Multiuser Detection in Fading Channels Using the EM Algorithm

by

Andrew W. Eckford

A thesis submitted in conformity with the requirements
for the Degree of Master of Applied Science,
Department of Electrical and Computer Engineering,
at the University of Toronto

Multiuser Detection in Fading Channels Using the EM Algorithm

Andrew William Eckford

Master of Applied Science

Department of Electrical and Computer Engineering

University of Toronto

1999

Abstract

Channels shared by multiple users, such as in DS-CDMA, frequently suffer from Multiple-Access Interference (MAI), which may be mitigated through the use of multiuser detection. In previous work, the EM algorithm was used in an interference cancellation multiuser detection scheme which used soft decisions. However, this method, along with most currently available multiuser detection algorithms, assumed accurate knowledge of users' transmitted amplitudes, which is impractical for use in a fading channel. In this work, the role of soft decisions in the EM algorithm and multiuser detection is carefully examined, with emphasis on the new idea of turbo detection. Adaptations of the EM-based multiuser detector for Rayleigh and Rician fading channels are also presented, both with and without knowledge of carrier phase. Simulation results for all these methods are also presented, demonstrating that the proposed algorithms improve somewhat on existing algorithms over wide ranges of SNR.

Acknowledgements

I wish to thank my supervisor, Prof. S. Pasupathy, for his valuable support, advice, and comments throughout the course of this project.

Ad Majorem Dei Gloriam.

Contents

List of Figures

Chapter 1

Introduction

1.1 Introduction to Multiuser Detection

Channels shared by multiple users are frequently plagued by multiple access interference (MAI), in which users' transmissions interfere with each other. In many cases of interest, the MAI is intentionally present in the channel, such as in direct sequence code division multiple access (DS-CDMA), where users are assigned signalling waveforms that are known to have some degree of correlation. In other cases, such as in high-speed networks, MAI is an unavoidable consequence of the channel conditions. Historically, receiver designers took the approach that this interference could be treated as noise and thus ignored. This assumption produced the conventional receiver, which was merely a matched-filter receiver equivalent to the optimum receiver in the single-user additive white Gaussian noise (AWGN) channel.

The conventional receiver is sufficient for some applications. However, it has two problems which motivate the search for alternatives. Firstly, if all users transmit with approximately the same power, then beyond some point increasing the power of every user will not produce a decrease in bit error rate, as the interference will have begun to dominate the noise in the signal and produce an error floor. Secondly, if the users transmit with widely different powers, the linear nature of the conventional receiver allows the interference from powerful users to overwhelm the signal from weaker users - referred to as the "near-far effect". A receiver design capable of taking interference into account could potentially mitigate these

problems, hence making such systems more practical and increasing their capacity. Such receivers are known as multiuser detectors.

The fundamental work in multiuser detection was done by Verdu in [1], in which a method optimal in the maximum-likelihood (ML) sense was proposed (i.e. selecting the set of symbols which maximized the likelihood function for a given observation). Unfortunately, that method was shown to be NP-complete, meaning that no known algorithm exists to solve it in polynomial time; in fact, the complexity of the proposed optimal solution was shown to be exponential in the number of users. Much research effort since that time has focused on obtaining a suboptimal method with realizable complexity but approaching optimal performance. One particular set of detectors that has attracted attention is the family of linear detectors, which as the name implies rely on linear combinations of the matched filter outputs to remove interference. These detectors were first outlined in [2] and [3], in which a decorrelating detector, a linear minimum mean squared error (MMSE) detector, and an optimal linear detector were proposed. Linear receivers have attracted much interest, since they have the desirable properties of being straightforward and relatively simple to implement. In particular, the decorrelator can be shown to be the optimal detector in the absence of any knowledge of the users' received amplitudes [13], which is an attractive property for wireless channels.

Although linear multiuser detectors solve the near-far problem, they fail to approach optimal performance. This is because they tend to enhance noise, in much the same way as a zero-forcing equalizer enhances noise in a channel with inter-symbol interference (ISI). A natural solution, as in the case of ISI channels, is to incorporate some sort of nonlinear decision feedback and interference cancellation. This approach was taken by [4], using a decision-feedback detector that iteratively cancelled interference in reverse order of users' strengths. Several alternative nonlinear approaches have been described (detailed surveys are available in [6] and [7]). Of particular interest to this research was a scheme proposed by Nelson and Poor [8], in which the Expectation-Maximization (EM) algorithm was used to estimate unknown users' transmissions and eliminate them from the signal of interest.

1.2 The EM Algorithm and Multiuser Detection

The EM algorithm existed in various ad-hoc forms before being formally described and explained in [10]. Essentially, the EM algorithm simplifies the procedure for finding the ML estimate of some parameter of a distribution. In general, difficult ML estimation problems can be simplified if some additional random variables, ordinarily hidden, may be observed. The EM algorithm allows a two-step iterative calculation of the ML parameter estimate using this hidden data as follows. The algorithm begins with some initial estimate of the parameter determined by whatever means is convenient. In the Expectation step (E-step), estimates are taken of the hidden data given the current value of the parameter. Subsequently, in the Maximization step (M-step), the ML estimate of the parameter is obtained using the hidden data estimates obtained in the E-step. More formally:

$$\begin{aligned} \mathrm{E-step}: \quad & Q(\theta;\, \bar{\theta}_n) = E[\log f_x(x|\theta) \mid y, \theta = \bar{\theta}_n] \\ \mathrm{M-step}: \quad & \bar{\theta}_{n+1} = \arg\max_{\theta} Q(\theta;\, \bar{\theta}_n) \end{aligned} \tag{1.1}$$

where x is the complete (observed and hidden) data, y is the observed data, and $\bar{\theta}_n$ is the estimate of the parameter at the nth iteration. The E- and M-steps iterate back and forth until the algorithm converges.

The notation in (1.1) is somewhat dense, so let us give a brief example to illustrate the algorithm. Suppose we have a deterministic, unknown parameter h, and an observation of h given by $y = u + h + n$, where u is an interfering random variable with known distribution, and n is a zero-mean Gaussian random variable with variance σ^2. Ordinarily, to calculate the ML estimate of h, we would be required to calculate the marginal likelihood function $f_y(y \mid h)$ and find the value of h which maximized this function. However, depending on the distribution of u, this might be a difficult task. Obviously, if we could observe u directly, the problem would become much simpler. The EM algorithm eases the ML estimation problem by allowing us to calculate the estimate as though u could be observed. In this case, we take y to be the incomplete data and $\{y, u\}$ to be the complete data. Assuming u is independent of h, the complete data log-likelihood function would simplify to

$$\begin{aligned} \log f_x(x \mid \theta) &= \log f_{y,u}(y, u \mid h) \\ &= \log f_y(y \mid u, h) + \log f_u(u) \end{aligned} \tag{1.2}$$

The expression $\log f_u(u)$ is independent of θ, thus it would have no bearing on the M-step and can be dropped. Again eliminating terms irrelevant to the M-step, it is straightforward to show that the E-step in this case reduces to

$$Q(h;\ \bar{h}) = -h^2 + 2h(y - \tilde{u}) \tag{1.3}$$

where $\tilde{u} = E[u \mid y, h = \bar{h}]$, and $\bar{h}$ is the previous estimate of h. Maximizing this function in the M-step, we have $h = y - \tilde{u}$, which allows us to restate the algorithm as a simple iteration in a single line:

$$\bar{h}_{n+1} = y - E[u \mid y, h = \bar{h}_n] \tag{1.4}$$

which is presumably a simpler calculation than calculating the marginal likelihood function. Note that we are not necessarily required to choose the complete data to be $\{y, u\}$ - concievably, a clever choice of some other random variable dependent on y might lead to an even easier iteration.

It was shown [10] (and restated in Appendix A) that the estimates $\bar{\theta}_n$ increase monotonically in likelihood with n, which implies that the algorithm should normally converge to the ML estimate, assuming that there are no other local maxima in the likelihood space. That work also allowed for Generalized EM algorithms, which were defined as any algorithm in which the M-step did not necessarily involve an outright maximization of $Q(\theta; \bar{\theta}_n)$, but instead where any $\bar{\theta}_{n+1}$ was chosen such that $Q(\bar{\theta}_{n+1}; \bar{\theta}_n) \geq Q(\bar{\theta}_n; \bar{\theta}_n)$. One such EM algorithm that has been of particular use in signal processing is the Space-Alternating Generalized EM (SAGE) algorithm, proposed by Fessler and Hero in 1994 [12]. This algorithm splits the parameter space under estimation into several subspaces, each with its own possibly overlapping complete data space. The EM algorithm iterates with the E- and M-steps being performed sequentially on each subspace in turn. The advantage of this algorithm was shown to be faster convergence over the standard EM algorithm.

In the communications and signal processing community, the EM algorithm has found particular applications in image processing and biomedical engineering [11]. However, there have been many instances of its application to communication problems. In [15], a method of maximum-likelihood sequence estimation using EM in channels with unknown timing errors was proposed. The question of EM-based data detection for Trellis-coded sequences was dealt with in [16]. The EM algorithm has also been used in certain problems involving multiuser detection. For example, in [17], Poor considered the specific case of a DS-CDMA system, and proposed an EM-based model for estimating the amplitudes of the received signals. However, a different method (which depended on knowledge of the signal amplitudes) was used for data detection. More recently, Wang and Blostein [18] used the EM algorithm for data detection in a receiver using adaptive beamforming. Fawer and Aazhang [19] used the EM algorithm for ML estimation of channel parameters, but used a separate method for symbol detection.

Since the optimal ML multiuser detector has impractical complexity, and since the EM algorithm is capable of reducing the complexity of ML estimation, the actual problem of symbol detection seemed a natural application for this algorithm. This particular problem was first examined by Nelson and Poor [8], who noted that the standard formulation of the EM algorithm did not lead to an efficient result - if other users' symbols were used as the hidden data, the EM-based detector could not use any information gained elsewhere in the system from the demodulation of other users' transmissions. To overcome this problem, a SAGE-based detector was proposed, along with a detector based on a new algorithm the authors proposed called the Missing Parameter EM (MPEM) algorithm. The MPEM algorithm was arrived at intuitively, without rigorous mathematical formulation - in particular, the authors were unable to show that the likelihood of consecutive estimates increased monotonically in likelihood, although it performed very well in simulations.

1.3 Current Work

The existing literature on multiuser detection in general, and EM-based multiuser detection in particular, leaves several questions unanswered. Firstly, Nelson and Poor noted similar-

ities between some results in their work and the recent interest in so-called soft decision feedback, in which symbol decisions to be used in later stages of an algorithm are given continuous values to indicate their expected reliability (rather than the discrete values of the actual symbols themselves, which are known as "hard decisions," and which destroy this reliability information). In particular, they noticed that the definition of symbols as hidden data under the EM algorithm leads to a soft-decision-like structure in the E-step [9]. This observation was the basis of the MPEM algorithm, which makes extensive use of soft decisions. The potential benefits of soft decision feedback have been noted in the literature in applications such as equalization and decoding, although a rigorous theoretical framework has thus far remained elusive. Some further analysis of soft decisions as they apply to the MPEM algorithm, and how this work relates to similar work in the literature, is warranted.

Secondly, although nonlinear decision-feedback-type detectors, in which interference is estimated, regenerated, and removed from the received signal, can produce near-optimal results in multiuser detection, such detectors require accurate knowledge of the users' received amplitudes and phases so that their interference may be properly cancelled. It is well known that the mobile wireless environment suffers from fading, in which the amplitude and phase of each user's transmission is random, which clearly causes a problem for interference cancellation approaches to multiuser detection. Although in some works, the amplitudes were assumed to be known or assumed to be accurately estimated elsewhere in the system, in other works an effort was made to estimate these quantities. One common approach is to use ML estimation of the channel state information, which treats these quantities as deterministic (but unknown) parameters. However, in fading channels, users' amplitudes may be modelled as random variables that obey known probability density functions (PDFs), such as the Rayleigh or Rician density functions. This knowledge may be exploited to achieve a performance better than that of the decorrelator, although thus far little research effort has focused on exactly how this might be done. There is also the possibility of using known correlations between the amplitudes in consecutive samples to further improve these estimates. The EM algorithm, which explicitly estimates hidden random variables, presents a promising framework for this sort of detection.

In this thesis, the above two problems will be examined, and novel receiver designs will

be outlined. The remainder of this work is orgnanized as follows. In Chapter 2, a detailed discussion of the theory of multiuser detection will be presented, along with more rigorous definitions of the problem and the system models. In Chapter 3, the theoretical framework of soft decisions in multiuser detection, along with its relationship to other works in the literature, will be discussed. In particular, links will be shown between the work in [8] and the field of "turbo detection". In Chapter 4, the application of the EM and related algorithms to fading channels will be discussed, and simulation results for proposed receiver designs will be presented. A summary and conclusion will be given in Chapter 5.

Chapter 2

Theoretical Aspects of Multiuser Detection

In this chapter, the theoretical basis for this thesis will be presented, along with the system models used. Multiuser detection will be discussed, and fundamental receiver types analyzed. Soft decisions, as they apply to the EM algorithm and multiuser detection, will be discussed. The problem of multiuser detection in fading will also be considered.

2.1 Multiuser Detection Fundamentals

In a DS-CDMA system, the received signal can be modelled as the following:

$$y(t) = \sum_{k=1}^{n} \sum_{m=-\infty}^{\infty} a_k(t) b_k(m) c_k(t - mT - \tau_k) + z(t) \tag{2.1}$$

where n is the number of users, $a_k(t)$ is the kth user's time-varying amplitude, $b_k(m)$ is the mth symbol transmitted by the kth user, $c_k(t)$ is the code waveform (or chip sequence) of the kth user (with period T), τ_k is the time delay of the kth user, and $z(t)$ is a Gaussian noise process. A common simplifying assumption used in multiuser detection algorithms is that the constants τ_k are equal to zero, that is, that the system is chip-synchronized. We may now obtain a set of n sufficient statistics by applying the conventional matched filters for each user's chip sequence to the received signal (2.1) [1]. This scheme is mathematically

defined as follows:

$$y_k(m) = \int_{t=(m-1)T}^{mT} y(t)c_k(t)dt \tag{2.2}$$

and graphically depicted in Figure 2.1.

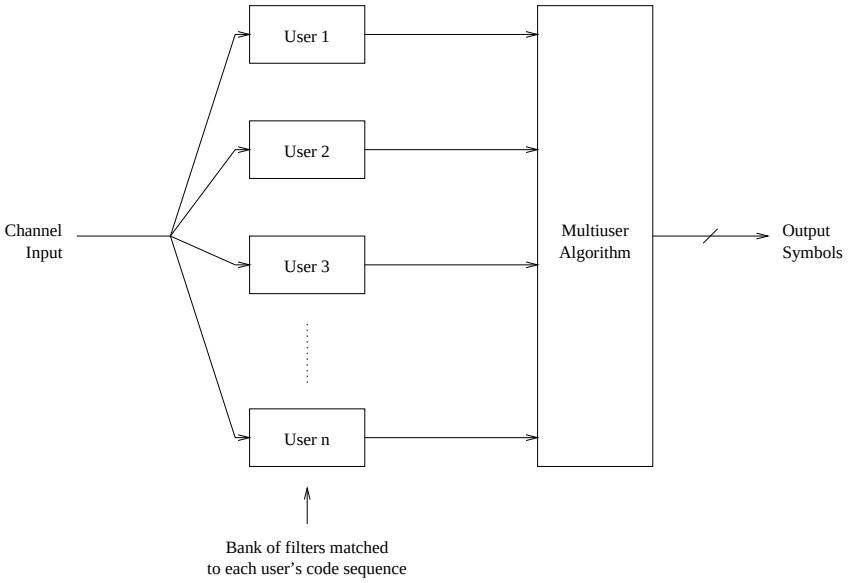

Figure 2.1: A generic multiuser detector with matched filter inputs

Using the synchronization assumption, the resultant signal may be expressed in vector form as follows:

$$\mathbf{y}(m) = \mathbf{R}\mathbf{A}(m)\mathbf{b}(m) + \mathbf{z}(m) \tag{2.3}$$

where $\mathbf{y}(m) = [y_1(m)\ y_2(m)\ \cdots\ y_n(m)]^T$ is the vector of matched filter outputs, $\mathbf{R}$ is the $n \times n$ Hermitian matrix of chip sequence cross correlations with elements defined as:

$$R_{jk} = \int_{t=0}^{T} c_j(t)c_k(t)dt \tag{2.4}$$

We assume that the amplitude envelope functions a_k remain constant over a single symbol interval, so $\mathbf{A}(m) = \text{diag}[a_1(m)a_2(m)\cdots a_n(m)]$ is the diagonal matrix of amplitude envelope functions over symbol interval m. Additionally, $\mathbf{b}(m) = [b_1(m)\ b_2(m)\ \cdots\ b_n(m)]^T$ is the vector of symbols. Finally, $\mathbf{z}(m) = [z_1(m)\ z_2(m)\ \cdots\ z_n(m)]^T$ is a vector of jointly Gaussian

random variables with cross-correlation matrix $\sigma^2\mathbf{R}$, which represent the thermal noise in the receiver that has acquired cross-correlation due to the correlations between chip sequences in matched filtering (analogous to correlated noise in an ISI channel). It is assumed that the chip sequences are normalized such that the diagonal elements of $\mathbf{R}$ are all equal to 1.

