m)}}\left[\mathbf{x}\mathbf{x}^{\mathrm{H}}\right]^{\mathrm{T}}\odot\left(\mathbf{\Phi}^{\mathrm{H}}\mathbf{\Phi}\right)\right)\right)
\end{array}\right\} \\
&=\frac{-1}{\sigma^2}\left\{\begin{array}{l}
\mathbf{y}^{\mathrm{H}}\mathbf{y}-\mathbf{y}^{\mathrm{H}}\mathbf{M}E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(m)}}\left[\mathbf{x}\right]-\mathbf{y}^{\mathrm{H}}\mathbf{\Phi}E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(m)}}\left[\mathbf{X}\right]\mathbf{F}\mathbf{a}-E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(m)}}\left[\mathbf{x}^{\mathrm{H}}\right]\mathbf{M}^{\mathrm{H}}\mathbf{y}- \\
\mathbf{a}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(m)}}\left[\mathbf{X}^{\mathrm{H}}\right]\mathbf{\Phi}^{\mathrm{H}}\mathbf{y}+tr\left(\mathbf{M}^{\mathrm{H}}\mathbf{M}E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(m)}}\left[\mathbf{x}\mathbf{x}^{\mathrm{H}}\right]\right)+ \\
\left[Diag\left(E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(m)}}\left[\mathbf{x}\mathbf{x}^{\mathrm{H}}\right]\mathbf{M}^{\mathrm{H}}\mathbf{\Phi}\right)\right]^{\mathrm{T}}\mathbf{F}\mathbf{a}+\mathbf{a}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}Diag\left(\mathbf{\Phi}^{\mathrm{H}}\mathbf{M}E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(m)}}\left[\mathbf{x}\mathbf{x}^{\mathrm{H}}\right]\right)+ \\
\mathbf{a}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\left(E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{w}}_{(m)}}\left[\mathbf{x}\mathbf{x}^{\mathrm{H}}\right]^{\mathrm{T}}\odot\left(\mathbf{\Phi}^{\mathrm{H}}\mathbf{\Phi}\right)\right)\mathbf{F}\mathbf{a}
\end{array}\right\}
\end{aligned}
\tag{4.14}
$$

where $\odot$ denotes the Hadamard product, the notation $Diag$ represents taking the diagonal elements of a matrix to form a column vector, $\mathbf{w}$ is replaced with $\mathbf{Fa}$, and $\mathbf{a}$ is a vector of channel slopes in time domain turning out to be the new channel information that should be estimated. Therefore, the E-step of (4.9) and the M-step of (4.10) are redefined by

E-step:

$$\Omega(\mathbf{a}\,|\,\mathbf{y},\hat{\mathbf{a}}_{(\mathrm{m}-1)}) = \frac{-1}{\sigma^2}\left\{\begin{array}{l} \mathbf{y}^{\mathrm{H}}\mathbf{y} - \mathbf{y}^{\mathrm{H}}\mathbf{M}E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{a}}_{(\mathrm{m})}}\left[\mathbf{x}\right] - \mathbf{y}^{\mathrm{H}}\mathbf{\Phi}E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{a}}_{(\mathrm{m})}}\left[\mathbf{X}\right]\mathbf{Fa} - E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{a}}_{(\mathrm{m})}}\left[\mathbf{x}^{\mathrm{H}}\right]\mathbf{M}^{\mathrm{H}}\mathbf{y} - \\ \mathbf{a}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{a}}_{(\mathrm{m})}}\left[\mathbf{X}^{\mathrm{H}}\right]\mathbf{\Phi}^{\mathrm{H}}\mathbf{y} + tr\left(\mathbf{M}^{\mathrm{H}}\mathbf{M}E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{a}}_{(\mathrm{m})}}\left[\mathbf{x}\mathbf{x}^{\mathrm{H}}\right]\right) + \\ \left[Diag\left(E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{a}}_{(\mathrm{m})}}\left[\mathbf{x}\mathbf{x}^{\mathrm{H}}\right]\mathbf{M}^{\mathrm{H}}\mathbf{\Phi}\right)\right]^{\mathrm{T}}\mathbf{Fa} + \mathbf{a}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}Diag\left(\mathbf{\Phi}^{\mathrm{H}}\mathbf{M}E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{a}}_{(\mathrm{m})}}\left[\mathbf{x}\mathbf{x}^{\mathrm{H}}\right]\right) + \\ \mathbf{a}^{\mathrm{H}}\mathbf{F}^{\mathrm{H}}\left(E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{a}}_{(\mathrm{m})}}\left[\mathbf{x}\mathbf{x}^{\mathrm{H}}\right]^{\mathrm{T}} \odot \left(\mathbf{\Phi}^{\mathrm{H}}\mathbf{\Phi}\right)\right)\mathbf{Fa} \end{array}\right\} \tag{4.15}$$

M-step: $\hat{\mathbf{a}}_{(\mathrm{m})} = \arg\max_{\mathbf{a}} \Omega(\mathbf{a}\,|\,\mathbf{y},\hat{\mathbf{a}}_{(\mathrm{m}-1)})$ (4.16)

4.1.2 Derivation of Estimated Channel Slopes

In order to find $\hat{\mathbf{a}}_{(\mathrm{m})}$ which maximizes $\Omega(\mathbf{a}\,|\,\mathbf{y},\hat{\mathbf{a}}_{(\mathrm{m}-1)})$, we have to differentiate (4.15) with respect to $\mathbf{a}^{\mathrm{H}}$ by using complex differentials introduced briefly in the following:

$$\begin{aligned} \nabla_{\mathrm{k}} J = 2\frac{\partial J}{\partial v_k^*}\bigg|_{v_k = g_k + jh_k} &= \frac{\partial J}{\partial g_k} + j\frac{\partial J}{\partial h_k} \\ \Rightarrow \frac{\partial J}{\partial v_k^*} &= \frac{1}{2}\left\{\frac{\partial J}{\partial g_k} + j\frac{\partial J}{\partial h_k}\right\} \end{aligned} \tag{4.17}$$

where a complex value J consists of a parameter v_k of which g_k and h_k are the real part and the imaginary part respectively and the superscript $(\cdot)^*$ stands for the complex conjugate of a value. Then the complex differentiation of vectors can be presented as

$$\frac{\partial(\mathbf{r}^{\mathrm{H}}\mathbf{s})}{\partial\mathbf{s}^{\mathrm{H}}}=\frac{1}{2}\mathbf{r}^{\mathrm{H}}\left[1+j\cdot j\right]=\mathbf{0} \tag{4.18}$$

$$\frac{\partial(\mathbf{s}^{\mathrm{H}}\mathbf{r})}{\partial\mathbf{s}^{\mathrm{H}}}=\frac{1}{2}\left[1+j\cdot(-j)\right]\mathbf{r}=\mathbf{r} \tag{4.19}$$

$$\frac{\partial(\mathbf{s}^{\mathrm{H}}\mathbf{Q}\mathbf{s})}{\partial\mathbf{s}^{\mathrm{H}}}=\frac{\partial(\mathbf{s}^{\mathrm{H}})}{\partial\mathbf{s}^{\mathrm{H}}}\cdot\mathbf{Q}\mathbf{s}+\mathbf{s}^{\mathrm{H}}\mathbf{Q}\frac{\partial\mathbf{s}}{\partial\mathbf{s}^{\mathrm{H}}}=\mathbf{Q}\mathbf{s} \tag{4.20}$$

where $\mathbf{r}$ and $\mathbf{s}$ denote complex vectors and $\mathbf{Q}$ is a matrix without the parameter $\mathbf{s}$ inside. Suppose that $f(\mathbf{a})$ represents the function of $\mathbf{a}$ in (4.15), and thus the maximization in the M-step can be derived as below,

$$\begin{aligned}\frac{\partial f(\mathbf{a})}{\partial\mathbf{a}^{\mathrm{H}}}=\mathbf{F}^{\mathrm{H}}E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{a}}_{(\mathrm{m})}}\left[\mathbf{X}^{\mathrm{H}}\right]\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{y}-\mathbf{F}^{\mathrm{H}}Diag\left(\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{M}E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{a}}_{(\mathrm{m})}}\left[\mathbf{x}\mathbf{x}^{\mathrm{H}}\right]\right)-\\ \mathbf{F}^{\mathrm{H}}\left(E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{a}}_{(\mathrm{m})}}\left[\mathbf{x}\mathbf{x}^{\mathrm{H}}\right]^{t}\odot\left(\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi}\right)\right)\mathbf{F}\mathbf{a}\end{aligned} \tag{4.21}$$

$$\frac{\partial f(\mathbf{a})}{\partial\mathbf{a}^{\mathrm{H}}}=0\ \Rightarrow\ \hat{\mathbf{a}}_{(\mathrm{m})}=\mathbf{C}^{-1}\mathbf{b} \tag{4.22}$$

where we define $\mathbf{C}=\mathbf{F}^{\mathrm{H}}\left(E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{a}}_{(\mathrm{m-1})}}\left[\mathbf{x}\mathbf{x}^{\mathrm{H}}\right]^{\mathrm{T}}\odot\left(\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi}\right)\right)\mathbf{F}$ which is an invertible matrix of size $N\times N$ and $\mathbf{b}=\mathbf{F}^{\mathrm{H}}E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{a}}_{(\mathrm{m-1})}}\left[\mathbf{X}^{\mathrm{H}}\right]\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{y}-\mathbf{F}^{\mathrm{H}}Diag\left(\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{M}E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{a}}_{(\mathrm{m-1})}}\left[\mathbf{x}\mathbf{x}^{\mathrm{H}}\right]\right)$ which is a column vector of length N. As a result, we can get the vector of estimated slopes $\hat{\mathbf{a}}_{(\mathrm{m})}$ at the mth EM iteration based on some given information such as the observed data $\mathbf{y}$ and the vector of previous estimated slopes $\hat{\mathbf{a}}_{(\mathrm{m-1})}$. As for the expected values of $\mathbf{X}^{\mathrm{H}}$ and $\mathbf{x}\mathbf{x}^{\mathrm{H}}$ needed in (4.22), we will obtain by applying Gibbs sampling method in next section.

4.2 Data Detection by Gibbs sampling

Now we consider the problem of computing the joint a posteriori probabilities of the transmitted signals

$$P\left(\mathbf{x} \mid \mathbf{y}, \hat{\mathbf{a}}_{(\text{m-1})}\right). \tag{4.23}$$

Based on Gibbs sampling, we estimate the PDF of (4.23) by using the probabilities of samples drawn uniformly from the marginal PDF illustrated as follows. The following case shows the way of drawing a sample for, x_1, one of the unknown parameters in the vector $\mathbf{x}$. By using Bayes' theorem and given the initial values of $\mathbf{x}^{(0)} = [x_1^{(0)} \cdots\cdots\cdots x_{\text{d}}^{(0)}]^{\text{T}}$, the marginal probability of x_1 is calculated as

$$\begin{aligned}
&P\left(x_1 = +1 \mid \mathbf{y}, \hat{\mathbf{a}}_{(\text{m-1})}, x_2, x_3, \ldots, x_{\text{d}}\right) \\
&= \frac{P\left(\mathbf{y} \mid x_1 = +1, \hat{\mathbf{a}}_{(\text{m-1})}, x_2, x_3, \ldots, x_{\text{d}}\right) P(x_1 = +1 \mid \hat{\mathbf{a}}_{(\text{m-1})}, x_2, x_3, \ldots, x_{\text{d}})}{P\left(\mathbf{y} \mid \hat{\mathbf{a}}_{(\text{m-1})}, x_2, x_3, \ldots, x_{\text{d}}\right)} \\
&= \frac{P\left(\mathbf{y} \mid x_1 = +1, \hat{\mathbf{a}}_{(\text{m-1})}, x_2, x_3, \ldots, x_{\text{d}}\right) P\left(x_1 = +1\right)}{\sum_{x_1} P\left(\mathbf{y} \mid x_1, \hat{\mathbf{a}}_{(\text{m-1})}, x_2, x_3, \ldots, x_{\text{d}}\right) P\left(x_1 \mid \hat{\mathbf{a}}_{(\text{m-1})}, x_2, x_3, \ldots, x_{\text{d}}\right)} \\
&= \frac{P\left(\mathbf{y} \mid x_1 = +1, \hat{\mathbf{a}}_{(\text{m-1})}, x_2, x_3, \ldots, x_{\text{d}}\right) P\left(x_1 = +1\right)}{\sum_{x_1} P\left(\mathbf{y} \mid x_1, \hat{\mathbf{a}}_{(\text{m-1})}, x_2, x_3, \ldots, x_{\text{d}}\right) P\left(x_1\right)} \\
&= \frac{P\left(\mathbf{y} \mid \mathbf{x}_{(x_1=+1)}, \hat{\mathbf{a}}_{(\text{m-1})}\right) P\left(x_1 = +1\right)}{\sum_{x_1} P\left(\mathbf{y} \mid \mathbf{x}, \hat{\mathbf{a}}_{(\text{m-1})}\right) P\left(x_1\right)}
\end{aligned} \tag{4.24}$$

where $P(x_1 = +1)$ is the prior symbol probability of x_1 and the subscript d stands for the size of $\mathbf{x}$, or precisely, the depth of the Gibbs sampler. When no prior information is available, it is assumed that $P(x_1 = +1) = \frac{1}{2}$, i.e., all symbols are equally likely. Then (4.24) can be written as

$$P\left(x_1=+1\,|\,\mathbf{y},\hat{\mathbf{a}}_{(\text{m-1})},x_2,x_3,\ldots,x_{\text{d}}\right)=\frac{P\left(\mathbf{y}\,|\,\mathbf{x}_{(x_1=+1)},\hat{\mathbf{a}}_{(\text{m-1})}\right)}{\sum_{x_1}P\left(\mathbf{y}\,|\,\mathbf{x},\hat{\mathbf{a}}_{(\text{m-1})}\right)}. \tag{4.25}$$

