การตรวจจับผู้ใช้หลายรายโดยวิธีการหาค่าต่ำสุดของเทอร์โบคูลแบ็ค-ลีเบลอร์สำหรับระบบซีดีเอ็มเอ

นาย ทัศนัย พลอยสุวรรณ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรดุษฎีบัณฑิต
สาขาวิชาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้า
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2551

MINIMUM KULLBACK-LEIBLER TURBO MULTIUSER DETECTOR FOR CDMA SYSTEM

Mr. Tuchsanai Ploysuwan

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy Program in Electrical Engineering
Department of Electrical Engineering
Faculty of Engineering
Chulalongkorn University
Academic Year 2008

Thesis Title	MINIMUM KULLBACK-LEIBLER TURBO MULTIUSER DETECTOR FOR CDMA SYSTEM
By	Mr. Tuchsanai Ploysuwan
Field of Study	Electrical Engineering
Advisor	Associate Professor Prasit Teekaput, Ph.D.
Co-Advisor	Sawat Tantiphanwadi, Ph.D.

Accepted by the Faculty of Engineering, Chulalongkorn University in Partial Fulfillment of the Requirements for the Doctoral Degree

....Dean of the Faculty of Engineering
(Associate Professor Boonsom Lerdhirunwong, Dr.Ing)

THESIS COMMITTEE

.... Chairman
(Professor Prasit Prapinmongkolkarn, D.Eng)

.... Advisor
(Associate Professor Prasit Teekaput, Ph.D.)

.... Co-Advisor
(Sawat Tantiphanwadi, Ph.D.)

.... Examiner
(Associate Professor Lunchakorn Wuttisittikulkij, Ph.D.)

.... Examiner
(Assistant Professor Nisachon Tangsangiumvisai, Ph.D.)

ทัศนัย พลอยสุวรรณ : การตรวจจับผู้ใช้หลายรายโดยวิธีการหาค่าต่ำสุดของเทอร์โบคูลแบ็ค-ลีเบลอร์สำหรับระบบซีดีเอ็มเอ (MINIMUM KULLBACK-LEIBLER TURBO MULTIUSER DETECTOR FOR CDMA SYSTEM) อ. ที่ปรึกษาวิทยานิพนธ์หลัก: รศ. ดร.ประสิทธิ์ ทีฆพุฒิ, อ. ที่ปรึกษาวิทยานิพนธ์ร่วม ดร.สวัสดิ์ ตันติพันธุ์วดี, 57 หน้า

ในปัจจุบันนี้ การแทรกสอดจากการเข้าถึงหลายทาง และ การแทรกสอดข้ามสัญลักษณ์ เป็นอุปสรรค์ที่มีปัจจัยสำคัญในด้านของการสื่อสารในช่องสัญญาณคลื่นหลายวิถีซีดีเอ็มเอ ในอดีตที่ผ่านมามีงานวิจัยหลายชิ้นเกี่ยวข้องกับเครื่องตรวจวัดหลายผู้ใช้ได้ถูกนำเสนอโดยอาศัยหลักการของ เทคนิคการขจัดการแทรกสอดจากการเข้าถึงหลายทางที่เป็นเชิงเส้น และ ไม่เป็นเชิงเส้น ซึ่งอาศัยหลักการ ของการขจัดการแทรกสอดแบบอ่อน และ วิธีการเชิงเส้นของค่าเฉลี่ยกำลังสองของค่าผิดพลาดทันทีทันใด ซึ่งงานวิจัยส่วนใหญ่ของเครื่องตรวจวัดหลายผู้ใช้จำเป็นต้องอาศัยค่าความรู้ของลำดับแผ่ , ค่าเริ่มต้นของการประวิง และ คลื่นหลายวิถีซีดีเอ็มเอของผู้ใช้บริการที่สนใจ ดังนั้น ค่าความรู้พารามิเตอร์ของช่องสัญญาณมีความสำคัญอย่างยิ่งสำหรับสมรรถนะในการตรวจจับสัญญาณที่ถูกส่งออกมา

จุดมุ่งหมายหลักของวิทยานิพนธ์ฉบับนี้ต้องการนำเสนอ วิธีการแก้ปัญหาซึ่งก้าวล้ำหน้าแบบใหม่ ในปัญหาของ การทำงานร่วมกันของเครื่องตรวจวัดหลายผู้ใช้, การประมาณค่าช่องสัญญาณ และ การตรวจจับสัญลักษณ์ สำหรับระบบ ไดเรคทซีเควนซีดีเอ็มเอ การพิสูจน์การทำงานของอัลกอริทึมอาศัยหลักการของวิธีการผันแปรของเบย์ และ ค่าต่ำสุดของวิธีการ คูลแบ็ค-ลีเบลอร์ สำหรับวิธีการค่าต่ำสุดของ คูลแบ็ค-ลีเบลอร์อาศัยการประมาณเชิงตัวเลข ของการกระจายความน่าจะเป็นแบบมีเงื่อนไขของพารามิเตอร์ซึ่งไม่ทราบค่า (ได้แก่ ข้อมูลสัญลักษณ์, ผลตอบสนองเชิงช่องสัญญาณของผู้ใช้แต่ละราย) จากเงื่อนไขการกระจายถดถอยและการคำนวณของการประมาณค่าเฉลี่ยแบบมีเงื่อนไข โดยวิธีการทำนำเสนอนั้นมีความคงทนสูงไม่เหมือนกับ วิธีการค่าคาดหวังสูงสุด ซึ่งต้องอาศัยค่าเริ่มต้นของเงื่อนไขและมีความคงทนต่อการเปลี่ยนแปลงของการสื่อสาร การนำเสนอ การตรวจจับผู้ใช้หลายรายของ ค่าต่ำสุดของ คูลแบ็ค-ลีเบลอร์ได้ใช้ ข้อมูลเข้าแบบอ่อน และ ข้อมูลออกแบบอ่อนซึ่งเหมาะสมสำหรับการทำงานแบบวนซ้ำในระบบการสื่อสารที่มีการเข้ารหัส

ผลการทดลอง ระบบสาธิตให้เห็นถึงคุณประโยชน์หลายประการของวิธีการนำเสนอที่เหนือกว่าของเครื่องตรวจวัดหลายผู้ใช้ซึ่งอาศัยวิธีการค่าต่ำสุดของคูลแบ็ค-ลีเบลอร์ ในช่องสัญญาณคลื่นหลายวิถีซีดีเอ็มเอซึ่งไม่ทราบค่าของค่าเริ่มต้นของการประวิง และ คลื่นหลายวิถีซีดีเอ็มเอของผู้ใช้บริการที่สนใจ สัญลักษณ์นำช่วยจำนวนที่น้อยของผู้ใช้ในแต่ละรายถูกใช้ในการหาค่าเริ่มต้น จากผลการจำลองระบบยืนยันประสิทธิภาพของ เครื่องตรวจวัดหลายผู้ใช้ซึ่งอาศัยวิธีการค่าต่ำสุดของ คูลแบ็ค-ลีเบลอร์ และแสดงให้เห็นถึงศักยภาพที่เหนือกว่าการออกแบบอื่นๆ

ภาควิชา	วิศวกรรมไฟฟ้า	ลายมือชื่อนิสิต	ทัศนัย
สาขาวิชา	วิศวกรรมไฟฟ้า	ลายมือชื่ออาจารย์ที่ปรึกษา	
ปีการศึกษา	2551	ลายมือชื่ออาจารย์ที่ปรึกษาร่วม	Sawat

##4771854021 : MAJOR ELECTRICAL ENGINEERING
KEY WORD: MINIMUM KULLBACK-LEIBLER / CHANNEL ESTIMATION / ITERATIVE DECODING AND DEMODULATION / CDMA / TURBO CODE

TUCHSANAI PLOYSUWAN : MINIMUM KULLBACK-LEIBLER TURBO MULTIUSER DETECTOR FOR CDMA SYSTEM. ADVISOR: ASSOC. PROF. PRASIT TEEKAPUT, Ph.D., COADVISOR DR. SAWAT TANTIPHANWADI, 57 pp.

The presence of both multiple-access interference (MAI) and intersymbol interference (ISI) constitutes a major impediment to reliable communications in multipath code-division multiple-access (CDMA) channels. Many multiuser detectors in the past have been developed based on novel linear and nonlinear interference suppression technique , which make use of both soft interference cancellation and instantaneous linear minimum mean-square error filtering. All of them are required knowledge of the spreading sequences , the initial delays and multipath channels of the desired users. therefore, the knowledge of channel parameters are essential to achieve an efficient detection of the transmitted signal.

The main purpose of this dissertation is to propose an advance novel solution to the problem of joint multiuser detection, channel estimation and data detection for the uplink of a multiuser DS-CDMA system exploiting. The devised algorithm from the application of ***Variational Bayes Methods*** and ***Minimum Kullback-Leibler*** (MKL) techniques. ***Minimum Kullback-Leibler*** (MKL) methods allow to efficiently numerical conditional distributional probability and parameter approximation of all the unknown parameters (i.e., data symbols, channel response of each users) from their conditional posterior distributions and then to compute their estimates by condition mean estimator. In addition, they are insensitive, unlike the ***Expectation maximization*** (EM) technique, to the choice of initial conditions and consequently perform robustly in quickly changing communication scenarios. The proposed MKL Bayesian multiuser dectector, being soft-input soft-output in its nature, can be exploited for iterative processing in a coded system.

The simulation results have demonstrated relevant merit of the optimum ***Minimum Kullback-Leibler*** (MKL) multiuser detector over multipath code-division multiple-access (CDMA) channels in the system with the unknown the initial delays and multipath channels of the desired users. A few pilot symbols are used by any users for channel parameters at initial. The simulation results confirm performance properties of proposed ***Minimum Kullback-Leibler*** (MKL) multiuser detector and show the advantages of the novel design over others.

Department	Electrical Engineering	Student's Signature	tuchsanai
Field of Study	Electrical Engineering	Advisor's Signature	P. Teekaput
Academic Year	2008	Co-advisor's Signature	Sawat

Acknowledgements

I would like to express my sincere gratitude to my advisor, Dr.Prasit Teekaput and my co advisor Dr. Sawat Tantiphanwadi. His encouragement, support, and expert guidance have been invaluable during my time at the Telecommunication System Research Laboratory (TSRL) Chulalongkorn University. He has contributed both on a professional and on a personal level to my life as a graduate student.

I would also like to acknowledge Thailand Graduate Institute of Science and Technology (TGIST) scholarship and the financial support from the Cooperation Project between Department of Electrical Engineering and Private Sector for Research and Development, Chulalongkorn University, Thailand.

My thanks also to the members of my dissertation committees, Professor Prasit Prapinmongkolkarn, Associate Professor Prasit Teekaput, Dr. Sawat Tantiphanwadi, Assist. Prof. Dr. Nisachon Tangsangiumvisai, and Associate Professor Lunchakorn Wuttisittikulkij for their suggestions and invaluable comments.

I would also like to thank the laboratory and staffs of communication division for providing good facilities. Many thanks go to my colleagues for their help, suggestion, and enjoyment during my time at the laboratory.

Finally, my life has been constantly fulfilled by love and support of my family. I am extremely grateful to my parents, Bantung Ploysuwan, Kimhung Ploysuwan, for their love, caring and sacrifices, my father and mother for educating and preparing me for my future life since I was a kid, and my brother, Bantoon Ploysuwan, Prajak Ploysuwan and my best friend, Aprium Kanjananon, for their love, caring and sacrifices. This work is dedicated to them.

Table of Contents

List of Tables

List of Figures

Figure | Page

CHAPTER I

INTRODUCTION

OVER THE PAST decade, a significant amount of research has addressed various multiuser detection methods for interference suppression in code-division multipleaccess (CDMA) communication systems [1]. Various detection methods have been reported in literatures for uncoded CDMA system, such as maximum-likelihood (ML) detector, sphere decoding algorithm, zero-forcing detector (ZF), minimum-mean-square-error detector (MMSE) detector, and successive cancellation detector. Most of the previous work on multiuser detection focused on uncoded CDMA systems, i.e., on the demodulation of multiuser signals. Since in practice, most CDMA systems employ error control coding and interleaving, recent work in this area has addressed multiuser detection for coded CDMA systems. In [2], [9] and [21], it is shown that the optimal decoding scheme for an asynchronous convolutionally coded CDMA system combines the trellises of both the asynchronous multiuser detector and the convolutional code, resulting in a prohibitive computational complexity $O(2^{K\nu})$, where K is the number of users in the channel, and ν is the code constraint length. For coded systems, an iterative turbo multiuser receiver can approach the optimal performance with an affordable receiver complexity [2]. Like turbo codes, the iterative turbo multiuser detector module exchanges the extrinsic information of transmitted symbols between individual channel decoders and its module [2]– [13]. Several papers have recently been devoted to an iterative turbo processing of the knowledge channel-state information (CSI) [13]– [24]. A good knowledge of channel parameter is essential to achieve an efficient detection of the transmitted signal. An iterative ("Turbo") processing techniques have received considerable attention followed by the discovery of the powerful Turbo codes [2-3]. The so-called Turbo-principle can be successfully applied to many detection/decoding problems such as serial concatenated decoding, equalization, coded modulation, multiuser detection and joint source and channel decoding [9]. In particular, a Turbo equalization scheme is proposed in [4] for convolutionally coded digital transmission over intersymbol interference channel. More recently, in [11] an optimal iterative multiuser detector for synchronous coded CDMA system is derived, based on iterative techniques for cross-entropy minimization. A practical suboptimal implementation is also presented. The computational complexity of this method, however, is $O(2^K + 2^\nu)$, which is still prohibitive for channels with medium to large number of users. A similar work has also appeared in [16].

Among of iterative turbo multiuser detectors, there are two well known difference extrinsic information calculation. First, linear minimum mean square error (LMMSE) and soft interference cancellation computes the extrinsic information feeding to individual channel decoders by iterative manner [13], [17]. Second, the Sequential Monte Carlo (SMC) algorithm (also referred to the particle filtering method) computes the extrinsic information feeding to individual channel decoders [18]– [23]. The sequential Monte Carlo methodology original emerged in the field of statistics and engineering has provided a promising new paradigm for design of signal processing algorithms with performance approaching the theoretical optimum for fast and reliable communication in highly severe wireless environments. All of the SMC techniques are aimed at building a recursive Bayesian filter, which estimates the extrinsic probability density function (pdf) based on Monte Carlo simulations. The high computational complexity of the optimal multiuser detectors (which is exponential in terms of the number of users in the channel) has motivated the study of a number of low-complexity suboptimal multiuser detectors. These low-complexity methods fall largely into two categories: linear detectors and nonlinear detectors. A linear detector is comprised of a linear filter applied to the received signal, followed by a scalar quantizer. The nonlinear detectors are based primarily on various techniques for successive cancellation of interference. In [7], some low-complexity receivers which perform multiuser symbol detection and decoding either separately or jointly are studied.

The theme of this dissertation is on the design of a iterative multiuser receiver for an uplink asynchronous coded CDMA system employing spreading sequences. It is assumed that the receiver has only the knowledge of the spreading sequences and the initial delays of the desired users within the cell. The multipath channels are unknown to the receiver. A few pilot symbols are used by any users. Some recentworks have addressed channel estimation in long CDMA systems [2], [3]. In these approaches, channel parameters are first estimated and receivers are then constructed based on the estimated channels. This is suboptimal due to the separation of channel estimation and data detection (as opposed to joint estimation of both channels and data). Moreover, these methods are primarily targeted at uncoded systems and they do not attempt to exploit the signal structures induced by channel coding existing in most communication systems. On the other hand, iterative processing has recently attracted vast attention. In [4]–[6], turbo multiuser detection schemes for coded CDMA systems are developed, which iterate between multiuser detection and channel decoding to successively improve the receiver performance. In these works, the user channels are assumed perfectly known at the receiver. In this dissertation, we address the problem of turbo multiuser detection in unknown multipath channels for asynchronous coded CDMA systems employing spreading sequences. A novel Bayesian multiuser detector is proposed, which computes the MAP estimates of the channel coded multiuser symbols, that are encoded before being sent to the channel. This technique is based on the Gibbs sampler [7], a Markov chain Monte Carlo (MCMC) technique for Bayesian computation. Although originated in the field of statistics, the Gibbs sampler and Minimum Kullback-Leibler (MKL) have recently been investigated for the optimal receiver design in various communication systems [8]–[10]. Another issue addressed in this dissertation is Minimum Kullback-Leibler (MKL) Bayesian multiuser detection in the presence of unknown multiple-access interference (MAI) and narrow-band interference (NBI), a scenario that occurs in CDMA overlay systems. Various techniques for interference suppression in CDMA overlay systems are reviewed in [11]. Existing methods include frequency-domain techniques [12], predictive techniques [13], [14], the linear MMSE estimation technique [15], [16] and the maximum-likelihood technique [17]. In this dissertation, we propose a approach to interference suppression. The Minimum Kullback-Leibler (MKL) is then used to calculate the Bayesian estimates of all unknowns.

In this dissertation, we propose a new technique for the channel estimation and the calculation of the extrinsic information, which is based on the Minimum Kullback-Leibler (MKL) algorithm for numerical Bayesian distributional approximation. To implement extrinsic probability by using the BCJR algorithm [25] without encoding, the measurement vector signals are decomposed into a signal components. Virtual trellis diagram, representing the ISI channels for each separation signal user, is designed to compute the extrinsic information by using the BCJR algorithm. Computer simulations are employed to assess the performance of the proposed schemes compare with iterative LMMSE turbo multiuser method [13].

1.1 Objectives

The primary objective of dissertation is on design a novel mathematical closed form joint iterative *Minimum Kullback-Leibler* (MKL) multiuser receiver and channel estimations for an uplink multipath asynchronous coded CDMA system employing spreading sequences. It is assumed that the receiver has only the knowledge of the spreading sequences and the initial delays of the desired users within the cell. The multipath channels are unknown to the receiver. A few pilot symbols are used by any users. In this approaches, channel parameters are first estimated and MKL multiuser receivers are then constructed based on the estimated channels. The Minimum Kullback-Leibler (MKL) multiuser receiver mitigate multiple-access interference (MAI) and intersymbol interference (ISI). Moreover, the implementation of the proposed *Minimum Kullback-Leibler* algorithm essentially designs virtual trellis corresponding to each user's symbols channel delay for generating extrinsic probability of the proposed optimization *Minimum Kullback-Leibler* algorithm. Simulation results making use of *Minimum Kullback-Leibler* algorithm can be considered.

1.2 Scope

As the *Minimum Kullback-Leibler* Multiuser detector for DS-CDMA systems under the presence of multiple-access interference (MAI) and intersymbol interference (ISI) investigated in the dissertation, the scope of the research works can be limited to the following:

1. Effects Bit error rate (BER), Channel estimation error and analyze computational results of proposed algorithm.
2. A novel MKL multiuser detector technique that is jointly considers both iterative decoding and channel estimation.
3. Parameters Estimation of the proposed optimization algorithm.
4. the limitation of proposed detector.
5. Performance of the proposed *Minimum Kullback-Leibler* detector and a comparative study of other techniques in the past. The concerned issue includes bit error rate performance of channel estimation error.

1.3 Outline

For the rest of this dissertation, chapter II reviews, the basic background and related Topics of iterative multiuser detector literatures including CDMA channel model, transmitters, receivers. Chapter III presents the proposed *Minimum Kullback-Leibler* Multiuser detector and its mathematical closed form channel estimations and iterative decoding will be proposed. In chapter IV, the effects of Bit error rate (BER), Channel estimation error and analyze computational results are demonstrated. Finally, chapter 5 will conclude this dissertation. The results and contributions will be summarized..

CHAPTER II

BASIC BACKGROUND AND RELATED TOPICS

Communication channels that involve both error-control coding and multiple-access signaling are of increasing interest in applications such as cellular telephony, wireless computer networks, and broadband local access. Optimal data detection and decoding in such channels generally requires a level of computational complexity that is prohibitive for these types of applications. Turbo multiuser detection (MUD) addresses this problem by applying the turbo principle of iteration among constituent decision algorithms, with intermediate exchanges of soft information (i.e., posterior probabilities) about tentative decisions. Here this principle is applied by considering MUD (which exploits the multiple-access signaling structure) and error-control decoding as the two consituent decision algorithms.

In this chapter reviews this area, outlining both the basic principles involved channel modeling and the basis for turbo multiuser detectors that require minimal increased complexity over that of the standard channel decoder.