Henceforth we will drop the notation (m) from (2.3) in cases where it is unambiguous to do so. It will also be assumed that the symbol vector is composed of equiprobable, antipodal elements, i.e., $\mathbf{b} \in \{-1, 1\}^n$. Let us first state the likelihood function for this set of random variables. Assume for now that the amplitudes $\mathbf{A}$ are known - then the likelihood function of $\mathbf{y}$ given $\mathbf{b}$ is merely a jointly Gaussian random vector, whose conditional PDF is given by

$$f_{\mathbf{y}}(\mathbf{y} \mid \mathbf{b}) = \frac{1}{\sqrt{(2\pi)^n \Delta(\sigma^2\mathbf{R})}} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{y}^T\mathbf{R}^{-1}\mathbf{y} - 2\mathbf{y}^T\mathbf{A}\mathbf{b} + \mathbf{b}^T\mathbf{A}\mathbf{R}\mathbf{A}\mathbf{b})\right) \tag{2.5}$$

where $\Delta(\cdot)$ represents the determinant of a matrix. Taking the natural log of the function in (2.5), eliminating all terms not dependent on $\mathbf{b}$, and removing multiplied constants applied to the entire expression (which are irrelevant to the maximization procedure), we find the log-likelihood function:

$$L(\mathbf{b}) = \mathbf{y}^T\mathbf{A}\mathbf{b} - \frac{1}{2}\mathbf{b}^T\mathbf{A}\mathbf{R}\mathbf{A}\mathbf{b} \tag{2.6}$$

From (2.6), we may now derive some important receiver structures. Firstly, we have the ML optimal receiver (from [1]), which selects the vector of bits $\tilde{\mathbf{b}}_{ML}$ that maximizes $L(\mathbf{b})$:

$$\tilde{\mathbf{b}}_{ML} = \underset{\mathbf{b}\in\{-1,1\}^n}{\arg\max} \left(\mathbf{y}^T\mathbf{A}\mathbf{b} - \frac{1}{2}\mathbf{b}^T\mathbf{A}\mathbf{R}\mathbf{A}\mathbf{b}\right) \tag{2.7}$$

It can be shown that the problem of solving (2.7) is NP-complete [1], which implies that no known algorithm can solve it in polynomial time (in fact the only known algorithms are brute-force, which have $O(2^n)$ complexity).

Now consider a second important receiver type which is constrained to have a linear structure to reduce complexity. Letting $\mathbf{h} = \mathbf{A}\mathbf{b}$, and substituting into (2.6), we have $L(\mathbf{h}) = \mathbf{y}^T\mathbf{h} - \mathbf{h}^T\mathbf{R}\mathbf{h}/2$. Let us consider the generic problem of finding the ML estimate of $\mathbf{h}$ if it is simply a continuous-valued parameter vector. The result is well known:

$$\nabla L(\mathbf{h}) = \mathbf{y} - \mathbf{R}\mathbf{h} = 0 \,|_{\mathbf{h}=\tilde{\mathbf{h}}_{ML}} \tag{2.8}$$

$$\tilde{\mathbf{h}}_{ML} = \mathbf{R}^{-1}\mathbf{y} \tag{2.9}$$

Substituting into (2.9) with the definition $\mathbf{h} = \mathbf{A}\mathbf{b}$, we have that

$$\mathbf{R}^{-1}\mathbf{y} = \mathbf{A}\mathbf{b} + \mathbf{u} \tag{2.10}$$

where $\mathbf{u}$ is a vector of jointly Gaussian random variables with cross-correlation matrix $\sigma^2\mathbf{R}^{-1}$. From (2.10), MAI has been cancelled from each element of the vector $\mathbf{R}^{-1}\mathbf{y}$, and thus a simple slicer decision may be performed to detect the symbols. This multiuser receiver was first proposed by Lupas and Verdu in [2], and is known as the **decorrelating detector**, or **decorrelator**. It can also be shown that this receiver is the optimal receiver in the absence of any knowledge of the received amplitudes, except the knowledge that they are positive [13]. To briefly sketch this idea, if the amplitudes $\mathbf{A}$ are unknown positive real numbers, then if $\mathbf{b}$ is antipodal, the unknown elements of $\mathbf{h}$, which are given by $h_k = a_k b_k$, can obviously take any value in $(-\infty, \infty)$. Thus, $\mathbf{h} = \mathbf{A}\mathbf{b}$ satisfies the conditions of an unknown parameter vector, which led to the ML estimation in (2.9). The covariance matrix of the noise in the resultant signal is coloured, given by $\sigma^2\mathbf{R}^{-1}$. This results in an important disadvantage for the decorrelating detector - it can easily be shown that the diagonal elements of $\sigma^2\mathbf{R}^{-1}$, which give the variance of the noise in each interference-cancelled observation, are all greater than or equal to σ^2. This implies some noise enhancement, which is analogous to the noise enhancement observed in a zero-forcing equalizer.

As in the problem of equalization, to achieve near-optimal performance with a suboptimal detector requires a nonlinear approach. A number of different receiver structures have been proposed, most of which can be categorized as either successive interference cancelling detectors or multistage detectors [6]. In a successive interference cancelling detector, decisions are made on users in sequence (normally in order from most to least powerful), and the interference arising from these decisions is subtracted in turn from the received signal. A common method for accomplishing this is the multiuser decision-feedback detector, outlined

in [4] and [5]. Here, the users are first sorted in increasing order of received amplitude. The Cholesky factorization $\mathbf{G}$ of the cross-correlation matrix $\mathbf{R} = \mathbf{G}\mathbf{G}^T$ is then obtained [14], where $\mathbf{G}$ is a lower-triangular matrix with the same dimensions as $\mathbf{R}$, and $\mathbf{y}$ is then multiplied on the left by $\mathbf{G}^{-1}$:

$$\mathbf{u} = \mathbf{G}^{-1}\mathbf{y} = \mathbf{G}^T\mathbf{A}\mathbf{b} + \mathbf{w} \tag{2.11}$$

where $\mathbf{w}$ is a jointly Gaussian noise vector with cross-correlation matrix $\sigma^2\mathbf{I}$. This procedure whitens the noise vector and makes the interference "causal" due to the triangular nature of $\mathbf{G}$. Detection then proceeds as follows. Noting that $\mathbf{G}^T$ is upper triangular, interference in the observation u_j only contains interference including b_k if $k > j$. Thus, a decision is made on user n, and the interference is subtracted from all other users; a decision is then made on user $n-1$, and the interference is cancelled from the remaining users; and so on. This method achieves good performance, but it can be shown that it is an unfair algorithm if the users have nearly the same amplitudes. Also, it is important to note that this method assumes accurate knowledge of the amplitudes.

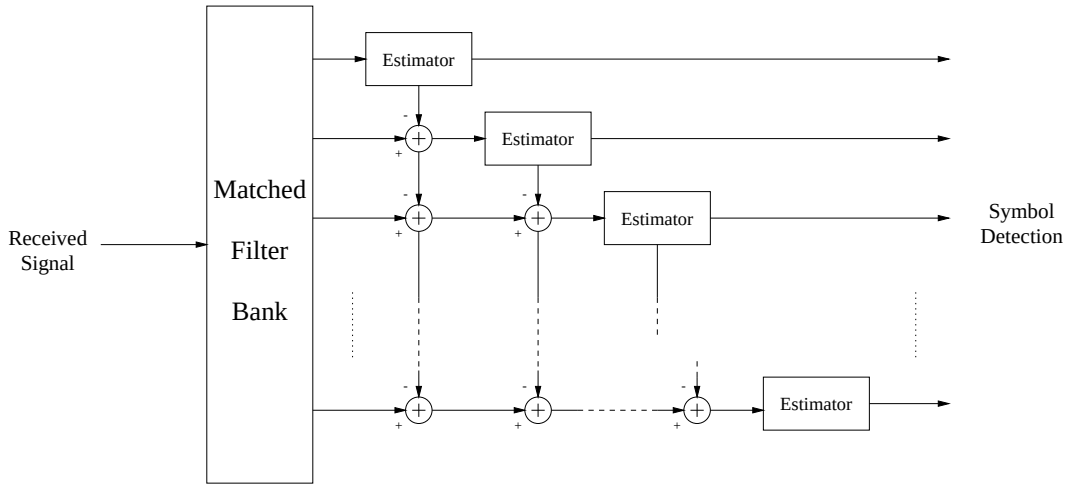

Figure 2.2: Block diagram representation of a successive interference cancellation multiuser detection scheme

In a multistage detector, some or all of the users' transmitted symbols are estimated at each stage of the algorithm, and these tentative results are fed forward to subsequent stages

for further interference cancellation. In one such method outlined in [20], the maximization problem in (2.7) is approached from the perspective of coordinate ascent. In each stage of this algorithm, the value for each symbol that maximizes the log-likelihood function is estimated in sequence, while all other symbols are held at their previously estimated values. This algorithm may be more succinctly represented by

$$\tilde{b}_k(n) = \underset{b_k \in \{-1,1\};\, b_j = \tilde{b}_j(n-1),\, j \neq k}{\arg\max} \left(\mathbf{y}^T \mathbf{A} \mathbf{b} - \frac{1}{2} \mathbf{b}^T \mathbf{A} \mathbf{R} \mathbf{A} \mathbf{b} \right) \tag{2.12}$$

Although there is no limitation on the number of stages in principle, such algorithms normally have a practical limit of two stages in order to reduce latency and complexity [6].

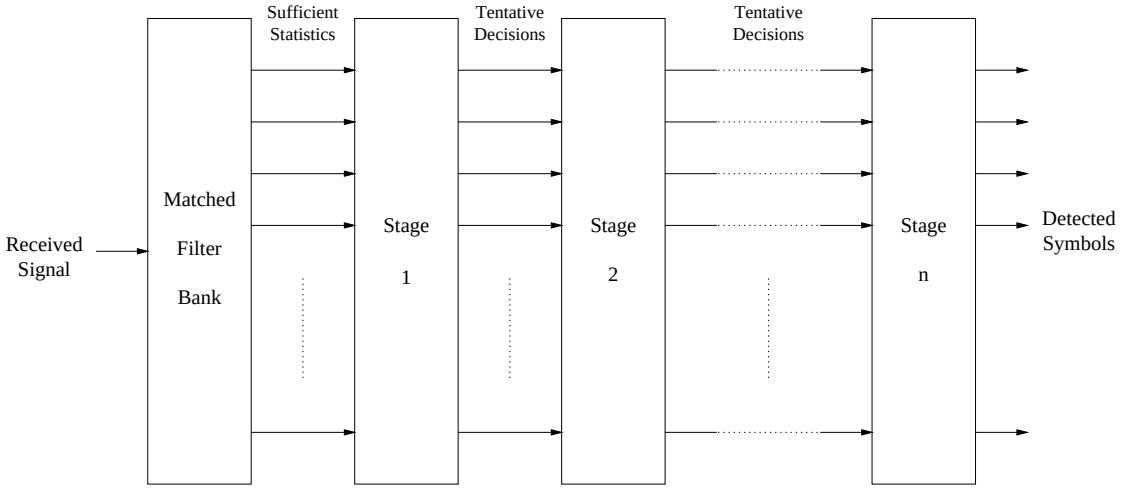

Figure 2.3: Block diagram representation of a multistage multiuser detection scheme

As mentioned previously, the particular multiuser detection algorithm of interest here was proposed by Nelson and Poor in [8]. This scheme was a multistage detector that used the EM algorithm to iteratively cancel interference. Again, this work assumed accurate knowledge of the received amplitudes. Three different receivers were proposed. In the first, the standard EM algorithm was used, and the interfering symbols were considered to be the hidden data, as in the E-step from (1.1). If b_i is the symbol of interest and $\mathbf{b}_{\tilde{i}}$ is the vector of interfering symbols, then the E-step is given by:

$$Q(b_i;\, \bar{b}_i) = E[\log f_{\mathbf{y},\mathbf{b}_{\tilde{i}}}(\mathbf{y}, \mathbf{b}_{\tilde{i}} \mid b_i) \mid \mathbf{y},\, \bar{b}_i] = E[\log f_{\mathbf{y}}(\mathbf{y} \mid \mathbf{b}) \mid \mathbf{y}, \bar{b}_i] + E[\log f_{\mathbf{b}_{\tilde{i}}}(\mathbf{b}_{\tilde{i}})] \tag{2.13}$$

The rightmost term in (2.13) is not dependent on b_i - it is thus irrelevant to the M-step and may be dropped. Now, we have:

$$Q(b_i;\, \bar{b}_i) = \frac{1}{2\sigma^2}\left(K_{b_i} + a_i^2 b_i^2 + 2a_i b_i \sum_{k\neq i} a_k \hat{b}_k R_{ik} - 2y_i a_i b_i\right) \tag{2.14}$$

where K_{b_i} are terms constant in b_i which may be removed prior to the M-step, a_j is the received amplitude of the jth user, and $\hat{b}_k = E[b_k \mid \mathbf{y},\, \bar{b}_i]$, which was given in [8] as

$$E[b_k \mid \mathbf{y},\, \bar{b}_i] = \tanh\left[\frac{a_k}{\sigma^2}(y_k - R_{ik} a_i \bar{b}_i)\right] \tag{2.15}$$

which was presented without derivation. From (2.14), the E- and M-steps may be combined into a single expression, as follows:

$$\bar{b}_i^{(n+1)} = \frac{1}{a_i}\left(y_i - \sum_{k\neq i} a_k R_{ik} \tanh\left[\frac{a_k}{\sigma^2}(y_k - R_{ik} a_i \bar{b}_i^{(n)})\right]\right) \tag{2.16}$$

Several comments may be made concerning the above derivation. Firstly, it is interesting to note from (2.15) that the interference is cancelled by a "soft decision" on the observation $\mathbf{y_i}$. This will be discussed in greater depth later in this chapter. Secondly, the estimate $\bar{b}_i^{(n+1)}$ is merely a parameter estimate; it does not take into account the fact that b must be in $\{-1, 1\}$. Nelson and Poor proposed using a slicer or other decision function to express this information. Thirdly, and most importantly, the result in (2.16) has no means to take advantage of information gleaned from the demodulation of other users' signals, which would presumably be occurring parallel to this symbol detection. Two schemes were proposed to improve this situation. One proposal was a detector based on the SAGE algorithm. Using this method, the symbols were separated such that each user's symbol occupied its own parameter space, and no hidden data was required. This detector reduced to simple coordinate ascent, similar to (2.12) above:

$$\bar{b}_i = G\left[\underset{b_i;\, b_j=\bar{b}_j,\, j\neq i}{\arg\max}\left(\mathbf{y}^T\mathbf{A}\mathbf{b} - \frac{1}{2}\mathbf{b}^T\mathbf{A}\mathbf{R}\mathbf{A}\mathbf{b}\right)\right] \tag{2.17}$$

where $G(\cdot)$ is some monotonic decision function, an example of which would be a slicer. The other proposal involved a new EM-like algorithm called the Missing Parameter EM (MPEM)

algorithm, which allowed conditioning of the E-step estimation based on the tentative demodulation results obtained for other users. The algorithm processes each user in sequence, with the E-step estimation replaced by

$$\hat{b}_k = \tanh\left[\frac{a_k}{\sigma^2}\left(y_k - \sum_{m\neq k} R_{mk} a_m \bar{b}_m)\right)\right] \tag{2.18}$$

which is evaluated for all $k \neq i$, where i is the bit of interest, and the M-step is then given by

$$\bar{b}_i = G\left(y_k - \sum_{k\neq i} R_{ik} a_i \hat{b}_i\right) \tag{2.19}$$

where $G(\cdot)$ is again a decision function. This algorithm, although not shown to increase monotonically in likelihood as the EM algorithm does, was shown to have excellent performance in simulation.

In a practical wireless cellular system, multiuser detection would normally be a candidate only for the reverse (mobile-to-base) link. There are two reasons for this. Firstly, synchronization is trivial in the forward link, and thus orthogonal spreading codes (such as Hadamard functions) may be used, making multiuser detection irrelevant (although multiuser detection might still be applicable in dealing with co-channel interference). Secondly, the computational complexity of even suboptimal multiuser detection algorithms would unacceptably increase the cost of a mobile receiver (which would be borne by the customer), whereas cellular service providers would generally be willing to pay more for a more complicated base receiver that would be capable of increasing the channel capacity.

2.2 Multiuser Detection and Soft Decisions

In many detection applications, decisions on certain symbols must be used in the detection of other symbols, such as in a decision-feedback equalizer, or as in some of the nonlinear multiuser detectors discussed above. It has been frequently observed in such situations that a "soft decision," incorporating some sort of reliability information, rather than a "hard decision" (e.g., a slicer output) can improve performance. Furthermore, above it was observed

that the EM algorithm leads to some natural soft decision structures. As was pointed out in the above section, Nelson and Poor [8] observed that feeding back detected symbols with a soft-decision function led to better performance than either hard decisions or no decisions. Research effort into soft decisions extends beyond multiuser detection into the related field of equalization, as well as in coding applications. For the purposes of this discussion, let us define a soft decision as follows:

Definition 2.1 *A **Soft Decision** is any value containing information regarding a corrupted observation of a symbol which incorporates both an estimate for the symbol and reliability information on the accuracy of this estimate.*

An important early work in this field was that of Taylor [28]. In that paper, the so-called "Estimate-Feedback Equalizer" was proposed, in which the minimum mean squared error (MMSE) estimate of each symbol was calculated for feedback, assuming that all symbols contributing to ISI were equal to their previous estimates (as opposed to the decision-feedback equalizer, in which symbols are assumed to be equal to previous hard decisions). The soft decisions are thus given by:

$$\begin{aligned} \tilde{b}_k &= E[b_k \,|y_k;\; b_i = \tilde{b}_i,\; k-n \leq i < k] \\ &= \tanh\left[\frac{1}{\sigma^2}\left(y_k - \sum_{j=1}^{n} c_j \tilde{b}_{k-j}\right)\right] \end{aligned} \tag{2.20}$$

where n is the channel memory length, and $\{c_j\}$ are the equalization coefficients. The $\tanh(\cdot)$ function satisfies Definition 2.1, for antipodal symbols, since the sign of the function conveys the estimated symbol value, while the distance from zero conveys the reliability (with greater distance indicating more reliability). In simulation, this method was shown to have better performance than hard decision feedback, largely owing to the fact that the effect of error propagation was minimized. It was observed that decisions in a hard decision scheme are weighted equally, whereas in a soft decision scheme an observation with low intrinsic reliability is given a smaller value, which implies that it will not have as major an effect on subsequent decisions. Obvious similarities exist between this method indicated by

(2.20) and the MPEM and SAGE multiuser algorithms proposed by [8] in (2.17), (2.18), and (2.19).