Since $\mathbf{z}$ is a Gaussian vector, $P(\mathbf{y}\,|\,\mathbf{x},\hat{\mathbf{a}}_{(\text{m-1})})$ can be derived as

$$P\left(\mathbf{y}\,|\,\mathbf{x},\hat{\mathbf{a}}_{(\text{m-1})}\right)\propto\exp\left(-\frac{1}{\sigma^2}\left\|\mathbf{y}-\mathbf{H}\mathbf{x}\right\|^2\right) \tag{4.26}$$

where we define $\sigma^2=\sigma_z^2+\sigma_{ICI}^2$ and σ_{ICI}^2 is the ICI power calculated in next chapter. Thus we can get the posterior probability of $x_1=+1$ by substituting (4.26) into (4.25), and $P(x_1=-1\,|\,\mathbf{y},\hat{\mathbf{a}}_{(\text{m-1})},x_2,x_3,\ldots,x_{\text{d}})$ is also obtained through the same method. With these posterior probabilities, we know about the probability of choosing one from all of the symbols. Let P_1 be $P(x_1=-1\,|\,\mathbf{y},\hat{\mathbf{a}}_{(\text{m-1})},x_2,x_3,\ldots,x_{\text{d}})$, and suppose that U is a uniform-distributed value bound in the interval of $[0,1]$. U combined with P_1 can be used to make a decision, e.g., a sample of $x_1=+1$ is chosen if U is smaller than P_1, whereas a sample of $x_1=-1$ is chosen when U is within the range of $(P_1,1]$.

After a sample of x_1 is drawn, x_2 is the next parameter to draw a sample with a given value of x_1. As above, the samples of $x_2,\ldots,x_{\text{d}}$ can be drawn from their a posteriori probabilities like the PDF in (4.25) and then a vector of samples is acquired. The vector of samples denoted as $\mathbf{x}_{(\text{m-1})}^{(\text{i-1})}$ is considered to be the given values for the ith iteration of Gibbs sampling, and the execution of drawing samples is presented in the following.

At the (m-1)th EM iteration:

- Draw a vector of samples $\hat{\mathbf{x}}_{(\text{m-1})}^{(1)}$ from $P\left(\mathbf{x}\,|\,\mathbf{y},\hat{\mathbf{a}}_{(\text{m-2})}\right)$ given $\hat{\mathbf{x}}_{(\text{m-2})}^{(\text{d})}$ and $\hat{\mathbf{a}}_{(\text{m-2})}$.
- Draw a vector of samples $\hat{\mathbf{x}}_{(\text{m-1})}^{(2)}$ from $P\left(\mathbf{x}\,|\,\mathbf{y},\hat{\mathbf{a}}_{(\text{m-2})}\right)$ given $\hat{\mathbf{x}}_{(\text{m-1})}^{(1)}$ and $\hat{\mathbf{a}}_{(\text{m-2})}$.

⋮

- Draw a vector of samples $\hat{\mathbf{x}}_{(m-1)}^{(d)}$ from $P\left(\mathbf{x}\,|\,\mathbf{y},\hat{\mathbf{a}}_{(m-2)}\right)$ given $\hat{\mathbf{x}}_{(m-1)}^{(d-1)}$ and $\hat{\mathbf{a}}_{(m-2)}$.

At the mth EM iteration: (Assume that m is the final number of iteration.)

- Draw a vector of samples $\hat{\mathbf{x}}_{(m)}^{(1)}$ from $P\left(\mathbf{x}\,|\,\mathbf{y},\hat{\mathbf{a}}_{(m-1)}\right)$ given $\hat{\mathbf{x}}_{(m-1)}^{(d)}$ and $\hat{\mathbf{a}}_{(m-1)}$.
- Draw a vector of samples $\hat{\mathbf{x}}_{(m)}^{(2)}$ from $P\left(\mathbf{x}\,|\,\mathbf{y},\hat{\mathbf{a}}_{(m-1)}\right)$ given $\hat{\mathbf{x}}_{(m)}^{(1)}$ and $\hat{\mathbf{a}}_{(m-1)}$.

⋮

- Draw a vector of samples $\hat{\mathbf{x}}_{(m)}^{(d)}$ from $P\left(\mathbf{x}\,|\,\mathbf{y},\hat{\mathbf{a}}_{(m-1)}\right)$ given $\hat{\mathbf{x}}_{(m)}^{(d-1)}$ and $\hat{\mathbf{a}}_{(m-1)}$.

The probabilities of the last drawing through EM iterations are utilized to be the joint a posteriori probabilities for estimating the channel information. The samples generated at the final EM iteration are formed into a sequence of vectors: $\left\{\hat{\mathbf{x}}_{(m)}^{(i)}:\ i\in\{1,\ldots,d\}\right\}$, and an average of the samples, denoted as $\hat{\mathbf{x}}^{(m)}$, represents the output of the EM detector where the superscript $(\cdot)^{(m)}$ stands for the number of EM iteration.

4.3 Initial Setting

The EM algorithm guarantees to converge to a local maximum instead of a global one, so the initialization of data and an initial channel estimate are significant for locating the global maximum. The initial channel estimation (CE) is accomplished by the use of pilot tones in the specific positions of an OFDM symbol, and it can be improved through the decided data symbols. Let $\mathbf{x}_P$ be a vector of J pilot tones situated at the indices $\{\varphi_0, \varphi_1, \ldots, \varphi_{J-1}\}$ of a symbol, and $d_1, d_2, \ldots, d_{\mathrm{L}}$ are defined as the channel delay of paths. Then the system model related to pilot symbols is expressed as

$$\mathbf{y}_P = \mathbf{X}_p \cdot \mathbf{F}_P \cdot \mathbf{h} + \mathbf{z}_p \tag{4.27}$$

where $\mathbf{y}_P = \left[y[\varphi_0], \cdots, y[\varphi_{J-1}] \right]^{\mathrm{T}}$ is the received data of size $J \times 1$ in the positions of pilot tones, $\mathbf{X}_P = diag\left\{ \left[x[\varphi_0], \cdots, x[\varphi_{J-1}] \right]^{\mathrm{T}} \right\}$ stands for a diagonal matrix of pilot symbols whose size is $J \times J$, and $\mathbf{F}_p = \begin{bmatrix} \mathbf{F}[\varphi_0, d_1] & \cdots & \mathbf{F}[\varphi_0, d_{\mathrm{L}}] \\ \vdots & \ddots & \vdots \\ \mathbf{F}[\varphi_{J-1}, d_1] & \cdots & \mathbf{F}[\varphi_{J-1}, d_{\mathrm{L}}] \end{bmatrix}$ is a matrix of size $J \times \mathrm{L}$ which composed of some elements of $\mathbf{F}$. Utilizing the MMSE-based CE method, we obtain an estimated channel impulse response [15] presented as

$$\begin{aligned} \hat{\mathbf{h}} &= \arg\min_{\mathbf{h}} \left\| \mathbf{y}_P - \mathbf{X}_P \mathbf{F}_P \mathbf{h} \right\|^2 \\ &= \left(\mathbf{F}_P^{\mathrm{H}} \mathbf{X}_P^{\mathrm{H}} \mathbf{X}_P \mathbf{F}_P + \left(\sigma_z^2 + \sigma_{ICI}^2 \right) \right)^{-1} \mathbf{F}_P^{\mathrm{H}} \mathbf{X}_P^{\mathrm{H}} \mathbf{y}_P \end{aligned} \tag{4.28}$$

where the variance of ICI, σ_{ICI}^2, equals $\left(2\pi f_{\mathrm{D}}\right)^2 / 12$ approximately by applying the central limit theorem [16]. Accordingly, the initial channel estimate in frequency domain is described as

$$\hat{\mathbf{m}} = \mathbf{F}\left(\mathbf{F}_P^{\mathrm{H}}\mathbf{X}_P^{\mathrm{H}}\mathbf{X}_P\mathbf{F}_P + \left(\sigma_z^2 + \sigma_{ICI}^2\right)\right)^{-1}\mathbf{F}_P^{\mathrm{H}}\mathbf{X}_P^{\mathrm{H}}\mathbf{y}_P \tag{4.29}$$

and thus the initial data $\hat{\mathbf{x}}^{(0)}$ can be derived from the one-tap equalizer.

As shown in Fig.4.1, making use of the decided data $\hat{\mathbf{x}}^{(0)}$, we can further generate an updated channel estimate $\bar{\mathbf{m}}$ via the similar formula in (4.29) and produce a vector of new decided data symbols denoted as $\bar{\mathbf{x}}^{(0)}$. Then $\bar{\mathbf{x}}^{(0)}$ and $\bar{\mathbf{M}} = diag\{\bar{\mathbf{m}}\}$ are exploited by the EM algorithm.

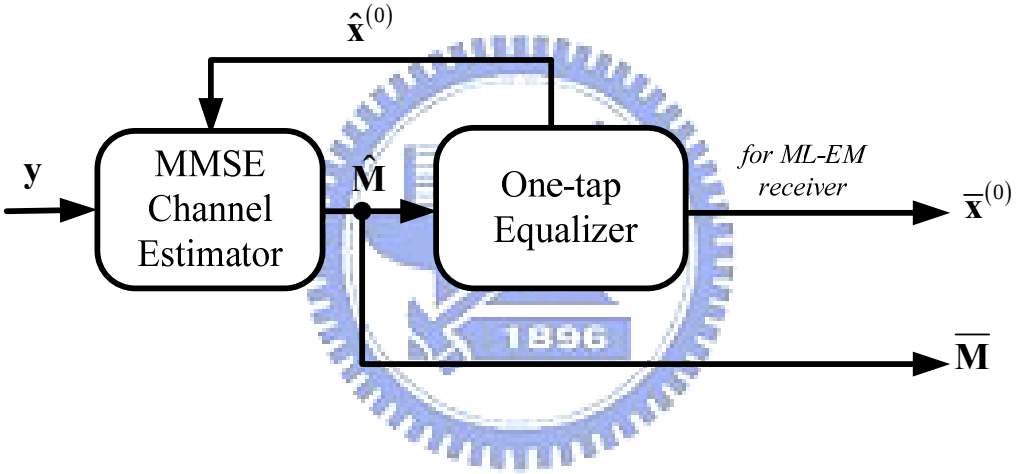


Fig. 4.1 Initialization for ML-EM receivers.

4.4 EM-based Channel Estimation Method

Applying the description in the preceding sections, the ML-EM algorithm for channel estimation can be summarized as follows.

Initialization:

Calculate $\bar{\mathbf{M}}$, produce an initial $\bar{\mathbf{x}}^{(0)}$ by zero-forcing (ZF), set $\hat{\mathbf{a}}_{(0)} = 0$, and let the subscript m be zero.

Procedure of the ML-EM algorithm:

- m=m+1
- Draw samples $\left\{\mathbf{x}_{(m)}^{(i)} : i \in \{1,\ldots,d\}\right\}$, compute the probabilities of symbols by using the samples, and then obtain $E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{a}}_{(m-1)}}\left[\mathbf{x}\mathbf{x}^{H}\right]$ and $E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{a}}_{(m-1)}}\left[\mathbf{X}^{H}\right]$.
- Estimate the channel information through the derivation $\hat{\mathbf{a}}_{(m)} = \mathbf{C}^{-1}\mathbf{b}$ where $\mathbf{b} = \mathbf{F}^{H}E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{a}}_{(m-1)}}\left[\mathbf{X}^{H}\right]\boldsymbol{\Phi}^{H}\mathbf{y} - \mathbf{F}^{H}Diag\left(\boldsymbol{\Phi}^{H}\mathbf{M}E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{a}}_{(m-1)}}\left[\mathbf{x}\mathbf{x}^{H}\right]\right)$ and $\mathbf{C} = \mathbf{F}^{H}\left(E_{\mathbf{x}|\mathbf{y},\hat{\mathbf{a}}_{(m-1)}}\left[\mathbf{x}\mathbf{x}^{H}\right]^{T} \odot \left(\boldsymbol{\Phi}^{H}\boldsymbol{\Phi}\right)\right)\mathbf{F}$.
- Initial setting for next iteration:

  Set $\hat{\mathbf{x}}^{(m)}$ be the initial data of the Gibbs sampler and let $\hat{\mathbf{a}}_{(m)}$ be the updated channel information.

Stopping criterion:

The algorithm stops only when the iteration number reaches to a predefined limit.