2.1 System Description and Channel Model

Let us consider an asynchronous multipath DS-CDMA system that has K active users, employing normalized pseudorandom spreading sequences and signaling through multipath channels with additive white Gaussian noise (AWGN). The transmitted signal due to the k^{th} user is given by

$$x_k(t) = \alpha_k \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} b_k[i] c_{k,i}[j] \varphi(t - iT - jT_c - d_k) \tag{2.1}$$

where M denotes the length of the data frame, N is the processing gain, T denotes the symbol interval, $\{c_{k,i}[j]\}_{j=0}^{N-1}$ is a signature sequence assigned to the k^{th} user for the i^{th} symbol, the α_k, $\{b_k[i]\}$ and $d_k\{0 \leq d_k \leq T\}$ denote, respectively, the amplitude, the symbol stream, and the delay of the kth user's signal, φ is a normalized chip waveform of duration $T_c = T/N$. The kth user's signal propagates through a multipath channel whose impulse response is given by

$$g_k(t) = \sum_{l=1}^{L} \beta_{k,l} \delta(t - \tau_{k,l}) \tag{2.2}$$

where L is the total number of resolvable paths in the channel, $\beta_{k,t}$ and $\tau_{k,l}$ are, respectively, the complex path gain and the delay of the k^{th} user's l^{th} path. The received continuous-time signal at the receiver is given by

$$\begin{aligned} r(t) &= \sum_{k=1}^{K} x_k(t) \star g_k(t) + v(t) \\ &= \sum_{k=1}^{K} \sum_{i=0}^{M-1} b_k[i] \sum_{l=1}^{L} \sum_{j=0}^{N-1} \bar{\beta}_{k,l} c_{k,l}[j] \varphi(t - iT - jT_c - d_k - \tau_{k,l}) + v(t) \end{aligned} \quad (2.3)$$

where denotes $\star$ convolution,$\bar{\beta}_{k,l} = \alpha_k \beta_{k,t}$ and $v(t)$ is the ambient noise. At the receiver, the received signal $r(t)$ is filtered by a chipmatched filter and sampled at the chip-rate. Let

$$\ell \triangleq \max_{1 \le k \le K} \left\{ \frac{[d_k + \tau_{k,L}]}{T} \right\} \quad (2.4)$$

be the maximum delay spread among the K users in terms of symbol intervals. The signal sample at the matched filter output at time $t = iT + qT_c$

$$\begin{aligned} r_q(i) &\triangleq \int_{iT+qT_c}^{iT+(q+1T_c)} r(t)\phi(t - iT - qT_c)dt \\ &= \int_{iT+qT_c}^{iT+(q+1T_c)} \phi(t - iT - qT_c) \sum_{k=1}^{K} \sum_{m=0}^{M-1} b_k[m] \sum_{l=1}^{L} \sum_{j=0}^{N-1} \bar{\beta}_{k,l} c_{k,l}[j] \varphi(t - mT - jT_c - d_k - \tau_{k,l})dt \\ &+ \ v_q(i) \\ &= \int_{iT+qT_c}^{iT+(q+1T_c)} \sum_{k=1}^{K} \sum_{m=i-\ell}^{i} b_k[m] \sum_{j=0}^{N-1} c_{k,l}[j] \underbrace{\int_0^{T_c} \sum_{\ell=1}^{L} \bar{\beta}_{k,l} \varphi(t + (i-m)T + (q-j)T_c - d_k - \tau_{k,l})dt}_{h_k[(i-m)N+(q-j)]} \\ &+ \ v_q(i) \\ &= \sum_{k=1}^{K} \sum_{m=i-\ell}^{l} \sum_{j=0}^{N-1} b_k[i-m] c_{k,i-m}[q-j] h_k[mN+j] + v_q(i) \end{aligned} \quad (2.5)$$

where $v_q(i) = \int_{iT+qT_c}^{iT+(q+1T_c)} v(t)\phi(t - iT - qT_c)dt$. Since the chip waveform has a duration T_c, $h_k(l)$ is nonzero only for $[(d_k + \tau_{k,1})/T_c] \le \ell \le [(d_k + \tau_{k,L})/T_c]$ For convenience,define $\ell_k \triangleq [(d_k + \tau_{k,1})/T_c] - 1$ as the initial delay in terms of number of chips for the k^{th} user's signal; define $P \triangleq \max_k [(\tau_{k,L} - \tau_{k,1})/T_c]$ as the maximum channel delay among all users; and define $h_k^{(0)} = h_k(\ell_k + 1)...h_k^{(L-1)} = h_k(\ell_k + L)$. as the channel response for the kth user. Throughout the dissertation,assume that both the maximum initial delay $max_k\{l_k\}$ and P are less than N . Hence, the maximum symbol delay satisfies $\ell_k \le 2$. It is convenient to express the signal model (2.5) in a vector form as

$$
\begin{aligned}
\mathbf{r}[t] &= \sum_{k=1}^{K} \underbrace{\left(b_k[t]C_k^{(0)} + \ldots + b_k[t-L+1]C_k^{(L-1)}\right)}_{D_k(\underline{b}_{k,t})} g_k + w[t] \\
&= \sum_{k=1}^{K} \left(b_k[t]h_k^{(0)} + \ldots + b_k[t-L+1]h_k^{(L-1)}\right) + w[t] \\
&= \mathbf{H}\mathbf{b}[t] + w[t]. \qquad (2.6)
\end{aligned}
$$

where user $k = 1, 2, \cdots, K$, the transmitted symbol $b_k[t]$ and $\{b_k[t-l]\}_{l=1}^{L-1}$ are modulation symbols and delay symbols assigned for k^{th} user at t^{th} interval (see section 2.4 and [21] for more details). The noise matrix term $w[t]$ is assumed to be white Gaussian vector with zero mean and variance $\sigma^2\mathbf{I}$. Spreading matrices $C_k^{(l)}, l = 0, ..., L-1$ are defined by $N \times P$ matrices [21].

$$
\begin{bmatrix} C_k^{(0)} \\ C_k^{(1)} \\ \vdots \\ C_k^{(L-1)} \end{bmatrix}
=
\begin{bmatrix}
0_{l_k \times 1} & 0 & \cdots & 0 \\
c_k[0] & & & \\
c_k[1] & c_k[0] & \ddots & \\
\vdots & c_k[1] & & c_k[0] \\
c_k[N-1] & \vdots & & c_k[1] \\
 & c_k[N-1] & \ddots & \vdots \\
 & & & c_k[N-1] \\
 & & & 0
\end{bmatrix}_{LN \times P} \qquad (2.7)
$$

N is processing gain and P is maximum channel delay. The channel vector g_k, channel matrix $\mathbf{H}$ and vector $\mathbf{b}[t]$ can be expressed as following $g_k = [\, g_{k,1}\, g_{k,2} \cdots \; g_{k,P} \,]^T$, $h_k^{(0)} = C_k^{(0)} g_k$, $h_k^{(1)} = C_k^{(1)} g_k$ and

$$\mathbf{H} = [h_1^{(0)} \; .. \, h_1^{(L-1)} \; \ldots h_K^{(0)} \, .. \, h_K^{(L-1)}] \qquad (2.8)$$

$$\mathbf{b}[t] = \left[b_1[t] \; .. \, b_1[t-L+1] \; .. \, b_K[t] \, .. \, b_K[t-L+1]\right]^T \qquad (2.9)$$

2.2 Multiuser detection (MUD)

MUD refers to the detection of data from multiple terminals in a communication network when observed in a nonorthogonal multiplex, that is, when derived from a multiple-access channel. This problem arises naturally, for example, in code-division multiple-access (CDMA) systems using nonorthogonal spreading codes. It also arises in orthogonally multiplexed wireless channels, such as time-division multiple-access channels, due to effects such as multipath or nonideal frequency channelization, and in wireline channels such as those arising in digital subscriber line (DSL) systems or powerline communications (PLC) in which crosstalk and other types of interference are major impairments. The basic idea of

MUD is to exploit the cross-correlations among the signals to be demodulated to improve the data detection process. Considerable progress has been made on this problem over the past two decades. (See, e.g., [13] and [1].) Among other things, it has been shown that the use of MUD can provide very significant performance advantages in interference-limited channels. There are many types of MUD techniques. Optimal techniques, based on maximum-likelihood (ML) or maximum a posteriori probability (MAP) criteria, can often achieve performance very close to that of a system that is free of interference. However, these methods tend to be quite complex, particularly when compared with the processing resources available in most communications receivers. Consequently, a considerable amount of effort has been devoted to the development of lower-complexity techniques that can achieve some of the benefits of the optimal procedures. One class of such methods are the linear multiuser detectors, which use linear processing to suppress interference, followed by simple memoryless quantization to perform data detection. Another class of lower-complexity multiuser detectors are the iterative multiuser detectors, which make use of tentative channel-symbol decisions (either soft or hard) to provide feedback that can improve the capabilities, in terms of complexity or performance, of optimal or linear MUD methods. When channel coding is considered in addition to nonorthogonal signaling, the complexity of optimal receiver processing is further exacerbated. In particular, the complexity of optimal (ML or MAP) joint MUD and channel decoding tends to be extremely high. However, this combination also lends itself very well to the use of iterative MUD methods in which the tentative channel-symbol decisions are produced by the channel decoders. Similarly, MUD can be used to provide tentative channel-symbol decisions to the channel decoders. Iteration between these two constituent processes, with intermediate exchanges of soft channelsymbol information, is known as turbo MUD. This idea was originally developed in the context of convolutionally encoded CDMA channels, but has since been applied in a number of other frameworks, including DSL, PLC, space-time coded CDMA, ultra-wideband (UWB), and turbo-coded CDMA channels.

2.2.1 Linear MUD

The basic difficulty with optimal multiuser detectors is their complexity. A considerable amount of research has been devoted to the development of suboptimal multiuser detectors that mitigate this complexity (see, e.g., [1]). One well-studied family of suboptimal multiuser detectors are the linear MMSE (Minimum Mean Square Error) multiuser detectors, which are of interest in their own right and which also form the basis for many iterative multiuser detectors, including the low-complexity turbo MUD. For each user ,k^{th} a soft interference cancellation is performed on the matched-filter output in (2.6), to obtain

$$y_k[i] = \mathbf{r}[i] - \sum_{k=1}^{K} \left(b_k[t]h_k^{(0)} + b_k[t-1]h_k^{(1)} + \ldots + b_k[t-L+1]h_k^{(L-1)} \right) \qquad (2.10)$$

Such a soft interference cancellation scheme was first proposed in [3]. Next, in order to further suppress the residual interference in $y_k[i]$, an instantaneous linear MMSE filter is w_k

applied to $y_k[i]$, to obtain The sufficient statistic $\mathbf{r}[t]$ of (2.6) and (2.10) obeys the linear model (2.11) and (2.12) ,

$$\begin{aligned} z_k[i] &= w_k^H y_k[i] \\ b_k[i] &= \operatorname{sgn}\{\Re(y_k[i])\} \end{aligned} \tag{2.11}$$

where the filter $w_k \in \Re^N$ is chosen to minimize the means quare error between the code bit $b_k[i]$ and $z_k[i]$ the filter output , i.e.,

$$\begin{aligned} w_k &= \underset{w_k \in \Re^N}{\mathbf{argmin}}\, E\left\{\left\|b_k[i] - w_k^H y_k[i]\right\|^2\right\} \\ &= \underset{w_k \in \Re^N}{\mathbf{argmin}}\, w_k^H E\left\{r[i]r[i]^H\right\} w_k - 2\Re\left(w_k^H E\left\{b_k[i]r[i]\right\}\right) \\ &= \underset{w_k \in \Re^N}{\mathbf{argmin}}\, w_k^H E\left\{r[i]r[i]^H\right\} w_k - 2\Re\left(w_k^H h_k^{(0)}\right) \end{aligned} \tag{2.12}$$

where

$$E\left\{r[i]r[i]^H\right\} = HH^H + \sigma^2 I \tag{2.13}$$

and

$$w_k^H E\left\{b_k[i]r[i]\right\} = w_k^H h_k^{(0)} \tag{2.14}$$

The solution to (2.12) is given by

$$w_k = \left[HH^H + \sigma^2 I\right]^{-1} h_k^{(0)} \tag{2.15}$$

MUD (and equalization as well) can be viewed as the fitting of this model to the observations. The complexity of these problems comes from the fact that the elements of the vector $y_k[i]$ and $b_k[i]$ take values in a finite alphabet. Without this constraint, the fitting of linear models such as (2.10) is of relatively low complexity. The basic idea of linear MUD is to take advantage of this relatively low complexity of unconstrained linear model-fitting by first estimating $b_k[i]$ in (2.11) as if it were a vector with real components, and then to project these real estimates onto the finite alphabet of the actual symbols. This, of course, will not yield ML or MAP symbol decisions, but it often works quite well. Note that the matched filter detector (2.6),(2.10) is a very simple example of a linear multiuser detector, in which the vector $y_k[i]$ itself is used to estimate $b_k[i]$ before quantization. As noted above, this choice is optimal against the white background noise in the absence of signal cross correlations. On the other hand, referring to (2.6),(2.10) and (2.15), we see that this choice essentially ignores the off-diagonal elements of the cross-correlation matrix H. A key alternative to the matched filter is the linear minimum-mean-squareerror (MMSE) detector, which detects $b_k[i]$ via where I denotes the NN identity matrix. This latter detector uses, as its linear estimation stage, the linear MMSE estimator of $b_k[i]$ given $y_k[i]$ in (2.10) under the assumption that the symbols have a prior distribution under which they are uncorrelated with zero means; namely, $b_k[i] \triangleq \operatorname{sgn}\left\{\Re\left[h_k^{(0)H}\left[HH^H + \sigma^2 I\right]^{-H} y_k[i]\right]\right\}$

2.3 Iterative MUD

Turbo MUD falls within the category of iterative MUD, in which tentative decisions are used iteratively to improve overall data detection. Aside from turbo MUD, iterative detectors include several varieties, including linear and nonlinear interference cancellers, and model-based techniques such as those based on the expectation-maximization (EM) algorithm. We now discuss these very briefly.

Note that the linear detectors discussed above typically require the inversion of a $N \times N$ matrix. The complexity of the matrix inversion is, in its worst case, $O(N^3)$. Although simpler in principle than the exponential complexity of ML or MAP MUD, this complexity can still be quite significant. Moreoever, this matrix inversion $\left[HH^H + \sigma^2 I\right]^{-1}$ cannot necessarily be amortized over more than one frame of data, since the channel and/or the signaling waveforms may vary from frame to frame. Thus, it is of interest to use lower-complexity methods for computing the estimates used in linear MUD.

Nonlinear interference cancellers are similar in spirit to linear interference cancellers, in that they use iterative methods to fit the model (2.6),(2.10). Unlike their linear counterparts, however, that exploit only the linearity of the model while iterating, nonlinear interference cancellers also exploit the discrete nature of the symbol $\mathbf{b}[t]$ at each iteration by making intermediate soft or hard decisions between iterations. As with linear interference cancellers, there are a number of such methods. (See [4] for a discussion of these detectors.)

As noted above, the basic problem of MUD is the accurate fitting of the model (2.6),(2.10). Linear interference cancellers perform this fitting by exploiting only the linear structure of the model, while nonlinear interference cancellers seek to improve on this fit by making use of further information about the model, namely that the symbols are elements of a known finite alphabet. Often, further information is known about the symbols, and this can also be exploited to provide further performance improvement. For example, the EM algorithm or Markov-chain Monte Carlo (MCMC) techniques can be used to exploit statistical information about b, leading to several soft-decision iterative nonlinear MUD algorithms. (See, e.g., [5].) Turbo MUD is a further example of such an exploitation, in which the information to be exploited is the set of constraints imposed by channel coding.

2.4 Iterative Joint MUD and Decoding

We now turn to the situation in (2.3) in which the symbols are constrained by having been produced by an error-correcting code. In principle, this constraint should strengthen our ability to fit the model (2.6) , as it reduces the number of sequences $\mathbf{b}[i]$ that are possible. However, the complexity of including such constraints is quite high, as we will see below. Essentially, turbo MUD is a technique for fitting (2.6) when the symbols satisfy coding constraints with dramatically lower complexity than optimal algorithms.

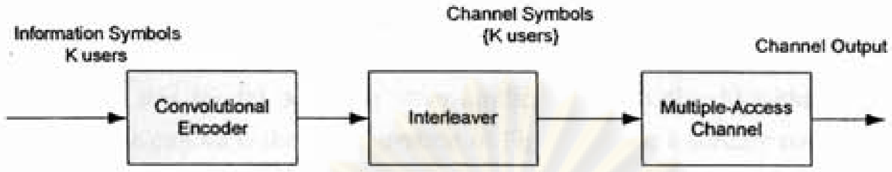


Figure 2.1 The multiple-access channel with convolutional coding

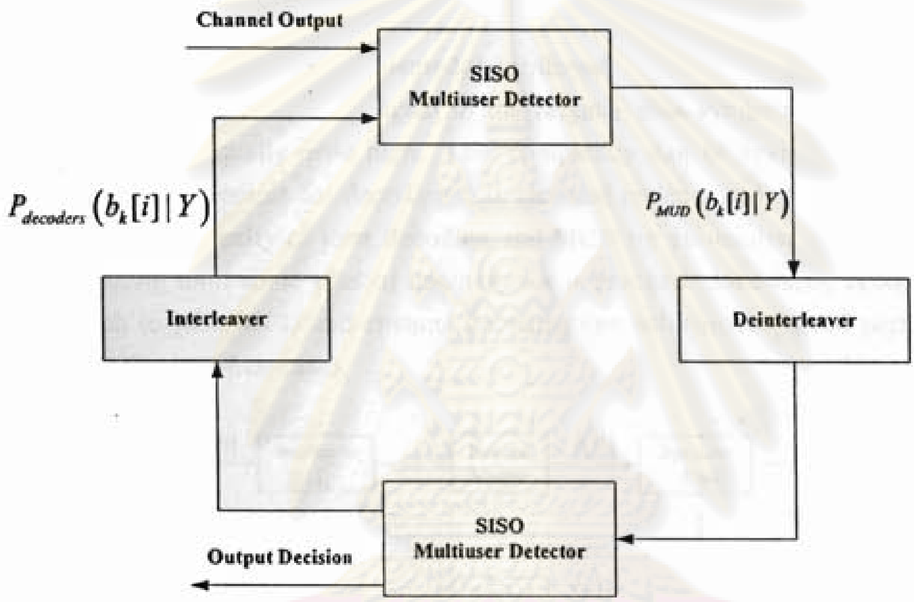


Figure 2.2 General structure of turbo multiuser detection

Error-control coding is, of course, ubiquitous in wireless and other impaired channels. Similarly to MUD, the decoding of error-control codes exploits the dependencies among successive channel symbols to improve the detection of a single stream of data symbols. Like MUD, channel decoding typically involves very complex optimal algorithms, and so complexity issues often dominate the study of these problems. Notable among coding techniques with this problem are parallel and serially concatenated codes separated by interleavers, which are known to offer considerable performance improvement over traditional codes, exhibiting near-Shannon-limit performance in many cases. However, although the optimal decoding of such codes is of particularly high complexity, iterative or turbo decoding algorithms that involve the iterative exchange of soft information between constituent decoders (separated by interleavers/de-interleavers) have been shown to be very effective approximations to optimal decoding. These well-known ideas are discussed, for example, in [6] and [7].

Many communication systems involve both errorcontrol coding and nonorthogonal multiplexing. A typical configuration is a convolutional encoder mapping data symbols into channel symbols, followed by an interleaver, and then a multiple-access channel, as shown in . We will focus on this model, although other applications can also fit within the formalism discussed here. One can view the configuration of Figure 2.1 as a serially concatenated code, in which the multiple-access channel (e.g., a CDMA spreading code) is the inner code, and the convolutional code is the outer code. A traditional way of decoding this concatenation is to first demodulate the multiple-access signals (using either a conventional matched-filter detector, or a multiuser detector) and then to follow this demodulator by a de-interleaver and a channel decoder. To seek optimality in such a situation, we could replace this traditional configuration with an overall optimal demodulator/decoder that uses an optimal (say ML or MAP) mapping from the received signal to the original data symbols. The complexity of such a system is potentially quite high. This complexity can be mitigated however, by appealing to the turbo principle for decoding concatenated codes noted above. In particular we can reduce the complexity of joint decoding and MUD by an iterative exchange of soft information, iterating until some kind of convergence is reached. Like turbo decoding, this iterative approach to joint MUD and channel decoding can achieve very good performance (close to the interference-free case).

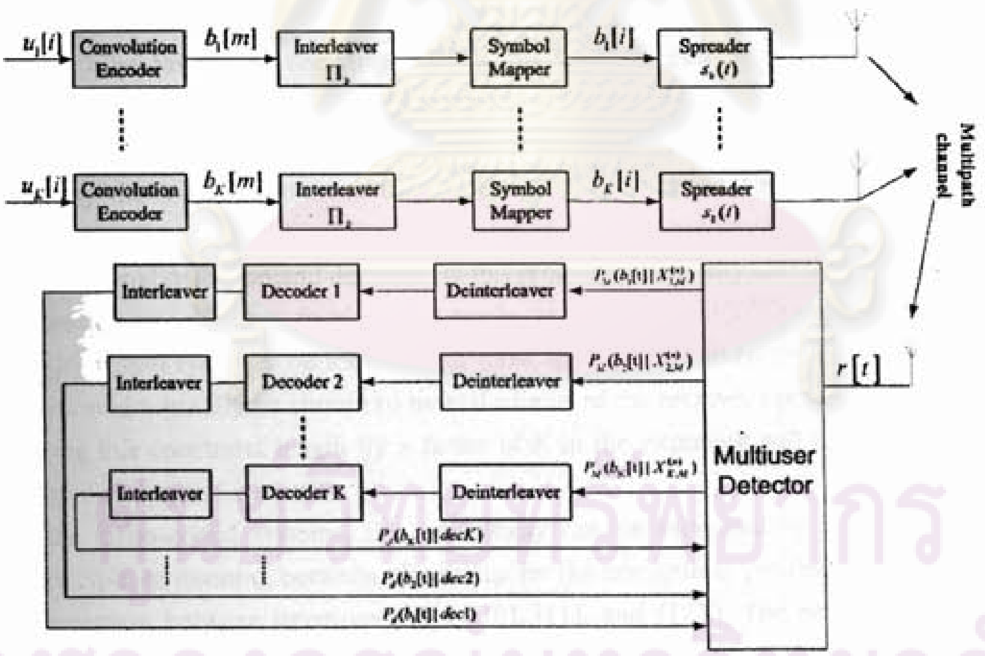


Figure 2.3 A coded CDMA system with iterative (Turbo) multiuser receiver.