Further applications have been found in the field of adaptive equalization. Nowlan and Hinton [29] showed that a minor modification in the development of the Least Mean Square (LMS) adaptive algorithm leads directly to soft decision feedback. In that work, rather than minimizing the mean squared error, the stochastic gradient procedure instead was configured to maximize the likelihood of the equalized data sequence. Using the straightforward assumption that the equalized observations are approximately Gaussian, which is similar to the assumption used to obtain (2.20), Nowlan and Hinton showed that the update procedure on the vector of equalization weights $\mathbf{w}_t$ given the vector of previous observations $\mathbf{x}_t$ and the equalized observation y_t was given by

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \epsilon \left[y_t - \tanh(y_t/\sigma^2) \right] \mathbf{x}_t \tag{2.21}$$

We shall show later that this decision function is related to a derivative of the score function. A similar method has since also been used in an adaptive equalization scheme with a joint entropy maximization update rule [30].

Also important to this discussion is the recent interest in soft-decision decoding. Hagenauer et al. [31] proposed an iterative soft-decision-decoding scheme whereby the reliabilities of the information and parity bits were repeatedly fed back to enhance each other. Consider the *a posteriori* log-likelihood ratio (LLR) of the symbol b, given by:

$$\begin{aligned} L(b) = \frac{P(b=1 \mid y)}{P(b=-1 \mid y)} &= \frac{f_y(y \mid b=1)}{f_y(y \;\; b=-1)} + \frac{P(b=1)}{P(b=-1)} \\ &= L_y + L_p \end{aligned} \tag{2.22}$$

where L_y is the LLR obtained from all channel observations, and L_p is the *a priori* LLR. Now, y represents all observations available at the receiver, so in the case of a systematic code where independent observations of b are available through parity bits, L_y may be broken down into $L_y = L_c + L_e$, where L_c is the "channel" LLR, the reliability based on the direct

observations of the information bits (which remains constant throughout the procedure), and L_e is the "extrinsic" LLR, the reliability obtained from all other observations. The final expression is then given by [31]

$$L(b) = L_p + L_c + L_e \tag{2.23}$$

The decoding algorithm given by Hagenauer involved calculating the extrinsic LLR, following which that value was used as the *a priori* LLR, L_p, in all subsequent stages and iterations of the decoding process. It is easy to see that the LLR satisfies the requirements of Definition 2.1.

Hagenauer also showed that this iterative decoding scheme was similar to that used in turbo decoding [40]. In turbo coding, information bits are normally encoded with two different codes, and the decoding algorithm iterates between these two codes, passing extrinsic LLR values gleaned from each code to the other. As we have seen, though, many types of communication problems can lend themselves to a soft decision structure. This observation has led to turbo detection (or turbo processing), in which one of the two codes is replaced by a soft-decision detector [38, 39]. The most common application of turbo detection is in equalization, though multiuser detection schemes have also been proposed [37]. In a later chapter, we will show that some of the multiuser detection schemes proposed by [8] can be interpreted as implementations of turbo detection in which no codes are used, which indicates that the addition of codes to these schemes should be straightforward.

2.3 Multiuser Detection and Fading Channels

In urban areas, the wireless channel frequently suffers from multipath fading, in which the received signal consists of the sum of a large number of reflections of the original transmitted signal, all with random amplitude and phase. Consider the complex envelope of such a signal:

$$y = \left(L + \sum_{u=0}^{w} \alpha_u e^{j\phi_u} \right) b + n \tag{2.24}$$

where L is the line-of-sight component (possibly complex), α_u and ϕ_u are the random am-

plitude and phase of the uth path respectively, w is the number of specular paths, and n is a complex Gaussian noise component. This signal may also be represented in the following equivalent forms:

$$y = (a_i + ja_q)b + n = abe^{j\theta} + n \tag{2.25}$$

As $w \to \infty$, if the random amplitudes α_u have finite variance, by the central limit theorem the distributions of the components a_i and a_q of (2.25) approach the Gaussian distribution, with means given by the line-of-sight component L.

Suppose that $L = 0$, in which case there is no line-of-sight component, a case that occurs frequently in densely built urban environments. In this case, the density of the signal amplitude a is given by the Rayleigh distribution:

$$f_a(a) = \frac{a}{\rho^2} \exp\left(-\frac{a^2}{2\rho^2}\right) \tag{2.26}$$

where ρ is a parameter related to the average power of the signal. If L is nonzero, then the received amplitude distribution is given by the Rician distribution:

$$f_a(a) = \frac{a}{\rho^2} \exp\left(-\frac{(a^2 + |L|^2)}{2\rho^2}\right) I_0\left(\frac{a|L|}{\rho}\right) \tag{2.27}$$

where $|L|$ is the magnitude of L, and $I_0(\cdot)$ is the zero order modified Bessel function of the first kind. This distribution is also useful in modeling the envelope of the received signal in satellite communications [21]. These fading models will be the ones principally used here; other more general fading models exist, but they are beyond the scope of this discussion.

In a mobile wireless channel, the phase and amplitude will tend to be time-variant, and thus we may examine the behavior of the fading variables as random processes. The short-term characteristics of these variables can be derived by considering the Doppler effect. If we consider the idealized case of a single transmitted frequency, isotropic antennas, and uniform distribution of the signal arrival angle, the power spectral density of the received signal can be shown to be [22]:

$$S_x(f) = \frac{K}{\sqrt{f_m^2 - (f - f_c)^2}},\ |f - f_c| < f_m;\ 0 \text{ elsewhere} \tag{2.28}$$

where f_c is the carrier frequency, $f_m = f_c v/c$ is the maximum Doppler frequency, and K is a constant proportional to the overall signal power. By the well known Wiener-Kinschine theorem, applying the inverse Fourier transform to (2.28) gives the autocorrelation of the signal envelope (at baseband), which here is given by [23]

$$R_x(\tau) = \frac{K}{2} J_0(\omega_m \tau) \tag{2.29}$$

where $J_0(\cdot)$ is the zero order Bessel function of the first kind. The correlation between consecutive fading coefficients is very important in incoherent detection, as we shall discuss later.

Consider the problem of detecting an antipodal signal in a fading channel with additive Gaussian noise, where the symbol alphabet is $b \in \{-1, 1\}$. From (2.25), if we can demodulate coherently with respect to the phase θ (i.e., if θ is perfectly known), then the maximum likelihood detector is the slicer, as in the non-fading case. Without an accurate estimate of θ, the problem is virtually impossible, since detection of the sign of y implies the necessity of observing the absolute phase of the signal [33]. This implies that θ must remain nearly constant over several symbols in order to permit tracking. In general, in fading channels with antipodal signalling where a separate phase reference signal is not transmitted, the value of θ can only be known modulo π, which implies that a differential coding scheme must be used to extract symbol information based on the relative phases of observations. In Rician channels, where the signal contains a constant additive component, phase estimation is less of a serious problem.

Now we may examine how fading affects the problem of multiuser detection. We recall from the previous section that the optimal detector in the absence of any knowledge of the amplitudes is the decorrelator. However, it is straightforward to show that implementation of the decorrelator in a fading channel requires the phase information from every user. Let us redefine (2.3) to incorporate the random phase element from (2.25):

$$\mathbf{y} = \mathbf{R}\boldsymbol{\Theta}\mathbf{A}\mathbf{b} + \mathbf{z} \tag{2.30}$$

where $\boldsymbol{\Theta}$ is a diagonal matrix of the normalized complex components of each user's envelope,

for which $\mathbf{\Theta}\mathbf{\Theta}^H = \mathbf{I}$, the identity matrix (i.e., the magnitude of each element of $\mathbf{\Theta}$ is 1). Clearly, multiplying in (2.30) on the left by $\mathbf{R}^{-1}$ is not sufficient for detection, owing to the unknown phases. Conversely, with accurate knowledge of $\mathbf{\Theta}$, we may simply multiply on the left in (2.30) by $[\mathbf{R}\mathbf{\Theta}]^{-1}$ to make a detection equivalent to the detection implied by (2.10). Heuristically, this is equivalent to decorrelating the interference terms (multiplying by $\mathbf{R}^{-1}$), followed by rotating the resultant signal to lie on the real axis (multiplying by $\mathbf{\Theta}^{-1}$), and making the appropriate slicer decision. Let us therefore make the following definitions for the purposes of the remainder of this discussion.

Definition 2.2 ***Phase-Coherent Multiuser Detection*** *is defined as multiuser detection in a fading channel under the assumption that accurate phase information is available at the receiver, but with unknown amplitudes.*

Definition 2.3 ***Incoherent Multiuser Detection*** *is defined as multiuser detection in a fading channel where the values of the phase and amplitude are unknown.*

It is clear from (2.30) that any multiuser detection scheme that examines only one symbol interval must be phase-coherent. The problem of phase-coherent multiuser detection has attracted some attention in the literature. Duel-Hallen [4] proposed a simple moving average FIR scheme for amplitude estimation. The problem is frequently approached from a maximum-likelihood standpoint, as mentioned in [17]. Another method, based on interference cancellation through estimating the correlation between users, was proposed by [24].

The problem of incoherent multiuser detection is significantly more difficult, and has attracted little attention in the literature. However, some recent work in incoherent single user detection in fading channels has shed light on how this might be accomplished. Owing to the correlations in fading coefficients implied by (2.29), the received signal has properties that may be exploited through Maximum Likelihood Sequence Estimation (MLSE). In 1995, Yu and Pasupathy [25, 26] proposed an innovations-based scheme whereby this task could be accomplished. In that work, a finite length vector of fading channel observations $\mathbf{r} = [r_1\ r_2\ \cdots\ r_n]^T$ was considered. These observations were given by $\mathbf{r} = \mathbf{A}\mathbf{c} + \mathbf{n}$, where $\mathbf{A}$ was a diagonal matrix representing the transmitted symbols, $\mathbf{c}$ was the vector of complex

fading coefficients, and $\mathbf{n}$ was a white noise vector with cross-correlation matrix $\sigma^2\mathbf{I}$. The cross-correlation matrix of $\mathbf{r}$ is therefore given by

$$\mathbf{R} = \mathbf{A}\mathbf{R}_c\mathbf{A}^H + \sigma^2\mathbf{I} \tag{2.31}$$

where $\mathbf{R}_c$ is the cross-correlation matrix of the fading coefficients.

It was observed that the cross-correlation matrix in (2.31) is unacceptable for sequence detection, since consecutive observations were not independent. This problem was alleviated by whitening the signal and finding its innovations process. As before, let $\mathbf{G}$ be the Cholesky factorization of $\mathbf{R}$, where $\mathbf{R} = \mathbf{G}\mathbf{G}^T$. Then we have

$$\mathbf{y} = \mathbf{G}^{-1}\mathbf{r} \tag{2.32}$$

Note that it is only possible to calculate $\mathbf{G}^{-1}$ conditioned on the transmitted sequence of symbols $\mathbf{A}$. Since $\mathbf{r}$ is a set of zero-mean Gaussian complex random variables, a linear combination of them will still be a zero mean complex Gaussian random variable. The probability density function of $\mathbf{r}$ conditioned on $\mathbf{A}$ is then shown to be (see also [27])

$$f_{\mathbf{r}}(\mathbf{r} \mid \mathbf{A}) = f_{\mathbf{y}}(\mathbf{y} \mid \mathbf{A})D(\mathbf{A}) = \pi^{-M} \prod_{j=1}^{M} \frac{1}{\sqrt{G_{jj}}} \exp\left(-\sum_{k=1}^{M} |y_k|^2\right) \tag{2.33}$$

where $D(\mathbf{A})$ is the reciprocal of the product of the square roots of the diagonal members of $\mathbf{G}$. A recursive logarithmic branch metric for the nth observation is straightforward to find from (2.33):

$$b_k = |y_k(a_1, \cdots, a_n)|^2 - \ln d_k(a_1, \cdots, a_n) \tag{2.34}$$

In [26], the MLSE algorithm was given as follows: first, apply whitening filters to the channel outputs and calculate the branch metric in (2.34) for every possible sequence of $[a_1 ... a_n]$. Secondly, apply these branch metrics to a trellis and find the data sequence corresponding to the shortest path through the trellis using the Viterbi algorithm [41]. To adapt this technique to multiuser detection, we need to modify this algorithm to handle multiple-input, multiple-output signals. This problem will be examined more closely in later chapters.

Chapter 3

On the Use of Soft Decisions in Multiuser Detection

A major contribution of the work of Nelson and Poor in [8] (and earlier in [9]) was to introduce the idea of soft decisions to the multiuser detection literature. In this chapter, we shall examine the general problem of the use of soft decisions, particularly as they apply to multiuser detection. We shall show the relationship between the Gaussian probability density function and soft decision functions based on MMSE criteria. Furthermore, we shall show methods for deriving multiuser detectors based on the Turbo principle, in combination with the EM algorithm. Analysis of some of the assumptions made will also be given, and an effort will be made to extend these results to the case of continuous random variables (such as the combined case of random amplitudes and symbols).

3.1 Soft Decisions and the Score Function

In [29], it was shown that a soft decision could be obtained directly from the likelihood function of a Gaussian mixture distribution. We will now show a more general result related to this, which will demonstrate that there is an intrinsic relationship between MMSE soft decisions and the likelihood function for the Gaussian PDF which may be observed through the score function. These results are in turn a special case of a result presented in more

general form in [10]. Consider a random vector $\mathbf{y}$ defined as follows:

$$\mathbf{y} = \mathbf{A}_\theta \theta + \mathbf{A}_z \mathbf{z} + \mathbf{x} \tag{3.1}$$

where θ is a k-dimensional vector of parameters to be estimated, $\mathbf{z}$ is a l-dimensional vector of interfering random variables with known PDF, $\mathbf{A}_\theta$ and $\mathbf{A}_z$ are, respectively, $n \times k$ and $n \times l$ matrices of constant coefficients, and $\mathbf{x}$ is an n-dimensional jointly Gaussian random vector with cross correlation matrix $\mathbf{R}_\mathbf{x}$. We wish to obtain the ML estimate for the vector θ, so we must obtain the log-likelihood function:

$$L(\theta) = \log f_\mathbf{y}(\mathbf{y} \mid \theta) \tag{3.2}$$

The score function [32] is obtained by taking the gradient of (3.2) with respect to θ. It is known that smaller scores correspond to better estimates, and in fact a score of zero corresponds to a critical point of the likelihood function, which is normally a maximum. Performing this operation, we have:

$$\nabla_\theta L(\theta) = \frac{1}{f_\mathbf{y}(\mathbf{y} \mid \theta)} \nabla_\theta f_\mathbf{y}(\mathbf{y} \mid \theta) \tag{3.3}$$

Furthermore, we know that

$$f_\mathbf{y}(\mathbf{y} \mid \theta) = \int_\mathbf{z} f_\mathbf{y}(\mathbf{y} \mid \mathbf{z}; \theta) f_\mathbf{z}(\mathbf{z}) d\mathbf{z} \tag{3.4}$$

Substituting (3.4) into (3.3) and interchanging the order of integration and differentiation, we have:

$$\nabla_\theta L(\theta) = \frac{1}{f_\mathbf{y}(\mathbf{y} \mid \theta)} \int_\mathbf{z} \nabla_\theta f_\mathbf{y}(\mathbf{y} \mid \mathbf{z}; \theta) f_\mathbf{z}(\mathbf{z}) d\mathbf{z} \tag{3.5}$$

From (3.1), we have that

$$f_\mathbf{y}(\mathbf{y} \mid \mathbf{z}; \theta) = K \exp\left(-\frac{1}{2} (\mathbf{y} - \mathbf{A}_\theta \theta - \mathbf{A}_z \mathbf{z})^T \mathbf{R}_\mathbf{x}^{-1} (\mathbf{y} - \mathbf{A}_\theta \theta - \mathbf{A}_z \mathbf{z}) \right) \tag{3.6}$$

where K is a normalization constant. It is straightforward to show that

$$\nabla_\theta f_{\mathbf{y}}(\mathbf{y} \mid \mathbf{z}; \theta) = \mathbf{A}_\theta^T \mathbf{R}_{\mathbf{x}}^{-1}(\mathbf{y} - \mathbf{A}_\theta \theta - \mathbf{A}_z \mathbf{z}) f_{\mathbf{y}}(\mathbf{y} \mid \mathbf{z}; \theta) \tag{3.7}$$

Substituting (3.7) into (3.5), we have

$$\nabla_\theta L(\theta) = \frac{1}{f_{\mathbf{y}}(\mathbf{y})} \int_{\mathbf{z}} \mathbf{A}_\theta^T \mathbf{R}_{\mathbf{x}}^{-1}(\mathbf{y} - \mathbf{A}_\theta \theta - \mathbf{A}_z \mathbf{z}) f_{\mathbf{y}}(\mathbf{y} \mid \mathbf{z}; \theta) f_{\mathbf{z}}(\mathbf{z}) d\mathbf{z} \tag{3.8}$$

By Bayes' rule, we have

$$\begin{aligned} \nabla_\theta L(\theta) &= \int_{\mathbf{z}} \mathbf{A}_\theta^T \mathbf{R}_{\mathbf{x}}^{-1}(\mathbf{y} - \mathbf{A}_\theta \theta - \mathbf{A}_z \mathbf{z}) f_{\mathbf{z}}(\mathbf{z} \mid \mathbf{y}; \theta) d\mathbf{z} \\ &= \mathbf{A}_\theta^T \mathbf{R}_{\mathbf{x}}^{-1}(\mathbf{y} - \mathbf{A}_\theta \theta - \mathbf{A}_z E[\mathbf{z} \mid \mathbf{y}; \theta]) \end{aligned} \tag{3.9}$$

To find the critical points, we set (3.9) to zero and solve for θ:

$$\theta = \mathbf{K}(\mathbf{y} - \mathbf{A}_z E[\mathbf{z} \mid \mathbf{y}; \theta]) \tag{3.10}$$

where $\mathbf{K} = (\mathbf{A}_\theta^T \mathbf{R}_{\mathbf{x}}^{-1} \mathbf{A}_\theta)^{-1} \mathbf{A}_\theta^T \mathbf{R}_{\mathbf{x}}^{-1}$.