4.5 Group-wise Method

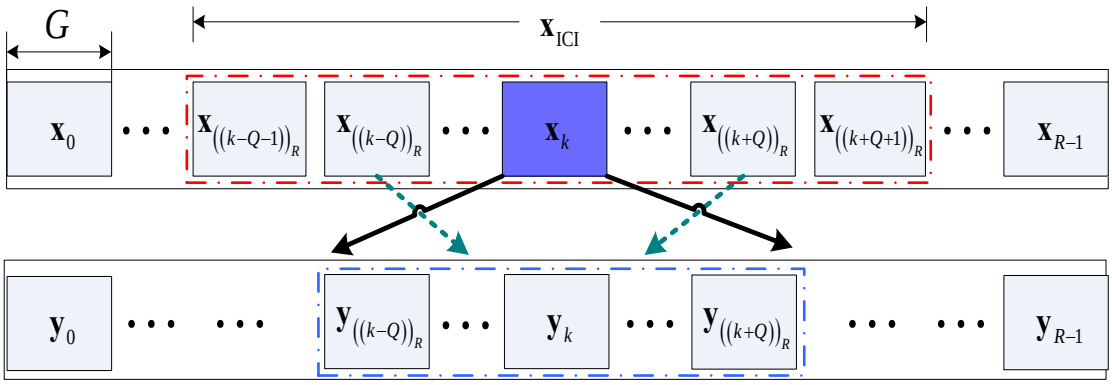


Fig. 4.2 An illustration of group-wise detection.

The aforementioned ML-EM algorithm has high computational complexity, and hence we use the group-wise method to provide a practical implementation for the ML-EM receiver. N_S subcarriers are partitioned into R groups, and each group contains G subcarriers. The j th group of subcarriers is given by $\mathbf{G}_j = \{jG,\ldots,(j+1)G-1\}$, for $j = 0,\ldots,R-1$. Define the j th data group as $\mathbf{x}_j = [x[jG],\ldots,x[(j+1)G-1]]^{\mathrm{T}}$, the j th observation group as $\mathbf{y}_j = [y[jG],\ldots,y[(j+1)G-1]]^{\mathrm{T}}$, two sets $\mathbf{B}_j = \{((j-Q-1))_R,\ldots,((j+Q+1))_R\}$ and $\mathbf{D}_j = \{((j-Q))_R,\ldots,((j+Q))_R\}$. In order to simplify the interpretation, the concept of grouping is demonstrated with one of the data groups e.g. the k th data group and illustrated in Fig.4.2. The energy of $\mathbf{x}_k$ spreads over the adjacent $2Q+1$ groups (involving $\mathbf{y}_k$ itself), so $\{\mathbf{y}_j : j \in \mathbf{D}_k\}$ is the corresponding set of observation groups. Moreover, the set $\{\mathbf{y}_j : j \in \mathbf{D}_k\}$ consists of the interference resulted from the spreading energy made by a set of neighboring data groups denoted as $\{\mathbf{x}_j : j \in \mathbf{B}_k \setminus \{k\}\}$.

Outside the EM detector, there is an **outer iteration** loop iterating between the

ML-EM detector and the ICI canceller as depicted in Fig.4.3. During an outer iteration, we diminish the ICI term by subtracting the leakage of other groups from the set of observation groups and then obtain the ICI-reduced signals:

$$\bar{\mathbf{y}}_j = \mathbf{y}_j - \sum_{i \in \mathbf{B}_k \setminus \{k\}} \bar{\mathbf{H}}_{j,i} \bar{\mathbf{x}}_i \text{, for } \; j \in \mathbf{D}_k \tag{4.30}$$

where the (r,s)th entry of $\bar{\mathbf{H}}_{j,i}$ is given by the $(jG+r, iG+s)$th entry of $\bar{\mathbf{H}}$ representing an estimate of $\mathbf{H}$ for $r,s \in \{1,\ldots,G-1\}$. $\bar{\mathbf{H}}$ and $\bar{\mathbf{x}}_i$ come from the output of the EM detector at the previous outer iteration or from the initialization in the beginning. After eliminating the ICI effect, the EM detector is performed by using EM-based channel estimation together with the group-wise method.

When Gibbs sampling is executed, samples are drawn by applying a ZF sampler. Every data group, which belongs to the sampler, takes turn by iteration loops to be calculated with the corresponding observation groups $\{\bar{\mathbf{y}}_j : j \in \mathbf{D}_k\}$, and thus a large number of samples are acquired. Then $\bar{\mathbf{x}}_k^{(\mathrm{m-1})}$ is the decided data group at the $(\mathrm{m}-1)$th EM iteration defined by an average of the collected samples drawn for the group $\mathbf{x}_k$. And $\bar{\mathbf{x}}_k^{(\mathrm{m-1})}$ is sent back to be an initial data group for the ZF sampler at the mth EM iteration, while $\bar{\mathbf{x}}_k^{(0)}$ given from the initial setting is an initial data group at the first EM iteration. Going through the iteration of Gibbs sampling, we obtain $d-n_0$ sets of detected data groups (for neglecting n_0 sets of burn-in samples).

With the obtainment of all the groups $\{\bar{\mathbf{x}}_k^{(\mathrm{m-1})} : k \in \{1,\ldots,R\}\}$, we can form $d-n_0$

vectors of N_S transmitted data symbols entirely and store up a sequence of probabilities of the final drawing within the $(\mathrm{m}-1)$th EM iteration denoted as $\mathbf{P}^{(\mathrm{m-1})}$. Because of the probabilities $\mathbf{P}^{(\mathrm{m-1})}$, the expected values required in (4.22) can be calculated, and hence derive the channel information $\hat{\mathbf{a}}_{(\mathrm{m}-1)}$ for the $(\mathrm{m}-1)$th EM iteration. $\hat{\mathbf{a}}_{(\mathrm{m}-1)}$ and $\bar{\mathbf{x}}^{(\mathrm{m-1})}$ are both provided for the given knowledge at next EM iteration. Besides, $\hat{\mathbf{a}}_{(\mathrm{m})}$ and $\bar{\mathbf{x}}^{(\mathrm{m})}$ are yielded as the way described above and become the output of the EM detector at the mth EM iteration. After cancelling the ICI effect, $\bar{\mathbf{M}}$ can be modified by substituting the updated data $\bar{\mathbf{x}}^{(\mathrm{m})}$ into (4.29) via the unit of channel estimation update shown in Fig.4.3. $\bar{\mathbf{M}}$ is replaced by $\underline{\mathbf{M}}$, and then $\bar{\mathbf{H}}$ is calculated as $\underline{\mathbf{M}}+\mathbf{\Phi F}\hat{\mathbf{a}}_{(\mathrm{m})}$. Therefore, another outer iteration starts with the given information $\bar{\mathbf{H}}$ and $\bar{\mathbf{x}}^{(\mathrm{m})}$ from the output of the previous outer iteration.

The intuition of combining our design with the group-wise method is presented as below. The data group $\mathbf{x}_k$ contributes most of energy to the ICI-reduced observation groups $\left\{\bar{\mathbf{y}}_j : j \in \mathbf{D}_k\right\}$; thus, we can obtain diversity gains and draw samples for the k th group data by performing only through $\mathbf{x}_k$ and $\left\{\bar{\mathbf{y}}_j : j \in \mathbf{D}_k\right\}$. And it is obvious that the full diversity gain is attainable when the ICI effect is completely eliminated as well as the value of Q is large enough. Because spreading energy of a group mostly affects the neighboring $2Q$ groups, i.e., the observation group $\mathbf{y}_{((k+Q))_R}$ is interfered by the data groups $\left\{\mathbf{x}_j : j \in \left\{k,\ldots,\left((k+2Q)\right)_R\right\}\right\}$, so the data groups used for the ICI cancellation should be a set of $(4Q+1)$ groups. However, we can use $(2Q+3)$ data groups to achieve the comparable performance by observing the

experimental trials.

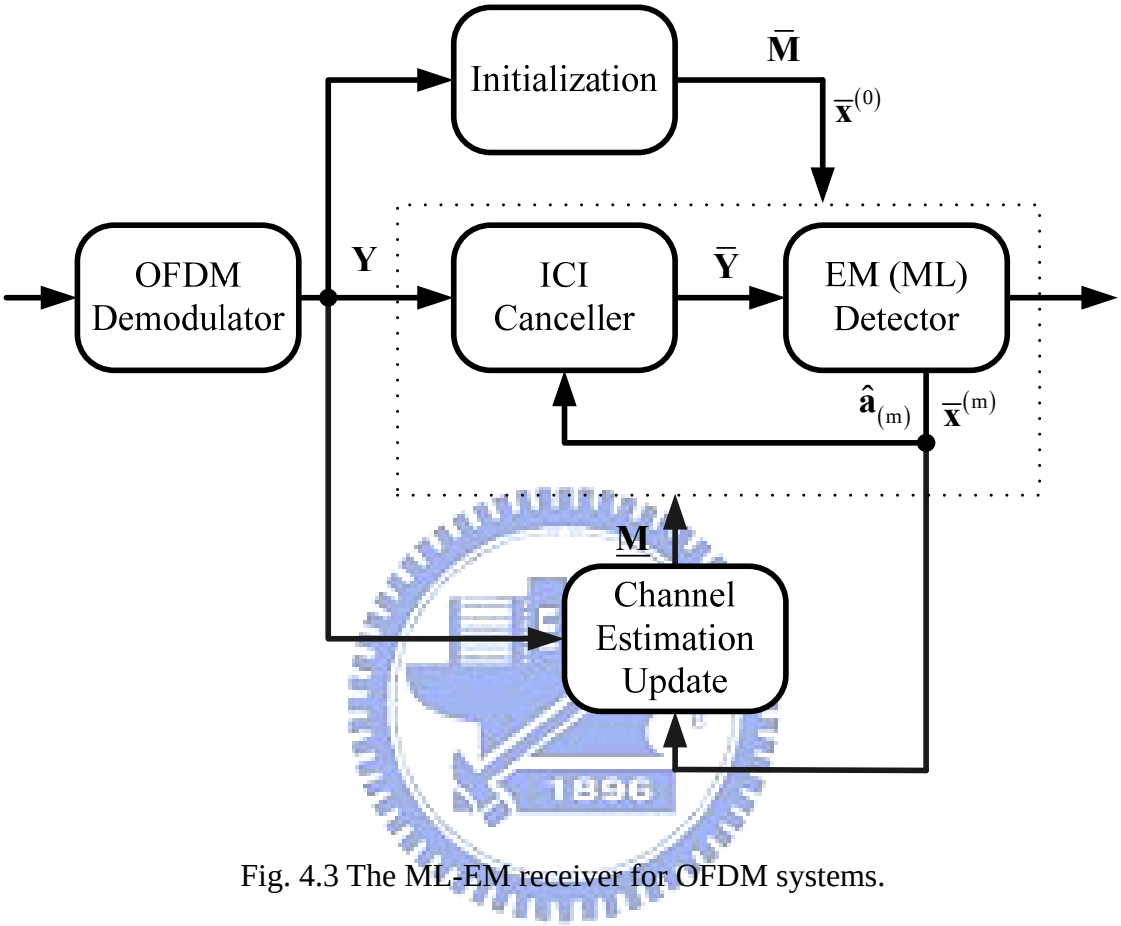


Fig. 4.3 The ML-EM receiver for OFDM systems.

Chapter 5 Residual ICI Power

When the received signals pass through a group-wise ICI canceller within an outer iteration, the ICI effect in those signals are cancelled by using the decided data at the previous outer iteration. If CE is perfect and the decided data are detected correctly, we assume that there is no ICI power after ICI cancellation. However, it is almost impossible to make all data correct; thus, some ICI power would be left and turns to be the so-called residual ICI power discussed as follows.

And accordingly, the quantity of the residual ICI power contained in the ICI-reduced signals varies with the correctness of the detected data used in the ICI canceller. Furthermore, the residual ICI power differs from the ICI power, σ_{ICI}^2, which also means the variance of ICI, because the latter represents the original power of the ICI effect without any subtraction of power from the received signals. The accuracy of the residual ICI power we estimate plays an important role in Gibbs sampling for the reason that the appropriate residual ICI power makes the a posteriori probabilities more reliable.