To consider this problem, we need to refer model of (2.3) to include coding. This can be done very simply, by writing the channel symbols $b_k[i]$ explicitly as functions of underlying data symbols; i.e., for a rate- R code, we have

$$r(t) = \sum_{k=1}^{K} \sum_{i=0}^{M-1} b_k[i] \sum_{l=1}^{L} \sum_{j=0}^{N-1} \bar{\beta}_{k,l} c_{k,l}[j] \varphi(t - iT - jT_c - d_k - \tau_{k,l}) + v(t) \qquad (2.16)$$

The block diagram of transmitter and receiver model is shown in Fig. 2.4. The binary information $\{u_k[i]\}$ for user $k, k = 1, .., K$ are convolutionally encoded by a single convolution encoder with code rate R_k. The code bits $\{b_k[m]\}$ are interleaved and mapped to BPSK symbols stream. Each data symbol $\{b_k[i]\}$ is modulated by a spreading waveform $s_k(t)$ [21] and transmitted through the multipath channel. The received signal vector r$[t]$ at the receiver is the superposition of K user signals plus additive white Gaussian noise, which is defined in (2.6). In the coded communication systems, the iterative receiver structure often makes use of turbo principle [2]- [13] to reduce the loss of information and performance. The receiver structure under consideration is a iterative receiver as shown in Fig 2.4. It consists of two states the *multiuser detector*, followed by K parallel *channel decoders*. The two states are separated by deinterleavers and interleavers.

We would like to make inferences about the set of data symbols $u_1[i], \ldots,$ $u_K[i]$. The observation vector $Y = y_k[1], .., y_k[M]$ of (2.10) is still a sufficient statistic for such inferences, and thus joint channel decoding and MUD is another problem of sequence detection. Like MUD, the decoding task in this situation can be simplified by dynamic programming. For example, in the single-user ($K = 1$) case, the per-symbol complexity of optimal decoding reduces to $O(2^v)$, with the corresponding dynamic program being specified by the Viterbi algorithm in the case of ML decoding and by the Bahl, Cocke, Jelinek, and Raviv (BCJR) algorithm in the case of MAP decoding [8]. With multiple users ($K > 1$), optimal detection and decoding in this problem essentially combines the complexity of the constituent problems, to yield a dynamic program with $O(2^{Kv})$ complexity [9]. This complexity would typically be too high for most applications, since the constraint length of the code would normally be chosen to meet the limits of the receiver's processing capabilities. Amplifying this constraint length by a factor of K in the exponent will push the processing capability well beyond its limits.

Like turbo-coded systems, this complexity can be mitigated by making use of the turbo principle of iterating between algorithms for the constituent problems, and exchanging soft information between iterations. (See [10], [11], and [12].) The basic building blocks of a turbo multiuser detector are a soft-input/soft-output (SISO) multiuser detector and a bank of single-user SISO channel decoders, as shown in Figure 2.2. The role of each of these algorithms is to compute posterior probabilities of the channel symbols based on given prior probabilities and on the corresponding signal structure. That is, the SISO multiuser detector uses prior symbol probabilities and the multiuser signaling structure to compute posterior symbol probabilities conditioned on the observations. Similarly, the SISO channel decoders use prior symbol probabilities and the structure imposed by the channel code to compute posterior symbol probabilities. (Of course, the SISO decoders also compute underlying data symbols; i.e., for a rate- R code, we have

$$r(t) = \sum_{k=1}^{K} \sum_{i=0}^{M-1} b_k[i] \sum_{l=1}^{L} \sum_{j=0}^{N-1} \bar{\beta}_{k,l} c_{k,l}[j] \varphi(t - iT - jT_c - d_k - \tau_{k,l}) + v(t) \qquad (2.16)$$

The block diagram of transmitter and receiver model is shown in Fig. 2.4. The binary information $\{u_k[i]\}$ for user $k, k = 1, .., K$ are convolutionally encoded by a single convolution encoder with code rate R_k. The code bits $\{b_k[m]\}$ are interleaved and mapped to BPSK symbols stream. Each data symbol $\{b_k[i]\}$ is modulated by a spreading waveform $s_k(t)$ [21] and transmitted through the multipath channel. The received signal vector $\mathbf{r}[t]$ at the receiver is the superposition of K user signals plus additive white Gaussian noise, which is defined in (2.6). In the coded communication systems, the iterative receiver structure often makes use of turbo principle [2]- [13] to reduce the loss of information and performance. The receiver structure under consideration is a iterative receiver as shown in Fig 2.4. It consists of two states the *multiuser detector*, followed by K parallel *channel decoders*. The two states are separated by deinterleavers and interleavers.

We would like to make inferences about the set of data symbols $u_1[i], \ldots,$ $u_K[i]$. The observation vector $Y = y_k[1], .., y_k[M]$ of (2.10) is still a sufficient statistic for such inferences, and thus joint channel decoding and MUD is another problem of sequence detection. Like MUD, the decoding task in this situation can be simplified by dynamic programming. For example, in the single-user ($K = 1$) case, the per-symbol complexity of optimal decoding reduces to $O(2^v)$, with the corresponding dynamic program being specified by the Viterbi algorithm in the case of ML decoding and by the Bahl, Cocke, Jelinek, and Raviv (BCJR) algorithm in the case of MAP decoding [8]. With multiple users ($K > 1$), optimal detection and decoding in this problem essentially combines the complexity of the constituent problems, to yield a dynamic program with $O(2^{Kv})$ complexity [9]. This complexity would typically be too high for most applications, since the constraint length of the code would normally be chosen to meet the limits of the receiver's processing capabilities. Amplifying this constraint length by a factor of K in the exponent will push the processing capability well beyond its limits.

Like turbo-coded systems, this complexity can be mitigated by making use of the turbo principle of iterating between algorithms for the constituent problems, and exchanging soft information between iterations. (See [10], [11], and [12].) The basic building blocks of a turbo multiuser detector are a soft-input/soft-output (SISO) multiuser detector and a bank of single-user SISO channel decoders, as shown in Figure 2.2. The role of each of these algorithms is to compute posterior probabilities of the channel symbols based on given prior probabilities and on the corresponding signal structure. That is, the SISO multiuser detector uses prior symbol probabilities and the multiuser signaling structure to compute posterior symbol probabilities conditioned on the observations. Similarly, the SISO channel decoders use prior symbol probabilities and the structure imposed by the channel code to compute posterior symbol probabilities. (Of course, the SISO decoders also compute posterior data symbol probabilities, which will ultimately yield the overall output of the combined algorithm.)

The turbo multiuser detector begins with a SISO multiuser detector applied to the frame of M channel symbols (M is assumed to be equal to the interleaver length). This detector particularly computes posterior probabilities, conditioned on the observations Y, for

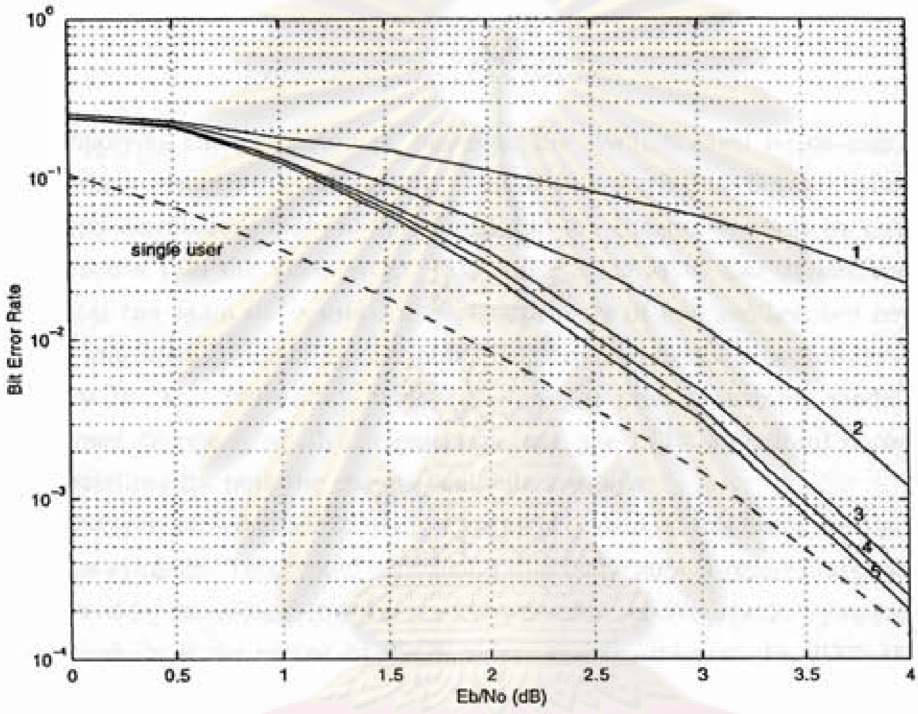


Figure 2.4: Performance of MMSE-based low-complexity turbo MUD: four users with equal power: rate-1/2 constraint-length-5 convolutional code.

each of the channel symbols of each of the users; that is, for each element of the vector $\mathbf{b}[i]$. This first set of posterior probabilities is based on the prior assumption that the channel symbols are drawn uniformly from $\{-1, +1\}^{KM}$; that is, that the channel symbols are i.i.d. equiprobably ±1 random variables. Although this assumption is not correct due to the channel coding (which correlates the channel symbols), it serves as a useful approximation for initializing the algorithm because the interleavers at the transmitter serve to decorrelate the symbols as they appear at the input to the channel.

The posterior probabilities computed by the SISO MUD will then be used as prior probabilities in the next step of the algorithm, which makes use of the bank of single-user channel decoders.

$$P_{MUD}(b_k[t]|Y) = \frac{P(z_k[t]|b_k[t], \mathbf{H})P_{decoders}(b_k[t]|Y)}{\sum\limits_{b_k[t]} P(z_k[t]|b_k[t], \mathbf{H})P_{decoders}(b_k[t]|Y)} \tag{2.17}$$

The output $z_k[t]$ can be well approximated by Gaussian distribution [2].

$$P(z_k[t]|b_k[t],\mathrm{H}) = \frac{1}{\gamma_{k,t}^2}\exp(-\frac{1}{\gamma_{k,t}^2}||z_k[t]-\mu_{k,t}b_k[t]||^2) \tag{2.18}$$

with $\mu_{k,t} = (C_k^{(0)}g_k)^H w_{k,t}$ and $\gamma_{k,t} = \mu_{k,t} - ||\mu_{k,t}||^2$ Keeping in mind that $z_k[t]$ is the function of $x_k[t]$ by$z_k[t] = (w_{k,t})^H y_k[t]$. $w_{k,t}$ is the function of MMSE criterion [2].

$$w_{k,t} = (\mathrm{HH}^H + \sigma_k^2 I)^{-1}(C_k^{(0)}g_k) \tag{2.19}$$

Before applying channel decoding, however, the symbols must be de-interleaved to return them to their correct order for decoding. This de-interleaving has the approximate effect of removing any correlations that are introduced into the channel symbols by conditioning on the observations y in the SISO MUD. Thus, after SISO MUD and de-interleaving, the channel symbols can again be assumed to be independent of one another, but now having marginal (i.e., individual) probability distributions determined by the probabilities computed by the SISO MUD. This probability model becomes the prior probability model used by the SISO channel decoders, which compute (via, say, the BCJR algorithm) corresponding posterior probabilities for both the channel and data symbols.

These posterior probabilities for the data symbols could, at this point, be used to MAP decode the data symbols. This would correspond to a conventional receiver approach based on MUD followed by decoding. However, a more powerful receiver results by re-interleaving the channel symbols at the output of the decoders and returning to the SISO MUD, now using as a prior distribution the posterior channel-symbol probabilities computed by the SISO decoders. The SISO MUD then refines its estimates of the posterior probabilities of the symbol probabilities and hands them back to the channel decoders after de-interleaving again. This process of softinformation exchange between the SISO MUD and the SISO decoders can continue until the posterior channelsymbol probabilities converge to stable values, at which point the data symbols can be MAP decoded via the data symbol posterior probabilities computed on the last application of the SISO decoding algorithm. The constituents of this process, namely MAP MUD and MAP decoding, are well known, and thus details are omitted for the sake of brevity. (Explicit equations can be found in [5].)

From this description, it can be seen that the interpretation of the multiuser detector as a posterior-probability calculator is an essential philosophical underpinning of this approach. Unlike the case with turbo decoding, however, in which the complexity of the constituent decoders is controlled by the system designer, the complexity of the SISO multiuser detector used in this turbo multiuser detector is dependent on the number of users in the channel and is thus beyond the designer's immediate control. Thus, although the $O(2^{Kv})$ complexity of optimal joint detection and decoding noted in [9] is reduced to $O(2^v) + O(2^K)$ via the turbo principle, the second term in this complexity order is prohibitive for most applications, as noted previously.

CHAPTER III

THE PROPOSED FRAMEWORKS

In this chapter, a mathematical closed form for computational probability density estimation ***Minimum Kullback-Leibler*** (MKL) were proposed for improving the iterative multiuser performance. This dissertation first proposed the frameworks MKL multiuser detector and channel estimator that is designed to combine the virtual trellis corresponding to each user's symbols channel delay.

3.1 *Variational Bayes* (VB) of the Distribution Approximation

In signal processing, as in all quantitative sciences, we are concerned with observation data, Y and we will model the data parametrically, so that a set, θ , of unknown parameters describes the data-generating system. In deterministic problems, knowledge of θ determines Y under some notional rule, $Y = g(\theta)$. This accounts for very few of the data contexts in which we must work. In particular, when Y is information-bearing, then we must model the uncertainty (sometimes called the randomness) of the process. The defining characteristic of Bayesian methods is that we use probabilities to quantify our beliefs amid uncertainty, and the calculus of probability to manipulate these quantitative beliefs [42]- [44]. Hence, our beliefs about the data are completely expressed via the parametric probabilistic observation model, $P(Y|\theta)$. In this way, knowledge of θ determines our beliefs about Y, not Y themselves.

In practice, the result of an observational experiment is that we are given Y, and our problem is to use them to learn about the system summarized by the unknown parameters, θ which generated them. This learning amid uncertainty is known as inductive inference [44], and it is solved by constructing the distribution $P(\theta|D)$, namely, the distribution which quantifies our a posteriori beliefs about the system, given a specific set of data, D. The simple prescription of Bayes'rule solves the implied inverse problem [45], allowing us to reverse the order of the conditioning in the observation model, $P(Y|\theta)$:

$$P(\theta|Y) \propto P(Y|\theta)P(\theta) \tag{3.1}$$

Bayes' rule specifies how our prior beliefs, quantified by the prior distribution, $P(\theta)$, are updated in the light of Y. Hence, a Bayesian treatment requires prior quantification of our beliefs about the unknown parameters, θ , whether or not θ is by nature fixed or randomly realized. The signal processing community, in particular, has been resistant to the philosophy of strong Bayesian inference [44], which assigns probabilities to fixed, as well as random, unknown quantities. Hence, they relegate Bayesian methods to inference problems involving only random quantities [46]- [47].

Tractability is a primary concern to any signal processing expert seeking to develop a parametric inference algorithm, both in the off-line case and, particularly, on-line. The Bayesian approach provides $P(\theta|Y)$ as the complete inference of θ, and this must be manipulated in order to solve problems of interest. For example, we may wish to concentrate the inference onto a subset, θ_1, by marginalizing over their complement, θ_2:

$$P(\theta_1|Y) = \int_{\Theta_1^*} P(\theta|Y) d\theta_2 \tag{3.2}$$

A decision, such as a point estimate, may be required. The mean a posteriori estimate may then be justified:

$$\bar{\theta}_1 = \int_{\Theta_1^*} \theta_1 P(\theta_1|Y) d\theta_1 \tag{3.3}$$

The integrations required in (3.2)—(3.3) will often present computational burdens that compromise the tractability of the signal processing algorithm.

An tractability example of parameters estimation, let we consider the moments of the posterior distribution i.e. the expected or mean value of known functions, $g(\theta)$, of the parameter will be denoted by

$$\bar{g}(\theta) = E_{P(\theta|Y)}[g(\theta)] = \int_{\Theta^*} g(\theta) P(\theta|Y) d\theta \tag{3.4}$$

In general, we will use the notation $\bar{g}(\theta)$ to refer to a posterior point estimate of $g(\theta)$. if $g(\theta) = \theta$, posterior mean is (3.3). The posterior mean (3.4) is only one of many decisions that can be made in choosing a point estimate, $\bar{g}(\theta)$, of $g(\theta)$. Bayesian decision theory [31] allows an optimal such choice to be made. The Bayesian model, $P(\theta, Y)$ is supplemented by a loss function, $L(g, \bar{g}) \in [0, \infty)$, quantifying the loss associated with estimating $g = g(\theta)$ by $\bar{g} = \bar{g}(\theta)$. The ***minimum Bayes risk*** estimate is found by minimizing the posterior expected loss,

$$\bar{g}(\theta) = \arg\min_{\bar{g}} E_{P(\theta|Y)}[L(g, \bar{g})] \tag{3.5}$$

The quadratic loss function , $L(g, \bar{g}) = (g(\theta) - \bar{g}(\theta))^H Q (g(\theta) - \bar{g}(\theta))$, Q positive definite, leads to the choice of the posterior mean (2.7). Other standard loss functions lead to other standard point estimates, such as the maximum and median a posteriori estimates [58]. The ***Maximum a Posteriori*** (MAP) estimate is defined as follows:

$$\theta_{MAP} = \arg\max_{\theta} P(\theta|Y) \tag{3.6}$$

In the special case where $P(\theta)$ = const., i.e. the improper uniform prior, then

$$\begin{aligned}\theta_{MAP} &= \arg\max_{\theta} P(\theta|Y) \\ &= \arg\max_{\theta} \frac{P(Y|\theta)P(\theta)}{P(Y)} \\ &= \arg\max_{\theta} \frac{P(Y|\theta)P(\theta)}{\int_{\Theta^*} P(Y|\theta)P(\theta)d\theta} \\ &\propto \arg\max_{\theta} P(Y|\theta) \\ &\propto \theta_{ML} = \arg\max_{\theta} l(\theta|Y)\end{aligned} \quad (3.7)$$

Here, θ_{ML} denotes the ***Maximum Likelihood*** (ML) estimate. ML estimation [59] is the workhorse of classical inference, since it avoids the issue of defining a prior over the space of possibilities. In this chapter, we will review some of the approximations which can help to address these problems, but the aim of this book is to advocate the use of the ***Variational Bayes*** (VB) approximation as an effective pathway to the design of tractable signal processing algorithms for parametric inference. These VB solutions will be shown, in many cases, to be novel and attractive alternatives to currently available Bayesian inference algorithms.

The central idea of the VB method is to approximate $P(\theta|Y)$, ab initio, in terms of approximate marginals:

$$P(\theta|Y) \cong \tilde{P}(\theta|Y) = \tilde{P}(\theta_1|Y)\tilde{P}(\theta_2|Y) \quad (3.8)$$

In essence, the approximation forces posterior independence between subsets of parameters in a particular partition of θ chosen by the designer. The optimal such approximation is chosen by minimizing a particular measure of divergence from . $\tilde{P}(\theta|Y)$ to $P(\theta|Y)$, namely, a particular ***Kullback-Leibler Divergence*** (KLD), which we will call KLD_{VB}

The ***Kullback-Leibler Divergence*** (KLD) provides objective statistical indicators for the difficulty in discriminating between two statistical hypotheses [29]. In addition to this role as a discrimination measure between probabilistic models, KLD is a fundamental quantity in information theory—all information quantities can be derived from it [34], [35], and it is fundamental for characterizing the rate function, which reflects the exponential rate of convergence of empirical measures to their probabilities (Sanov's theorem), in large deviations [36]. On the application side, ***Kullback-Leibler Divergence*** (KLD) has been widely used to compare probabilistic models from a discrimination point of view, and to globally evaluate the inherent discrimination complexity [37] and the feature space quality in pattern recognition. These are some of the reasons that explain its wide use in the context of classification based on statistical decision theory [38]– [41].

$$\tilde{P}(\theta|Y) = \arg\min_{\tilde{P}(\theta_1|Y),\ \tilde{P}(\theta_2|Y)} KL\left(\tilde{P}(\theta_1|Y)\tilde{P}(\theta_2|Y)||P(\theta|Y)\right) \quad (3.9)$$

In practical terms, functional optimization of (3.9) yields a known functional form for $\tilde{P}(\theta_1|Y)$ and $\tilde{P}(\theta_2|Y)$, which will be known as the ***Variational Bayes*** (VB)-marginals. However,the shaping parameters associated with each of these ***Variational Bayes*** (VB)-marginals are expressed via particular moments of the others. Therefore, the approximation is possible if all moments required in the shaping parameters can be evaluated. Mutual interaction of VB-marginals via their moments presents an obstacle to evaluation of its shaping parameters, since a closed-form solution is available only for a limited number of problems. However, a generic iterative algorithm for evaluation of VB moments and shaping parameters is available for tractable VB-marginals (i.e. marginals whose moments can be evaluated). This algorithm.reminiscent of the classical ***Expectation-Maximization*** (EM) algorithm will be called the ***Iterative Variational Bayes (IVB)*** algorithm in this book. Hence, the computational burden of the VB-approximation is confined to iterations of the IVB algorithm. The result is a set of moments and shaping parameters, defining the VB-approximation (3.8).

The ***Variational Bayes*** (VB) method of approximation is one of many techniques for approximation of probability functions. In the VB method, the approximating family is taken as the set of all possible distributions expressed as the product of required marginals, with the optimal such choice made by minimization of a ***Kullback-Leibler Divergence*** (KLD). The following are among the many other approximations deterministic and stochastic that have been used in signal processing:

- Point-based approximations: examples include the Maximum a Posteriori (MAP) and ML estimates. These are typically used as ***certainty equivalents*** [48] in decision problems, leading to highly tractable procedures. Their inability to take account of uncertainty is their principal drawback.

- Local approximations: the ***Laplace approximation*** [49], for example, performs a Taylor expansion at a point, typically the ML estimate. This method is known to the signal processing community in the context of criteria for model order selection, such as the Schwartz criterion and Bayes' Information Criterion (BIC), both of which were derived using the Laplace method [49]. Their principal disadvantage is their inability to cope with multimodal probability functions.

- Spline approximations: tractable approximations of the probability function may be proposed on a sufficiently refined partition of the support. The computational load associated with integrations typically increases exponentially with the number of dimensions.