From (3.10), we have that the ML estimate of θ is calculated by using the MMSE estimates of the interfering random variables to cancel the interference that they generate. Furthermore, from (3.7), the soft decision structure seems to be a distinct property of the Gaussian distribution. The solution to (3.10) requires us to solve a set of nonlinear simultaneous equations, but the difficulty is mitigated through iterative techniques such as the EM algorithm. In fact, (3.10) bears notable similarities to the EM solutions derived in chapter 2, which is not a surprise since the EM algorithm is merely an iterative procedure that seeks the maximum of a likelihood space.

Let us now return to the problem of multiuser detection. We must first redefine (3.1) as follows:

$$\mathbf{y} = \mathbf{R}\mathbf{A}\mathbf{b} + \mathbf{z} \tag{3.11}$$

To remain consistent with the above, and without loss of generality, define the first k elements of $\mathbf{b}$ to be the binary parameters $\theta \in \{-1, 1\}$, and define the remaining l elements to be

binary symmetric random variables $\mathbf{z} \in \{-1, 1\}$. Following the development in [8], let us set $k = 1$, i.e., we are only detecting one user's symbol, while estimating the others. The elements of the MMSE estimate vector from (3.10), $E[\mathbf{z} \mid \mathbf{y}; \theta]$ are then given by

$$\begin{aligned} [E[\mathbf{z} \mid \mathbf{y}; \theta]]_k &= \sum_{\mathbf{z}} b_k f_{\mathbf{z}}(\mathbf{z} \mid \mathbf{y}, \theta) \\ &= \frac{\sum_{\tilde{\mathbf{b}} \in \{-1,1\}^{n-1}} b_k f_{\mathbf{y}}(\mathbf{y} \mid \mathbf{b}) P(\tilde{\mathbf{b}})}{\sum_{\tilde{\mathbf{b}} \in \{-1,1\}^{n-1}} f_{\mathbf{y}}(\mathbf{y} \mid \mathbf{b}) P(\tilde{\mathbf{b}})} \end{aligned} \tag{3.12}$$

where $\tilde{\mathbf{b}}$ represents the vector of all symbols excluding the symbol to be detected, and $f_{\mathbf{y}}(\mathbf{y} \mid \mathbf{b})$ is the likelihood function defined above in (2.5). Substituting $\bar{b}_1$ for b_1 in (3.12), this expression can be used in the E-step of the EM algorithm as the estimate of any bit b_k. However, it is straightforward to show that this expression is not equivalent to (2.15) (the three-user case provides a simple counterexample), which implies that the expression in [8] is an approximation. The problem with the exact expression in (3.12), and the motivation for an approximate solution, is that it is difficult to express it in closed form and hence difficult to calculate concisely.

3.2 A Turbo Detection Interpretation

The MPEM algorithm from [8] was given as an ad-hoc modification to the EM algorithm to improve the accuracy of the E-step. In this section, we shall show that the MPEM E-step may also be understood from the perspective of "turbo detection," discussed in [37, 38, 39].

Let us first give some relations that will aid in this analysis. Firstly, we will show a simple relation between the MMSE estimate of a binary random variable and its *a posteriori* probability. Let $b \in \{-1, 1\}$ be a binary random variable, and let y be some corrupted observation of b. From [27], we know that the minimum mean squared error estimate of b is given by the expected value conditioned on y, $E[b|y]$. Further, we know that

$$E[b|y] = P(b = 1|y) - P(b = -1|y) \tag{3.13}$$

Let $\hat{b} \in \{-1, 1\}$ be a decision on b. We may now substitute into (3.13) as follows:

$$E[b|y] = \hat{b}[2P(b = \hat{b}|y) - 1] \tag{3.14}$$

Furthermore, since $\hat{b}^2 = 1$, the above expression may also be given by

$$P(b = \hat{b}|y) = \frac{1}{2}[1 + \hat{b}E(b|y)] \tag{3.15}$$

Using the above, we can also show an equally straightforward relationship between the MMSE estimate and the Log-Likelihood Ratio (LLR). We begin with the following identity, which is straightforward to obtain:

$$\log \frac{1+x}{1-x} = 2 \operatorname{arctanh} x \tag{3.16}$$

From the above, we have that $P(b = 1 \mid y) = [1 + E(b|y)]/2$, while $P(b = -1 \mid y) = [1 - E(b|y)]/2$. Therefore, letting $x = E(b|y)$, and assuming that the symbols $b \in \{-1, 1\}$ are equiprobable, we have that

$$L = \log \frac{1+x}{1-x} \tag{3.17}$$

where L is the LLR, which is the same expression as is on the left in (3.16). Therefore, it is simple to see that

$$E(b \mid y) = \tanh\left(\frac{L}{2}\right) \tag{3.18}$$

which is true for every density of y. Identity (3.16) was used to obtain related results in [31].

3.2.1 The EM Algorithm and Turbo detection

As pointed out in Chapter 2, turbo detection has already been applied to the combined problem of equalization and decoding. While the original turbo arrangement relied on iterations through two codes applied simultaneously, the theory of turbo detection replaces the extrinsic information normally gained from one of these codes with the soft outputs of the equalization algorithm. It is straightforward to adapt the EM-based multiuser detection

scheme proposed by [8] to turbo detection. Suppose our bit of interest is b_k. In development related to (3.10), (2.14), and (2.16), the EM algorithm which solves this problem is an iteration defined by

$$\bar{b}_k^{(n+1)} = G\left(y_k - \mathbf{r}_{\bar{k}}\mathbf{A}E[\mathbf{b} \mid \mathbf{y}, \bar{b}_k^{(n)}]\right) \tag{3.19}$$

where $\mathbf{r}_{\bar{k}}$ is defined as the kth row of the cross-correlation matrix $\mathbf{R}$ where the kth element is replaced by zero, $\mathbf{b}_k^{(n)}$ is the estimate of the vector $\mathbf{b}_k$ at iteration n, and $G(\cdot)$ is a decision function, ideally a slicer. The iteration is expressed in the above form to be consistent with (3.12), although equivalently it may be expressed as

$$\bar{b}_k^{(n+1)} = G\left(y_k - \sum_{j \neq k} R_{jk} a_j \tilde{b}_j\right) \tag{3.20}$$

where $\tilde{b}_j = E[b_j \mid \mathbf{y}, \bar{b}_k^{(n)}]$ are the estimates of the interfering users from the E-step. From (3.18), we know that the expectation in (3.19) and (3.20) may be replaced with a straightforward expression involving the LLR of b_k, which allows simple addition of extrinsic information gleaned from decoding. Suppose an arbitrary code is applied to each user's transmission in a multiuser detection system, and suppose that the observations of the parity bits give us the extrinsic information L_{e,b_j} on the received symbol b_j, as well as some prior information L_{p,b_j}, as in the method from [31]. From the above analysis, a straightforward change may be made in (3.19) and (3.20) to reflect the extrinsic code information. The estimates $\tilde{b}_j$ may be replaced by

$$\tilde{b}_j = \tanh\left(\operatorname{arctanh}(E[b_j \mid \mathbf{y}, b_k^{(n)}]) + \frac{L_{e,b_j} + L_{p,b_j}}{2}\right) \tag{3.21}$$

Furthermore, with judicious choice of a decision function, the LLR of the new estimates $b_k^{(n+1)}$ may be extracted (or approximated) and fed back to further decoding iterations. Since the primary focus of this work is not on coding, we will leave the more serious theoretical issues to future work.

3.2.2 Uncoded multiuser detection by the Turbo principle

Alternatively, apart from coding, we may consider the MPEM algorithm itself to be an implementation of turbo detection, with certain minor modifications. For the purposes of illustration, consider the two-user case with unit amplitudes. We have two correlated outputs corresponding to the outputs of filters matched to each user's code sequence. Rather than treating the entire task of multiuser detection to be equivalent to a single decoding step under turbo detection, let us instead distinguish each observation at the output of the matched filter bank as an effective "code" that must be iteratively evaluated. In Chapter 2, a model for iterative decoding was given that relied on separating the LLR into terms expressing intrinsic (channel) information, prior information (from previous iterations and initial knowledge), and extrinsic information (from parity bits). Following (2.22), we may express this quantity as follows:

$$L = L_y + L_p = L_c + L_e + L_p \tag{3.22}$$

where L_c is the LLR from channel observations, L_p is the *a priori* LLR, and L_e is the extrinsic LLR. From a coding perspective, Hagenauer et al. [31] gave several coding schemes that could be applied to turbo decoding by reducing them to the form in (3.22).

Initially, one might consider a scheme that operates exclusively on the observations $\mathbf{y}$. However, this would be impractical because the elements of $\mathbf{y}$ are not independent given the vector of symbols $\mathbf{b}$, and thus it would be impossible to express the multiuser detection problem in an additive form, as in (3.22). In turbo decoding, the problem of correlated observations is dealt with by interleaving. Unfortunately, this recourse is not available in multiuser detection, since every user has some degree of correlation with every other user, and the effective block size dealt with by the receiver is a fixed quantity, equal to the number of users. We will instead apply two whitening operations to the vector $\mathbf{y}$ to obtain a conditionally independent set of observations:

$$\mathbf{y}_W = G^{-1}\mathbf{y} \tag{3.23}$$

$$\hat{\mathbf{y}}_W = G^{-1}\hat{\mathbf{y}} \tag{3.24}$$

where $\hat{\mathbf{y}} = [y_2 y_1]^T$, i.e., the same elements as $\mathbf{y}$ but in the reverse order. This formulation gives

$$\begin{aligned} y_{W,1} &= G_{1,1}^T b_1 + G_{1,2}^T b_2 + n_1 \\ y_{W,2} &= G_{2,2}^T b_2 + n_2 \end{aligned} \tag{3.25}$$

$$\begin{aligned} \hat{y}_{W,1} &= G_{1,1}^T b_2 + G_{1,2}^T b_1 + \hat{n}_1 \\ \hat{y}_{W,2} &= G_{2,2}^T b_1 + \hat{n}_2 \end{aligned} \tag{3.26}$$

As in turbo decoding, it is straightforward to separate (3.25) and (3.26) into channel information and extrinsic information. Specifically, we see that $\hat{y}_{W,2}$ depends on symbol b_1 alone, while $y_{W,2}$ depends on symbol b_2 alone. This is analogous to the observation of the systematic information bits in a code sequence, and we may therefore use $\hat{y}_{W,2}$ and $y_{W,2}$ to calculate the channel LLR for b_1 and b_2, respectively:

$$L_{c,b_1} = \frac{G_{2,2}^T}{\sigma^2} \hat{y}_{W,2} \tag{3.27}$$

$$L_{c,b_2} = \frac{G_{2,2}^T}{\sigma^2} y_{W,2} \tag{3.28}$$

It is shown in Appendix B that these LLR values are equal to the LLR values for each user at the output of the decorrelator, which justifies the use of the decorrelator as an initial stage detector in [8]. Now, from (3.25) and (3.26), we see that additional information concerning b_1 and b_2 is available in $\hat{y}_{W,1}$ and $y_{W,1}$, so these observations will be used to calculate the extrinsic information. Because $\hat{y}_{W,1}$ and $\hat{y}_{W,2}$ are conditionally independent as a result of the whitening operation, we will use $\hat{y}_{W,1}$ to calculate the extrinsic information for b_1 - and by a similar argument we will use $y_{W,1}$ to calculate the extrinsic information for b_2. Using b_2 as an example, we must now calculate

$$L_{e,b_2} = \log \frac{f_{y_{W,1}}(y_{W,1} \mid b_2 = 1)}{f_{y_{W,1}}(y_{W,1} \mid b_2 = -1)} \tag{3.29}$$

The above quantity is relatively straightforward to calculate for two users, though when this concept is extended to several users we will require a simpler approximation. In turbo decoding, parity bits give information concerning their information bits by inverting the

modulo-2 addition used to form them. Analogously, we may approximately calculate L_{e,b_2} by subtracting an estimate for the interfering symbol. Furthermore, from equation (3.10) in section 3.1, it is observed that for the purposes of log-likelihood calculations it is reasonable to approximate unknown parameters with their MMSE estimates. A reasonable approximation for the distribution $f_{y_{W,1}}(y_{W,1} \mid b_2)$ is then

$$f_{y_{W,1}}(y_{W,1} \mid b_2) \simeq f_{y_{W,1}}\left(y_{W,1} \mid b_2; b_1 = \tanh\left(\frac{L_{b_1}}{2}\right)\right) \tag{3.30}$$

The reliability of this assumption will be evaluated quantitatively in a later section. We know that $f_{y_{W,1}}(y_{W,1} \mid b_1, b_2)$ is Gaussian. Therefore, the extrinsic information is given by

$$L_{e,b_2} = \log \frac{f_{y_{W,1}}(y_{W,1} \mid b_2 = 1)}{f_{y_{W,1}}(y_{W,1} \mid b_2 = -1)} = \frac{2G_{1,2}^T}{\sigma^2}\left(y_{W,1} - G_{1,1}^T\left(\frac{L_{b_1}}{2}\right)\right) \tag{3.31}$$

where L_{b_1} is the LLR of b_1, given by $L_{b_1} = L_{c,b_1} + L_{p,b_1}$ - initially, L_{p,b_1} is equal to zero, but as the detection algorithm proceeds, it takes the value of the extrinsic information calculated in the previous iteration. Making a few appropriate substitutions, it is straightforward to calculate the extrinsic information for b_1:

$$L_{e,b_1} = \log \frac{f_{\hat{y}_{W,1}}(\hat{y}_{W,1} \mid b_2 = 1)}{f_{\hat{y}_{W,1}}(\hat{y}_{W,1} \mid b_2 = -1)} = \frac{2G_{1,2}^T}{\sigma^2}\left(\hat{y}_{W,1} - G_{1,1}^T\left(\frac{L_{b_2}}{2}\right)\right) \tag{3.32}$$

Once the extrinsic information has been calculated for a symbol, it is fed back in subsequent iterations to be used as the *a priori* LLR for that bit, and used in the interference estimates in (3.31) and (3.32). This detection algorithm has strong links with algorithms from [8]. Consider b_1 - taking the sum of the extrinsic LLR from (3.31) and the channel LLR from (3.27), we have

$$L_{b_1} = \frac{2}{\sigma^2}[y_2 - R_{1,2}\tanh(L_{b_2}/2)] \tag{3.33}$$

and a similar argument may be made concerning L_{b_2}. Assuming that the feedback function $G(x)$ is given by $\tanh(x/2)$, (3.33) expresses the MPEM algorithm from (2.18) and (2.19) for two users. Figure 3.1 illustrates the basic principle of operation of the turbo detectors discussed thus far.