Following the group-wise EM-based data detection described in section 4.5, the calculation of the residual ICI power can be performed through the group-wise method as well and then illustrated with the subsequent case by employing the detected data groups $\left\{\bar{\mathbf{x}}_j : j \in \mathbf{B}_k \setminus \{k\}\right\}$ and the observation groups $\left\{\mathbf{y}_j : j \in \mathbf{D}_k\right\}$ that are shown in Fig.4.2. Each observation group consists of *G* subcarriers which suffer from different levels of residual ICI

power. For a subcarrier in the *k*th observation group, we calculate its residual ICI power by summing up the interference power caused by the incorrectness of the adjacent subcarriers in the set $\left\{\bar{\mathbf{x}}_j : j \in \mathbf{B}_k \setminus \{k\}\right\}$. The observation groups of received signals are expressed as

$$\mathbf{y}_j = \sum_{i \in \mathbf{B}_k} \left(\mathbf{M}_{j,i} + \mathbf{\Phi}_{j,i}\mathbf{W}_{j,i}\right)\mathbf{x}_i + \mathbf{z}_j \tag{5.1}$$

where $j \in \mathbf{D}_k$ and $\mathbf{W} = \mathbf{F}\mathbf{a}$. Now it is assumed that an initial channel estimate $\bar{\mathbf{M}}$ and the channel information $\hat{\mathbf{W}}$ are perfectly estimated, and the ICI-reduced signals are given by

$$\begin{aligned}
\bar{\mathbf{y}}_j &= \mathbf{y}_j - \sum_{i \in \mathbf{B}_k \setminus \{k\}} \bar{\mathbf{H}}_{j,i}\bar{\mathbf{x}}_i \\
&= \sum_{i \in \mathbf{B}_k} \left(\mathbf{M}_{j,i} + \mathbf{\Phi}_{j,i}\mathbf{W}_{j,i}\right)\mathbf{x}_i + \mathbf{z}_j - \left(\sum_{i \in \mathbf{B}_k \setminus \{k\}} \left(\bar{\mathbf{M}}_{j,i} + \mathbf{\Phi}_{j,i}\hat{\mathbf{W}}_{j,i}\right)\bar{\mathbf{x}}_i\right) \\
&= \sum_{i \in \mathbf{B}_k} \left(\mathbf{M}_{j,i} + \mathbf{\Phi}_{j,i}\mathbf{W}_{j,i}\right)\mathbf{x}_i + \mathbf{z}_j - \left(\sum_{i \in \mathbf{B}_k \setminus \{k\}} \left(\mathbf{M}_{j,i} + \mathbf{\Phi}_{j,i}\mathbf{W}_{j,i}\right)\bar{\mathbf{x}}_i\right) \\
&= \left(\mathbf{M}_{j,k} + \mathbf{\Phi}_{j,k}\mathbf{W}_{j,k}\right)\mathbf{x}_k + \underbrace{\sum_{i \in \mathbf{B}_k \setminus \{k\}} \left(\mathbf{M}_{j,i}\left(\mathbf{x}_i - \bar{\mathbf{x}}_i\right) + \mathbf{\Phi}_{j,i}\mathbf{W}_{j,i}\left(\mathbf{x}_i - \bar{\mathbf{x}}_i\right)\right)}_{\text{The term causes the residual ICI.}} + \mathbf{z}_j
\end{aligned} \tag{5.2}$$

The first two terms of (5.2) represent the signal energy of the *k*th data group, and the third term and the fourth term are caused by the mismatch between the transmitted data and the decided data and induce the residual ICI.

Supposing that $\bar{\mathbf{x}}$ are the same as the transmitted signals, the ICI-reduced signals are free from the ICI effect and presented as

$$\bar{\mathbf{y}}_j = \mathbf{M}_{j,k}\mathbf{x}_k + \mathbf{\Phi}_{j,k}\mathbf{W}_{j,k}\mathbf{x}_k + \mathbf{z}_j \tag{5.3}$$

where we find that the *k*th data group produces the remaining energy spreading over the ICI-reduced signals. In case that some of the detected signals are wrong, there is not only the

energy of the *k*th data group but also the residual ICI power resulted from the incorrect cancelling in the ICI-reduced signals. According to (5.2), the residual ICI power is defined by σ_{RIP}^2 and calculated as below.

At the first outer iteration:

We have no information about the channel variable $\hat{\mathbf{W}}$ at the first outer iteration, and thus (5.2) is rewritten as

$$\bar{\mathbf{y}}_j = \mathbf{M}_{j,k}\mathbf{x}_k + \mathbf{\Phi}_{j,k}\mathbf{W}_{j,k}\mathbf{x}_k + \sum_{i\in\mathbf{B}_k\setminus\{k\}} \mathbf{M}_{j,i}\left(\mathbf{x}_i - \bar{\mathbf{x}}_i\right) + \sum_{i\in\mathbf{B}_k\setminus\{k\}} \mathbf{\Phi}_{j,i}\mathbf{W}_{j,i}\mathbf{x}_i + \mathbf{z}_j . \qquad (5.4)$$

σ_{RIP}^2 in the ICI-reduced signals is expressed as

$$\sigma_{RIP}^2 = \sum_{i\in\mathbf{B}_k\setminus\{k\}} \mathbf{M}_{j,i}\mathbf{M}_{j,i}^* E\left[\left(\mathbf{x}_i - \bar{\mathbf{x}}_i\right)\odot\left(\mathbf{x}_i - \bar{\mathbf{x}}_i\right)^*\right] + \sum_{i\in\mathbf{B}_k\setminus\{k\}} E\left[\mathbf{\Phi}_{j,i}\mathbf{\Phi}_{j,i}^*\mathbf{W}_{j,i}\mathbf{W}_{j,i}^*\right] E\left[\mathbf{x}_i\odot\mathbf{x}_i^*\right] \qquad (5.5)$$

$E\left[\left(x-\bar{x}\right)^2\right] = P\left(x=1\right)\left(1-\bar{x}\right)^2 + P\left(x=-1\right)\left(-1-\bar{x}\right)^2$ and $E\left[x^2\right]=1$ for $x\in\mathbf{x}_i$.

Besides, $P\left(x=1\right)$ and $P\left(x=-1\right)$ can be regarded as $1/2$ if no prior probability is given.

At other outer iterations:

After executing the previous outer iteration, we obtain $P\left(x=1\right)$ and $P\left(x=-1\right)$ from the last drawing in Gibbs sampling and have the estimated channel information $\hat{\mathbf{W}}$ from the output of the EM detector. (5.2) shows the ICI-reduced signals in this case and σ_{RIP}^2 is provided by

$$\sigma_{RIP}^2 = \sum_{i \in \mathbf{B}_k \setminus \{k\}} \mathbf{M}_{j,i} \mathbf{M}_{j,i}^* E\left[\left(\mathbf{x}_i - \bar{\mathbf{x}}_i \right) \odot \left(\mathbf{x}_i - \bar{\mathbf{x}}_i \right)^* \right] + \sum_{i \in \mathbf{B}_k \setminus \{k\}} E\left[\mathbf{\Phi}_{j,i} \mathbf{\Phi}_{j,i}^* \mathbf{W}_{j,i} \mathbf{W}_{j,i}^* \right] E\left[\left(\mathbf{x}_i - \bar{\mathbf{x}}_i \right) \odot \left(\mathbf{x}_i - \bar{\mathbf{x}}_i \right)^* \right] \quad (5.6)$$

where $E\left[\mathbf{\Phi}_{j,i} \mathbf{\Phi}_{j,i}^* \mathbf{W}_{j,i} \mathbf{W}_{j,i}^* \right]$ can be defined by μ and approximated by the percentage of spreading energy of one subcarrier through the computer simulation. And (5.6) is rewritten as

$$\sigma_{RIP}^2 = \sum_{i \in \mathbf{B}_k \setminus \{k\}} \mathbf{M}_{j,i} \mathbf{M}_{j,i}^* E\left[\left(\mathbf{x}_i - \bar{\mathbf{x}}_i \right) \odot \left(\mathbf{x}_i - \bar{\mathbf{x}}_i \right)^* \right] + \sum_{i \in \mathbf{B}_k \setminus \{k\}} \mu \cdot E\left[\left(\mathbf{x}_i - \bar{\mathbf{x}}_i \right) \odot \left(\mathbf{x}_i - \bar{\mathbf{x}}_i \right)^* \right]$$

A subcarrier spreads the energy over the neighboring subcarriers, and the spreading energy becomes smaller when the distance is long between two subcarriers. The percentages of spreading energy resulted from one subcarrier are generated according as the normalized maximized Doppler frequencies are 0.05 and 0.1, and depicted in Fig.5.1 and Fig.5.2 respectively. As shown in these two figures, it is found that the percentages are very much alike. Varying with the accuracy of the detected data, σ_{RIP}^2 is used to be the adaptive ICI power for the a posteriori probabilities in Gibbs sampling.

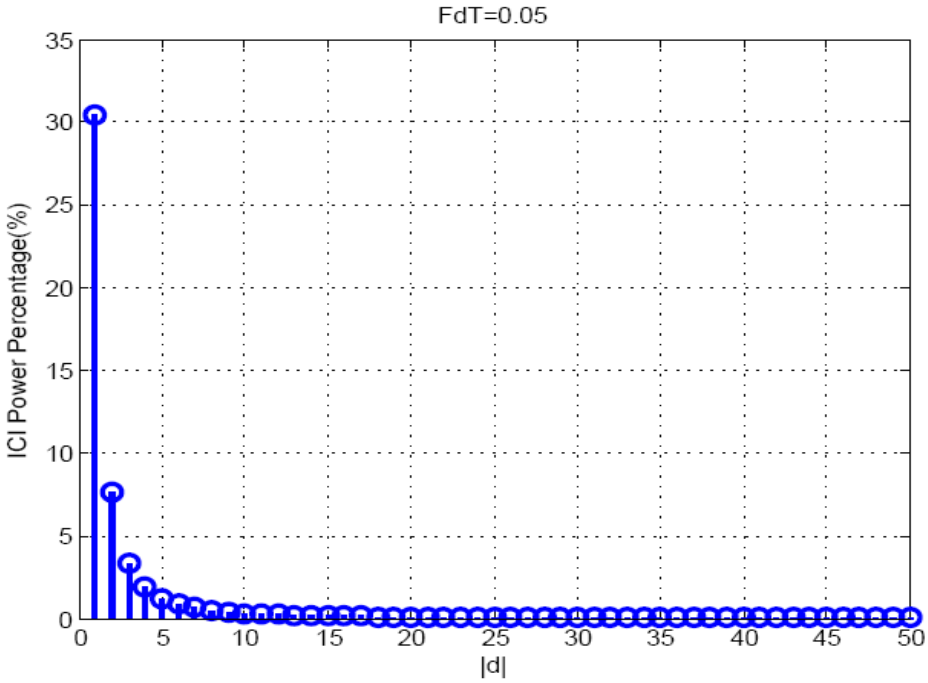


Fig. 5.1 The ICI power percentage for the normalized MDF=0.05.

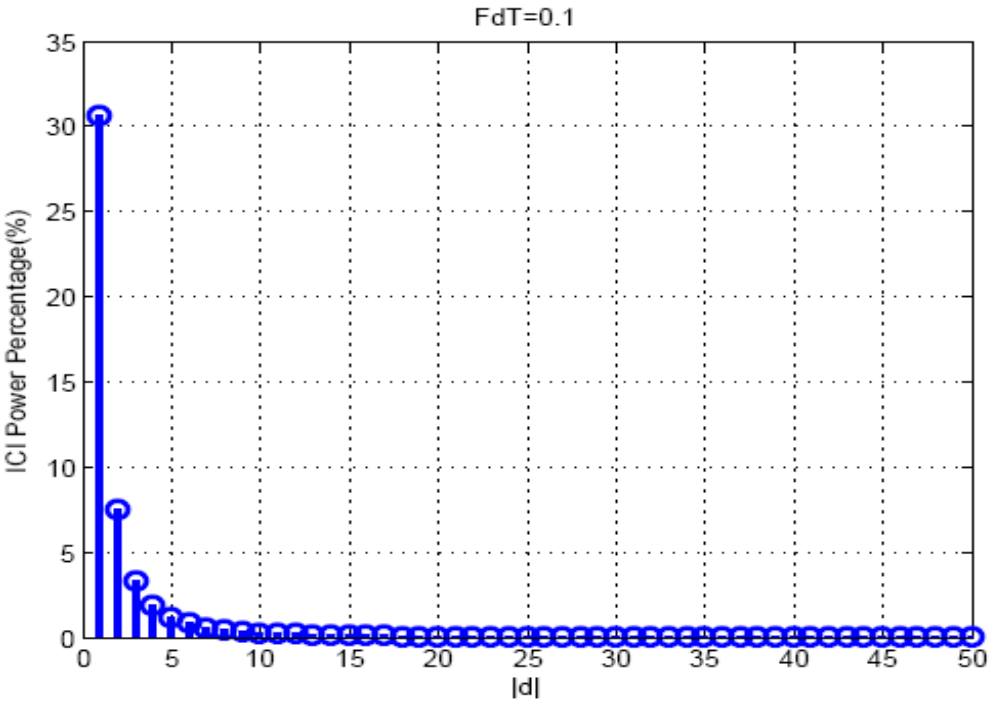


Fig. 5.2 The ICI power percentage for the normalized MDF=0.1.

Chapter 6 Computer Simulation

6.1 System Parameters

Results of computer simulation in this section demonstrate the performance of the ML-EM receiver. Based on the parameters defined in the 802.16e OFDM standard [12], we know that the system occupies a bandwidth of 5*MHz* and the carrier frequency is 2.3*GHz*. The entire bandwidth is divided into 256 sub-bands for N = 256 subcarriers among which J = 8 subcarriers carry the pilot tones, N_S = 192 subcarriers transmit data and the remaining 56 subcarriers are used as virtual subcarriers. Those pilot subcarriers transmit the pilot tones adopting the BPSK modulation scheme and each having the same power as the data carried by a data subcarrier. The length of guard interval is N_G = 64 (i.e. one quarter of 256 for the cyclic prefix). Each OFDM frame consists of N_F = 40 OFDM data symbols and one OFDM symbol used for the CP-added preamble. Besides, the parameter of Q is set to 4 through the observation from experimental trials. The numbers of the EM iteration and the outer iteration are selected as N_{EM} = 2 and N_{OL} = 4.

A two-path channel and an International Telecommunication Union (ITU) Veh-A channel are simulated with the path delays uniformly distributed from 0 to 50 sample periods. The relative path power profiles are set as 0, 0 (*dB*) for the two-path channel and set as 0, -1, -9, -10, -15, -20 (*dB*) for the ITU Veh-A channel, where the fading channel can be generated with Jake's Model by setting the normalized MDF equal to 0.05 and 0.1. And it is assumed

that both symbol synchronization and carrier synchronization are perfect and the receivers have some of the statistical information like noise power, power delay profiles and Doppler frequency. In addition, the parameter E_b/N_0 represents a ratio of received bit energy to the power spectral density of noise.