- ***Maximum Entropy*** (MaxEnt) and moment matching: the approximating distribution may be chosen to match a selected set of the moments of the true distribution [50]. Under the MaxEnt principle [33], the optimal such moment-matching distribution is the one possessing maximum entropy subject to these moment constraints.

- Empirical approximations: a random sample is generated from the probability function, and the distributional approximation is simply a set of point masses placed at these independent, identically-distributed (i.i.d.) sampling points. The key technical challenge is efficient generation of i.i.d. samples from the true distribution. In recent years, stochastic sampling techniques [51] particularly the class known as ***Markov Chain Monte Carlo*** (MCMC) methods [52] have overtaken deterministic methods as the golden standard for distributional approximation. They can yield approximations to an arbitrary level of accuracy, but typically incur major computational overheads. It can be instructive to examine the performance of any deterministic method— such as the VB method— in terms of the accuracy-vs-complexity trade-off achieved by these stochastic sampling techniques.

The VB method has the potential to offer an excellent trade-off between computational complexity and accuracy of the distributional approximation. This is suggested in Fig. 3.1. The main computational burden associated with the VB method is the need to solve iteratively via the IVB algorithm a set of simultaneous equations in order to reveal the required moments of the VB-marginals. If computational cost is of concern, VB-marginals may be replaced by simpler approximations, or the evaluation of moments can be approximated, without, hopefully, diminishing the overall quality of approximation significantly. This pathway of approximation is suggested by the dotted arrow in Fig. 3.1 , and will be traversed in some of the signal processing applications presented in chapter. Should the need exist to increase accuracy, the VB method is sited in the flexible context of Mean Field Theory, which offers more sophisticated techniques that might be explored.

3.2 How to Choose a Distributional Approximation

It will be convenient to classify all distributional approximation methods into one of two types:

First : Deterministic distributional approximations: the approximation, $\tilde{P}(\theta|Y)$ (3.8), is obtained by application of a deterministic rule; i.e $\tilde{P}(\theta|Y)$ is uniquely determined by $P(\theta|Y)$. The following are deterministic methods of distributional approximation: (i) ***certainty equivalence*** [48], which includes maximum likelihood and Maximum a posteriori (MAP) point inference [53] as special cases; (ii) the ***Laplace approximation*** [54] (iii) the ***Maximum Entropy*** (MaxEnt) approximation [55], [56] and (iv) fixed-form minimization [50].

Second : Stochastic distributional approximations: the approximation is developed via a random sample of realizations from $P(\theta|Y)$. The fundamental distributional approximation in this class is the empirical distribution from nonparametric statistics [57]. The main focus of attention is on the numerically efficient generation of realizations

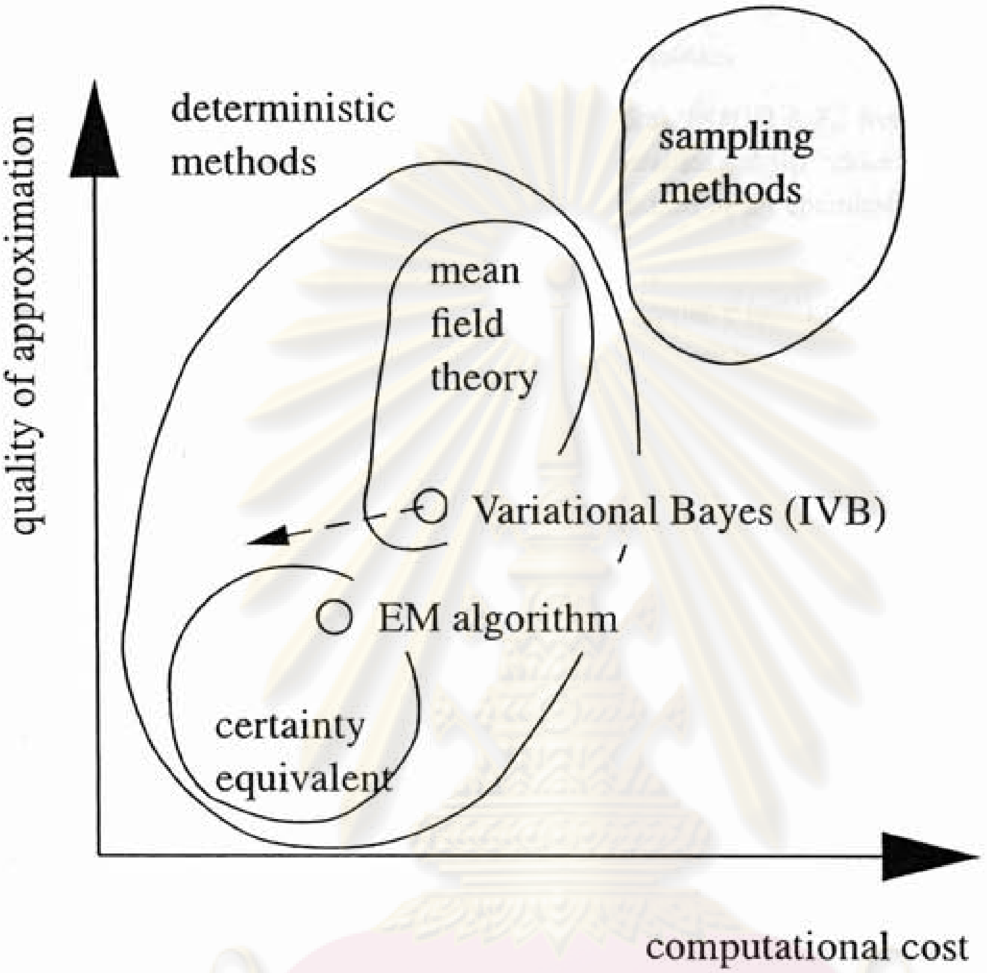


Figure 3.1 An transmitter and receiver structure

from $P(\theta|Y)$. An immediate consequence of stochastic approximation is that .

$$\theta^{(i)} \sim P(\theta|Y) \tag{3.10}$$

$$\{\theta\}_n = \{\theta^{(1)}, \ldots, \theta^{(n)}\} \tag{3.11}$$

$$\tilde{P}(\theta|Y) = \frac{1}{n}\sum_{i=1}^{n} \delta\left(\theta - \theta^{(i)}\right) \tag{3.12}$$

$\tilde{P}(\theta|Y)$ will vary with repeated use of the method. We briefly review this class of approximations in this chapter. Our main focus of attention will be the ***Variational Bayes*** (VB) approximation, which as we will see in dissertation is a deterministic, free-form distributional approximation.

3.2.1 Distributional Approximation as an Optimization Problem

In general, the task is to choose an optimal distribution, $\tilde{P}(\theta|Y) \in \mathbb{F}_c$, from the space, $\mathbb{F}$, of all possible distributions. $\tilde{P}(\theta|Y)$ should be (i) tractable, and (ii) "close" to the true posterior ,$P(\theta|Y)$, in some sense. The task can be formalized as an optimization problem requiring the following elements:

1. A subspace of distributions, $\mathbb{F}_c \subset \mathbb{F}$, such that all functions,$\tilde{P}(\theta|Y) \in \mathbb{F}_c$, are regarded as tractable Here, $\tilde{P}(\theta|Y)$ denotes a "wildcard" or *candidate tractable distribution* from the space $\mathbb{F}_c$.
2. A proximity measure,$\Delta\left(P(\theta|Y)\|\,\tilde{P}(\theta|Y)\right)$, between the true distribution and any tractable approximation. $\Delta\left(P(\theta|Y)\|\,\tilde{P}(\theta|Y)\right)$ must be defined on $\mathbb{F} \times \mathbb{F}_c$, such that it accept two distributions, $P(\theta|Y) \in \mathbb{F}$, and $\tilde{P}(\theta|Y) \in \mathbb{F}_c$, as input arguments, yield a positive scalar as its value, and have $\tilde{P}(\theta|Y) = P(\theta|Y)$ as its (unique) minimizer.Then, the optimal choice of the approximating function must satisfy

$$\breve{P}(\theta|Y) = \underset{\tilde{P}(\theta|Y)\in\mathbb{F}_c}{\arg\min}\ \Delta(P(\theta|Y)\|\,\tilde{P}(\theta|Y)) \tag{3.13}$$

where we denote the optimal distributional approximation by $\breve{P}(\theta|Y)$

3.2.2 The Bayesian Approach to Distributional Approximation

From the Bayesian point of view, choosing an approximation , $\tilde{P}(\theta|Y)$ (3.9), can be seen as a decision-making problem (3.4). Hence, the designer chooses a loss function (3.13) measuring the loss associated with choosing each possible $\tilde{P}(\theta|Y) \in \mathbb{F}_c$, when the true distribution is $P(\theta|Y)$. In [60], a logarithmic loss function was shown to be optimal if we wish to extract maximum information from the data. Use of the logarithmic loss function leads to the ***Kullback-Leibler*** (KL) divergence [29] (also known as the ***cross-entropy***) as an appropriate assignment for Δ in (3.13):

$$\Delta\left(P(\theta|Y)\|\,\tilde{P}(\theta|Y)\right) = KL\left(P(\theta|Y)\|\,\tilde{P}(\theta|Y)\right) \tag{3.14}$$

The ***Kullback-Leibler*** (KL) divergence from $P(\theta|Y)$ to $\tilde{P}(\theta|Y)$ is defined as:

$$\begin{aligned} KL\left(P(\theta|Y)\|\,\tilde{P}(\theta|Y)\right) &= \int_{\Theta^*} P(\theta|Y)\ln\frac{P(\theta|Y)}{\tilde{P}(\theta|Y)}d\theta \\ &= E_{P(\theta|Y)}\left[\ln\frac{P(\theta|Y)}{\tilde{P}(\theta|Y)}\right] \end{aligned} \tag{3.15}$$

It has the following properties:

1. $KL\left(P(\theta|Y)\|\,\tilde{P}(\theta|Y)\right) \geq 0$.

2. $KL\left(P(\theta|Y)\|\tilde{P}(\theta|Y)\right)=0$ iff $P(\theta|Y)=\tilde{P}(\theta|Y)$ almost everywhere.

3. $KL\left(P(\theta|Y)\|\tilde{P}(\theta|Y)\right)=\infty$ iff on a set of a positive measure $P(\theta|Y)>0$ and $\tilde{P}(\theta|Y)=0$.

4. $KL\left(P(\theta|Y)\|\tilde{P}(\theta|Y)\right)\neq KL\left(\tilde{P}(\theta|Y)\middle\|P(\theta|Y)\right)$ in general, and the KL divergence does not obey the triangle inequality.

Given 4, care is needed in the syntax describing $KL(\cdot)$. We say that (3.15) is from $P(\theta|Y)$ to $\tilde{P}(\theta|Y)$. This distinction will be important in what follows. For future purposes, we therefore distinguish between the two possible orderings of the arguments in the KL divergence:

KL divergence for Minimum Risk (MR) calculations, as defined in (3.15):

$$KLD_{MR}=KL\left(P(\theta|Y)\|\tilde{P}(\theta|Y)\right) \tag{3.16}$$

KL divergence for Variational Bayes (VB) calculations:

$$KLD_{VB}=KL\left(\tilde{P}(\theta|Y)\middle\|P(\theta|Y)\right) \tag{3.17}$$

The notations KLD_{MR} and KLD_{VB} imply the order of their arguments, which are, therefore, not stated explicitly.

3.2.3 The Variational Bayes (VB) Method of Distributional Approximation

The Variational Bayes (VB) method of distributional approximation is an optimization technique with the following elements: The space of tractable distributions Fc is chosen as the space of conditionally independent distributions:

$$\mathbb{F}_c\equiv\{P(\theta_1,\theta_2|Y):P(\theta_1,\theta_2|Y)=P(\theta_1|Y)P(\theta_2|Y)\} \tag{3.18}$$

A necessary condition for applicability of the VB approximation is therefore that Θ be multivariate. The proximity measure is assigned as (3.17):

$$\Delta\left(P(\theta|Y)\|\tilde{P}(\theta|Y)\right)=KL\left(\tilde{P}(\theta|Y)\middle\|P(\theta|Y)\right)=KLD_{VB} \tag{3.19}$$

Since the divergence, KLD_{MR} (3.15),(3.16), is not used, the VB approximation, $\tilde{P}(\theta|Y)$, defined from (3.2) and (3.8), is not the minimum Bayes risk distributional approximation. A schematic illustrating the VB method of distributional approximation is given in 3.2.

3.3 Other Deterministic Distributional Approximations

3.3.1 The Certainty Equivalence Approximation

In many engineering problems, full distributions $P(\theta|Y)$ are avoided. Instead, a point estimate, $\tilde{\theta}$, is used to summarize the full state of knowledge expressed by the posterior

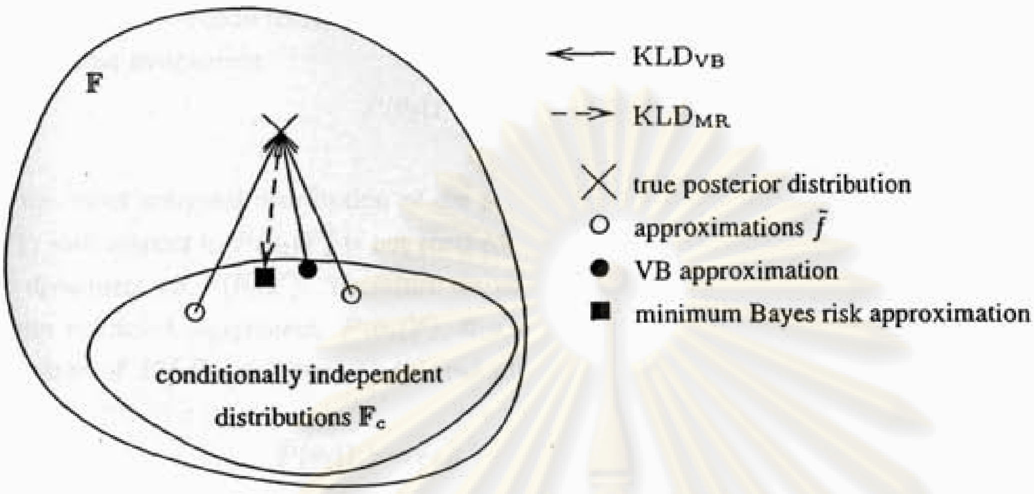


Figure 3.2: Schematic illustrating the VB method of distributional approximation. The minimum Bayes' risk approximation is also illustrated for comparison.

distribution The point estimate, $\tilde{\theta} = \tilde{\theta}(Y)$, can be interpreted as an extreme approximation of the posterior distribution, replacing $P(\theta|Y)$ by a suitably located Dirac δ function:

$$P(\theta|Y) = \tilde{P}(\theta|Y) = \delta\left(\theta - \tilde{\theta}(Y)\right) \tag{3.20}$$

where $\tilde{\theta}$ is the chosen point estimate of parameter θ. The approximation (3.20) is known as the certainty equivalence principle [30], and we have already encountered it in the QB (***Quasi-Bayes***) and EM (***Expectation-Maximization***) approximations for next section. It remains to determine an optimal assignment for the point estimate. The Bayesian decision-theoretic framework for design of point estimates was mentioned, where, also, popular choices such as the MAP, ML and mean a posteriori estimates were reviewed.

3.3.1.1 The Quasi-Bayes (QB) Approximation

Let we rewrite the KL divergence in (3.17) of the model parameters $\theta = [\theta_1, \theta_2]$ as follows :

$$\begin{aligned} KL\left(\tilde{P}(\theta|Y)\middle\| P(\theta|Y)\right) &= \int_{\Theta^*} \tilde{P}(\theta_1|Y)\tilde{P}(\theta_2|Y)\ln\frac{\tilde{P}(\theta_1|Y)\tilde{P}(\theta_2|Y)}{P(\theta_1|\theta_2,Y)P(\theta_2|Y)}d\theta \\ &= \int_{\Theta^*} \tilde{P}(\theta_1|Y)\tilde{P}(\theta_2|Y)\ln\frac{\tilde{P}(\theta_1|Y)}{P(\theta_1|\theta_2,Y)}d\theta + \\ &\quad + \int_{\Theta^*} \tilde{P}(\theta_2|Y)\ln\frac{\tilde{P}(\theta_2|Y)}{P(\theta_2|,Y)}d\theta \end{aligned} \tag{3.21}$$

We note that the second term in (3.21) is $KL\left(\tilde{P}(\theta|Y)\middle\| P(\theta|Y)\right)$, which is minimized for the restricted assignment

$$\tilde{P}(\theta_2|Y) = \int_{\Theta_1^*} P(\theta|Y)d\theta_1 \tag{3.22}$$

i.e. the exact marginal distribution of the joint posterior $P(\theta|Y)$. The global minimum of (3.21) with respect to $\tilde{P}(\theta_2|Y)$ is not reached for this choice, since the first term in (3.21) is also dependent on $\tilde{P}(\theta_2|Y)$. Therefore we consider (3.22) to be the best analytical choice for the restricted assignment, $\tilde{P}(\theta_2|Y)$, that we can make. It is also consistent with the minimizer of KLD_{MR}. From (3.49) and (3.22), the ***Quasi-Bayes*** (QB) approximation is therefore

$$\tilde{P}(\theta_1|Y) = \exp\left(E_{\tilde{P}(\theta_2|Y)}\left[\ln\left(P(\theta, Y)\right)\right]\right) \tag{3.23}$$

The name Quasi-Bayes (QB) was first used in the context of finite mixture models [32], to refer to this type of approximation. In [50], the marginal for θ_1 was approximated by conditioning the joint posterior distribution on $\bar{\theta}_2$, which was assigned as the true posterior mean of θ_2:

$$\bar{\theta}_2 = E_{P(\theta_2|Y)}\left[\theta_2\right] \tag{3.24}$$

In this case, therefore, the approximation is the following conditional distribution:

$$\tilde{P}(\theta_1|Y) \equiv P(\theta_1|\,\bar{\theta}_2, Y) \tag{3.25}$$

3.3.1.2 The Expectation-Maximization (EM) Algorithm

The ***Expectation-Maximization*** (EM) algorithm is a well known algorithm for Maximum Likelihood (ML) estimation and extension for MAP estimation of subset 2 of the model parameters $\theta = [\theta_1, \theta_2]$. Here, we follow an alternative derivation of EM via distributional approximations. The task is to estimate parameter θ_2 by maximization of the (intractable) marginal posterior distribution:

$$\hat{\theta}_2 = \arg\max_{\theta_2} P(\theta_2|Y) \tag{3.26}$$

This task can be reformulated as an optimization problem for the constrained distributional family:

$$\mathbf{F}_c \equiv \left\{P(\theta_1, \theta_2|Y) : P(\theta_1, \theta_2|Y) = P(\theta_1|\theta_2, Y)\delta(\theta_2 - \hat{\theta}_2)\right\} \tag{3.27}$$

Here, $\delta(\cdot)$ denotes the Dirac $\delta-$function,

$$\int_X \delta(x - x_0)g(x)dx = g(x_0) \tag{3.28}$$

if $x \in \mathrm{X}$ is a continuous variable, and the Kronecker function,

$$\delta(x) = \begin{cases} 1, & if \;\; x = 0 \\ 0, & otherwise \end{cases} \tag{3.29}$$

if x is integer. We optimize over the family $\mathbb{F}_c$ with respect to KLD_{VB}. Hence, we recognize that this method of distributional approximation is a special case of the VB approximation, with the functional restrictions $\tilde{P}(\theta_1|Y) = P(\theta_1|\theta_2, Y)$ and $\tilde{P}(\theta_2|Y) = \delta(\theta_2 - \tilde{\theta}_2)$. The functional optimization for $\tilde{P}(\theta_1|Y)$ is trivial since in all moments and expectations with respect to $\delta(\theta_2 - \tilde{\theta}_2)$ simply result in replacement of θ_2 by $\tilde{\theta}_2$ The resulting distributional algorithm is then a cyclic iteration (alternating algorithm) of two steps:

E-step: Compute the approximate marginal distribution of θ_2, at iteration i:

$$\tilde{P}^{[i]}(\theta_1|Y) = P(\theta_1|\theta_2^{[i-1]}, Y). \tag{3.30}$$

M-step: Use the approximate marginal distribution of θ_1 from the E-step to update the certainty equivalent for θ_2 :

$$\tilde{\theta}_2^{[i]} = \arg\max_{\theta_2} \int_{\Theta_1^*} \tilde{P}^{[i]}(\theta_1|Y) \ln P(\theta_1, \theta_2, Y) d\theta_1 \tag{3.31}$$

In the context of uniform priors (i.e. ML estimation), it was proved in [61] that this algorithm monotonically increases the marginal likelihood, $P(Y|\theta_2)$, of θ_2, and therefore converges to a local maximum [61].

3.3.2 The Laplace Approximation

This method is based on local approximation of the posterior distribution, $P(\theta|Y)$, around its MAP estimate, $\tilde{\theta}$, using a Gaussian distribution. Formally, the posterior distribution is approximated as follows:

$$P(\theta|Y) \approx N\left(\tilde{\theta}, H^{-1}\right) \tag{3.32}$$

$\tilde{\theta}$ is the MAP estimate (3.4) of $\theta \in \mathbb{R}^p$ and $H \in \mathbb{R}^{p\times p}$ is the (negative) Hessian matrix of the logarithm of the joint distribution, $P(\theta, Y)$, with respect to θ evaluated at $\theta = \tilde{\theta}$:

$$H = -\left[\frac{\partial^2 \ln(\theta, Y)}{\partial\theta_i \partial\theta_j}\right]_{\theta=\tilde{\theta}} \qquad i, j = 1, \ldots p \tag{3.33}$$

The asymptotic error of approximation was studied in [49].