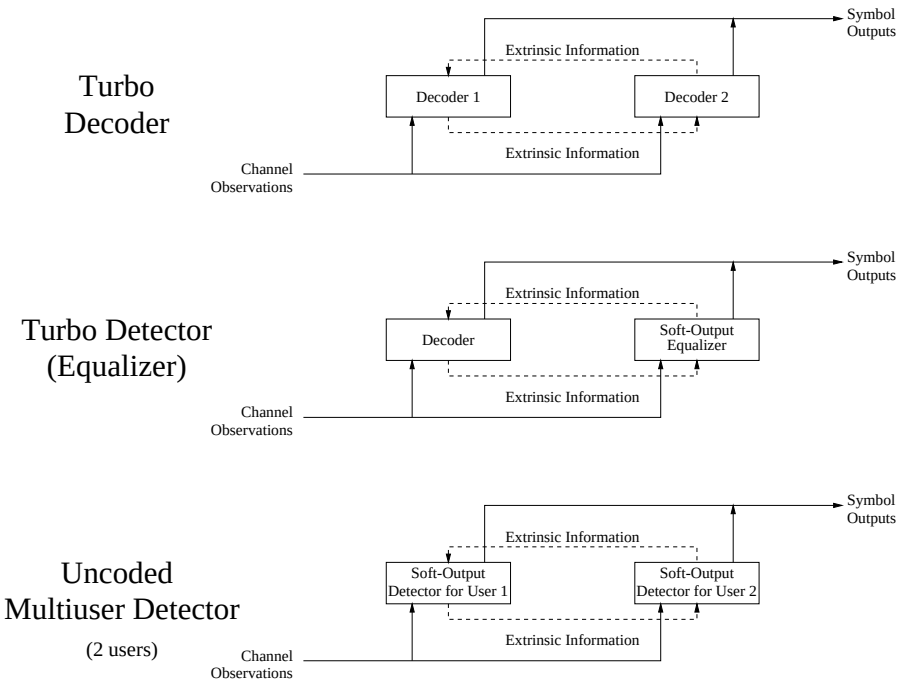

Figure 3.1: Block diagram representations of turbo decoders and variants

Above, we have shown how a two-user scheme can be reformulated as a turbo detection problem. Let us now generalize this discussion to deal with the n-user detection problem, with arbitrary but known amplitudes. In [39], the definition of a turbo detector is generalized to deal with the case of several stages. Individually, for each user's symbol b_k, we must calculate the quantity L_{b_k}, the LLR of that symbol. Equivalently, we must individually calculate for each user the quantity $L_{c,b_k} + L_{e,b_k}$, which is the total of the sum of the channel and extrinsic probabilities. We begin with the likelihood function for the whitened set of observations $\mathbf{y}_W$:

$$f_{\mathbf{y}_W}(\mathbf{y}_W \mid \mathbf{b}) = K \exp\left(-\frac{1}{2\sigma^2}(\mathbf{y}_W - \mathbf{G}^T\mathbf{A}\mathbf{b})^T(\mathbf{y}_W - \mathbf{G}^T\mathbf{A}\mathbf{b})\right) \tag{3.34}$$

where K is a normalization constant. We may separate the term $\mathbf{G}^T\mathbf{A}\mathbf{b}$ into two additive expressions, one consisting of all terms in b_k and one consisting of all terms not in b_k, as follows:

$$\mathbf{G}^T\mathbf{A}\mathbf{b} = \mathbf{h}_k a_k b_k + \mathbf{H}_k \mathbf{A}\mathbf{b} \tag{3.35}$$

where $\mathbf{h}_k$ is the kth column of $\mathbf{G}^T$, and $\mathbf{H}_k$ is $\mathbf{G}^T$ with the kth column replaced by zeros. Substituting back into (3.34), and calculating the log-likelihood ratio, we now have

$$L_{b_k} = \frac{2a_k}{\sigma^2}\mathbf{y}_W^T\mathbf{h}_k - \frac{2a_k}{\sigma^2}\mathbf{b}^T\mathbf{A}\mathbf{H}_k^T\mathbf{h}_k \tag{3.36}$$

Recall that $\mathbf{y}_W = \mathbf{G}^{-1}\mathbf{y}$. Since $\mathbf{h}_k$ is the kth column of $\mathbf{G}^T$, and since the order of transposition and inversion may be exchanged in this case, we have that $\mathbf{y}_W^T\mathbf{h}_k = \mathbf{y}$. Furthermore we may express $\mathbf{b}^T\mathbf{A}\mathbf{H}_k^T\mathbf{h}_k$ as $\mathbf{b}^T\mathbf{A}(\mathbf{G}-\mathbf{U}_k)\mathbf{h}_k$, where the elements of $\mathbf{U}_k$ are all zero except for the kth row, which are given by $\mathbf{h}_k^T$. We know that $\mathbf{G}\mathbf{h}_k$ gives the kth column of the cross-correlation matrix $\mathbf{R}$, while $\mathbf{U}_k\mathbf{h}_k$ gives a vector whose elements are all zero except for the kth element, which is equal to R_{kk}. Thus, L_{b_k} is given by

$$L_{b_k} = \frac{2a_k}{\sigma^2}\left(y_k - \sum_{j\neq k} R_{jk}a_j b_j\right) \tag{3.37}$$

for which the special case of two users with unit amplitudes is given by (3.33). Furthermore, if as before we substitute all b_j with their MMSE estimates $\tanh(L_{b_j}/2)$, this function becomes

$$L_{b_k} = \frac{2a_k}{\sigma^2}\left(y_k - \sum_{j\neq k} R_{jk}a_j \tanh\left(\frac{L_{b_j}}{2}\right)\right) \tag{3.38}$$

It is interesting to note that the form of the final expression is the same regardless of the value of k, which implies that the initial ordering of the users is irrelevant.

Let us now define a complete multiuser detection algorithm based on the development up to (3.38). Following the turbo detection scheme, it is natural to define an algorithm as follows. First calculate the intrinsic information for each symbol by calculating $L_{c,b_k} = 2d_k/\sigma_d^2$, where d_k are obtained from the decorrelator. Define an index $k = 1$, and calculate L_{b_k} using (3.38). Subsequently let $k = [(k+1) \bmod n] + 1$, where n is the number of users, and repeat the calculations of L_{b_k} until the algorithm converges. With minor modifications, this algorithm is identical to the MPEM algorithm outlined in Chapter 2 and in [8]. The advantage to this approach is that it is more straightforward to adapt the derivation given here to turbo coding, following work done above in Subsection (3.2.1).

3.3 Analysis of Soft Decision Methods

In the previous section, we have shown that a multiuser detection scheme can be derived by considering turbo detection, which is similar to some EM-based approaches in [8]. It is also possible to interpret the final result above from (3.38) as a successive interference cancellation scheme. We will now discuss some of the assumptions used in the above derivation. Letting $\tilde{b}_j = \tanh(L_{b_j}/2)$, we may introduce a quantity u_k as an observation of the channel with interference cancelled:

$$u_k = y_k - \sum_{j \neq k} a_j R_{jk} \tilde{b}_j \tag{3.39}$$

Knowing that $y_k = a_k b_k + \sum_{i \neq k} a_i R_{ik} b_i + n_k$, where b_i are the actual values of the bits, and n_k is a Gaussian random variable with zero mean and variance $\sigma^2 R_{kk}$, equation (3.39) may be re-expressed as follows:

$$u_k = a_k b_k + \sum_{j \neq k} a_j R_{jk} \left(b_j - \tilde{b}_j \right) + n_k = a_k b_k + I_k + n_k \tag{3.40}$$

where I_k is referred to as the residual interference. The assumption that the Gaussian approximation leading to (3.30) is valid was discussed and demonstrated in [43].

It is also of interest to determine in an analytical manner the effects of *A Posteriori* Probability (APP) feedback. Ideally, such analysis would examine the probability of error, although the analysis for that performance measure is exceedingly difficult. In this case, it is simpler to analyze the resultant variance after a stage is complete. It is easy to see that the variance of a single component of the residual interference is given by:

$$V[i_{k,i}] = \int_{i_{k,i}} i_{k,i}^2 f(i_{k,i}) di_{k,i} \tag{3.41}$$

where $f(i_{k,i})$ is the density of a single component of the residual interference. Equivalently, the following expression will also calculate this variance:

$$V[i_{k,i}] = V[R_{ik}(b_k - \tilde{b}_k)] = R_{ik}^2 \left(E[b_k^2] - 2E[b_k \tilde{b}_k] + E[\tilde{b}_k^2] \right) \tag{3.42}$$

The first term in the above calculation, $E[b_k^2]$, is clearly equal to 1. Since $\tilde{b}_k$ is a function of the observation u, we can calculate the variance of the latter two terms as follows:

$$E[b_k\tilde{b_k}] = \sum_{b_k} b_k f_{b_k}(b_k) \int_u \tilde{b}_k(u) f_u(u|b_k) du \tag{3.43}$$

$$E[b_k^2] = \sum_{b_k} f_{b_k}(b_k) \int_u \tilde{b}_k^2(u) f_u(u|b_k) du \tag{3.44}$$

For the first stage of the algorithm, and approximately for subsequent stages, the distribution of the observations is Gaussian, and $\tilde{b}_k = \tanh(u/\sigma^2)$. Calculating the variance with the above equations using this knowledge is very difficult. However, since from [27] we know that the estimate $\tilde{b}_k = E[b_k|u]$ is the minimum-variance estimate, we may simply choose any convenient function to calculate this quantity, and use the result as an upper bound on the variance. One such function is the linear clipper, given as follows:

$$f(x) = \begin{bmatrix} -1, & x < -\sigma^2 \\ u/\sigma^2, & -\sigma^2 \leq x \leq \sigma^2 \\ 1, & x > \sigma^2 \end{bmatrix} \tag{3.45}$$

which corresponds to a first-order Taylor series expansion of $\tanh(u/\sigma^2)$, hard limited to -1 and 1. If we use the linear clipper in place of the estimates $\tilde{b}_k$ in (3.43) and (3.44), we find that (3.43) becomes relatively straightforward to calculate, while (3.44) remains quite difficult. It is easy to show, however, that the function in (3.44) is always less than 1, and in fact that 1 is a reasonable approximation for this quantity under conditions of high SNR. Thus, if 1 is substituted for (3.44), it is straightforward to show that the variance upper bound calculated from the linear clipper in (3.45) is given by:

$$\begin{aligned} V[i_{k,i}] \leq & \\ & 2R_{ik}^2 \Big[\frac{1}{\sigma\sqrt{2\pi}} \Big[\exp\Big(\frac{-(\sigma^2-1)^2}{2\sigma^2}\Big) - \exp\Big(\frac{-(\sigma^2+1)^2}{2\sigma^2}\Big) \Big] \\ & + \frac{1}{2}\Big(1 - \frac{1}{\sigma^2}\Big) \Big[\mathrm{erf}\Big(\frac{1+\sigma^2}{\sigma\sqrt{2}}\Big) - \mathrm{erf}\Big(\frac{1-\sigma^2}{\sigma\sqrt{2}}\Big) \Big] \Big] \end{aligned} \tag{3.46}$$

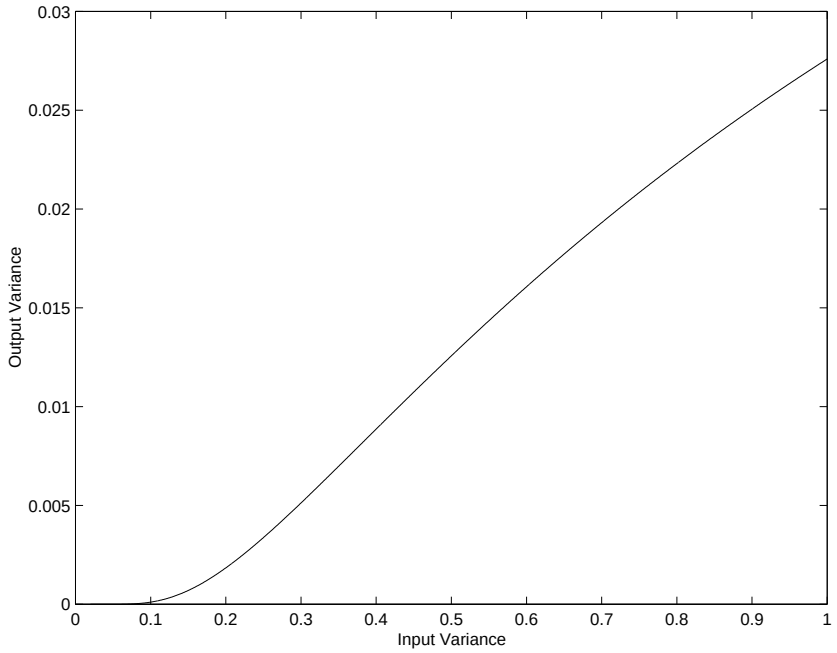

Figure 3.2: Plot of the output residual interference variance versus input noise variance

where $\mathrm{erf}(\cdot)$ is the error function. We continue to assume that individual residual interference terms are independent. Therefore, the overall variance of all corrupting influences in the observation u is given as follows:

$$V[I_k] = \sum_i V[i_{k,i}] + \sigma_n^2 \tag{3.47}$$

where σ_n^2 is the thermal noise already present in the signal. The variance upper bound for (3.47) is easily obtained by substituting the expression in (3.46) for each term $V[i_{k,i}]$.

In figure 3.2, a plot of the variance upper bound is given for five users, where $R_{ij} = 0.2$ for all $i \neq j$, and the corresponding log-log plot is given in figure 3.3. We remark that the variance upper bound in (3.46) is always smaller than the input variance, so from (3.47) we expect that the overall variance will decline to some value slightly greater than that of the existing thermal noise, which is verified by simulation. Furthermore, we note that for relatively high values of SNR ($\geq$ 10 dB), the resultant interference is basically negligible, and thus the algorithm can be considered to converge in one iteration. It is reasonable to assume that there is a monotonic relationship between variance and bit error rate, though

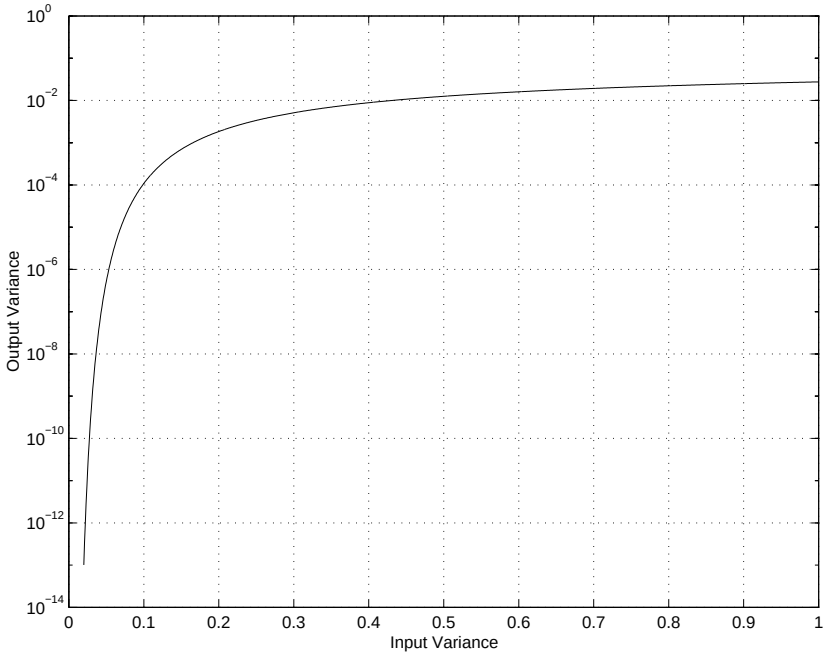

Figure 3.3: Log-log plot of the output residual interference variance versus input noise variance

we shall leave the determination of the precise relationship to future work.

3.4 Adaptation to Random Parameters with Continuous Values

In this analysis, we have discussed the detection of discrete random parameters (i.e., users' transmitted symbols) using turbo detection and modifications to the EM algorithm. In later chapters, we will be examining the problem of estimating continuous random parameters (i.e., users' symbols and unknown fading amplitudes) within the same context of a multiuser detection system. This new requirement complicates the problem considerably.

Let us consider a generic case where we have some random parameter ϕ, with known PDF $f_\phi(\phi)$, and two corrupted observations of ϕ:

$$\begin{aligned} u_1 &= k_1\phi + n_1 \\ u_2 &= k_2\phi + n_2 \end{aligned} \tag{3.48}$$

where k_1 and k_2 are known constants, and where n_1 and n_2 are independent (which implies that u_1 and u_2 are independent when conditioned on ϕ). The MMSE estimate of ϕ given u_1 and u_2 is given by

$$E[\phi \mid u_1, u_2] = \int_\phi \phi f_\phi(\phi \mid u_1, u_2) d\phi = \frac{\int_\phi \phi f_{u_1}(u_1 \mid \phi) f_{u_2}(u_2 \mid \phi) f_\phi(\phi) d\phi}{\int_\phi f_{u_1}(u_1 \mid \phi) f_{u_2}(u_2 \mid \phi) f_\phi(\phi) d\phi} \tag{3.49}$$

It is immediately observed that a general formulation for arbitrary PDFs of u_1, u_2, and ϕ would be difficult to find. However, for the special case where u_1 and u_2 are Gaussian, interesting results are obtained. It is noted that, since u_1 and u_2 are independent and Gaussian when conditioned on ϕ, then

$$\begin{aligned} & f_{u_1}(u_1 \mid \phi) f_{u_2}(u_2 \mid \phi) \\ &= K \exp\left(-\frac{1}{2\sigma^2}(u_1 - k_1\phi)^2 - \frac{1}{2\sigma^2}(u_2 - k_2\phi)^2 \right) \\ &= K \exp\left[-\frac{1}{2\sigma^2} \left(\frac{k_1 u_1}{\sqrt{k_1^2 + k_2^2}} + \frac{k_2 u_2}{\sqrt{k_1^2 + k_2^2}} - \phi\sqrt{k_1^2 + k_2^2} \right)^2 + h_u \right] \end{aligned} \tag{3.50}$$

where h_u is some function of u_1 and u_2 alone. Substituting this expression back into (3.49), we find that the h_u term and the constant K will cancel, leaving an equivalent Gaussian PDF formed by the weighted sum of u_1 and u_2. It is straightforward to show that this idea extends to n observations, as follows:

$$\prod_{i=1}^{n} f_{u_i}(u_i|\phi) = K \exp\left[-\frac{1}{2\sigma^2} \left(\sum_{i=1}^{n} \frac{k_i}{D} u_i - D\phi \right)^2 + h_u \right] \tag{3.51}$$

where $D = \sqrt{\sum_{i=1}^{n} k_i^2}$. Suppose that we have a normalized estimator $M_\phi(\hat{u}, \hat{\sigma}^2)$ that is defined as the MMSE estimator for ϕ given $\hat{u} = \phi + n$, where the variance of n is $\hat{\sigma}^2$. It is straightforward to rearrange (3.51) such that the MMSE estimate of ϕ given the set of observations $\{u_i\}$, $i \in [1,\ 2,\ \cdots,\ n]$, is given by

$$\tilde{\phi} = M_{\phi}\left(\sum_{i=1}^{n} \frac{k_i}{D^2} u_i \, , \, \frac{\sigma^2}{D^2}\right) \tag{3.52}$$

We remark that this derivation does not depend on the PDF of ϕ. Therefore, this idea also applies to the discrete antipodal case with known amplitudes considered in previous sections. Above, we considered the whitened vector of observations, conditioned that other random parameters were equal to their MMSE estimates, as independent observations of a given symbol b_i. Here, let us substitute $\mathbf{p} = \mathbf{A}\mathbf{b}$, and since $\mathbf{A}$ is a diagonal matrix, the elements of $\mathbf{p}$ are given by $p_k = a_k b_k$. In a fashion similar to the development leading to (3.38), if we are to estimate p_k rather than b_k, it is straightforward to show that $D = 1$, and furthermore that

$$\sum_{i=1}^{n} \frac{k_i}{D} u_i = y_k - \sum_{j \neq k} R_{jk} \tilde{p}_j \tag{3.53}$$

and, following (3.52), the MMSE estimate follows easily, given the appropriate function $M_{p_k}(\cdot, \cdot)$ for the distribution of p_k. The similarity of form between this expression and (3.38) indicates that an estimation algorithm may be defined similarly to the turbo detection scheme proposed for random discrete symbols, where estimates rather than LLR values are fed back. In the subsequent chapter we shall propose receivers based on such an algorithm.