6.2 Simulation Results

The following three cases are used in the simulation for comparison.

(A) **CSI and data known:** The curves with ideal CSI initialization and perfect initial data can be regarded as a performance lower bound.

(B) **CSI known:** This kind of curves is generated by using ideal CSI initialization and initial data given from the one tap equalizer.

(C) **CSI est:** The curves labeled as "CSI est" are made by setting initial CSI estimated and initial data given from the one tap equalizer.

The ways to produce initial data make difference between (A) and (B), and the modes of CSI initialization make (B) perform better than (C) if other conditions remain the same. Most of the figures come from the simulation in the two-path channel, while Fig.6.9 and Fig.6.10 are given by the simulation in both the two-path channel and the ITU Veh-A channels.

The group size of the ML-EM receiver must be decided first of all, and the BER

performance curves are compared with each other in the case of “CSI and data known”. As depicted in Fig.6.1 and Fig.6.2 for the normalized MDF=0.05 and 0.1 respectively, joint detection of more subcarriers improves the performance, so the receiver with group sizes of 1 and 2 are worse than with group sizes of 4 and 8 which have nearly identical performance, and we choose the group size of 4 for the reason that a smaller group size takes less computational time. Next, the BER performance makes progress by applying the residual ICI power described in chapter 5 to the Gibbs sampling. Based on the case of “CSI known”, it can be shown in both Fig.6.3 and Fig.6.4 for comparison between the performance of the ML-EM receiver with ICI power update and without ICI power update, and then the former is proved to be better. We subsequently develop the receiver combined with ICI power update.

Fig.6.5 and Fig.6.6 demonstrate the BER performance of the ML-EM receiver with ICI power update in the cases of “CSI known” and “CSI est”. It is observed that the curves with perfect CSI initialization perform better than the curves with CSI initialization estimated by zero forcing criteria and the performance improves as the outer loop iterates more times. Moreover, Fig.6.7 and Fig.6.8 show that the ML-EM receiver with CE refinement has better performance than without CE refinement in the case of “CSI est”. And the “CSI est” curve with CE refinement is quite close to the “CSI known” curve, which means the receiver with CE refinement decrease the gap between the two modes of CSI initialization. CE refinement can obviously make the performance better for a channel with more paths, i.e. the ITU Veh-A

channel, and the BER performance is illustrated in Fig.6.9 and Fig.6.10.

Fig.6.11 and Fig.6.12 demonstrate the BER performance of the ML-EM receiver with the group size of 4 and CE refinement. Compared with the “CSI and data known” curve for the normalized MDF=0.05, the “CSI and data known” curve for the normalized MDF=0.1 has lower BER at the same E_b/N_0, because time-variant channels introduce more diversity gains for higher speed when the initial CSI and data are both perfect. And it is seen from Fig.6.11 that the three curves in different cases are rather close; however, there is a gap between the performance lower bound and the two curves else due to the error propagation effect observed from Fig.6.12. Finally, the number of samples required for Gibbs sampling affects the BER performance shown in Fig.6.13 and Fig.6.14, and thus we find that the receiver with more samples attain better performance, and the improvement is gradually saturated as the number of samples increases.

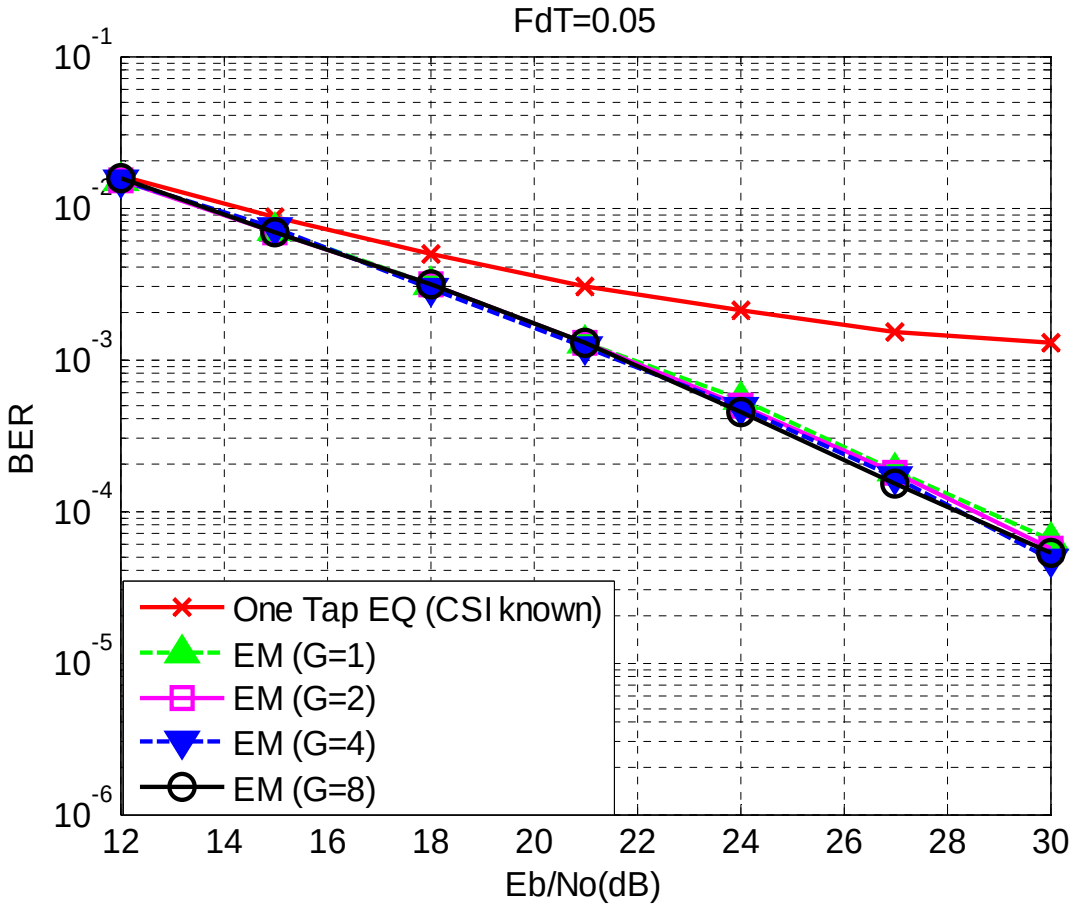


Fig. 6.1 BER performance of the ML-EM receiver with different group sizes for the normalized MDF=0.05.

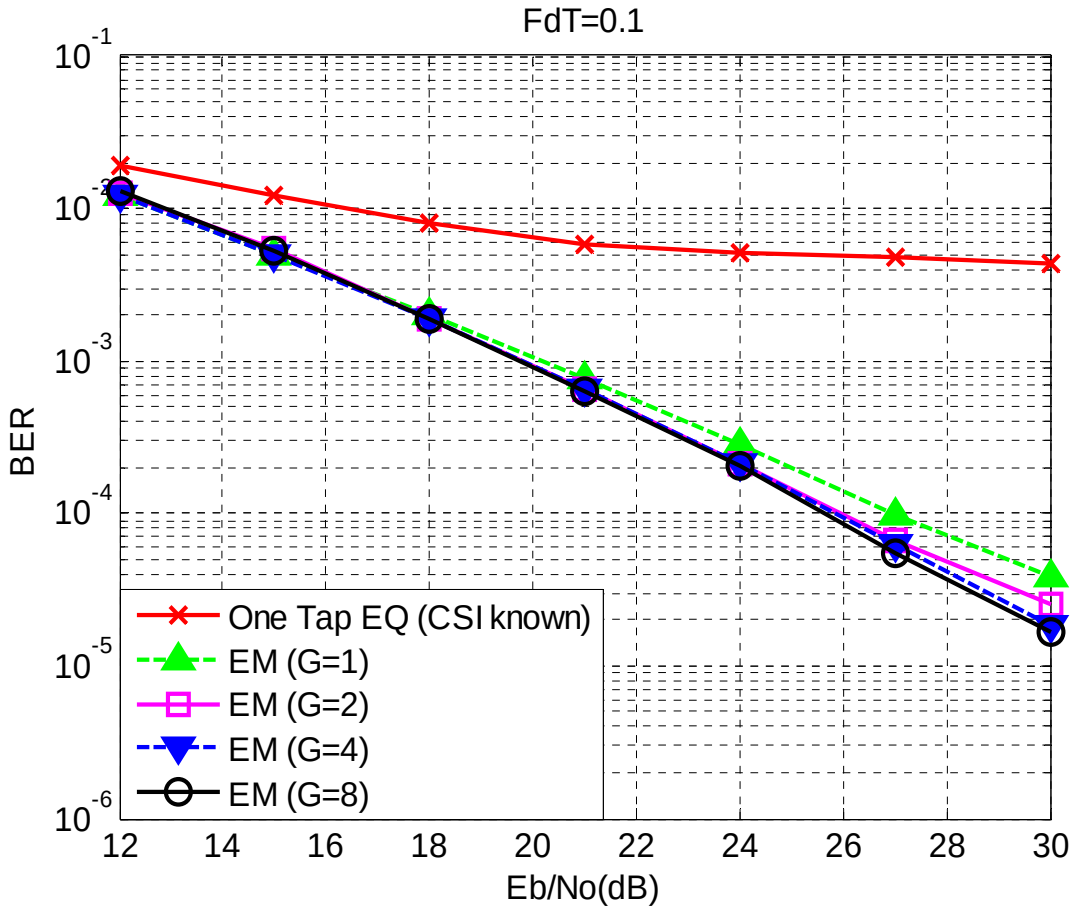


Fig. 6.2 BER performance of the ML-EM receiver with different group sizes for the normalized MDF=0.1.

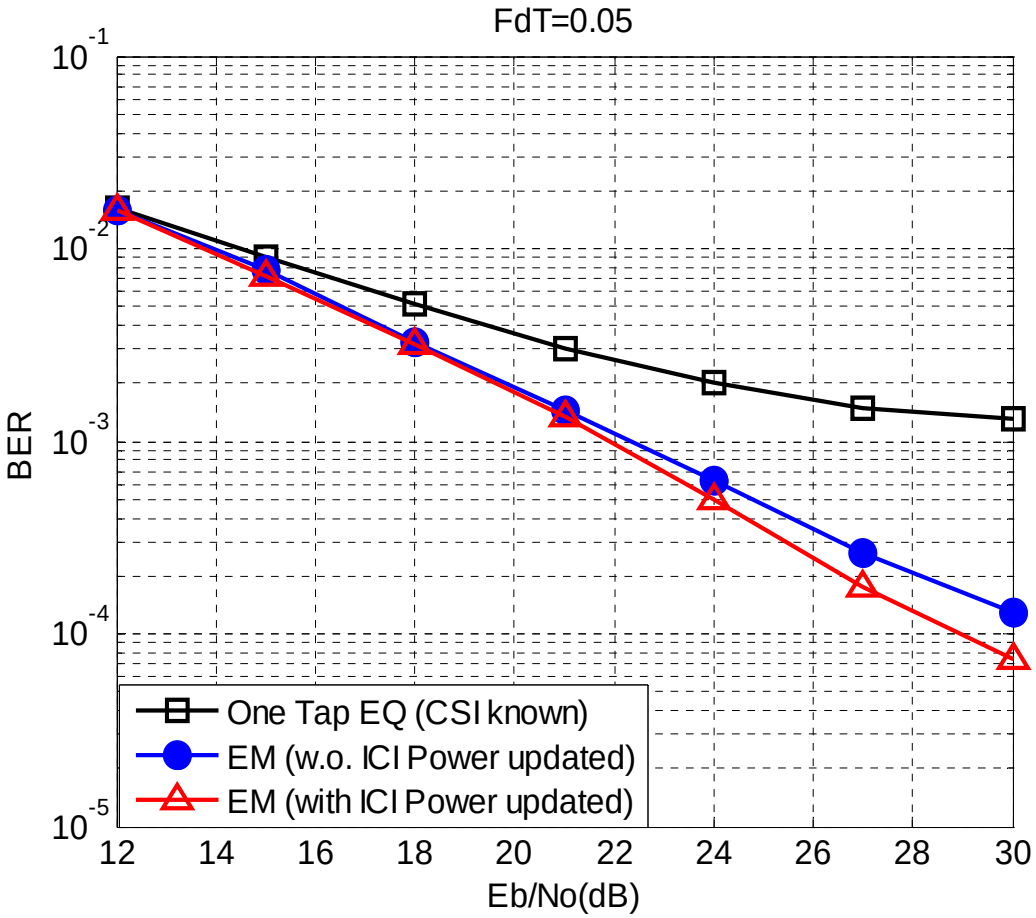


Fig.6.3 BER performance of the ML-EM receiver with/w.o. ICI power update for the normalized MDF=0.05.