3.3.3 The Maximum Entropy (MaxEnt) Approximation

The ***Maximum Entropy*** Method of distributional approximation is a freeform method in common with the VB method, since a known distributional form is not stipulated a priori. Instead, the approximation $\tilde{P}(\theta|Y) \in \mathbb{F}_c$ is chosen which maximizes the entropy,

$$\begin{aligned} H &= -\int_{\Theta_1^*} P(\theta|Y) \ln P(\theta|Y) d\theta \\ &= -E_{P(\theta|Y)}\left[\ln P(\theta|Y)\right] \end{aligned} \tag{3.34}$$

constrained by any known moments

$$m_i = \bar{g}(\theta) = E_{P(\theta|Y)}[g(\theta)] = \int_{\Theta^*} g(\theta)P(\theta|Y)d\theta \tag{3.35}$$

In the context of MaxEnt, (3.35) are known as the mean constraints. The MaxEnt distributional approximation is of the form

$$\tilde{P}(\theta|Y) \propto \exp\left(-\sum_i \alpha_i(Y)g_i(\theta)\right) \tag{3.36}$$

where the α_i are chosen using, for example, the method of Lagrange multipliers for constrained optimization.to satisfy the mean constraints (3.35) and the normalization requirement for $\tilde{P}(\theta|Y)$. Since its entropy (3.3.3) has been maximized, (3.36) may be interpreted as the smoothest (minimally informative) distribution matching the known moments of f (3.35). The MaxEnt approximation has been widely used in solving inverse problems [62]– [63] notably in reconstruction of non-negative data sets, such as in Burg's method for power spectrum estimation [64] and in image reconstruction [65].

3.3.4 Stochastic Distributional Approximations

A stochastic distributional approximation maps f to a randomly-generated approximation, $\tilde{P}(\theta|Y)$, in contrast to all the methods we have reviewed so far, where $\tilde{P}(\theta|Y)$ is uniquely determined by $P(\theta|Y)$ and the rules of the approximation procedure. The computational engine for stochastic methods is therefore the generation of an independent, identically-distributed (i.i.d.) sample set (i.e. a random sample),

$$\begin{aligned} \theta^{(i)} &\sim P(\theta|Y) \\ \{\theta\}_n &= \{\theta^{(1)}, \ldots, \theta^{(n)}\} \\ \tilde{P}(\theta|Y) &= \frac{1}{n}\sum_{i=1}^{n} \delta\left(\theta - \theta^{(i)}\right) \end{aligned} \tag{3.37}$$

The posterior moments (3.35) of $P(\theta|Y)$ under the empirical approximation (3.37) are therefore

$$E_{\tilde{P}(\theta|Y)}[g_j(\theta)] = \frac{1}{n}\sum_{i=1}^{n} g_j\left(\theta^{(i)}\right) \tag{3.38}$$

Note that marginal distributions and measures are also generated with ease under approximation (3.37) via appropriate summations.

For low-dimensional θ it may be possible to generate the i.i.d. set $\{\theta\}_n$ using one of a vast range of standard stochastic sampling methods. The real challenge being addressed by modern stochastic sampling techniques is to generate a representative random sample (3.37) for difficult notably high dimensional distributions. ***Markov Chain Monte Carlo*** (MCMC) methods refer to a class of stochastic sampling algorithms that generate a correlated sequence of samples $\left\{ \theta^{(0)} \quad \theta^{(1)} \quad \ldots \quad \theta^{(k)} \quad \ldots \right\}$, from a first-order (Markov) kernel, $P(\theta^{(k)}|\theta^{(k-1)}, Y)$. For mild regularity conditions on $P(\cdot|\cdot)$ then $\theta^{(k)} \sim P_s(\theta|Y)$ and $k \longrightarrow \infty$ where $\theta^{(k)} \sim P_s(\theta|Y)$ is the stationary distribution of the Markov process with this kernel. This convergence in distribution is independent of the initialization ,$\theta^{(0)}$, of the Markov chain. Careful choice of the kernel can ensure that $P_s(\theta|Y) = P(\theta|Y)$. Hence, repeated simulation from the Markov process, $i = 1, 2, \ldots$, with n sufficiently large, generates the required random sample (3.37) for construction of the empirical approximation (3.37). Typically, the associated computational burden is large, and can be prohibitive in cases of high-dimensional θ. Nevertheless, the very general and flexible way in which MCMC methods have been defined has helped to establish them as the golden standard for (Bayesian) distributional approximation. In the online scenario, sequential Monte Carlo techniques, such as particle filtering, have been developed for recursive updating of the empirical distribution (3.37) via MCMC-based sampling.

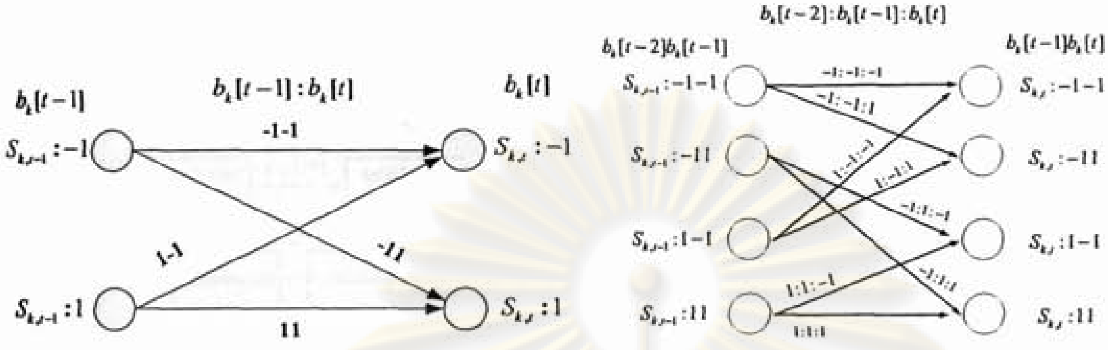
3.4 Signal Decomposition Model

The problem of parameter estimation for separable signals is encountered in a wide range of signal processing, sonar, and communication applications. Most such problems can be posed using the following measurement model $r = \sum_{k=1}^{K} x_k$. In [16, pp.2670−2671] and [28], the SAGE algorithm for the special case of separable signal in Gaussian noise suggests a specific way to decompose observation vector (2.6) into a number of signal components. The set of each decomposed signal components is so called admissible hidden data space. Following [28], we assume the received signal is separable and comprises of a set of variables of an admissible hidden data space. Specifically we formulate the following equation.

$$\begin{aligned} \mathbf{r}[t] &= \sum_{k=1}^{K} x_k[t] \ , \ \ x_k[t] = D_k(\underline{b}_{k,t}) g_k + w_k[t] \ , \\ w[t] &= \sum_{k=1}^{K} w_k[t] \ , \ \ w_k[t] \sim \mathcal{N}(0, \sigma_k^2). \end{aligned} \tag{3.39}$$

The noise vectors $w_k[t]$ are mutually independent circular Gaussian, with covariance $\sigma_k^2 = \beta_k \sigma^2$. where β_k ,that is defined in [28], is a scalar probability. The β_k is a free parameter satisfying $0 \leq \beta_k \leq 1$, such that $\sum_{k=1}^{K} \beta_k = 1$. In principle SAGE algorithm [16] and [28], the separable signal can be estimated by

$$x_k[t] = D_k(\underline{b}_{k,t}) g_k + \beta_k \left[\mathbf{r}[t] - \sum_{l=1}^{K} D_k(\underline{b}_{l,t}) g_l \right]. \tag{3.40}$$



(a) Virtual trellis representing ISI (L=2)

(b) Virtual trellis representing ISI (L=3)

Figure 3.3 Virtual Trellis Diagram Based on Signal Decomposition

3.5 Virtual Trellis Diagram Based on Signal Decomposition

To generate the extrinsic information by using the BCJR algorithm [25] for the separable signal, we first focus on a particular k^{th} separable signal $x_k[t]$ in (3.40). The informations $\underline{b}_{k,t} = (b_k[t-L+1], \ldots, b_k[t-1], b_k[t])$ bits of $D_k(\underline{b}_{k,t})$ are composed of the signal $x_k[t]$. Without encoding, the information $\underline{b}_{k,t}$ of $D_k(\underline{b}_{k,t})$ are modeled as corresponding output of virtual trellis that cause the state translation from $S_{k,t-1} = (\ b_k[t-L+1]b_k[t-L+2]\ldots b_k[t-1]\ )$ to $S_{k,t} = (\ b_k[t-L+2]b_k[t-L+3]\ldots b_k[t]\ )$ as shown in Fig 3.3. In this technique, Markov model representing the ISI channel with $2^{(L-1)}$ states is combined with each one of separable signal $x_k[t]$. The calculation of the extrinsic information utilized for each separable signal $x_k[t]$ is obtained by applying the BCJR algorithm via the virtual trellis representing the ISI channel.

3.6 Iterative Receiver Structure

The block diagram of transmitter and receiver model is shown in Fig. 3.4. The binary information $\{u_k[i]\}$ for user $k, k = 1, .., K$ are convolutionally encoded by a single convolution encoder with code rate R_k. The code bits $\{b_k[m]\}$ are interleaved and mapped to BPSK symbols stream. Each data symbol $\{b_k[i]\}$ is modulated by a spreading waveform $s_k(t)$ [21] and transmitted through the multipath channel. The received signal vector $\mathbf{r}[t]$ at the receiver is the superposition of K user signals plus additive white Gaussian noise, which is defined in section 2.1. In the coded communication systems, the iterative receiver structure often makes use of turbo principle [2]- [13] to reduce the loss of information and performance. The receiver structure under consideration is a iterative receiver as shown in Fig 3.4. It consists of two states the ***Minimum Kullback-Leibler (MKL)*** **multiuser detector**, followed by K parallel *channel decoders*. The two states are separated by deinterleavers and interleavers. The MKL multiuser detector computes both iterative channel estimation

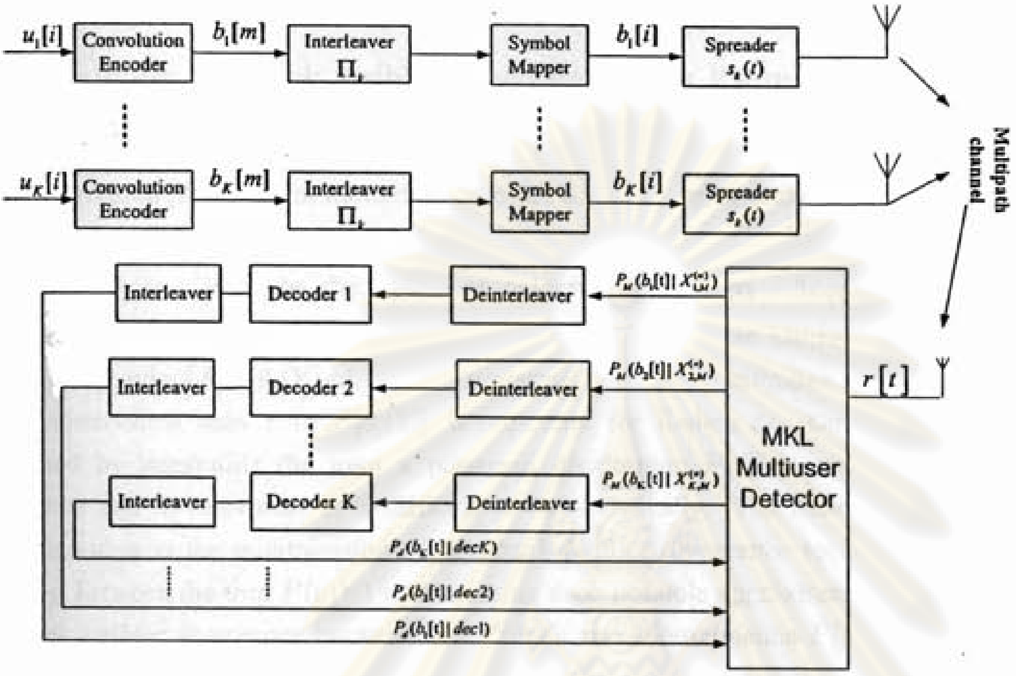


Figure 3.4 An transmitter and receiver structure

and extrinsic probability approximation for each separable signal $x_k[t]$ by using MKL algorithm and virtual trellis model. The extrinsic probability $P_M\left(b_k[t]|X_{k,M}^{(n)}\right)$ employed in MKL multiuser detector is the modified BCJR algorithm [25] for numerical Minimum Kullback-Leibler solution (please see section 5. and appendix A).

$$\begin{aligned}
P_M(b_k[t]|X_{k,M}^{(n)}) &= \frac{\sum\limits_{(S_{k,t-1}:S_{k,t}):b_k[t]} \alpha_{t-1}(S_{k,t-1})\gamma_{e,t}(S_{k,t},S_{k,t-1})\beta_t(S_{k,t})}{\sum\limits_{S_{k,t}}\sum\limits_{S_{k,t-1}} \alpha_{t-1}(S_{k,t-1})\gamma_{e,t}(S_{k,t},S_{k,t-1})\beta_t(S_{k,t})} \\
\gamma_{e,t}(S_{k,t},S_{k,t-1}) &= P(x_k^{(n)}[t]|\underline{b}_{k,t})\prod_{l=1}^{L-1} P_d(b_k^{(n-1)}[t-l]\,|\,dec\,k) \\
P(x_k^{(n)}[t]|\underline{b}_{k,t}) &= \frac{1}{\sigma_k^{2N(n-1)}}\exp\left(\frac{-||x_k^{(n)}[t]-D_k(\underline{b}_{k,t})g_k^{(n-1)}||^2}{\sigma_k^{2(n-1)}}\right)
\end{aligned} \tag{3.41}$$

The extrinsic information $P_M\left(b_k[t]|X_{k,M}^{(n)}\right)$ at the current $(n)^{th}$ iteration is deinterleaved and delivered to the k^{th} channel decoder. Each user's channel decoder base on MAP decoding algorithm [13], [25]. It computes the extrinsic information $P_d\left(b_k^{(n)}[t]\,|\,dec\,k\right)$ for each k^{th} user encode bits and the information is fed to the MKL multiuser detector as as priori information for $(n+1)^{th}$ iteration. Then, the next iteration the MKL multiuser detector will receive data with more accurate prior information about the transmitted symbols.

3.7 Minimum Kullback-Leibler Based Parameter Estimation

The MKL multiuser detector developed in this paper base on ***Minimum Kullback-Leibler*** algorithm [29], [30] for numerical Bayesian distributional approximation. In our batch processing, we denote the observation corresponding to M time slot $Y = \{\mathbf{r}[1], \ldots, \mathbf{r}[M]\}$. Let $\theta = \left[\theta_1, \theta_2, \cdots, \theta_d\right]^T$ be the vector of unknown parameters. Suppose that we are interested in finding estimation of $\bar{\theta}_j$. Based on the condition mean estimator [31], the $\bar{\theta}_j$ is given by $E[\theta_j|y]=\int \theta_j P[\theta_j|Y]d\theta_j$. The a posteriori marginal distribution of θ_j, conditional on the observation data Y i.e.$P[\theta_j|Y]$, is important for finding $\bar{\theta}_j$. Direct calculation is performed by integrating the joint a posteriori distribution $P[\theta|Y]$ over the rest of the parameters and in most case it is difficult to do analytically. The basic idea behind the MKL algorithm is the minimization of ***Kullback-Leibler divergence*** for variational Bayes KLD_{VB} between the true $P(\theta|Y)$ and choosing each possible approximation $\bar{P}(\theta|Y)$. The ***Kullback-Leibler divergence*** between true $P(\theta|Y)$ and approximation $\bar{P}(\theta|Y)$ (also known as the cross-entropy) is defined by non-negative cost function as follows:

$$KL(\bar{P}(\theta|Y)||P(\theta|Y)) = \int_\theta \bar{P}(\theta|Y) \ln \frac{\bar{P}(\theta|Y)}{P(\theta|Y)} d\theta \tag{3.42}$$

A strategy for the MKL algorithm is to choose possible $\bar{P}(\theta|Y)$ with a suitable form close to the true a posteriori $P(\theta|Y)$. The task can be formalized as an optimization problem. A logarithmic loss function (3.42) was shown to be optimal ,i.e., $\bar{P}(\theta|Y) \triangleq P(\theta|Y)$. If we wish to minimum information of loss data (3.42). Then, the minimum of KLD_{VB} is

$$\bar{P}(\theta|Y) = \arg\min_{\bar{P}(\theta|Y)} KL(\bar{P}(\theta|Y)||P(\theta|Y)) \tag{3.43}$$

However, in some case the true a posteriori distribution $P(\theta|Y)$ is often computationally intractable, especially in higher dimensions θ. This problem can be overcome by approximating the true a posterior distribution $P(\theta|Y)$ by a distribution that is computationally tractable $\bar{P}(\theta|Y)$.

For the a posteriori distribution approximation $\bar{P}(\theta|Y)$ discussed above, the task is to choose an optimal distribution $\bar{P}(\theta|Y)$ by minimizing KLD_{VB} in (3.43). Let distribution approximation $\bar{P}(\theta|Y)$ restrict to the set of conditionally independent for $\theta_1, \theta_2, ..., \theta_d$.

$$\bar{P}(\theta|Y) = \bar{P}(\theta_1, \theta_2, ..., \theta_d|Y) = \prod_{j=1}^{d} \bar{P}(\theta_j|Y). \tag{3.44}$$

Then, the minimum of KLD_{VB} for a posteriori marginal distribution $\bar{P}(\theta_j|Y)$ is

$$\bar{P}(\theta_j|Y) = \arg\min_{\bar{P}(\theta_j|Y)} KL(\bar{P}(\theta|Y)||P(\theta|Y)) \tag{3.45}$$

Using this strategy in (3.45), we can compute $\tilde{P}(\theta_j|Y)$ by minimizing the KLD_{VB} cost function.

$$
\begin{aligned}
KL(\tilde{P}(\theta|Y)||P(\theta|Y)) &= \int_{\Theta^*} \tilde{P}(\theta_j|Y)\tilde{P}(\theta_{\setminus j}|Y) \ln \frac{\tilde{P}(\theta_j|Y)\tilde{P}(\theta_{\setminus j}|Y)P(Y)}{P(\theta|Y)P(Y)} d\theta \\
&= \int_{\Theta^*} \tilde{P}(\theta_j|Y)\tilde{P}(\theta_{\setminus j}|Y) \ln \tilde{P}(\theta_j|Y) d\theta \\
&- \int_{\Theta^*} \tilde{P}(\theta_j|Y)\tilde{P}(\theta_{\setminus j}|Y) \ln P(\theta, Y) d\theta \\
&+ \int_{\Theta^*} \tilde{P}(\theta_j|Y)\tilde{P}(\theta_{\setminus j}|Y) [\ln \tilde{P}(\theta_{\setminus j}|Y) + \ln P(Y)] d\theta
\end{aligned}
$$

$$
\begin{aligned}
KL(\tilde{P}(\theta|Y)||P(\theta|Y)) &= \int_{\Theta_j^*} \tilde{P}(\theta_j|Y) \ln \tilde{P}(\theta_j|Y) \int_{\Theta_{\setminus j}^*} \tilde{P}(\theta_{\setminus j}|Y) d\theta_{\setminus j} d\theta_j + \ln P(Y) \\
&+ \underbrace{E_{\tilde{P}(\theta_{\setminus j}|Y)} \left(\ln \tilde{P}(\theta_{\setminus j}|Y) \right)}_{Const.} \\
&- \int_{\Theta_j^*} \tilde{P}(\theta_j|Y) \underbrace{\int_{\Theta_{\setminus j}^*} \tilde{P}(\theta_{\setminus j}|Y) \ln P(\theta, Y) d\theta_{\setminus j}}_{E_{\tilde{P}(\theta_{\setminus j}|Y)}[\ln P(\theta,Y)]} d\theta_j \\
&= \int_{\Theta_j^*} \tilde{P}(\theta_j|Y) \left[\ln \tilde{P}(\theta_j|Y) - \ln \left\langle \exp E_{\tilde{P}(\theta_{\setminus j}|Y)} [\ln P(\theta, Y)] \right\rangle \right] d\theta_j \\
&+ Const. \\
&= KL\left(\tilde{P}(\theta_j|Y) \,||\, \exp E_{\tilde{P}(\theta_{\setminus j}|Y)} [\ln P(\theta, Y)] \right) + Const. \quad (3.46)
\end{aligned}
$$

Only the first term on the right-hand side of (3.46) is dependent on $\tilde{P}(\theta_j|Y)$. Hence, minimization of (3.46) with respect to $\tilde{P}(\theta_j|Y)$, keeping $\tilde{P}\left(\theta_{\setminus j}|Y\right)$ fixed, is achieved by minimization of the first term. Invoking non-negativity property [30] of KLD_{VB} on the first term of (3.46), the minimizer is almost surely that

$$
\tilde{P}(\theta_j|Y) \propto \exp\left(E_{\tilde{P}(\theta_{\setminus j}|Y)} [\ln P(\theta, Y)] \right). \quad (3.47)
$$

where $\theta_{\setminus j}$ denotes the complement of θ_j in θ. Note that $\setminus$ denote the exclusive operator. To estimate any a posteriori marginal distribution using the MKL algorithm, let we given the samples at iteration (l)

$$
\tilde{P}(\theta^{(l)}|Y) = \tilde{P}(\theta_1^{(l)}, \theta_2^{(l)}, \ldots \theta_d^{(l)}|Y) = \prod_{i=1}^{d} \tilde{P}(\theta_i^{(l)}|Y) \quad (3.48)
$$

The $\bar{P}(\theta^{(l+1)}|Y)$ is estimated from the following step

$$\begin{aligned}
\bar{P}(\theta_1^{(l+1)}|Y) &= \exp(E_{\bar{P}(\theta_2^{(l)},\ldots,\theta_d^{(l)}|Y)}[\ln P(\theta,Y)]) \\
\bar{P}(\theta_2^{(l+1)}|Y) &= \exp(E_{\bar{P}(\theta_1^{(l+1)},\theta_3^{(l)},\ldots,\theta_d^{(l)}|Y)}[\ln P(\theta,Y)]) \\
&\vdots \\
\bar{P}(\theta_d^{(l+1)}|Y) &= \exp(E_{\bar{P}(\theta_1^{(l+1)},\theta_3^{(l+1)},\ldots,\theta_{d-1}^{(l+1)}|Y)}[\ln P(\theta,Y)]) \\
\bar{P}(\theta^{(l+1)}|Y) &= \bar{P}(\theta_1^{(l+1)},\theta_2^{(l+1)},\ldots\theta_d^{(l+1)}|Y) = \prod_{i=1}^{d} \bar{P}(\theta_i^{(l+1)}|Y).
\end{aligned} \tag{3.49}$$

3.8 Prior Distribution

In principle, the prior distributions are used to incorporate the prior knowledge about the unknown parameters. Prior should be employed when such knowledge is limited. For MKL algorithm, we need to be able to compute a posteriori marginal probability distribution from prior and conditional posteriori distributions. A well-known strategy for MKL computation is to choose the prior distribution with a suitable form so that the posteriori belong to be the same function family as prior. The prior and posteriori are then said to be conjugated [33]. The choice of functional family depends on likelihood. We will use this conjugate prior strategy throughout the paper.