Chapter 4

Receiver Design for Multiuser Detection in Fading

In this chapter, we shall examine the issue of receiver design for multiuser detectors in fading channels, with particular emphasis on the EM algorithm and extensions of the work in [8]. Some general bounds on performance in fading channels will first be presented, followed by receiver designs for phase coherent multiuser detection in Rayleigh and Rician channels. Receiver designs for incoherent multiuser detection will also be presented based on MLSD criteria for Rayleigh and Rician channels. All proposed receiver designs are accompanied by simulation results.

4.1 Bounds on Algorithm Performance

We continue with the matched filtered received signal model from (2.30). The particular design problem that this chapter will tackle involves the cancellation of interference based on knowledge of the PDF of the fading parameters, and possibly the knowledge of the signal phase. Prior to delving into the issue of designing receivers, it is helpful to have some idea of the performance and limitations that we can expect from them, particularly concerning their performance compared to existing receiver designs. Calculation of the bit error rates of an arbitrary receiver is in general a difficult problem, so let us instead calculate some lower

bounds on the error performance under two simplifying assumptions: that the interference is perfectly known, and that the channel state is perfectly known.

4.1.1 Perfectly known interference bound

Let us assume that the interference generated by all users other than the user of interest is perfectly known to the receiver. In this case, the interference may be exactly cancelled and the channel reduces to the single-user AWGN case:

$$y = a\theta b + n \tag{4.1}$$

where a is the fading amplitude, θ is the phase, b is the transmitted symbol, and n is a circularly symmetric complex Gaussian noise component with variance σ^2. To simplify detection, let us consider the phase coherent case, i.e., that θ is known. If we have an expression for the Bit Error Rate (BER) conditioned on the amplitudes, by Bayes' rule the marginal bit error rate is given by

$$P_e = \int_a P(\text{err} \mid a) f_a(a) da \tag{4.2}$$

The BER conditioned on knowledge of the amplitude is well known:

$$P(\text{err} \mid a) = Q\left(\frac{a}{\sigma}\right) \tag{4.3}$$

where $Q(x) = (1/\sqrt{2\pi}) \int_x^\infty \exp(-y^2/2) dy$, which is related to the complementary error function. In the case where $f_a(a)$ is the Rayleigh distribution with parameter ρ, as in (2.26), the bit error rate of this system is given by [33]:

$$P_e = \frac{1}{2}\left(1 - \sqrt{\frac{\gamma}{2+\gamma}}\right) \tag{4.4}$$

where γ represents the average signal-to-noise ratio, i.e., $\gamma = 2\rho^2/\sigma^2$. As $\gamma \to \infty$, it is straightforward to show that

$$P_e \simeq \frac{1}{2\gamma} \mid_{\gamma \to \infty} \tag{4.5}$$

Recall that the optimal phase-coherent detector in the absence of any knowledge of the amplitudes is the decorrelator. Detection with a decorrelator may be modelled as detection in a single-user AWGN channel where the noise variance is $[\mathbf{R}^{-1}]_{ii}\sigma^2$, where $[\mathbf{R}^{-1}]_{ii}$ is the ith diagonal component of the inverse of the autocorrelation matrix, $\mathbf{R}$, and $[\mathbf{R}^{-1}]_{ii} \geq 1$ always. The probability of error for such a detector in a Rayleigh fading channel may be found by substituting the new γ into (4.4). In particular, in the high SNR domain, from (4.5) we have that $\lim_{\gamma\to\infty} P_e = [\mathbf{R}^{-1}]_{ii}/2\gamma$, where γ is the equivalent single user average SNR. This implies that in the high SNR domain, using phase coherent detection, the performance of the decorrelator and the single user bound are separated by a constant ratio. Expressing the performance difference in terms of decibels, if S_d is the required decorrelator average SNR for a certain error probability, and S_s is the required single user lower bound average SNR, then we have that:

$$10(\log_{10} S_d - \log_{10} S_s) = 10\log_{10}[\mathbf{R}^{-1}]_{ii} \tag{4.6}$$

In general, in DS-CDMA, the cross correlations between users are very small, which implies that $[\mathbf{R}^{-1}]_{ii}$ is only slightly greater than 1. In practical cases, the difference between the decorrelator and the bound from (4.6) is less than 1 dB, or a factor of 1.26.

Although the analysis is somewhat more complicated, results are also available for this bound when we substitute into (4.2) with the Rician PDF, (2.27). From [34], the bit error rate of a single user where a from 4.1 is Rician is given by

$$P_e = Q(u,w) - \frac{1}{2}\left(1 + \sqrt{\frac{\gamma}{2+\gamma}}\right)\exp\left(-\frac{u^2+w^2}{2}\right) I_0(uw) \tag{4.7}$$

where

$$Q(a,b) = \int_b^\infty z\exp\left(-\frac{z^2+a^2}{2}\right) I_0(az)dz$$

$$u = \sqrt{\frac{\alpha\left[1+\gamma-\sqrt{\gamma(2+\gamma)}\right]}{2+\gamma}}$$

$$w = \sqrt{\frac{\alpha\left[1+\gamma+\sqrt{\gamma(2+\gamma)}\right]}{2+\gamma}}$$

The $Q(\cdot,\cdot)$ function in the above is referred to as the Marcum Q function. As well, $I_0(\cdot)$ is the zero order modified Bessel function of the first kind, and α is the ratio between the energy in the direct component and the average energy in the scattered components: $\alpha = |L|^2/2\rho^2$. Finally, γ is no longer the average total SNR, but is the average SNR of the scattered components, and is calculated as in the Rayleigh case: $\gamma = 2\rho^2/\sigma^2$. Asymptotic analysis of this system as SNR approaches ∞ is somewhat difficult, but we may make some intuitive assumptions. If $\alpha \to 0$, the fading approaches the Rayleigh distribution and the constant ratio bound from (4.6) should apply. If $\alpha \to \infty$, the fading can be modelled as a nonzero mean Gaussian random variable with small variance, which implies that over certain ranges of SNR the single user bound should outperform the decorrelator dramatically. Figure 4.1 illustrates the performance gap between the decorrelating detector and the single user bound.

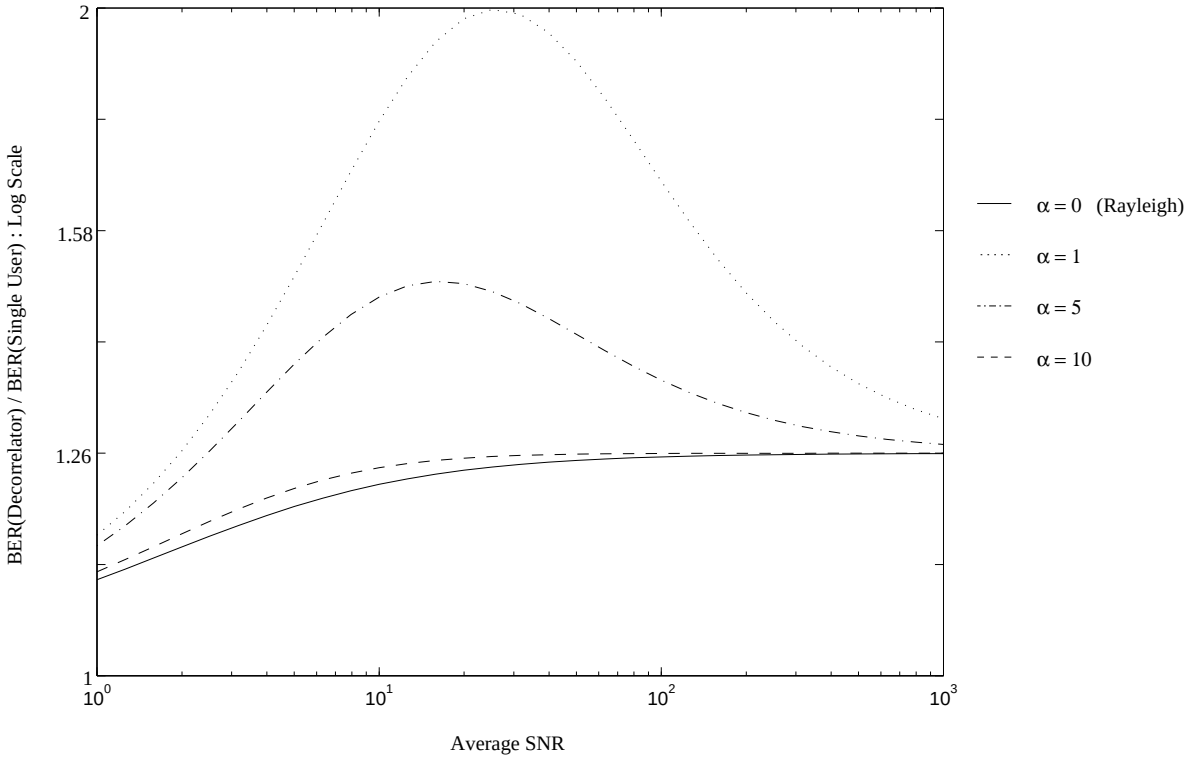

Figure 4.1: Plot of the ratio of decorrelator BER and single user BER versus SNR

Figure 4.1 was generated with $[\mathbf{R}^{-1}]_{ii} = 1.26$ for all i, which corresponds to the expected ratio between the decorrelator BER and the single user BER for a high-SNR Rayleigh

channel. We shall make a brief comment concerning the interesting and somewhat counter-intuitive behavior observed in the figure. For constant α, we know that $\gamma \to \infty$ as the average SNR $S \to \infty$. For $\gamma \gg 1$, it is straightforward to show that

$$\begin{aligned} u &\simeq 0 \\ w &\simeq \sqrt{\tfrac{\alpha}{2}} \end{aligned} \tag{4.8}$$

Given these conditions, it is straightforward to show that the value of the Marcum Q function in (4.7) is given by

$$Q(u,w) \simeq \int_{\sqrt{\alpha/2}}^{\infty} z \exp\left(-\frac{z^2}{2}\right) dz = \exp\left(-\frac{\alpha}{4}\right) \tag{4.9}$$

Substituting back into (4.7), we have that

$$P_e \simeq \frac{1}{2} \exp\left(-\frac{\alpha}{4}\right) \left(1 - \sqrt{\frac{\gamma}{2+\gamma}}\right) \tag{4.10}$$

We know that the rightmost term in (4.10) approaches $1/2\gamma$ for $\gamma \gg 1$. Thus, for very large γ, the value of P_e approaches a constant times $1/\gamma$, and thus in the limit as $\gamma \to \infty$, the factor separating the performance of the decorrelator from the single user bound is constant, regardless of the value of α. However, as we see from Figure 4.1, at lower SNR, much greater improvement may be expected from algorithms operating in Rician channels with high α.

From the above work, we have established that the BER of the optimal multiuser detector in Rayleigh fading is lower bounded by a function proportional to the phase coherent decorrelator, and that the constant of proportionality is close to 1 in practical cases. We have also illustrated that the BER of the optimal multiuser detector in Rician fading has a lower bound several decibels below the decorrelator over specific ranges of SNR. From this analysis we would expect a greater advantage, in general, from applying interference cancellation and amplitude estimation techniques to Rician fading channels.

4.1.2 Perfectly known channel state bound

Another bound that may be used to study the performance of the optimal multiuser receiver in a fading channel is given based on the assumption that the channel state (amplitude and

phase) is perfectly known to the receiver. Historically, this has been an extremely difficult quantity to calculate, but a new expression for the $Q(\cdot)$ function [35, 36] has mitigated some of the difficulty. Nonetheless, we will only solve this case for a two-user system with Rayleigh fading under a high-SNR assumption.

In this case, the receiver at each sampling instant reduces to the well-known maximum likelihood detector proposed by [1]. Intuitively, we expect that this bound should represent a performance worse than the single-user bound. We start with the signal definition, as before:

$$\mathbf{y} = \mathbf{RAb} + \mathbf{z} \tag{4.11}$$

where $\mathbf{y}$ is the observed vector, $\mathbf{R}$ is the matrix of normalized cross-correlation coefficients, $\mathbf{A}$ is the diagonal matrix of random amplitudes, $\mathbf{b}$ is a random antipodal vector of equiprobable bits, and $\mathbf{z}$ is a vector of jointly Gaussian noise components with cross-correlation matrix $\sigma^2 R$. We assume that the known phase values are removed from the signal prior to detection.

Again, we consider the high-SNR case, where σ is much greater than ρ for each component of the signal. Following a development similar to that found in [2], we know that the probability of error for a given user at very high SNR is equivalent to the probability of error between the two hypotheses with the least Euclidean distance between them that differ in the bit of interest. Even with this simplification, this is still a difficult problem to solve - in fact [2] shows that it is NP-complete. However, a relatively simple solution is available for the 2-user case. Here, the minimum distance for user 1 is given by [2]:

$$\begin{aligned} d_1 = a_1, \ \frac{a_2}{a_1} &\geq |r\cos\phi| \\ d_1 = \sqrt{a_1^2 + a_2^2 - |r\cos\phi| a_1 a_2}, \ \frac{a_2}{a_1} &< |r\cos\phi| \end{aligned} \tag{4.12}$$

where ϕ is the phase angle between the signals for user 1 and user 2. Let $s = |r\cos\phi|$. This implies that the probability of error for user 1 given the amplitudes is given by:

$$P(e_1|a_1, a_2) \simeq \begin{bmatrix} Q\left[\frac{a_1}{\sigma}\right], & \frac{a_2}{a_1} \geq s \\ Q\left[\frac{\sqrt{a_1^2 + a_2^2 - s a_1 a_2}}{\sigma}\right], & \frac{a_2}{a_1} < s \end{bmatrix} \tag{4.13}$$

Clearly, the overall probability of error is then given by:

$$P_e = \int_{a_1=0}^{\infty} \int_{a_2=0}^{\infty} P(e_1|a_1, a_2) f_{a_1}(a_1) f_{a_2}(a_2) da_2 da_1 \tag{4.14}$$

Using (4.13), this is broken down into the following:

$$\begin{aligned} P_e = & \\ & \int_{a_1=0}^{\infty} \int_{a_2=0}^{a_1 s} Q\left[\frac{\sqrt{a_1^2 + a_2^2 - s a_1 a_2}}{\sigma}\right] f_{a_2}(a_2) f_{a_1}(a_1) da_2 da_1 \\ & + \int_{a_1=0}^{\infty} \int_{a_2=a_1 s}^{\infty} Q\left[\frac{a_1}{\sigma}\right] f_{a_2}(a_2) f_{a_1}(a_1) da_2 da_1 \end{aligned} \tag{4.15}$$

Let us consider the first term of the above expression. Using the standard definition for $Q[\cdot]$, and substituting the Rayleigh distributions for the fading PDFs, we have:

$$P_{e,1} = \frac{1}{\rho_1^2 \rho_2^2 \sqrt{2\pi}} \int_{a_1=0}^{\infty} \int_{a_2=0}^{a_1 s} \int_{\tau=\frac{\sqrt{a_1^2+a_2^2-s a_1 a_2}}{\sigma}}^{\infty} a_1 a_2 e^{-\frac{\tau^2}{2} - \frac{a_1^2}{2\rho_1^2} - \frac{a_2^2}{2\rho_2^2}} d\tau da_1 da_2 \tag{4.16}$$