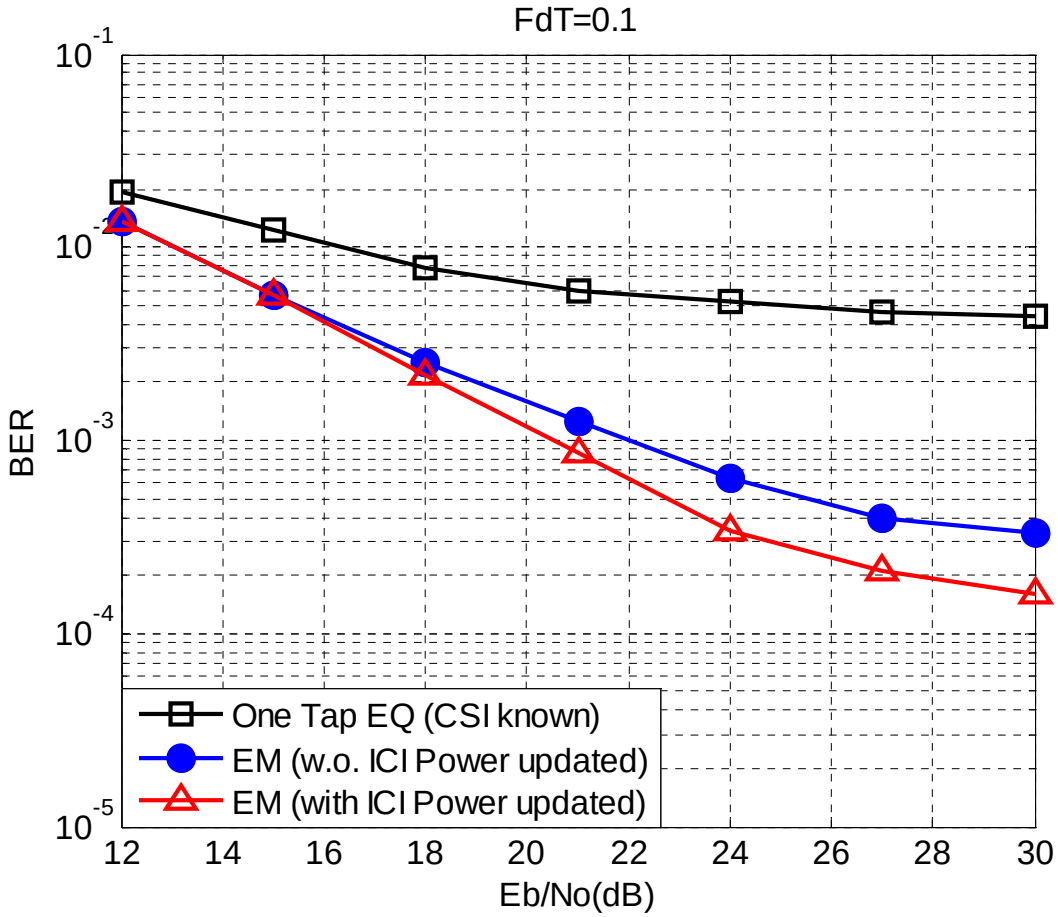


Fig.6.4 BER performance of the ML-EM receiver with/w.o. ICI power update for the normalized MDF=0.1.

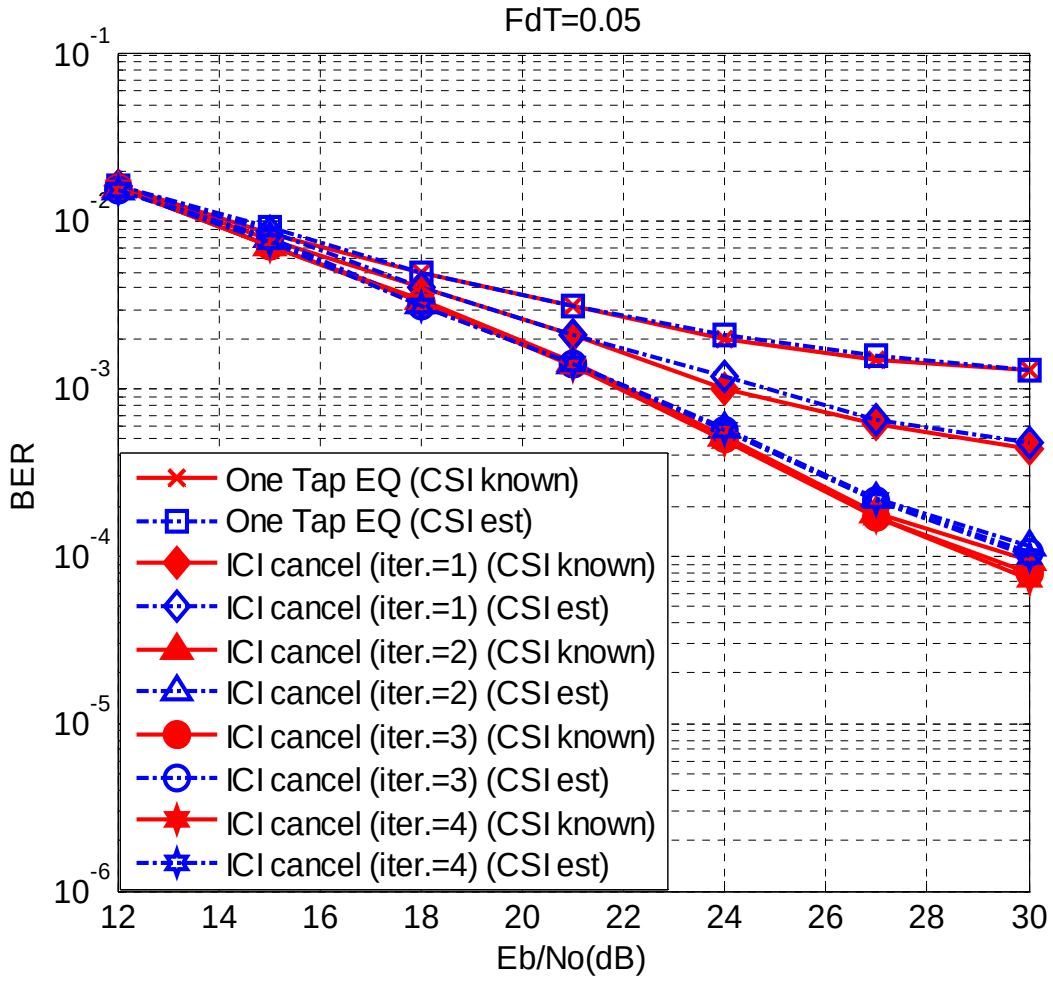


Fig.6.5 BER performance of the ML-EM receiver for the normalized MDF=0.05.

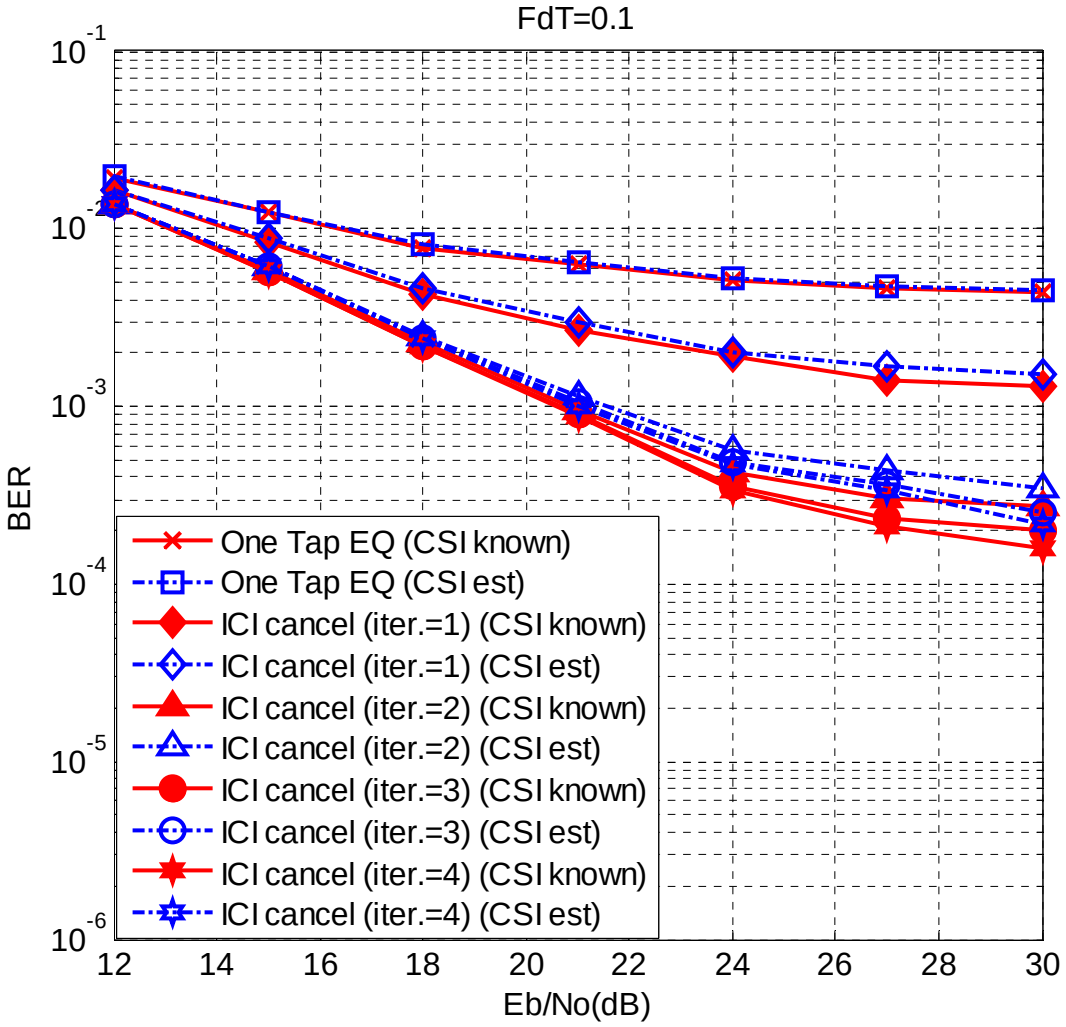


Fig.6.6 BER performance of the ML-EM receiver for the normalized MDF=0.1.

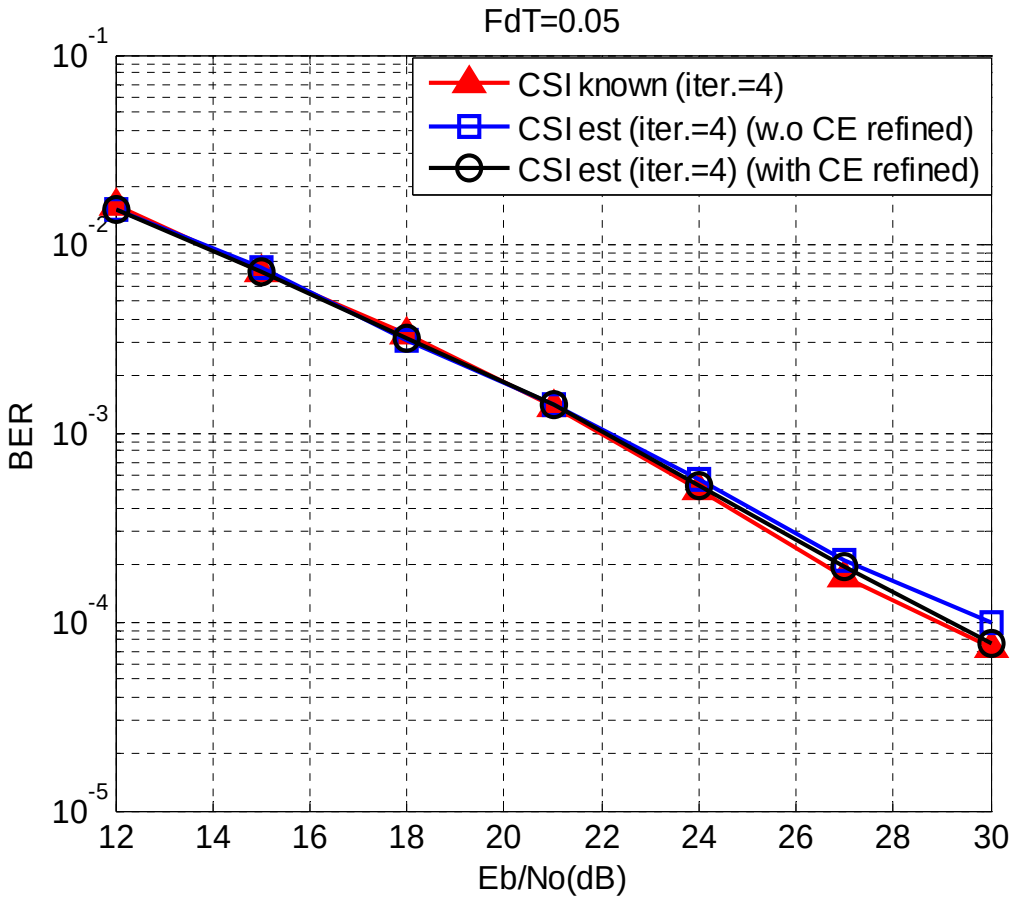


Fig.6.7 BER performance of the ML-EM receiver with/w.o. CE refinement for the normalized MDF=0.05.