1. For unknown channel g_k, a complex Gaussian prior distribution is assumed $P[g_k] \sim \mathcal{N}(g_k^{(0)}, \Sigma g_k{}^{(0)}))$.

2. For the noise variance σ_k^2, an inverse chi-square prior distribution is assumed $P[\sigma_k^2] \sim (\frac{1}{\sigma_k^2})^{(\alpha_{(k,0)}-1)} \exp(\frac{-\beta_{(k,0)}}{\sigma_k^2}) \sim \chi^{-2}\left(\alpha_{(k,0)}, \beta_{(k,0)}\right)$.

3.9 The Turbo Minimum Kullback-Leibler Multiuser Detector

Let us further defining the parameters $\theta = [\theta_1, \theta_2, \ldots, \theta_K]$ with set $\theta_k = \{g_k, B_k, \sigma_k^2\}$,$B_k = \{b_k[t]\}_{t=0}^{M}$ and $\sigma_k^2 = \beta_k\sigma^2$. The complement of $b_k[t]$ in B_k is $B_{k\backslash b_k[t]} = B_k\backslash b_k[i]$. The MKL multiuser detector is an efficient method to find a posteriori distribution $P[\theta^{(n)}|Y_M]$ given by observation $Y_M = \{\mathrm{r}[1], \ldots, \mathrm{r}[M]\}$ and previous estimation $\theta^{(n-1)} = [\theta_1^{(n-1)}, \ldots, \theta_K^{(n-1)}]$ parameters . In [16], the ***SAGE algorithm*** suggests the special way to update parameters. First rather than updating all parameters simultaneously at iteration n, only a subset θ_k of θ is updated while keeping the parameter in the complement $\theta_{\backslash\theta_k}$ fixed. Second the concept of admissible hidden data space [16] X_M which depend only on θ_k is selected for our algorithm. The incomplete data Y_M is related by mean of a possible nondeterministic mapping by $X_M \rightarrow Y_M(X_M)$ for each θ_k. Let we define admissible hidden data space for k^{th} parameter at iteration n by (3.40)

$$\begin{aligned} x_k^{(n)}[t] &= D_k(\underline{b}_{k,t}^{(n-1)})g_k^{(n-1)} \\ &+ \beta_k \left[r[t] - \sum_{l=1}^{k-1} D_k(\underline{b}_{k,t}^{(n)})g_k^{(n)} - \sum_{l=k}^{K} D_k(\underline{b}_{k,t}^{(n-1)})g_k^{(n-1)} \right] \end{aligned} \quad (3.50)$$

where observation $X_{k,M}^{(n)}=\{x_k^{(n)}[1], x_k^{(n)}[2], \ldots, x_k^{(n)}[M]\}$. Based on the Minimum ***Kullback-Leibler*** (MKL) notation in eq.(3.49), the a posteriori distribution approximation for the k^{th} admissible hidden data space $X_{k,M}^{(n)}$ at iteration n^{th} is given by setting $\theta_k^{(n)} = \theta_k^{(n-1)} = \{B_k^{(n-1)}, g_k^{(n-1)}, \sigma_k^{2(n-1)}\}$ and following operations

$$\begin{aligned} \tilde{P}_M(b_k^{(n)}[t]|X_{k,M}^{(n)}) &= \exp(E_{\tilde{P}(\theta^{(n)}_{k\backslash b_k^{(n-1)}[t]}|X_{k,M}^{(n)})}[\ln P(\theta_k, X_k^{(n)})]) \\ \tilde{P}(B_k^{(n)}|X_{k,M}^{(n)}) &= \prod_{t=1}^{M} \tilde{P}_M(b_k^{(n)}[t]|X_{k,M}^{(n)}) \\ Update \quad & \theta_k^{(n)} = \{g_k^{(n-1)}, B_k^{(n)}, \sigma_k^{2(n-1)}\} \\ \tilde{P}(g_k^{(n)}|X_{k,M}^{(n)}) &= \exp(E_{\tilde{P}(\theta^{(n)}_{k\backslash g_k^{(n-1)}}|X_{k,M}^{(n)})}[\ln P(\theta_k, X_k^{(n)})]) \\ Update \quad & \theta_k^{(n)} = \{g_k^{(n)}, B_k^{(n)}, \sigma_k^{2(n-1)}\} \\ \tilde{P}(\sigma_k^{(n)}|X_{k,M}^{(n)}) &= \exp(E_{\tilde{P}(\theta^{(n)}_{k\backslash \sigma_k^{2(n-1)}}|X_{k,M}^{(n)})}[\ln P(\theta_k, X_k^{(n)})]) \\ Update \quad & \theta_k^{(n)} = \{g_k^{(n)}, B_k^{(n)}, \sigma_k^{2(n)}\} \\ \tilde{P}(\theta_k^{(n)}|X_{k,M}^{(n)}) &= \tilde{P}(B_k^{(n)}|X_{k,M}^{(n)})\tilde{P}(g_k^{(n)}|X_{k,M}^{(n)})\tilde{P}(\sigma_k^{(n)}|X_{k,M}^{(n)}) \end{aligned} \quad (3.51)$$

where, the details for $P_M(b_k^{(n)}[t]|X_{k,M}^{(n)})$, $\tilde{P}(g_k^{(n)}|X_{k,M}^{(n)})$ and $\tilde{P}(\sigma_k^{(n)}|X_{k,M}^{(n)})$ in (3.51) are found in section 3.11 , 3.12 and 3.13 respectively.

We next propose a combined ***Minimum Kullback-Leibler*** (MKL) algorithm in (3.51) and channel decoding, based on BCJR algorithm in Table 1. As shown in Fig. 3.4, the new detector is called the ***Minimum Kullback-Leibler***-Based turbo multiuser detector. The MKL multiuser detector module is functionally as notation in eq.(3.51).

3.10 Initialization of the MKL Algorithm

The performance of the MKL algorithm is closely depends on the quality of the initial values of conjugate prior distribution.

$$\begin{aligned} P[g_k] &\sim \mathcal{N}(g_k^{(0)}, \Sigma {g_k}^{(0)})) \\ P[\sigma_k^2] &\sim (\frac{1}{\sigma_k^2})^{(\alpha_{(k,0)}-1)} \exp(\frac{-\beta_{(k,0)}}{\sigma_k^2}) \sim \chi^{-2}\left(\alpha_{(k,0)}, \beta_{(k,0)}\right) \end{aligned} \quad (3.57)$$

To estimate prior distribution, the data sequence is divided into a sequence of frame length M and pilot overhead of each frame known by user length $M_b = 5$. The initial conjugate priors are compute by the aid of pilot symbols. First, By assuming the perfect known pilot symbols, the initial conjugate priors $P[g_k]$ is computed by substituting known pilot symbols in (3.54). Second, the result of channel estimation $g_k^{(0)}$ and the known aid of pilot symbols are obtained to estimate conjugate priors $P[\sigma_k^2]$ in (3.55). The initial data sequence $B_k^{(0)}$ is simplified by linear MMSE estimation which is based on knowledge of $g_k^{(0)}$ and $\sigma_k^{2(0)}$. Finally, the initial conjugate priori algorithm,based on (3.54)-(3.55), is summarized in table 2.

3.11 Prove algorithm of probability distribution $P_M(b_k[t]|X_{k,M}^{(n)})$

In some engineering problems, the full stochastic distributions approximation are avoided due to intractable chosen $P(\theta_k, X_{k,M}^{(n)})$ bayesian model. Instead, the estimation points $\theta_k^{(n-1)}$ are used to summarize the full state of knownledge expressed by the joint distribution $P(\theta_k, X_{k,M}^{(n)})$. The points estimation ,$\bar{\theta}_k = \theta_k^{(n-1)}$, can be interpreted as an extreme approximation replacing by Dirac δ-function form.

$$\begin{aligned} P(\theta_k, X_{k,M}^{(n)}) &= P(\theta_k|X_{k,M}^{(n)})P(X_{k,M}^{(n)}) \\ &\propto P(\theta_k|X_{k,M}^{(n)}) = \delta(\theta_k - \theta_k^{(n-1)}) \end{aligned} \tag{3.58}$$

The approximation in eq. (3.58) is known as *certainty equivalence principle* [30]. In order to exploit extrinsic information $P_M(b_k[t]|X_{k,M}^{(n)})$ for marginal distribution using MKL algorithm, we can calculate by using the following expression

$$P_M(b_k[t]|X_{k,M}^{(n)}) = \exp(E_{\bar{P}(\theta_{k\backslash b_k[t]}^{(n-1)}|X_{k,M}^{(n)})}[\ln P(\theta_k, X_{k,M}^{(n)})]) \tag{3.59}$$

Based on the ***certainty equivalence principle***, we can utilize estimation points $\theta_{k\backslash b_k[t]}^{(n-1)}$ for the restricted marginal density function (3.59). Then, the density function for joint admissible hidden data space X_M and θ_k can be expressed as

$$\begin{aligned} P(\theta_k, X_{k,M}^{(n)}) &= P(B_k, g_k, \sigma_k^2, X_{k,M}^{(n)}) \\ &= P(b_k[t], B_{k\backslash b_k[t]}, g_k, \sigma_k^2, X_{k,M}^{(n)}) = P(b_k[t], \theta_{k\backslash b_k[t]}, X_{k,M}^{(n)}) \\ &= P(b_k[t]|\theta_{k\backslash b_k[t]}, X_{k,M}^{(n)})P(\theta_{k\backslash b_k[t]}|X_{k,M}^{(n)})P(X_{k,M}^{(n)}) \\ &\propto P(b_k[t]|\theta_{k\backslash b_k[t]}, X_{k,M}^{(n)})\delta(\theta_{k\backslash b_k[t]} - \theta_{k\backslash b_k[t]}^{(n-1)}) \\ &\propto P(b_k[t]|\theta_{k\backslash b_k[t]}^{(n-1)}, X_{k,M}^{(n)}) \end{aligned} \tag{3.60}$$

Substituting (3.60) back into (3.59), an expectation of function $\ln P(b_k[t]|\theta_{k\backslash b_k[t]}^{(n-1)} X_{k,M}^{(n)})$ with respect to $\bar{P}(\theta_{k\backslash b_k[t]}^{(n-1)}|X_{k,M}^{(n)})$ simply result in replacement $\ln P(b_k[t]|\theta_{k\backslash b_k[t]}^{(n-1)} X_{k,M}^{(n)})$. Then, we obtain

$$
\begin{aligned}
P_M(b_k[t]|X_{k,M}^{(n)}) &= P(b_k[t]|\theta_{k\backslash b_k[t]}^{(n-1)}X_{k,M}^{(n)}) \\
&= \frac{P(b_k[t],\theta_{k\backslash b_k[t]}^{(n-1)}, X_{k,M}^{(n)})}{P(\theta_{k\backslash b_k[t]}^{(n-1)}, X_{k,M}^{(n)})} \\
&= \frac{\sum\limits_{(S_{k,t-1}:S_{k,t}):b_k[t]} P(S_{k,t-1} = s', S_{k,t} = s, \theta_{k\backslash b_k[t]}^{(n-1)}, X_{k,M}^{(n)})}{\sum\limits_{S_{k,t}}\sum\limits_{S_{k,t-1}} P(S_{k,t-1} = s', S_{k,t} = s, \theta_{k\backslash b_k[t]}^{(n-1)}, X_{k,M}^{(n)})}
\end{aligned} \tag{3.61}
$$

where $S_{k,t-1} = s'$ and $S_{k,t} = s$ are the states of the virtual trellis at time t-1 and t with corresponding output $\underline{b}_{k,t}$ of $D_k(\underline{b}_{k,t})$. Incorporating the modulation symbol $b_k[t]$ in virtual trellis in eq.(3.61), the summation of all ordered pairs (s', s) ,caused by $b_k[t]$, is similar define $b_k[t]$. According to the approach eq.(3.61), the BCJR algorithm [25] is revisited.

$$
P(S_{k,t-1}{=}s', S_{k,t}{=}s, \theta_{k\backslash b_k[t]}^{(n-1)}, X_{k,M}^{(n)}) = \alpha_{t-1}(S_{k,t-1})\gamma_t(S_{k,t}, S_{k,t-1})\,\beta_t(S_{k,t}) \tag{3.62}
$$

From our analysis, the recursive forward probability $\alpha_t(S_{k,t}) = P(S_{k,t} = s, X_{k,1:t}^{(n)})$ and recursive backward probability $\beta_t(S_{k,t}) = P(X_{k,t+1:M}^{(n)}|S_{k,t} = s)$ can be formulated as follows [25]

$$
\alpha_t(S_{k,t}) = \frac{\sum\limits_{S_{k,t-1}} \alpha_{t-1}(S_{k,t-1})\gamma_t(S_{k,t}, S_{k,t-1})}{\sum\limits_{S_{k,t-1}}\sum\limits_{S_{k,t-1}} \alpha_{t-1}(S_{k,t-1})\gamma_t(S_{k,t}, S_{k,t-1})} \tag{3.63}
$$

$$
\beta_t(S_{k,t}) = \frac{\sum\limits_{S_{k,t+1}} \beta_{t+1}(S_{k,t+1})\gamma_{t+1}(S_{k,t+1}, S_{k,t})}{\sum\limits_{S_{k,t}}\sum\limits_{S_{k,t+1}} \alpha_t(S_{k,t})\gamma_{t+1}(S_{k,t+1}, S_{k,t})} \tag{3.64}
$$

with initial condition $\alpha_0(S_{k,0}) = 1/2^{(L-1)}$ and boundary condition $\beta_M(S_{k,M}) = 1/2^{(L-1)}$. The probability $\gamma_t(S_{k,t}, S_{k,t-1})$ is defined by

$$
\begin{aligned}
\gamma_t(S_{k,t}, S_{k,t-1}) &= P(S_{k,t} = s, x_k^{(n)}[t], \theta_{k\backslash b_k[t]}^{(n-1)}|S_{k,t-1} = s') \\
&= P(x_k^{(n)}[t]|S_{k,t} = s, S_{k,t-1} = s', \theta_{k\backslash b_k[t]}^{(n-1)})P(S_{k,t} = s|S_{k,t-1} = s') \\
&= P(x_k^{(n)}[t]|\underline{b}_{k,t}, \theta_{k\backslash b_k[t]}^{(n-1)})\prod_{l=0}^{L-1} P_d(b_k^{(n-1)}[t-l]\,|\,dec\,k) \\
&= \gamma_{e,t}(S_{k,t}, S_{k,t-1})P_d(b_k^{(n-1)}[t]\,|\,dec\,k)
\end{aligned} \tag{3.65}
$$

The likelihood probability of admissible hidden data space $x_k^{(n)}[t]$ is defined by

$$
P(x_k^{(n)}[t]|\underline{b}_{k,t}, \theta_{k\backslash b_k[t]}^{(n-1)}) = \frac{1}{\sigma_k^{2N(n-1)}}\exp\left(\frac{-\left\|x_k^{(n)}[t] - D_k(\underline{b}_{k,t})g_k^{(n-1)}\right\|^2}{\sigma_k^{2(n-1)}}\right) \tag{3.66}
$$

Like turbo code [26], the modified extrinsic information probability $P_M(b_k[t]|X_{k,M}^{(n)})$ on virtual trellis structure involves the information about $b_k[t]$ without knowledge from priori information $P_d(b_k^{(n-1)}[t]|dec\ k)$. Combining eq. [(3.62)-(3.65)] into (3.61), we obtain

$$P_M(b_k[t]|X_{k,M}^{(n)}) = \frac{\sum\limits_{(S_{k,t-1}:S_{k,t}):b_k[t]} \alpha_{t-1}(S_{k,t-1})\gamma_{e,t}(S_{k,t},S_{k,t-1})\beta_t(S_{k,t})}{\sum\limits_{S_{k,t}}\sum\limits_{S_{k,t-1}} \alpha_{t-1}(S_{k,t-1})\gamma_{e,t}(S_{k,t},S_{k,t-1})\beta_t(S_{k,t})} \tag{3.67}$$

3.12 Prove algorithm of probability distribution $P(g_k^{(n)}|X_{k,M}^{(n)})$

The density function for joint admissible hidden data space X_M and θ_k can be expressed as

$$P(\theta_k, X_{k,M}^{(n)}) = \prod_{t=1}^{M}\left[P(x_k^{(n)}[t]|\underline{b}_{k,t},\theta_k)P(\underline{b}_{k,t})\right]P(g_k)P\left(\sigma_k^2\right) \tag{3.68}$$

Expanding (3.68) and taking the natural logarithm, we obtain

$$\begin{aligned}\ln P(\theta_k, X_{k,M}^{(n)}) &= \sum_{t=1}^{M}[\ln P(x_k^{(n)}[t]|\underline{b}_{k,t},\theta_k)+\ln P\left(\underline{b}_{k,t}\right)]+\ln P\left(g_k\right)+\ln P\left(\sigma_k^2\right)\\ &= M\ln(\frac{1}{\sigma_k^{2)}})^N+\sum_{t=1}^{M}[-\frac{\left\|x_k^{(n)}[t]-D_k(\underline{b}_{k,t})g_k\right\|^2}{\sigma_k^{2(n-1)}}+\ln P\left(\underline{b}_{k,t}\right)]\\ &- \ (g_k-g_k^{(n-1)})^H\Sigma_{g_k}^{-1(n-1)}(g_k-g_k^{(n-1)})+\ln(\frac{1}{\sigma_k^2})^{(\alpha_{(n-1)}-1)}\frac{-\beta_{(n-1)}}{\sigma^2}\end{aligned} \tag{3.69}$$

Computational conditional probability of g_k by

$$\begin{aligned}P(g_k^{(n)}|X_{k,M}^{(n)}) &= \exp(E_{P(\theta^{(n)}_{k\backslash g_k^{(n-1)}}|X_{k,M}^{(n)})}[\ln P\left(\theta_k, X_k^{(n)}\right)])\\ &\propto \exp(-g_k^H[\Sigma_{g_k}^{-1(n-1)}+\sum_{t=1}^{M}D_k(\underline{b}_{k,t}^{(n)})^H D_k(\underline{b}_{k,t}^{(n)})]g_k)\\ &\cdot \ \exp(2\Re\{g_k^H[\Sigma_{g_k}^{-1(n-1)}g_k^{(n-1)}+\frac{1}{\sigma_k^{2(n-1)}}\sum_{t=1}^{M}D_k(\underline{b}_{k,t})^H x_k^{(n)}[t]]\})\\ &\propto \exp((g_k-g_k^{(n)})^H\Sigma_{g_k}^{-1(n)}(g_k-g_k^{(n)}))\end{aligned} \tag{3.70}$$

With

$$\begin{aligned}\Sigma_{g_k}^{(n)} &= \Sigma_{g_k}^{(n-1)}-\frac{\Sigma_{g_k}^{(n-1)}}{\sigma_k^{2(n-1)}}\left(I+\frac{1}{\sigma_k^{2(n-1)}}\Sigma_{k0}\Sigma_{g_k}^{(n-1)}\right)^{-1}\Sigma_{k0}\Sigma_{g_k}^{(n-1)}\\ g_k^{(n)} &= g_k^{(n-1)}+\frac{\Sigma_{g_k}^{(n)}}{\sigma_k^{2(n-1)}}\sum_{t=1}^{M}D_k(\underline{b}_{k,t}^{(n)})^H\left(x_k^{(n)}[t]-D_k(\underline{b}_{k,t}^{(n)})g_k^{(n-1)}\right)\\ \Sigma_{k0} &= \sum_{t=1}^{M}D_k(\underline{b}_{k,t}^{(n)})^H D_k(\underline{b}_{k,t}^{(n)})\end{aligned} \tag{3.71}$$

3.13 Prove algorithm of probability distribution $P(\sigma_k^{2(n)}|X_{k,M}^{(n)})$

Computational conditional probability of σ_k^2 by

$$
\begin{aligned}
P(\sigma_k^{2(n)}|X_{k,M}^{(n)}) &= \exp(E_{P(\theta^{(n)}_{k\setminus\sigma_k^{2(n-1)}}|X_{k,M}^{(n)})}[\ln P(\theta_k, X_k^{(n)})]) \\
&\propto \left(\frac{1}{\sigma_k^2}\right)^{NM+\alpha_{(n-1)}-1}\exp\left(-\frac{\sum_{t=1}^{M}\left\|x_k^{(n)}[t]-D_k(\underline{b}_{k,t}^{(n)})g_k^{(n)}\right\|^2+\beta_{(n-1)}}{\sigma_k^2}\right) \\
&\propto \chi^{-2}(\alpha_{(k,n)}, \beta_{(k,n)}) \qquad (3.72)
\end{aligned}
$$

With

$$
\begin{aligned}
\alpha_{(k,n)} &= NM+\alpha_{(k,n-1)} \\
\beta_{(k,n)} &= \sum_{t=1}^{M}\left\|x_k^{(n)}[t]-D_k(\underline{b}_{k,t}^{(n)})g_k^{(n)}\right\|^2+\beta_{(k,n-1)} \qquad (3.73)
\end{aligned}
$$

Table 3.1: Minimum Kullback-Leibler-Based Turbo Multiuser Detection For Separable DS-CDMA Signal algorithm

Initialization part : Set initial channel parameters by section 6.