This integral is extremely difficult to solve, largely due to the complicated limits of integration for the inner integral. It may be simplified to the following expression using an alternative form of the $Q[\cdot]$ function described in [35, 36]:

$$P_{e,1} = \frac{1}{\rho_1^2 \rho_2^2 \pi} \int_{\theta=0}^{\pi/2} \int_{a_1=0}^{\infty} \int_{a_2=0}^{a_1 s} a_1 a_2 \exp\left(-\frac{a_1^2 + a_2^2 - s a_1 a_2}{2\sigma^2 \sin^2\theta} - \frac{a_1^2}{2\rho_1^2} - \frac{a_2^2}{2\rho_2^2}\right) d\theta da_1 da_2 \tag{4.17}$$

We know that $0 \leq \sin^2\theta \leq 1$. Thus, under the assumption that $(\rho_1, \rho_2) >> \sigma$, the expression in (4.17) can be further simplified to:

$$P_{e,1} \simeq \tilde{P}_{e,1} = \frac{1}{\rho_1^2 \rho_2^2 \pi} \int_{\theta=0}^{\pi/2} \int_{a_1=0}^{\infty} \int_{a_2=0}^{a_1 s} a_1 a_2 \exp\left(-\frac{a_1^2 + a_2^2 - s a_1 a_2}{2\sigma^2 \sin^2\theta}\right) d\theta da_1 da_2 \tag{4.18}$$

For convenience, for solving the inner two integrals, let $u = \sigma \sin\theta$. Using integral relations derived in Appendix C, solving the innermost integral in (4.18) results in:

$$\tilde{P}_{e,1} = \frac{s}{\rho_1^2 \rho_2^2 \sqrt{2\pi}} \int_{\theta=0}^{\pi/2} \int_{a_1=0}^{\infty} u a_1^2 \exp\left(-\frac{(1 - s^2/4)}{2u^2} a_1^2\right) \mathrm{erf}\left(\frac{s a_1}{2\sqrt{2} u}\right) d\theta da_1 \tag{4.19}$$

The inner integral in (4.19) is intractable. However, if we make a change of variables such that $z = a_1/u$, the expression becomes:

$$\begin{aligned}\tilde{P}_{e,1} &= \frac{s}{\rho_1^2\rho_2^2\sqrt{2\pi}}\int_{\theta=0}^{\pi/2}\int_{a_1=0}^{\infty} u^4 z^2 \exp\left(-\frac{(1-s^2/4)}{2}z^2\right)\mathrm{erf}\left(\frac{sz}{2\sqrt{2}}\right)d\theta dz \\ &= \frac{s\sigma^4}{\rho_1^2\rho_2^2\sqrt{2\pi}}\int_{\theta=0}^{\pi/2}\int_{a_1=0}^{\infty} z^2 \sin^4\theta \exp\left(-\frac{(1-s^2/4)}{2}z^2\right)\mathrm{erf}\left(\frac{sz}{2\sqrt{2}}\right)d\theta dz \quad (4.20)\end{aligned}$$

After integration, the expression under the integral signs in (4.20) is a function of r alone, and the structure of the integral is such that it would not be difficult to calculate numerically. Importantly, however, we have managed to extract σ, ρ_1, and ρ_2 from inside the integrals, so the resulting expression is straightforward to evaluate in terms of average SNR of the users. Since s is a function of the random variable ϕ, the expression (4.20) should, strictly speaking, be averaged with respect to ϕ, but as we shall show later, this calculation is not necessary at very high SNR.

We may now return to the consideration of the second term in (4.15). Here we have:

$$P_{e,2} = \int_{a_1=0}^{\infty}\frac{a_1}{\rho_1^2\rho_2^2}Q\left[\frac{a_1}{\sigma}\right]e^{-a_1^2/2\rho_1^2}\int_{a_2=a_1 s}^{\infty} a_2 e^{-a_2^2/2\rho_2^2}da_2 da_1 \quad (4.21)$$

Solving for the inner integral in (4.21) is straightforward, and results in:

$$P_{e,2} = \int_{a_1=0}^{\infty}\frac{a_1}{\rho_1^2}Q\left[\frac{a_1}{\sigma}\right]e^{-a_1^2\left(\frac{1}{2\rho_1^2}+\frac{s^2}{2\rho_2^2}\right)}da_1 \quad (4.22)$$

Let $S = \sqrt{(\rho_1^2\rho_2^2)/(\rho_2^2 + s^2\rho_1^2)}$. Now, (4.22) becomes:

$$P_{e,2} = \frac{\rho_2^2}{\rho_2^2 + s^2\rho_1^2}\int_{a_1=0}^{\infty}\frac{a_1}{S^2}Q\left[\frac{a_1}{\sigma}\right]e^{-a_1^2/2S^2}da_1 \quad (4.23)$$

The solution for this integral was given in (4.4). We remain interested in its asymptotic performance, which was given in (4.5). For $\rho_1 >> \sigma$, we now have:

$$P_{e,2} \simeq \frac{\rho_2^2}{\rho_2^2 + s^2\rho_1^2}\frac{\sigma^2}{4S^2} = \frac{\sigma^2}{4\rho_1^2} \quad (4.24)$$

Returning to (4.15), and using the expression obtained in (4.20), we find that the overall probability of error is given by:

$$P_e = \frac{\sigma^4}{\rho_1^2 \rho_2^2} f(s) + \frac{\sigma^2}{4\rho_1^2} \tag{4.25}$$

where $f(r)$ is the function of r observed in (4.15). Interestingly, if $\sigma^2 \to 0$, or if both $\rho_1^2 \to \infty$ and $\rho_2^2 \to \infty$, the left term in (4.25) vanishes and the result is the single-user bound derived above - a result independent of s. This result is somewhat counter-intuitive. However, experimental results (shown in Figure 4.2) verify these conclusions.

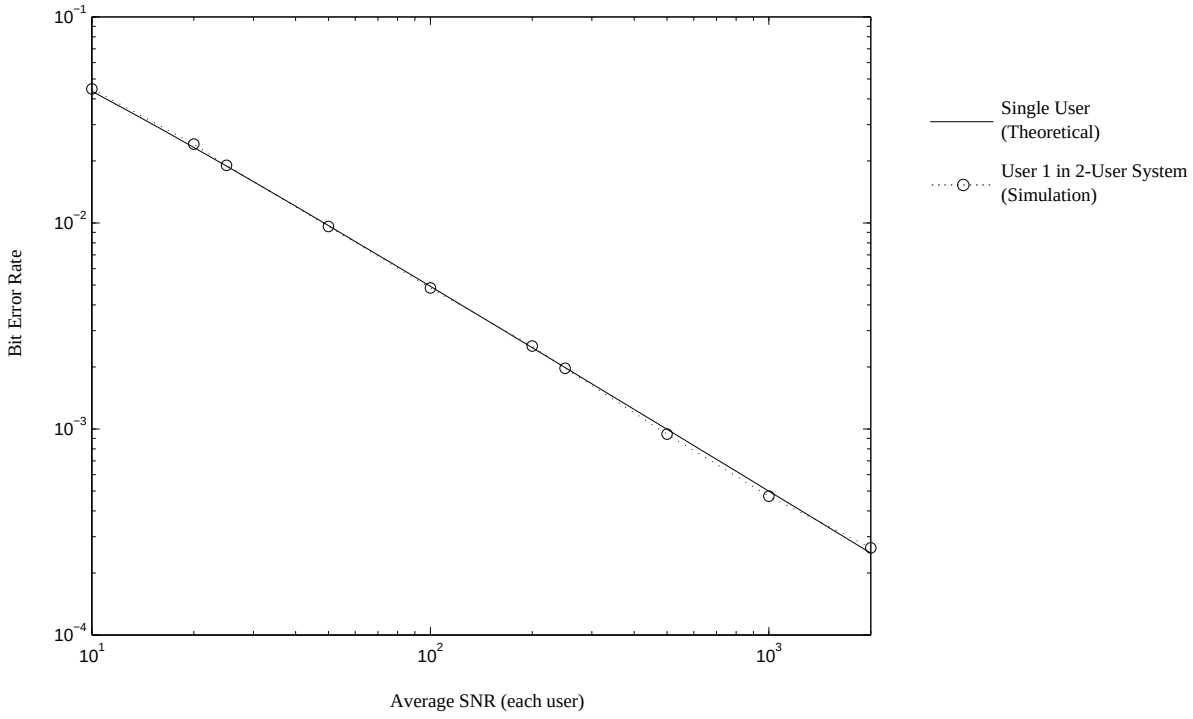

Figure 4.2: Plot of theoretical single user performance and simulated performance of one user in a two-user system versus average SNR

In this section we have gained an understanding of the limitations of multiuser detection in a fading channel. We expect that an arbitrary receiver for the Rayleigh channel will probably not improve greatly on the decorrelator for phase-coherent detection. Significantly greater improvements are expected for Rician channels over specific ranges of SNR. With these facts in mind, we may proceed to the task of designing EM-based detectors.

4.2 Phase-Coherent EM-Based Detectors

In this section, receiver structures based on the EM algorithm for multiuser detection in fading channels will be proposed. An alternative design will also be proposed based on certain approximations related to the MPEM algorithm, as well as work done in Section 3.4. Let us first consider the ML optimal phase coherent receiver in a general fading channel. As in the non-fading case, we have:

$$\tilde{\mathbf{b}}_{ML} = \underset{\mathbf{b}\in\{-1,1\}^n}{\arg\max} f_{\mathbf{y}}(\mathbf{y} \mid \mathbf{b}) \tag{4.26}$$

However, now we have that

$$f_{\mathbf{y}}(\mathbf{y} \mid \mathbf{b}) = \int_{\mathbf{a}} f_{\mathbf{y}}(\mathbf{y} \mid \mathbf{a},\ \mathbf{b}) f_{\mathbf{a}}(\mathbf{a}) d\mathbf{a} \tag{4.27}$$

where $\mathbf{a}$ is the vector of amplitudes, and $f_{\mathbf{a}}(\mathbf{a})$ is the PDF of $\mathbf{a}$, in this case either the Rayleigh PDF given by (2.26) or the Rician PDF given by (2.27). In general, quite apart from the computational complexity, which increases exponentially as in the non-fading case, the quantity in (4.27) may only be calculated numerically. As a result, the ML optimal receiver is not practically implementable in any sense.

We also remark that the form of the likelihood function conditioned on both knowledge of symbols and amplitudes must change from (2.5) in order to accommodate a complex received signal. We now have:

$$f_{\mathbf{y}}(\mathbf{y} \mid \mathbf{a},\ \mathbf{b}) = K \exp\left(-\frac{\xi}{2\sigma^2}\right) \tag{4.28}$$

where K is a normalization constant. Assuming that the noise is circularly symmetric [42], ξ is given by

$$\begin{aligned} \xi &= [\Re(\mathbf{y} - \mathbf{R\Theta Ab})]^T \mathbf{R}^{-1} [\Re(\mathbf{y} - \mathbf{R\Theta Ab})] \\ &\quad + [\Im(\mathbf{y} - \mathbf{R\Theta Ab})]^T \mathbf{R}^{-1} [\Im(\mathbf{y} - \mathbf{R\Theta Ab})] \end{aligned} \tag{4.29}$$

where $\Re$ and $\Im$ represent the real and imaginary components, respectively. The above results will be useful in subsequent receiver designs.

4.2.1 Detectors for Rayleigh channels

Let us now consider the task of designing an EM receiver for the Rayleigh channel. The observed vector $\mathbf{y}$ is the incomplete data by definition. The most straightforward receiver uses $\mathbf{y}$ along with the fading random variables $\mathbf{a}$ as the complete data, and $\mathbf{b}$ as the parameter vector under estimation. From (1.1), we have that the E-step is given by

$$Q(\mathbf{b};\, \bar{\mathbf{b}}) = E[\log f_{\mathbf{y},\mathbf{a}}(\mathbf{y}, \mathbf{a} \mid \mathbf{b}) \mid \mathbf{y}, \bar{\mathbf{b}}] \tag{4.30}$$

and the M-step is given by selecting the vector $\mathbf{b}$ that maximizes $Q(\mathbf{b}; \bar{\mathbf{b}})$. Eliminating terms irrelevant to the M-step, and applying Bayes' rule, the expression in (4.30) becomes

$$Q(\mathbf{b};\, \bar{\mathbf{b}}) = \int_{\mathbf{a}} \log f_{\mathbf{y}}(\mathbf{y} \mid \mathbf{a}, \mathbf{b}) \frac{f_{\mathbf{y}}(\mathbf{y} \mid \mathbf{a}, \bar{\mathbf{b}}) f_{\mathbf{a}}(\mathbf{a})}{f_{\mathbf{y}}(\mathbf{y} \mid \bar{\mathbf{b}})} d\mathbf{a} \tag{4.31}$$

The expression $f_{\mathbf{y}}(\mathbf{y} \mid \mathbf{a}, \mathbf{b})$ is the jointly Gaussian PDF, while $f_{\mathbf{a}}(\mathbf{a})$ is the fading PDF. We observe that the marginal distribution $f_{\mathbf{y}}(\mathbf{y} \mid \bar{\mathbf{b}})$ is a multiplicative constant independent of both $\mathbf{a}$ and $\mathbf{b}$, and thus it may be dropped. Making the appropriate substitutions, we have

$$Q(\mathbf{b};\, \bar{\mathbf{b}}) = CK \int_{\mathbf{a}} \left[\prod_{i=1}^{n} a_i\right] \left[\log K - \frac{\xi}{2\sigma^2}\right] \exp\left(-\frac{\bar{\xi}}{2\sigma^2} - \sum_{i=1}^{n} \frac{a_i^2}{2\rho_i^2}\right) d\mathbf{a} \tag{4.32}$$

where C and K are normalization constants, a_i is the ith element of the vector $\mathbf{a}$, and $\bar{\xi}$ is given by the expressions in (4.29) where $\bar{\mathbf{b}}$ is substituted everywhere for $\mathbf{b}$. In (4.29), we may make the expansion

$$\begin{aligned}
&[\Re(\mathbf{y} - \mathbf{R}\boldsymbol{\Theta}\mathbf{A}\mathbf{b})]^T \mathbf{R}^{-1} [\Re(\mathbf{y} - \mathbf{R}\boldsymbol{\Theta}\mathbf{A}\mathbf{b})] \\
&= \mathbf{y}_{\Re}^T \mathbf{R}^{-1} \mathbf{y}_{\Re} - 2\mathbf{y}_{\Re}^T \boldsymbol{\Theta}_{\Re} \mathbf{A}\mathbf{b} + \mathbf{b}^T \mathbf{A} \boldsymbol{\Theta}_{\Re} \mathbf{R} \boldsymbol{\Theta}_{\Re} \mathbf{A}\mathbf{b} \\
&= \mathbf{y}_R e^T \mathbf{R}^{-1} \mathbf{y}_{\Re} - 2\sum_{i=1}^{n} \Re(y_i) a_i \cos\theta_i b_i \\
&\quad + 2\sum_{j \neq k} b_i b_j a_i a_j \cos\theta_i \cos\theta_j \mathbf{R}_{jk} + \sum_{i=1}^{n} b_i^2 a_i^2 cos^2 \theta_i
\end{aligned} \tag{4.33}$$

where the subscript $\Re$ represents the real components of the appropriate vector. Similarly,

$$
\begin{aligned}
&[\Im(\mathbf{y}-\mathbf{R\Theta Ab})]^T\mathbf{R}^{-1}[\Im(\mathbf{y}-\mathbf{R\Theta Ab})] \\
&= \mathbf{y}_{\Im}^T\mathbf{R}^{-1}\mathbf{y}_{\Im} - 2\sum_{i=1}^{n}\Im(y_i)a_i\sin\theta_i b_i \\
&\quad +2\sum_{j\neq k}b_ib_ja_ia_j\sin\theta_i\sin\theta_j\mathbf{R}_{jk} + \sum_{i=1}^{n}b_i^2a_i^2\sin^2\theta_i
\end{aligned} \tag{4.34}
$$

Substituting the above into (4.32) and again eliminating terms irrelevant to the M-step, we have:

$$
\begin{aligned}
Q(\mathbf{b};\,\bar{\mathbf{b}}) &= 2\sum_{i=1}^{n}[\Re(y_i)\cos\theta_i + \Im(y_i)\sin\theta_i)\Psi(a_i)b_i \\
&\quad -2\sum_{j\neq i}\Psi(a_ia_j)cos(\theta_i-\theta_j)b_ib_j[\mathbf{R}]_{jk} \\
&\quad -\sum_{i=1}^{n}\Psi(a_i^2)b_i^2
\end{aligned} \tag{4.35}
$$

where, following the development above, $\Psi[x(\mathbf{a})]$ is given by

$$
\begin{aligned}
&\Psi[x(\mathbf{a})] = \\
&\quad \int_{\mathbf{a}} x(\mathbf{a})\left[\prod_{i=1}^{n}a_i\right]\exp\left(-\frac{\xi}{2\sigma^2}-\sum_{i=1}^{n}\frac{a_i^2}{2\rho_i^2}\right)d\mathbf{a}
\end{aligned} \tag{4.36}
$$

We note that, for our purposes, $x(\mathbf{a})$ is of the form $a_i^u a_j^v$, where $u,v\in[0,1,2]$; $1\leq i,j\leq n$; and $i\neq j$. The solution to the integral in (4.36) is equivalent to solving an integral of the form

$$
\int_{\mathbf{a}}\left[\prod_{i=1}^{n}a_i^{u(i)}\right]\exp\left(\sum_{i=1}^{n}(\alpha_ia_i^2+\gamma_ia_i)+\sum_{j\neq k}\beta_{jk}a_ia_k+\phi\right)d\mathbf{a} \tag{4.37}
$$

where ϕ is a constant. A closed-form solution exists only for the innermost integral, for which expressions are given in Appendix C, and otherwise the solutions must be found by series expansion or by other appropriate approximations.