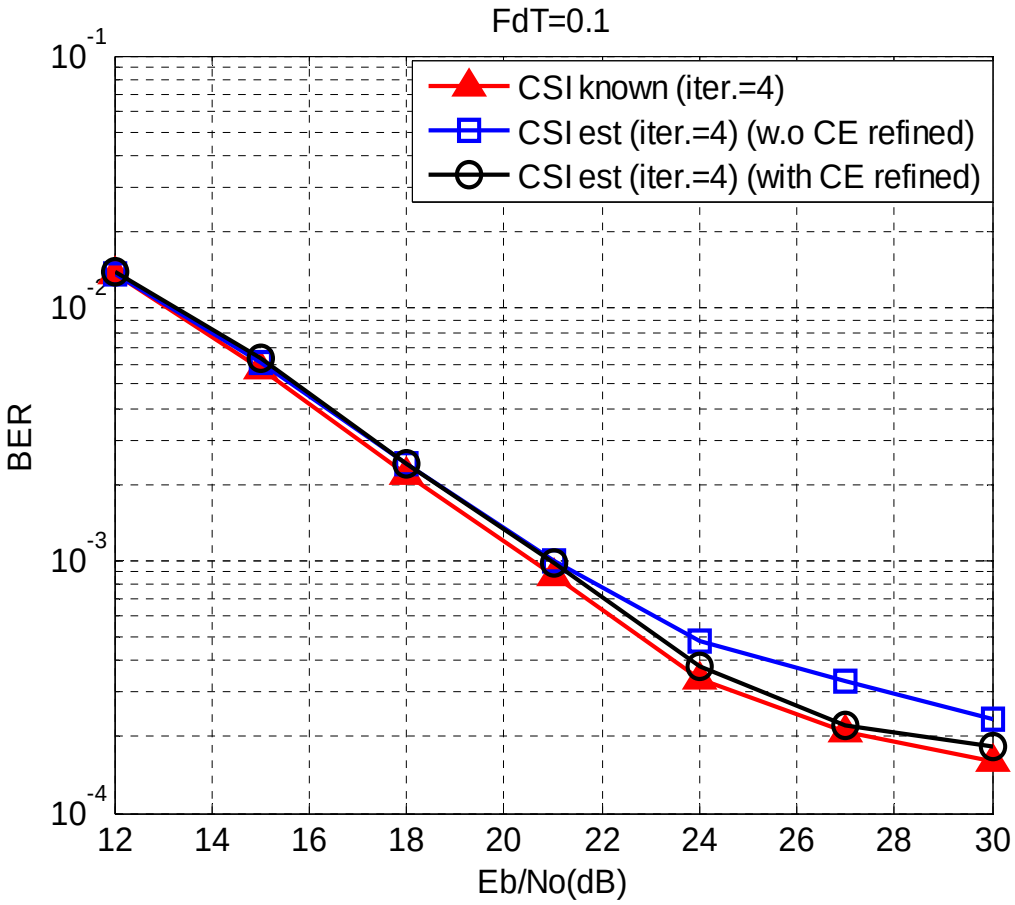


Fig.6.8 BER performance of the ML-EM receiver with/w.o. CE refinement for the normalized MDF=0.1.

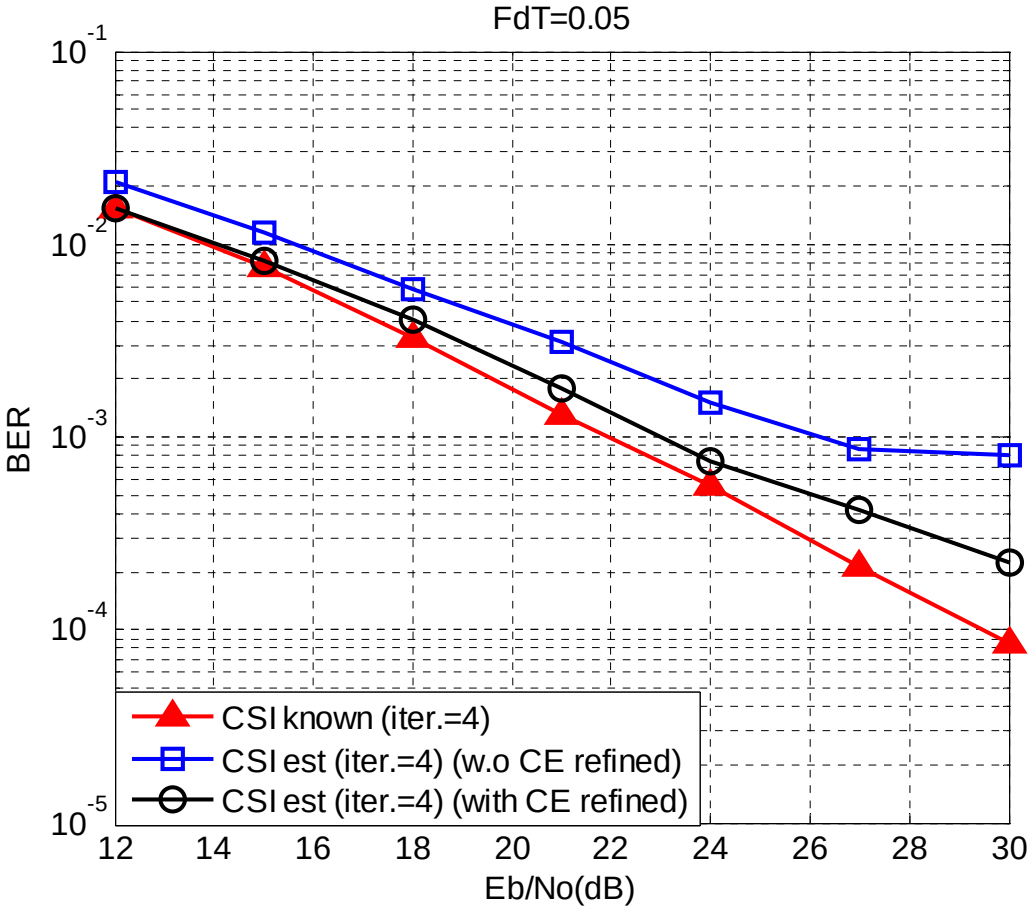


Fig.6.9 BER performance of the ML-EM receiver with different cases in the ITU Veh-A channel for the normalized MDF=0.05.

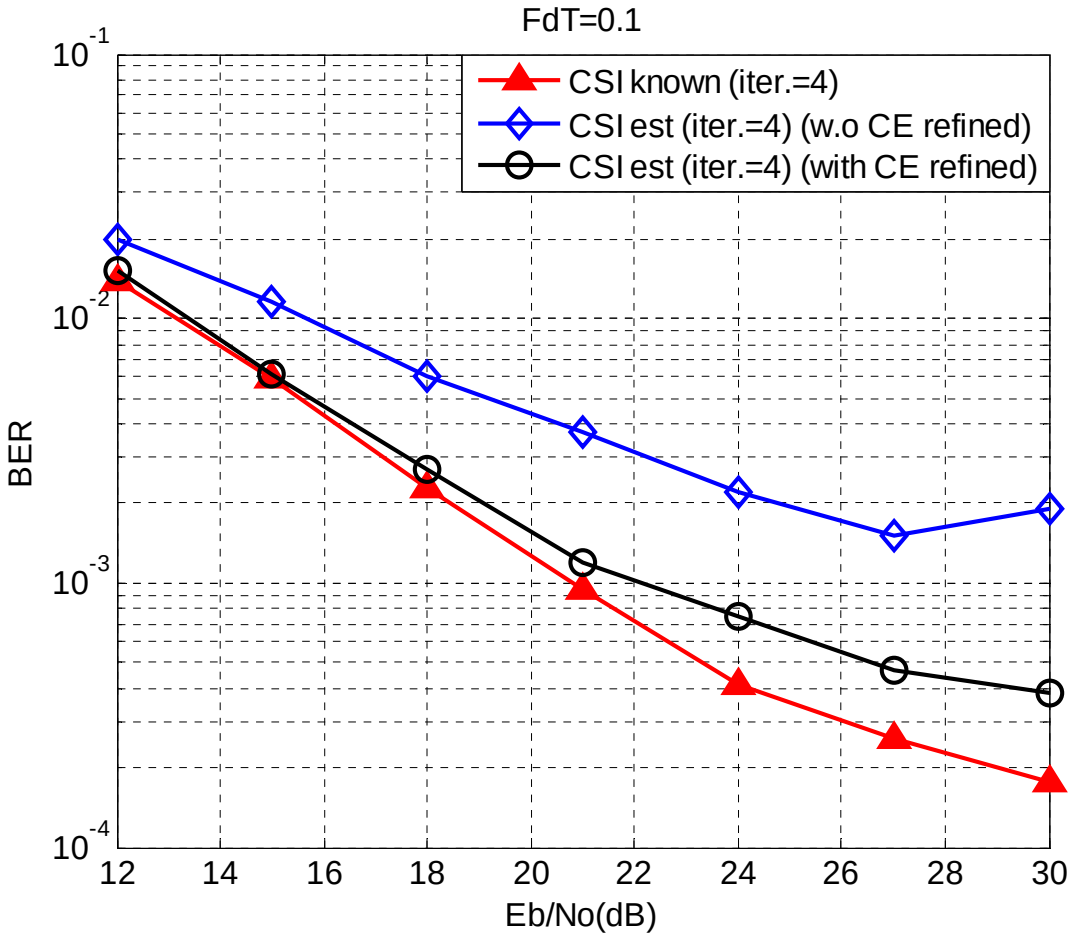


Fig.6.10 BER performance of the ML-EM receiver with different cases in the ITU Veh-A channel for the normalized MDF=0.1.

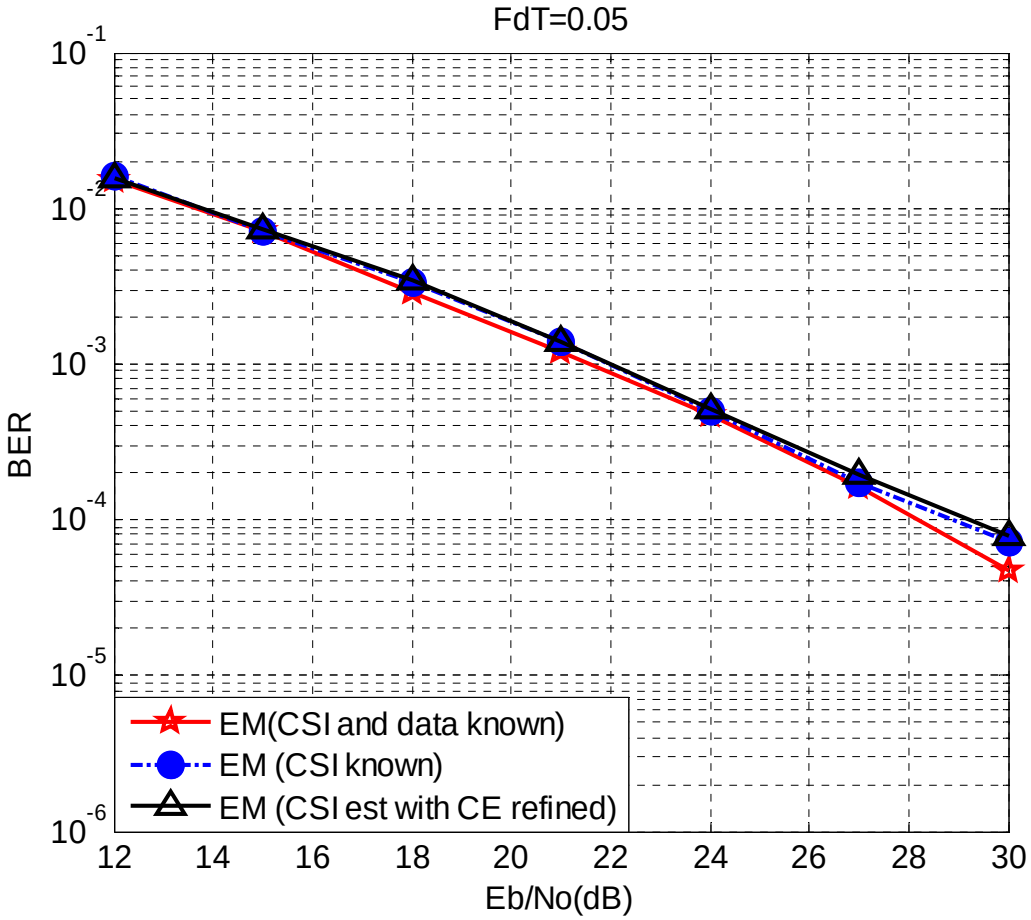


Fig.6.11 BER performance of the ML-EM receiver in different cases for the normalized MDF=0.05.