$P[g_k] \sim \mathcal{N}(g_k^{(0)}, \Sigma g_k{}^{(0)}))$

$P[\sigma_k^2] \sim (\frac{1}{\sigma_k^2})^{(\alpha_{(k,0)}-1)} \exp(\frac{-\beta_{(k,0)}}{\sigma_k^2}) \sim \chi^{-2}\left(\alpha_{(k,0)}, \beta_{(k,0)}\right)$

Computation part :

For Iteration index : n = 1, 2,3,. . .

For User index : $k = 1, 2, \ldots, K$

• Compute admissible hidden data space $X_{k,M}^{(n)}$ by collection the observation $X_{k,M}^{(n)} = \{x_k^{(n)}[1], x_k^{(n)}[2], \ldots, x_k^{(n)}[M]\}$

$$\begin{aligned} x_k^{(n)}[t] &= D_k(\underline{b}_{k,t}^{(n-1)})g_k^{(n-1)} \\ &+ \beta_k \left[r[t] - \sum_{l=1}^{k-1} D_k(\underline{b}_{k,t}^{(n)})g_k^{(n)} - \sum_{l=k}^{K} D_k(\underline{b}_{k,t}^{(n-1)})g_k^{(n-1)} \right] \end{aligned} \quad (3.52)$$

• Set $\theta_k^{(n)} = \theta_k^{(n-1)} = \{B_k^{(n-1)}, g_k^{(n-1)}, \sigma_k^{(n-1)}\}$

• The MKL multiuser detector in Fig. 3.4 takes a priori informations $\{P_d(b_k^{(n-1)}[t] \mid dec\, k)\}_{t=1}^{M}$ from k^{th} channel decoder and $\theta_k^{(n-1)}$ in previous iteration (n-1) to compute a posterior probability $P_M(b_k^{(n)}[\mathrm{t}]|X_{k,M}^{(n)})$ see (section 3.11).

$$P_M(b_k^{(n)}[\mathrm{t}]|X_{k,M}^{(n)}) = \exp(E_{P(\theta^{(n)}_{k \setminus b_k^{(n-1)}[t]}|X_{k,M}^{(n)})}[\ln P\left(\theta_k, X_k^{(n)}\right)]) \quad (3.53)$$

• Deinterleaving and feeding $\{P_M(b_k^{(n)}[\mathrm{t}]|X_{k,M}^{(n)})\}_{t=1}^{M}$ as the extrinsic information to k^{th} channel decoder (see Fig. 3.4), the k^{th} convolutional channel decoder computes BCJR MAP decoding probability $\{P_d(b_k^{(n)}[t] \mid dec\, k)\}_{t=1}^{M}$ [25] at current iteration (n).

• The extrinsic information $\{P_d(b_k^{(n)}[t] \mid dec\, k)\}_{t=1}^{M}$ from k^{th} convolutional channel decoder (see Fig. 3.4) is then interleaved and fed back to MKL Multiuser Detector as priori information at next iteration (n+1).

• Updating $B_k^{(n)} = \{b_k^{(n)}[t]\}_{t=1}^{M}$ base on knowledge $\{P_d(b_k^{(n)}[t] \mid dec\, k)\}_{t=1}^{M}$ and setting $\theta_k^{(n)} = \{g_k^{(n-1)}, B_k^{(n)}, \sigma_k^{2(n-1)}\}$.

• Computing conditional distribution g_k by (section 3.12)

$$P(g_k^{(n)}|X_{k,M}^{(n)}) = \exp(E_{P(\theta_{k\backslash g_k^{(n-1)}}^{(n)}|X_{k,M}^{(n)})}[\ln P\left(\theta_k, X_k^{(n)}\right)]) \tag{3.54}$$

• Updating $g_k^{(n)}$ base on the condition mean estimator $g_k^{(n)} \simeq \int g_k P(g_k|X_{k,M}^{(n)})dg_k$ and setting $\theta_k^{(n)} = \{g_k^{(n)}, B_k^{(n)}, \sigma_k^{2(n-1)}\}$.

• Finally, Computing noise variance conditional distribution σ_k^2 see (section 3.13).

$$P(\sigma_k^{2(n)}|X_{k,M}^{(n)}) = \exp(E_{P(\theta_{k\backslash\sigma_k^{2(n-1)}}^{(n)}|X_{k,M}^{(n)})}[\ln P(\theta_k, X_k^{(n)})]) \tag{3.55}$$

• Updating $\sigma_k^{2(n)}$ base on the condition mean estimator by $\sigma_k^{2(n)} \simeq \frac{\beta_{(n)}}{\alpha_{(n)}-1} \simeq \int \sigma_k^2 P(\sigma_k^2|X_{k,M}^{(n)})d\sigma_k^2$ and setting $\theta_k^{(n)} = \{g_k^{(n)}, B_k^{(n)}, \sigma_k^{2(n)}\}$.

• Updating $\theta^{(n)} = [\theta_1^{(n)}, \dots \theta_k^{(n)}\theta_{k+1}^{(n-1)}, \dots \theta_K^{(n-1)}]$.

$$\begin{aligned}\bar{P}(\theta_k^{(n)}|X_{k,M}^{(n)}) &= \bar{P}(B_k^{(n)}|X_{k,M}^{(n)})\bar{P}(g_k^{(n)}|X_{k,M}^{(n)})\bar{P}(\sigma_k^{(n)}|X_{k,M}^{(n)}) \\ \bar{P}(\theta^{(n)}|Y_M) &= \prod_{l=1}^{k} \bar{P}(\theta_l^{(n)}|X_{l,M}^{(n)}) \prod_{l'=k+1}^{K} \bar{P}(\theta_{l'}^{(n-1)}|X_{l',M}^{(n-1)})\end{aligned} \tag{3.56}$$

End : User index : k

End : Iteration index : n

Table 3.2 Initial Conjugate Priori Algorithm

Initial value $g_{k*}^{(0)} = 0$, $\Sigma_{g_{k*}}^{(0)} = \mathrm{I}$, $\sigma_{k*}^{2(0)} = 1$, $\alpha^*_{(k,0)} = 2$, $\beta^*_{(k,0)} = 0.1$

For Iteration index : n = 1, 2, . . .,20

For User index : $k = 1, \ldots, K$

• Compute admissible hidden data space $X_{k,M_b}^{(n)}$ by collection the observation $X_{k,M_b}^{(n)} = \{x_k^{(n)}[1], x_k^{(n)}[2], \ldots, x_k^{(n)}[M_b]\}$

$$x_k^{(n)}[t] = D_k(\underline{b}_{k,t}^{(n-1)})g_k^{(n-1)} + \beta_k\left[r[t] - \sum_{l=1}^{k-1} D_k(\underline{b}_{k,t}^{(n)})g_k^{(n)} - \sum_{l=k}^{K} D_k(\underline{b}_{k,t}^{(n-1)})g_k^{(n-1)}\right]$$

• Compute initial of channel parameters priori $P[g_k]$ and $P[\sigma_k^2]$

$$\Sigma_{k0*} = \sum_{t=1}^{M_b} D_k(\underline{b}_{k,t})^H D_k(\underline{b}_{k,t})$$

$$\Sigma_{g_{k*}}^{(n)} = \Sigma_{g_{k*}}^{(n-1)} \frac{\Sigma_{g_{k*}}^{(n-1)}}{\sigma_{k*}^{2(n-1)}}\left(I + \frac{1}{\sigma_{k*}^{2(n-1)}}\Sigma_{k0*}\Sigma_{g_{k*}}^{(n-1)}\right)^{-1}\Sigma_{k0*}\Sigma_{g_{k*}}^{(n-1)}$$

$$g_{k*}^{(n)} = g_{k*}^{(n-1)} + \frac{\Sigma_{g_{k*}}^{(n)}}{\sigma_{k*}^{2(n-1)}}\sum_{t=1}^{M_b} D_k(\underline{b}_{k,t})^H\left(x_{k*}^{(n)}[t] - D_k(\underline{b}_{k,t})g_{k*}^{(n-1)}\right)$$

$$\alpha^*_{(k,n)} = NM_b + \alpha^*_{(k,n-1)}$$

$$\beta^*_{(k,n)} = \sum_{t=1}^{M_b}\left\|x_k^{(n)}[t] - D_k(\underline{b}_{k,t})g_{k*}^{(n)}\right\|^2 + \beta^*_{(k,n-1)}$$

$$\sigma_{k*}^{2(n)} = \frac{\beta^*_{(k,n)}}{\alpha^*_{(k,n)} - 1}$$

End;

End;

Setting initial conjugate priori (3.57) by $g_k^{(0)} = g_{k*}^{(20)}$, $\Sigma_{g_k}^{(0)} = \Sigma_{g_{k*}}^{(20)}$, $\sigma_k^{2(0)} = \sigma_{k*}^{2(20)}$, $\alpha_{(k,0)} = \alpha^*_{(k,20)}$, $\beta_{(k,0)} = \beta^*_{(k,20)}$

CHAPTER IV

SIMULATION RESULTS

4.1 Simulation Results

This section provides the simulation results to illustrate the performance of ***Minimum Kullback-Leibler*** (MKL) turbo multiuser detector in multipath fading channel. The multipath lengths for each user are L=2 and L=3. The fading coefficients are generated according to $g_{k,l} \sim \mathcal{N}(0, 1/P)$. The fading coefficients are assumed to remain constant over block of M=300 symbols. The pilot symbols overhead for each frame are $M_b = 5$. All users' spreading sequence are chose as a shot sequence with m-sequences N=15 and their shifted versions are employed. The maximum channel delay P is generated randomly with restriction $P < N$. The number of user is K=7. The channel code is a rate of $1/2$ constraint length-5 convolution code (with generators 23 , 35 in octal notation). The interleaver is generated randomly and fixed for all simulation. In comparison with other similar receiver, the joint turbo (LMMSE) iterative multiuser [13] is selected with the same channel estimation algorithm (3.54)-(3.55). In this section, the three differen simulations are tested in order to demonstrate the performance of our proposed schemes.

In Fig. 4.1 and 4.2, the performance results of iterative receiver in term of BER are presented for symbol channel delay L=2 and 3. In this case, initial channel estimation and iterative channel estimation algorithm are (3.54)-(3.55). The 20 iteration of turbo (LMMSE) multiuser detector are simulated and The 10 iteration Tubo MKL multiuser detector are compared. BER performance of our proposed detector reaches the convergence with a few iteration faster than turbo (LMMSE) multiuser detector.

In Fig. 4.3, 4.4 and 4.5, the quality of channel estimations are measured by the average normalized mean square error (NMSE) for both L=2 and L=3. This measurements are compared with the estimation and true channel response given by $\mathbf{NMSE}_g = E[\frac{||g_k^{(n)} - g_k||^2}{||g_k||^2}]$ and $\mathbf{NMSE}_{\sigma^2} = E[\frac{||\sigma_k^{2(n)} - \sigma^2||^2}{||\sigma^2||^2}]$

In Fig. 4.6 and 4.7, we demonstrate the performance of Turbo MKL multiuser detector and Turbo LMMSE multiuser detector by variation number of L. Growing number ISI (L), the performance of LMMSE and soft interference cancelation will degradation. However, the increasing (L) make the growing number of virtual Markov states and the ISI channel representing parity check bits in virtual trellis. Thus, BER reaches convergence with a few iteration.

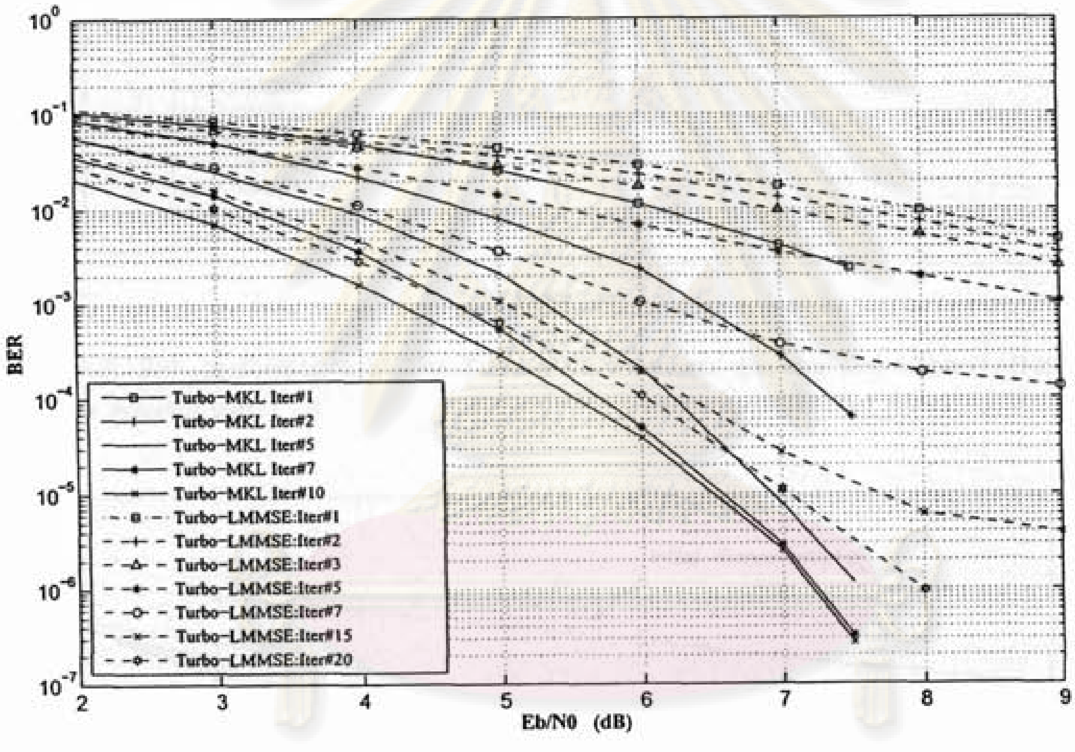


Figure 4.1: BER performance of The Tubo MKL multiuser detector and turbo (LMMSE) multiuser detector with L=2

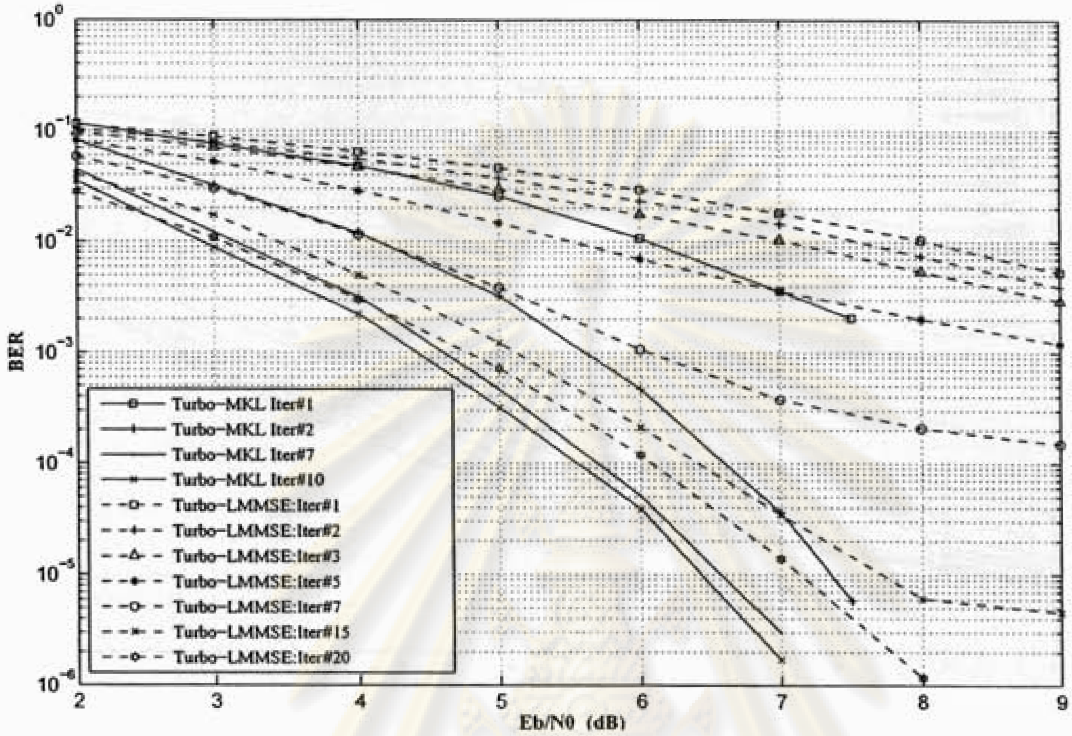


Figure 4.2: BER performance of The Tubo MKL multiuser detector and turbo (LMMSE) multiuser detector with L=3

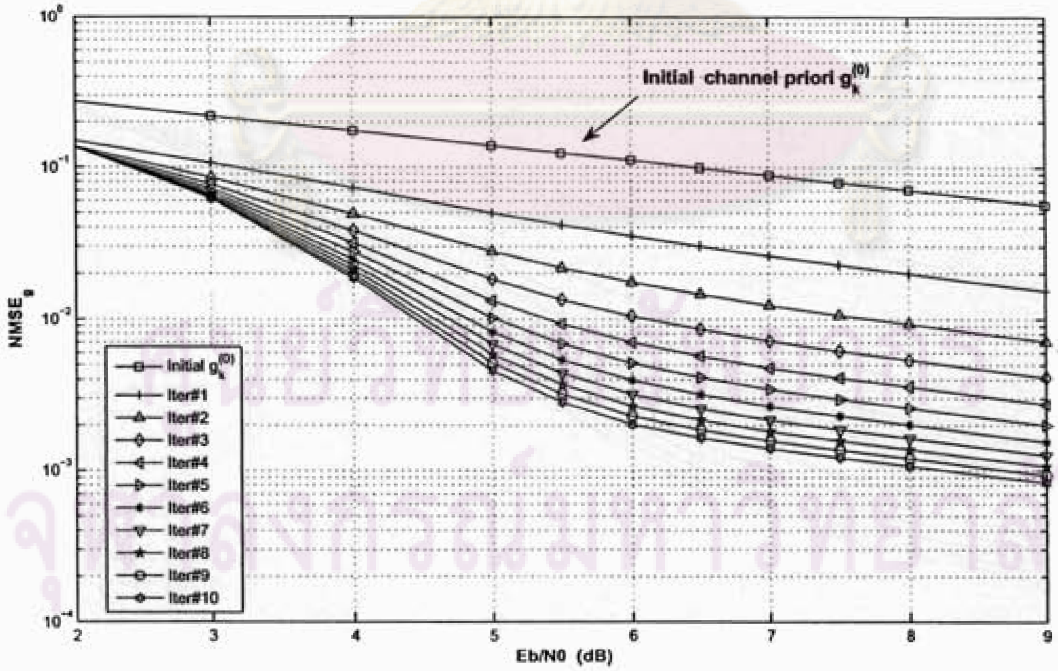


Figure 4.3: The normalized mean square error channel estimation of Turbo-MKL algorithm with various numbers of iteration

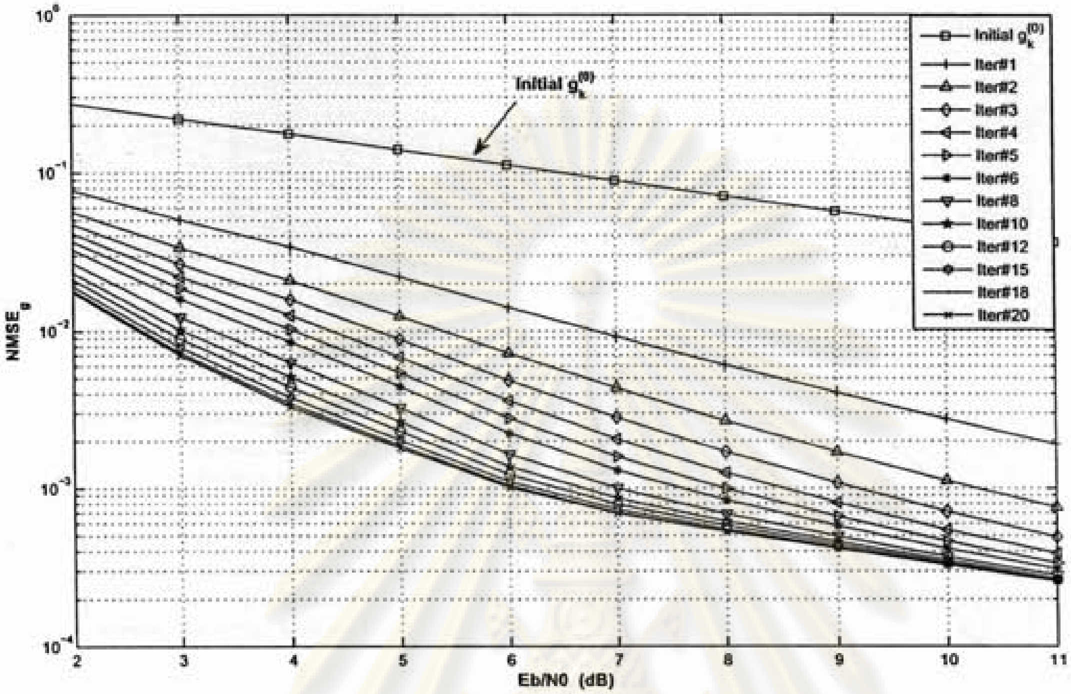


Figure 4.4: The normalized mean square error channel estimation of Turbo-LMMSE algorithm with various numbers of iteration

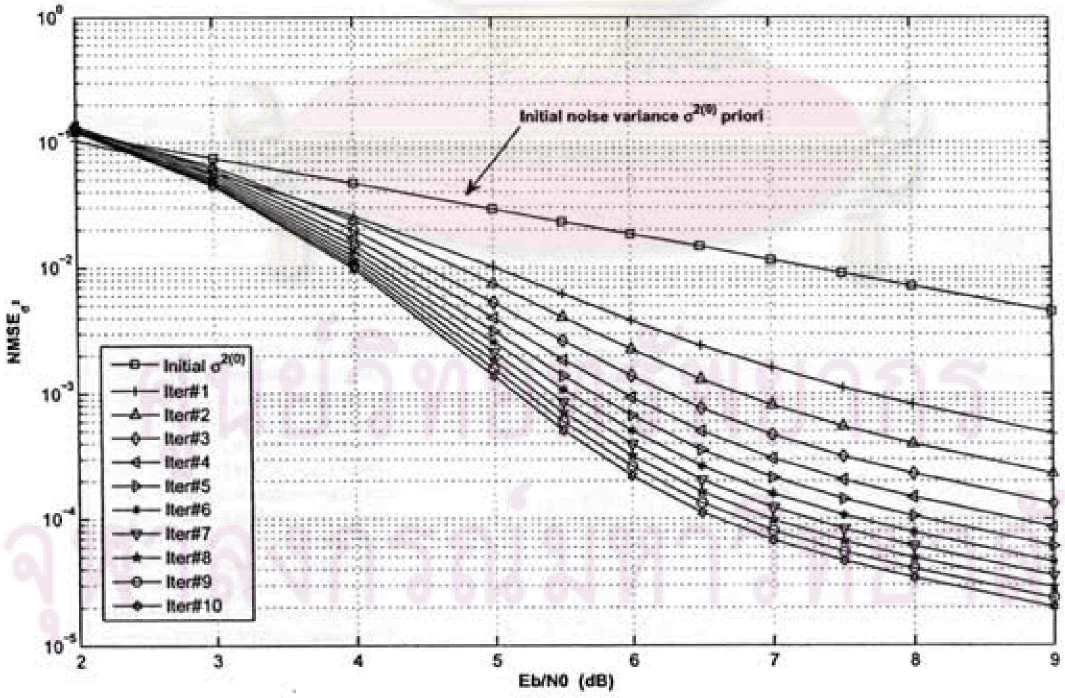


Figure 4.5: The normalized mean square error of noise variance for Turbo-MKL algorithm with various numbers of iteration

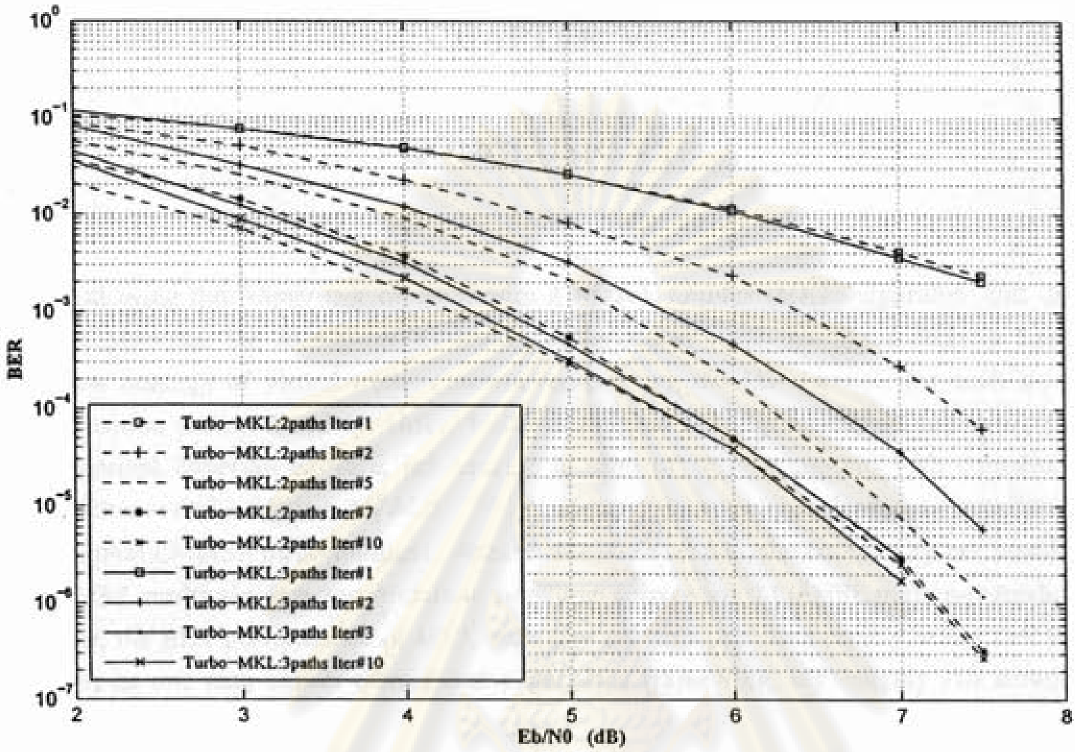


Figure 4.6 BER performance of Turbo MKL multiuser with L=2 and L=3

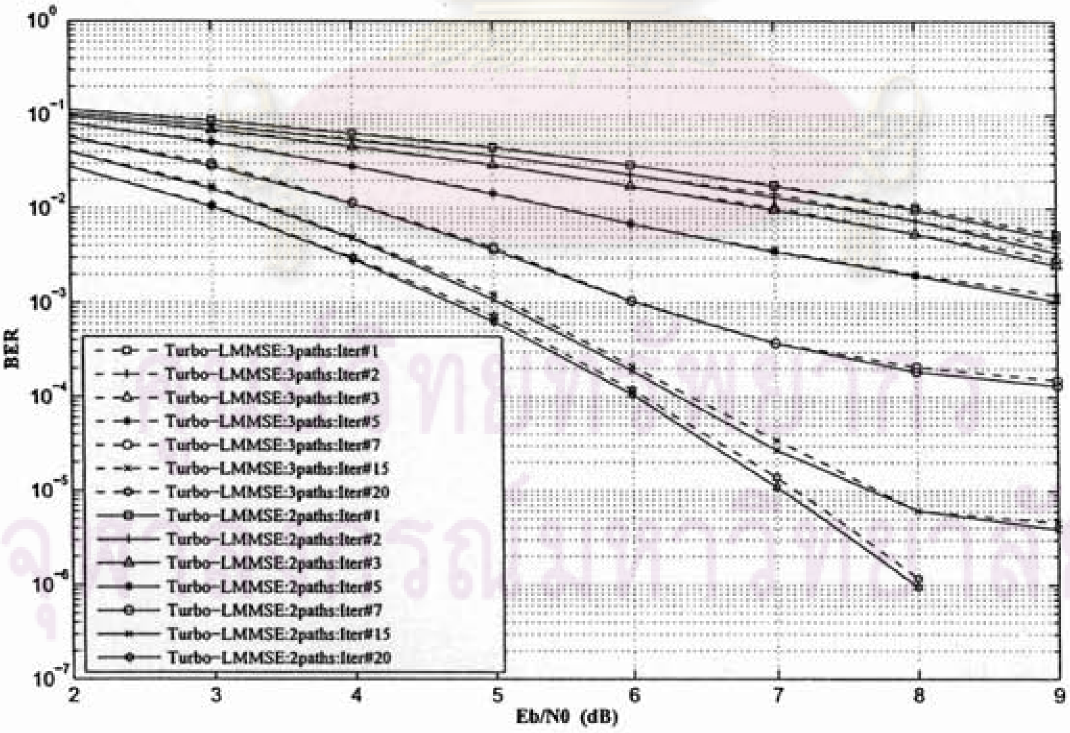


Figure 4.7 BER performance of Turbo LMMSE multiuser with L=2 and L=3

CHAPTER V

CONCLUSIONS

Following the framework of ***Minimum Kullback-Leibler*** (MKL) algorithm, this dissertation proposes the ***Minimum Kullback-Leibler-Based*** turbo multiuser detector based on a new concept of virtual trellis diagram representing the ISI channels for each signal decomposition users. Based on virtual trellis model, the MKL multiuser module computes the numerical Bayesian extrinsic information in term of BCJR algorithm with complexity per iteration $\approx O(8KM2^L + 4KM2^{(L-1)})$. Compared with the complexity per iteration of conventional LMMSE turbo multiuser detector [13] $\approx O(N^2M)$, the complexity per iteration of our MKL multiuser detector module is more than conventional LMMSE multiuser module. However, the BER performance of MKL multiuser detector improves rapidly in the first few iterations as one may expect from iterative decoding framework, say roughly $7th$ iteration for $L = 2$ and $3th$ iteration for $L = 3$ respectively. By integrating the iterative MKL channel estimation algorithm, the average complexity per iteration for channel vector g_k is $\approx O(KMPL^2)$ and for noise variance is $\approx O(KNM)$.

References

[1] Verdu, S., Multiuser Detection. Cambridge Univ. Press: 1998.

[2] Poor, H., Iterative multiuser detection. IEEE Signal Processing Magazine. 21 (January 2004): 81–88.

[3] Hagenauer, J., Foward error correcting for CDMA systems. Proceedings of IEEE Fourth Int. Symp. on Spread Spectrum Techniques and Applications (September 1996): 566–569.

[4] Dai, H. and Poor, H., Iterative space-time processing for multiuser detection in multipath CDMA channels. IEEE Trans. Signal Processing 50 (Sept 2002): 2116–2127.

[5] Wang, X. and Poor, H., Wireless Communication Systems: Advanced Techniques for Signal Reception. Upper Saddle River NJ: Prentice-Hall 2004.

[6] Hagenauer, J., The turbo principle: Tutorial introduction and state of the art. Proceedings of. Int. Symp. Turbo Codes and Related Topics (1997):1–11.

[7] Heegard, C. and SWicker B., Turbo Coding. Boston MA: Kluwer. 1995.

[8] Proakis, J.G., Digital Communications, 4th ed. New York: McGraw-Hill. 2000.

[9] Giallorenzi, T.R. and Wilson, S.G. Multiuser ML sequence estimator for convolutional coded asynchronous CDMA. IEEE Trans. Communication. 44 (1996): 997–1008.

[10] Boutros, J. and Caire, G., Iterative multiuser joint decoding: Unified framework and asymptotic analysis. IEEE Trans. Inform. Theory 48 (July 2002): 1772–1773.

[11] Moher, M., An iterative multiuser decoder for near-capacity communications IEEE Trans. Commun.. 46 (1998): 870–880 .

[12] Poor, H., Turbo multiuser detection: A primer. J. Commun. Networks 3 (2001): 196–201 .

[13] Wang, X. and Poor, H., Iterative (turbo) soft interference cancellation and decoding for coded CDMA. IEEE Trans. Communication 47 (1999): 1046–1061.

[14] Tuchler, M., Singer, A., and Koetter, R., Minimum mean squared error equalization using a priori information. IEEE Trans. Signal Processing 50 (March 2002) 673–683.

[15] Abe, T. and Matsumoto, T. Space-time turbo equalization in frequency-selective MIMO channels. IEEE Trans. Vehicular Technology vol. 52, no.3, pp. 469–475, May 2003.

[16] Yu, Q. and Lambotharan, S. Iterative (Turbo) Estimation and Detection Techniques for Frequency-Selective Channels With Multiple Frequency Offsets IEEE Trans. Signal Processing 14 (April 2007): 236–239.

[17] Roy, S.and Duman, T., Soft input soft output Kalman equalizer for MIMO frequency selective fading channels. IEEE Trans. Wireless Communications 6 (Feb 2007): 506–514.

[18] Chen, R., Wang, X. and J. Liu, Adaptive joint detection and decoding in flat-fading channels via mixture Kalman fltering IEEE Trans. Inf. Theory, 46 (September 2000): 2079–2094.

[19] Dong, B., Wang, X. and Doucet, A. A new class of soft MIMO demodulation algorithms. IEEE Trans. Signal Processing 51 (November 2003): 2752–2763.

[20] Guo, D. and Wang, X., Blind detection in MIMO systems via sequential Monte Carlo IEEE J. Sel. Areas Communication, 21 (Apr 2003): 464–473.

[21] Yang, Z. and Wang, X. Blind turbo multiuser detection for long-code multipath CDMA IEEE Trans. Communication 50 (January 2002): 112–125.

[22] Qian, Y., Guoan, B. and Wan, C., SMC-based blind detection for DS-CDMA systems over multipath fading channels IEEE Trans. Communication 54 (January 2006):971–974.

[23] Yong, S., Xian, Z. and Xiao, Z., A low-complexity sequential Monte Carlo algorithm for blind detection in MIMO systems IEEE Trans. Signal Processing 54(March 2006): 2485–2496.

[24] Ploysuwan, T., Tantiphanwadi S., and Teekaput, P., Minimum Kullback-Leibler-Based Turbo Multiuser Detection For Separable DS-CDMA Signal. Proceeding of IEICE SISB. (Nov 2007).

[25] Bahl, L. R,. Cocke, J., Jelinek, F. and Raviv, J., Optimal decoding of linear codes for minimizing symbol error rate. IEEE Trans. Inform. Theory 20 (Mar 1974) : 284–287.

[26] Berrou, C., Glavieux, A. and Thitimajshima, P., Near Shannon limit error Correcting coding and decoding Turbo codes. In Proc. ICC, (May 1993): 1064–1070.

[27] Fessler, J. A. and Hero, A. O. Space alternating generalized expectation maximization algorithm IEEE Trans. Communication 42 (Oct 1994): 2664–2677.

[28] Iltis, R. and Sunwoo, K., Geometric derivation of expectation-maximization and generalized successive interference cancellation algorithms with applications to CDMA channel estimation. IEEE Trans. Signal Processing 51 (May 2003): 1367–1377 .

[29] Kullback, S. and Leibler, R. On information and sufficiency Annals ofMathematical Statistics. 22 (1951): 79–87.

[30] Šmídl, V. and Quinn, A., The Variational Bayes Method in Signal Processing. Springer-Verlag Berlin Heidelberg. 2006.

[31] Haykin, S. Adaptive Filter Theory 4rd ed. Upper Saddle Rive, NJ:Prentice-Hall 2002.

[32] Neal, R. M. and Hinton, G. E. A New View of the EM Algorithm that Justifies Incremental Sparse and Other Variants NATO Science Series Dordrecht Kluwer Academic Publishers. 1998.

[33] Bernardo, J. M. and Smith, A. F. M. Bayesian Theory. New York Wiley, 1994.

[34] Gray, R. M. Entropy and Information Theory. New York: Springer- Verlag 1990.

[35] Cover, T. M. and Thomas, J. A. Elements of Information Theory. New York:Wiley Interscience 1991.

[36] Hollander, F. den Large Deviations. Providence, RI: Amer. Math. Soc. 2000.

[37] Silva, J. and Narayanan, S., Average divergence distance as a statistical discrimination measure for hidden Markov models IEEE Trans. Audio Speech Lang. Processing. 14 (May 2006) 890–906.

[38] Do, M. N. and Vetterli, M., Rotation invariant texture characterization and retrieval using steerable wavelet-domain hidden Markov model IEEE Trans. Multimedia 4 (Dec. 2002): 517–527.

[39] o, M. N. and Vetterli, M. Wavelet-based texture retrieval using generalized Gaussian density and Kullback–Leibler distance. IEEE Trans. Image Processing, 11 (Feb 2002): 146–158.

[40] asconcelos, N. On the efficient evaluation of probabilistic similarity functions for image retrieval. IEEE Trans. Inf. Theory 50 (July 2004): 1482–1496.

[41] Smyth, P. Clustering sequences using hidden Markov models in Advances in Neural Information Processing. Cambridge, MA: MIT Press. 9 (1997): 648–654.

[42] Cox, R. T., Probability, frequency and reasonable expectation Am. J. Phys. 14 1946.

[43] de, B. Finneti, Theory of Probability: A Critical Introductory Treatment. New York: J. Wiley, 1970.

[44] Quinn, A. P. Bayesian Point Inference in Signal Processing. PhD thesis Cambridge University Engineering Dept. 1992.

[45] Jaynes, E. T., Bayesian methods: General background. The Fourth Annual Workshop on Bayesian/Maximum Entropy Methods in Geophysical Inverse Problems 1984.

[46] Kay, S. M. Modern Spectral Estimation Theory and Application. Prentice-Hall 1988.

[47] Marple Jr., S. L., Digital Spectral Analysis with Applications. Prentice-Hall 1987.

[48] Pratt, J., Raiffa, H. and Schlaifer, R. Introduction to Statistical Decision Theory. MIT Press 1995

[49] Kass, R. E. and Raftery, A. E. Bayes factors Journal of American Statistical Association 90 (1995): 773–795 .

[50] Titterington, D., Smith, A. and Makov, U. Statistical Analysis of Finite Mixtures. New York: John Wiley 1985.

[51] Tanner, M. Tools for statistical inference. New York: Springer-Verlag 1993.

[52] Ruanaidh, J. J. K. O and Fitzgerald, W. J. Numerical Bayesian Methods applied to Signal Processing. Springer 1996.

[53] Box, G. E. P. and Tiao, G. C. Bayesian Statistics. Oxford 1961.

[54] Kass, R. E. and Raftery, A. E. Bayes factors and model uncertainty tech. rep. University of Washington 1994.

[55] Gull, S. F. Bayesian inductive inference and maximum entropy in MaximumEntropy and Bayesian Methodsin Science and Engineering 1988.

[56] Skilling, J., The axioms of maximum entropy in Maximum Entropy and Bayesian Methods in Science and Engineering 1988.

[57] Bosq, D. Nonparametric Statistics for Stochastic Processes: estimation and prediction. Springer. 1998.

[58] Bernardo, J. and Smith, A., Bayesian Theory John Wiley and Sons 2 ed 1997.

[59] Edwards, A. W. F. Likelihood. Cambridge Univ. Press 1972.

[60] Bernardo, J. M. Expected infromation as expected utility The Annals of Statistics 7 (1979): 686–690.

[61] Wu, C.F.J., On the convergence properties of the EM algorithm. The Annals of Statistics (1983): 95–103 .

[62] Demoment, G. and Idier, J., Problmes inverses et dconvolution Journal de Physique IV, (1992): 929–936.

[63] Nikolova, M. and Mohammad-Djafari, A., Maximum entropy image reconstruction in eddy current tomography in Maximum Entropy and Bayesian Methods 1993.

[64] Kay, S. M. Modern Spectral Estimation: Theory and Application. Prentice-Hall. 1988.

[65] Gull, S. F. and Skilling, J. Maximum entropy method in image processing. Proceeding of IEE October 1984.

Appendices

Appendix A
List of Abbreviations

AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BW	Bandwidth
CDMA	Code Division Multiple Access
CM	Channel Model
DS	Direct Sequence
EM	Expectation-Maximization
IVB	Iterative Variational Bayes
KL	Kullback-Leibler
KLD	Kullback-Leibler Divergence
KLD_{VB}	Kullback-Leibler divergence for Variational Bayes
KLD_{MR}	Kullback-Leibler divergence for Minimum Risk
LMMSE	Linear Minimum Mean Square Error
MAI	multiple access interference
MAP	Maximum a Posteriori
MaxEnt	Maximum Entropy
MCMC	Markov Chain Monte Carlo
MKL	Minimum Kullback-Leibler
ML	Maximum Likelihood
MMSE	Minimum Mean Square Error
MR	Minimum Risk
MUD	Multiuser Detector
QB	Quasi-Bayes
SISO	Soft Input Soft Output
SMC	Sequential Monte Carlo
VB	Variational Bayes
ZF	Zero Forcing

Appendix B
Publications and Presentations

Tuchsanai Ploysuwan, Sawat Tantiphanwadi, Prasit Teekaput

"Minimum Kullback-Leibler-Based Turbo Multiuser Detector over Decomposition CDMA Signal ," IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences , Vol. E91-A, No. 10, pp. 2963–2972, October. 2008.

Tuchsanai Ploysuwan, Sawat Tantiphanwadi, Prasit Teekaput

"Iterative Turbo MKL Multiuser Detection and Channel Estimation for DS-CDMA Signals," Proceeding on IEEE International Conference on Information, Communication and signal Processing (ICICS 2007), Singapore, 10–13 December 2007.

Tuchsanai Ploysuwan, Sawat Tantiphanwadi, Prasit Teekaput

"Minimum Kullback-Leibler-Based Turbo Multiuser Detection For Separable DS-CDMA Signal," Proceeding on IEICE International Workshop on Smart Info-Media Systems in Bangkok (SISB 2007), Bangkok Thailand, 1–2 November 2007.

Tuchsanai Ploysuwan, Prasit Teekaput

"Blind Turbo Multiuser Detector With Unknown Intercell Interferences," Proceeding on IEEE International Symposium on Wireless Pervasive Computing(ISWPC 2006), Phuket Thailand, 16–18 January 2006.

Vitae

Tuchsanai Ploysuwan was born in Nakhonsawan, Thailand, in 1981. He received the B.Eng. from Department of Electrical Engineering Kasetsart University, Thailand, in 2002 and the M.Eng. degree in Electrical Engineering from Department of Electrical Engineering Chulalongkorn University, Thailand, in 2004. He is now working toward the Ph.D. degree in electrical engineering from Telecommunication System Research Laboratory (TSRL) Chulalongkorn University, Thailand. His research interests are in area of statistical signal processing, sequential Monte Carlo method, iterative decoding/demodulation and signal processing for communication.