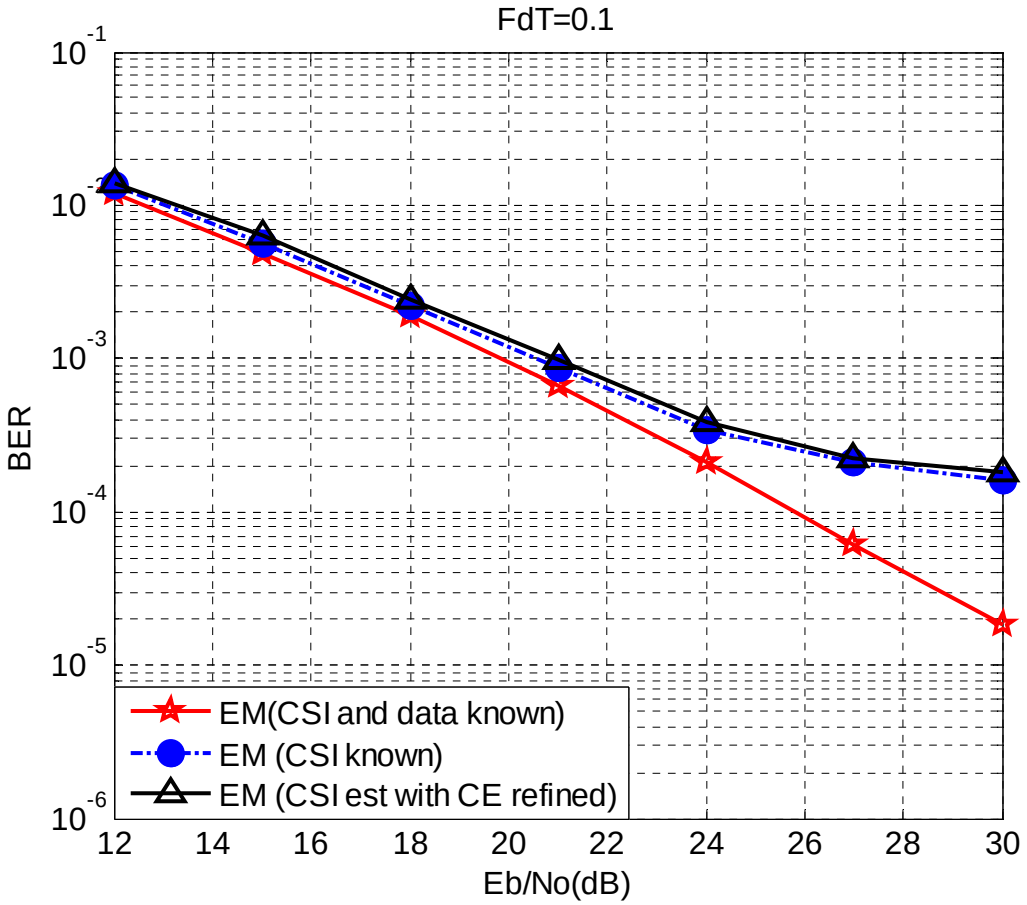


Fig.6.12 BER performance of the ML-EM receiver in different cases for the normalized MDF=0.1.

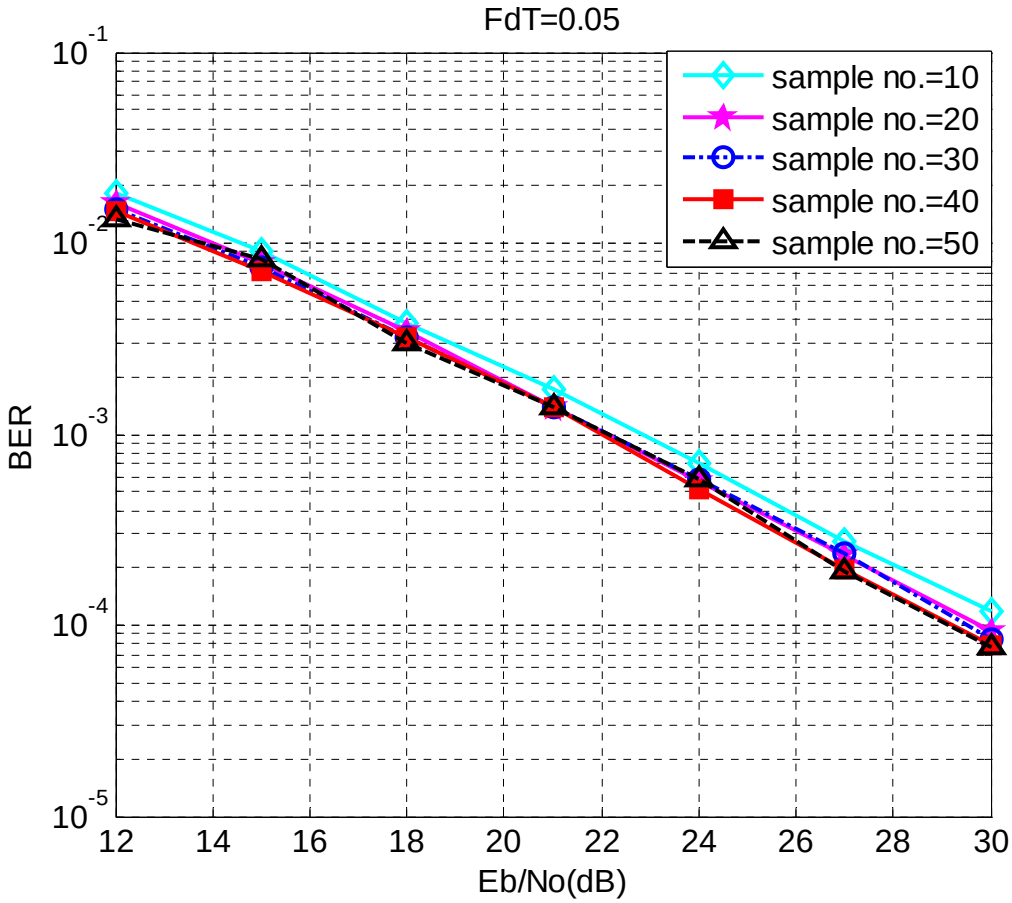


Fig.6.13 BER performance of the ML-EM receiver with various numbers of samples for the normalized MDF=0.05.

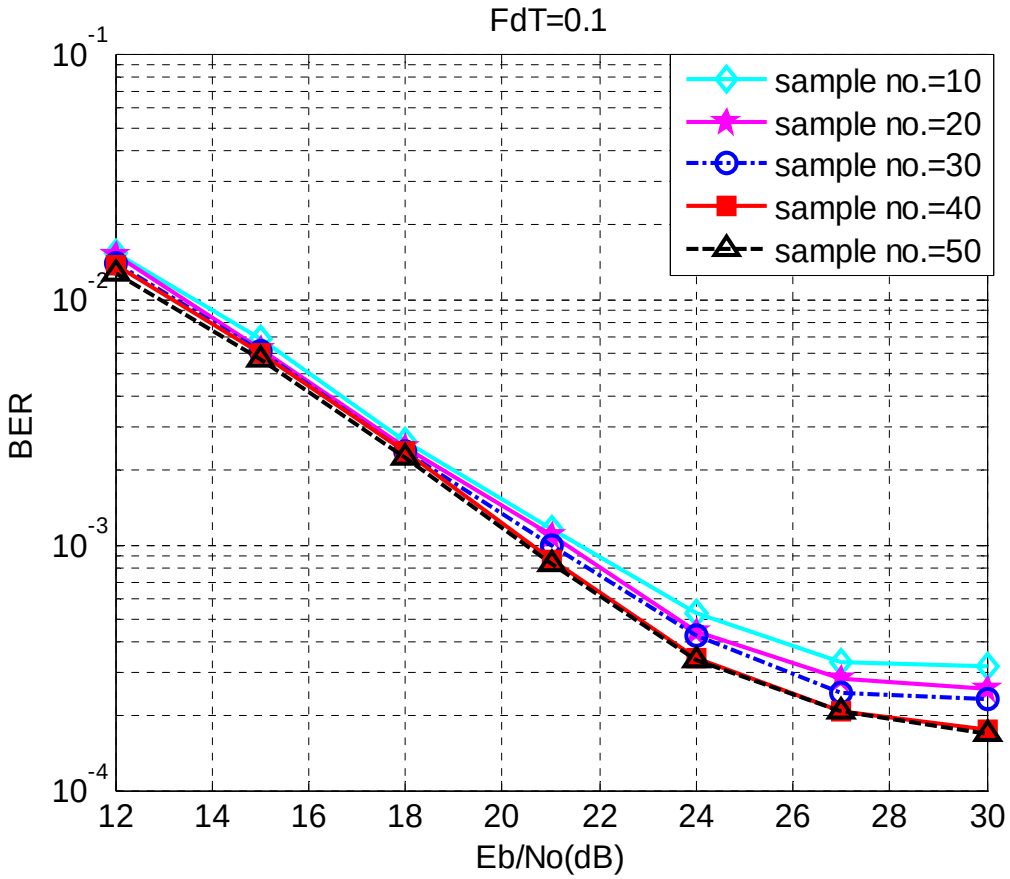


Fig.6.14 BER performance of the ML-EM receiver with various numbers of samples for the normalized MDF=0.1.

Chapter 7 Conclusions

In this paper, we investigate an EM-based iterative receiver under the ML criterion for OFDM systems in doubly selective channels. The receiver consists of an initialization unit, an ICI canceller, a CE update unit and an ML-EM detector. First, the initial setting of CSI and data can be executed through the use of the MMSE-based CE method and a decision-directed approach with the one-tap equalizer. Next, the ML-EM algorithm is proposed for channel variable estimation by using the samples given from Gibbs sampling. Incorporated with the group-wise processing, the ICI cancellation is developed to reduce the computational complexity and to exploit the time diversity inherent in time-variant channels. And the CE update unit provides better CSI to improve the performance at high E_b/N_0 especially. Simulation results indicate that the ML-EM receiver significantly outperforms the one-tap equalizer.

References

[1] T. S. Rappaport : *Wireless Communications – Principles and Practice*, 2nd Ed., Prentice-Hall, 2002.

[2] Juha Heiskala and John Terry : *OFDM Wireless LANs* : A Theoretical and Practical Guide, 1st Ed, Sams, 2003.

[3] Y.-S. Choi, P. J. Voltz, and F.A Cassara, ”On channel estimation and detection for multicarrier signals in fast and selective Rayleigh channels,” *IEEE Trans. On Wireless Communications.*, vol.49, pp. 1375-1387, Aug. 2001.

[4] J. Armstrong, P. M. Grant, and G. Povey, ”Polynomial cancellation coding of OFDM to reduce intercarrier interference due to Doppler spread,” in *Proc. IEEE Golbal Telecommunications Conf.*, vol.5, pp.2771-2776, 1998.

[5] A. Seyedi, and G.J. Saulnier, ”General self-cancellation scheme for mitigation of ICI in OFDM systems,” *IEEE Communications Conf.*, vol.5, pp. 2653-2657, June 2004.

[6] Y. Mostofi and D. C. Cox, ”ICI Mitigation for Pilot-Aided OFDM Mobile Systems,” *IEEE Trans. On Communications*, vol.4, No.2, pp. 765-774, Mar. 2005.

[7] Chorong-Ren Sheu, Ming-Chien Tseng, Ching-Yung Chen and Hsin-Piao Lin , ”A Low-Complexity Concatenated ICI Cancellation Scheme for High-Mobility OFDM Systems,” *IEEE Communications Society*, pp. 1525-3511, July 2007.

[8] Hongmei Wang, Xiang Chen, Shidong Zhou and Yan Yao, ”A Low-Complexity ICI cancellation Scheme in Frequency Domain for OFDM in Time-varying Multipath Channels,” *IEEE*, 2005, pp. 1234-1237.

[9] Xiaodong Cai and Georgios B. Giannakis, ”Bounding Performance and Suppressing

Intercarrier Interference in Wireless Mobile OFDM," *IEEE Trans. on Communications*, vol.51, No.12, pp. 2047-2056, Dec. 2003.

[10] Meir Feder and Ehud Weinstein, "Parameter Estimation of Superimposed Signals Using the EM Algorithm," *IEEE Tran.*, vol.36, No.4, pp. 477-489, April 1998.

[11] Xiaoqiang Ma, Hisashi Kobayashi and Stuart C. Schwartz, "EM-Based Channel Estimation Algorithms for OFDM," *Euraship Journal on Applied Signal Processing*, pp. 1460-1477, Dec. 2004.

[12] IEEE Std 802.16e-2005 and IEEE 802.16-2004/Cor1-2005, "Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems," IEEE-SA Standards Boards, Tech. Rep., 2006.

[13] P. Schniter, "Low-Complexity Equalization of OFDM in Doubly Selective Channels," *IEEE Trans. Signal Processing*, vol. 52, No. 4, pp. 1002-1011, Apr. 2004.

[14] Xiaoqiang Ma, Hisashi Kobayashi and Stuart C. Schwartz, "EM-Based Channel Estimation Algorithm for OFDM," *Euraship Journal on Applied Signal Processing,* , pp. 1460-1477, Dec. 2004.

[15] O. Edfors, M. Sandell, J.van de Beek, S. K. Wilson, and P. O. Borjesson, "Analysis of DFT-based Channel Estimators for OFDM," *Wireless Pers. Commun.*, vol. 12, pp. 55-70, Jan. 2000.

[16] Y. Li and L. J Cimini, Jr., "Bounds on the interchannel interference of OFDM in time-varying impairments," *IEEE Tran. Commun.*, vol. 49, pp. 401-404, Mar. 2001.

[17] 陳文娟,"雙選擇性通道中基於期望值最大化之正交分頻多工系統接收機的設計與模擬," 交通大學電信工程學系碩士論文.

[18] Xiaodong Wang and Rong Chen, "Adaptive Bayesian Multiuser Detection for

Synchronous CDMA with Gaussian and Impulsive Noise,” IEEE Tran. Signal Process. , vol. 47, No.3, July 2000.

[19] Behrouz Farhang-Boroujeny, Haidong Zhu and Zhenning Shi, “Markov Chain Monte Carlo Algorithm for CDMA and MIMO Communication Systems,” IEEE Tran. Signal Process. , vol. 54, No. 5, May 2006.

[20] Christian P. Robert and George Casella: *Monte Carlo Statistical Methods*, New York: Springer-Verlag, 1999.

[21] Steven M. Kay: *Fundamentals of Statistical Signal Processing Estimation Theory*, Prentice Hall, 1998.