We may now discuss the M-step of this algorithm. By definition, from (1.1), the M-step of any EM algorithm is given by

$$\bar{\theta}_{n+1} = \underset{\theta}{\arg\max}\, Q(\theta; \bar{\theta}_n) \tag{4.38}$$

Further, from [10], we have the definition of a Generalized EM Algorithm (GEM), which is any algorithm where

$$Q(\bar{\theta}_{n+1}, \bar{\theta}_n) \geq Q(\bar{\theta}_n, \bar{\theta}_n) \tag{4.39}$$

which also increases monotonically in likelihood. To satisfy the requirements of (4.38) as it applies to the parameter vector $\mathbf{b} \in \{-1, 1\}^n$, we may choose an M-step based on the optimal receiver from [1]:

$$\bar{\mathbf{b}}_{n+1} = \underset{\mathbf{b}\in\{-1,1\}^n}{\arg\max}\, Q(\mathbf{b}; \bar{\mathbf{b}}_n) \tag{4.40}$$

which satisfies the requirements of the EM algorithm, although it suffers from the disadvantage of $O(2^n)$ complexity. As an alternative, we may select a coordinate ascent M-step, as in the SAGE detector from [8], which is a GEM:

$$\begin{aligned} &\text{Step 1}: && k = (k+1) \bmod n \\ &\text{Step 2}: && \bar{b}_{k,n+1} = \underset{b_k\in\{-1,1\}}{\arg\max}\, Q(\mathbf{b}; \bar{\mathbf{b}}_{k,n}) \end{aligned} \tag{4.41}$$

which may be implemented with much less complexity. The use of this M-step actually produces a "greedy" SAGE algorithm [12].

An alternative algorithm based on work done in Chapter 3 may also be proposed. Whereas in the above development, only the fading amplitudes were treated as hidden data, we shall instead return to the case where only the symbol of the user of interest is the parameter under estimation, and use all unknown interfering terms (symbols and amplitudes) as hidden data. The development presented here may be considered as the MPEM algorithm adapted to random amplitudes. We first define the vector $\mathbf{s} = \mathbf{A}\mathbf{b}$, such that $\mathbf{s}$ is a column vector with elements $s_i = a_i b_i$, which includes all the unknown interference variables. Now, analogously to (2.18), as well as (3.51) and (3.52), we have in the E-step:

$$\hat{s}_k = \Phi\left(\left| y_i - \sum_{j\neq i}^{n} H_{ij}\bar{s}_j \right| \cos\phi_i\right) \tag{4.42}$$

where $\Phi(x) = E[s|x]$ is the estimation function for the quantity s, $H_{ij} = R_{ij}\cos(\theta_i - \theta_j)$ is the cross-correlation coefficient adjusted for different transmitted phase angles, and ϕ_i is the difference in phase between the phase of y_i and the known signal phase angle θ_i. Following from work done in Appendix C, $\Phi(x)$ is given by

$$\begin{aligned}
\Phi(x) &= \int_a \int_b ab f_{a,b}(a,b \mid x) dbda \\
&= \frac{\sum_{b\in\{-1,1\}} \int_{a=0}^{\infty} a^2 b \exp[-(x-ab)^2/2\sigma^2 - a^2/2\rho^2] da}{\sum_{b\in\{-1,1\}} \int_{a=0}^{\infty} a \exp[-(x-ab)^2/2\sigma^2 - a^2/2\rho^2] da} \\
&= \frac{1}{w} \cdot \frac{\sum_{b\in\{-1,1\}} b\left[(2w+v^2)\sqrt{\pi/w}\exp(v^2/4w)\mathrm{erfc}(-v/2\sqrt{w}) + 2v\right]}{\sum_{b\in\{-1,1\}} v\sqrt{\pi/w}\exp(v^2/4w)\mathrm{erfc}(-v/2\sqrt{w}) + 2}
\end{aligned} \tag{4.43}$$

where $w = 1/2\sigma^2 + 1/2\rho^2$, and $v = -bx/\sigma^2$. The $\Phi(\cdot)$ function may be deemed analogous to the $\tanh(\cdot)$ function as a "soft decision" function (or, more correctly, an estimation function) where the amplitudes are unknown. We now specify the M-step, analogous to (2.19). Combining both steps on one line, we have

$$\bar{s}_i = \left| y_i - \sum_{j\neq i} H_{ij}\Phi(\bar{s}_j) \right| \cos\phi_i \tag{4.44}$$

We use the E-step and M-step in the same manner as the MPEM algorithm for known amplitudes described in Chapter 2, with the stipulation that the estimation functions in the E-step and M-step are identical. The algorithm iterates through each user in sequence for the desired number of iterations. Plots of the estimation function $\Phi(\cdot)$ are given in Figure 4.3. Henceforth we shall refer to this algorithm simply as the "modified EM Algorithm". It is interesting to note that the function simplifies to $\Phi(x) = x$ as the average SNR $\to \infty$. Simulation results depicting the performance of both the EM-based and modified-EM-based multiuser receivers in Rayleigh channels will be presented in Section 4.4. Block diagram representations of the proposed receivers are given in Figures 4.4 and 4.5.

4.2.2 Detectors for Rician channels

In this section, receiver structures based on the EM algorithm for multiuser detection in Rician channels will be proposed. The development will largely parallel the development

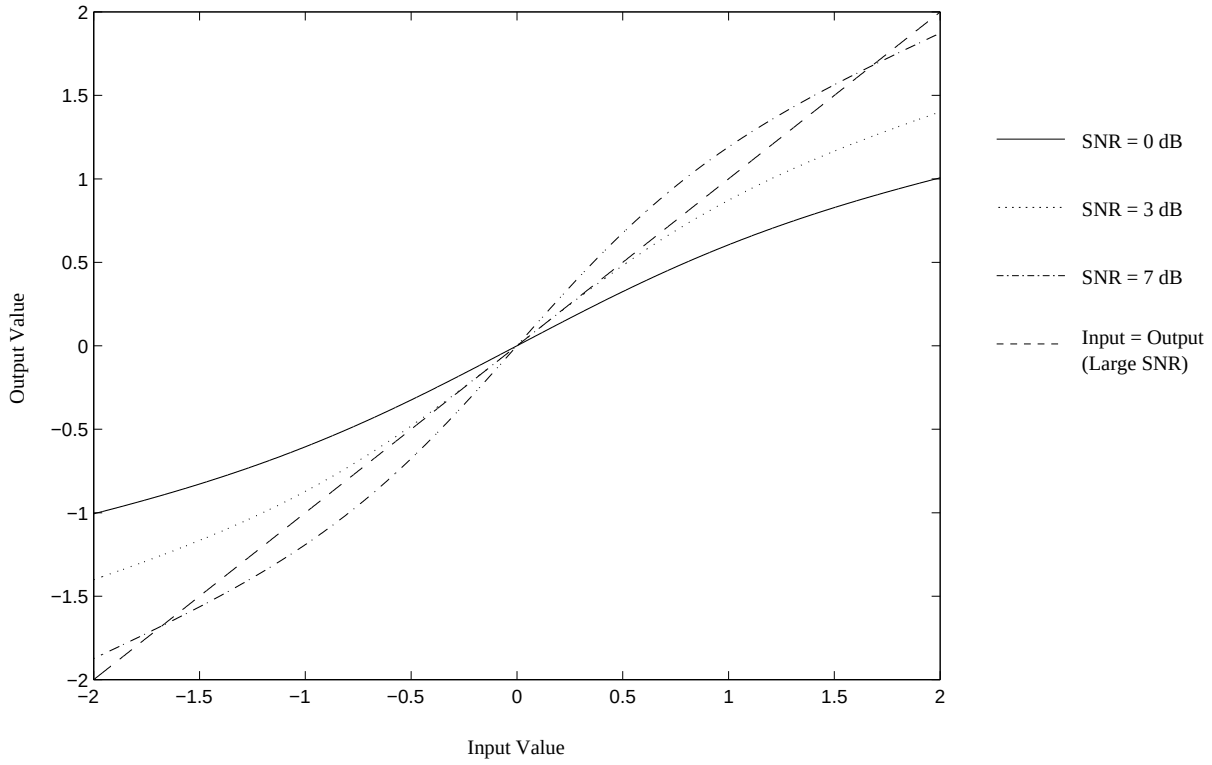

Figure 4.3: Plot of the input-output characteristic for the feedback function in Rayleigh fading for various values of average SNR

of receivers for Rayleigh channels in the previous section. The development of the Rician receiver is identical to the above development up to (4.31). However, in this case, the density $f_{\mathbf{a}}(\mathbf{a})$ is given by the Rician PDF from (2.27). We may thus expand (4.31) as follows:

$$Q(\mathbf{b};\,\bar{\mathbf{b}}) = CK \int_{\mathbf{a}} \left[\prod_{i=1}^{n} a_i I_0\left(\frac{a_i|L_i|}{\rho_i}\right)\right] \left[\log K - \frac{\xi}{2\sigma^2}\right] \exp\left(-\frac{\bar{\xi}}{2\sigma^2} - \sum_{i=1}^{n} \frac{a_i^2 + |L_i|^2}{2\rho_i^2}\right) d\mathbf{a} \quad (4.45)$$

where $|L_i|$ is the magnitude of the ith user's direct component, and ξ is defined as before. The expansion of the modified Bessel function $I_0(\cdot)$ is well known:

$$I_0(x) = \sum_{m=0}^{\infty} \frac{1}{(m!)^2} \left(\frac{x}{2}\right)^{2m} \quad (4.46)$$

Analogously to the derivation for the Rayleigh channel, the E-step function $Q(\mathbf{b};\,\bar{\mathbf{b}})$ is given by the expression in (4.35), in which the function $\Psi(\cdot)$ is replaced by $\Psi_R(\cdot)$, where

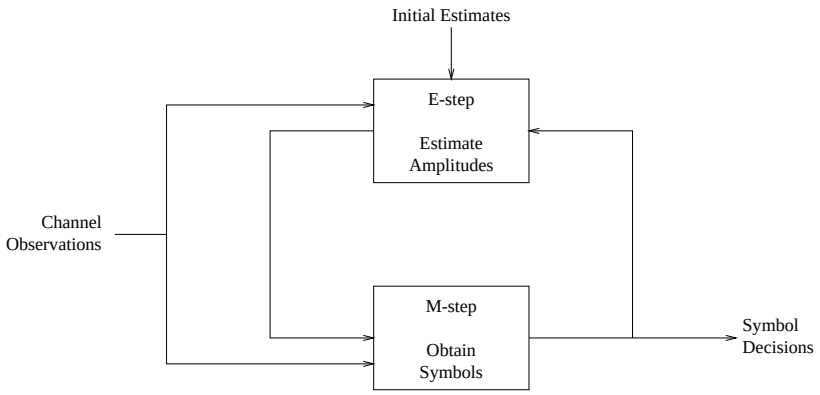

Figure 4.4: The EM Algorithm method: Block diagram

$$\Psi_R[x(\mathbf{a})] = \int_{\mathbf{a}} x(\mathbf{a}) \left[\prod_{i=1}^{n} a_i I_0 \left(\frac{a_i |L_i|}{\rho_i} \right) \right] \exp \left(-\frac{\xi}{2\sigma^2} - \sum_{i=1}^{n} \frac{a_i^2 + |L_i|^2}{2\rho_i^2} \right) d\mathbf{a} \tag{4.47}$$

Since the modified bessel function may be expanded as in (4.46), the inner integral in (4.47) may be solved through methods discussed in Appendix C. However, the complexity of the required numerical calculations are tremendous, motivating a search for an approximate solution.

We may again appeal to an MPEM-like structure to solve this problem. Using a geometric argument, illustrated in Figure 4.6, the problem may be reduced to nearly the same as the Rayleigh case if the absolute phase of the line-of-sight component, in addition to the absolute phase of the entire signal, is known. The multiuser signal is given by

$$\mathbf{y} = \mathbf{R}(\mathbf{L}\mathbf{b} + \mathbf{A}\mathbf{b}) + \mathbf{z} \tag{4.48}$$

where $\mathbf{L}$ and $\mathbf{A}$ are, respectively, complex diagonal matrices containing the in-phase and scattered amplitude components of the signal. In this case, the projection of the single-user (interference cancelled) Rician received signal on the vector in the direction of the transmitted signal may be described as

$$|y_i| \cos\theta_i = |L_i| b_i \cos\phi_L + a_i b \cos\phi_a + z \tag{4.49}$$

where $|L_i|$ and a_i are the magnitude of the line-of-sight and scattered components, respec-

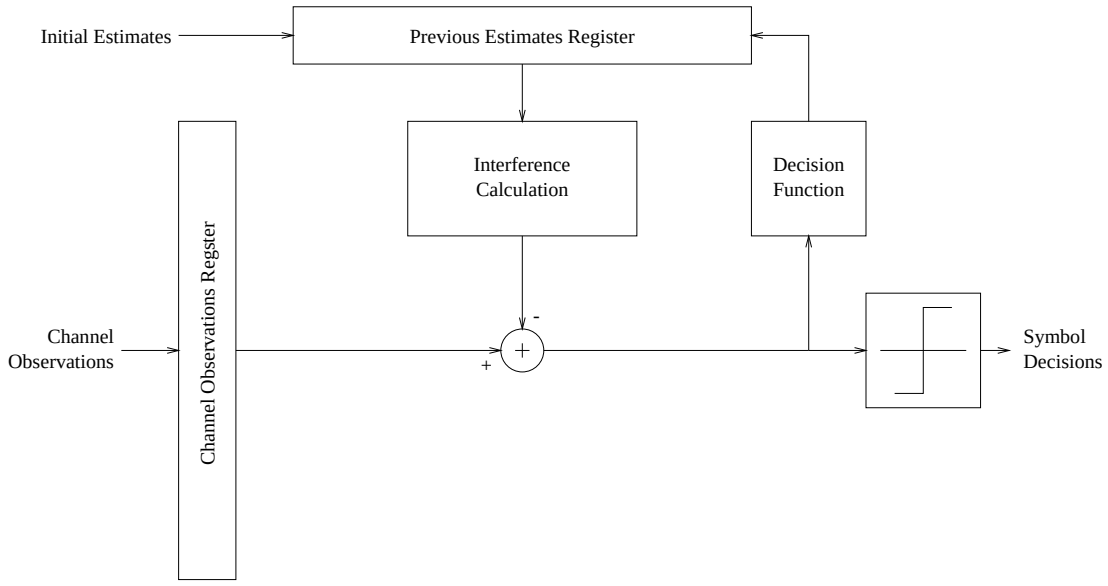

Figure 4.5: The modified EM Algorithm method: Block diagram

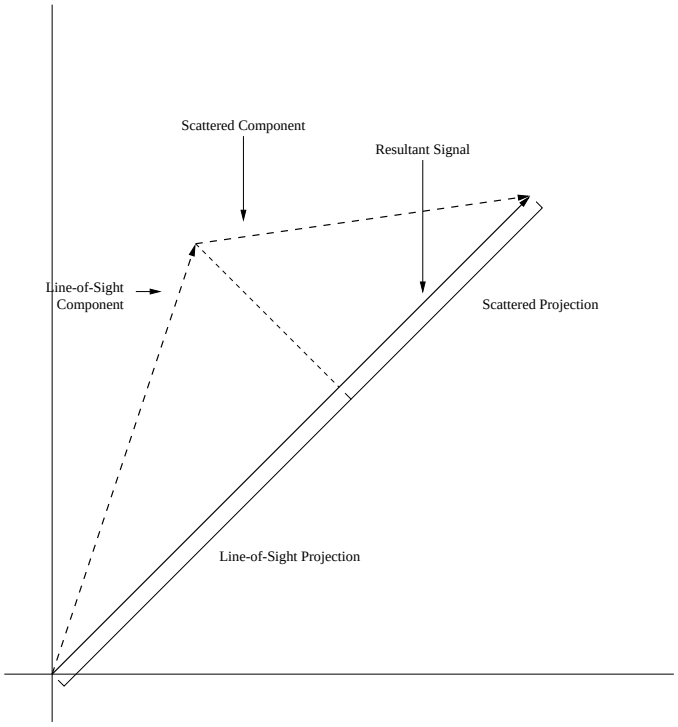

Figure 4.6: Geometric interpretation of Rician fading model for coherent detection

tively; ϕ_L and ϕ_a are the angles between the signal phase angle and, respectively, the line-of-sight component phase angle and the scattered component phase angle; b_i is the transmitted symbol; and z is Gaussian noise with variance σ^2. We remark that the quantity a has the Rayleigh distribution, since it remains the magnitude of two orthogonal, independent Gaussian random processes. From (4.48), to accomplish interference cancellation in the E-step, we must cancel both the line-of-sight components, which requires an estimate of b given y, as well as the scattered components, which requires an estimate of ab given y. Only minor modifications on (4.43) are required to implement this task. It is straightforward to show that the estimate of ab given y is given by

$$\begin{aligned}
&\Phi_R(x) \\
&= \frac{\sum_{b\in\{-1,1\}} \int_{a=0}^{\infty} a^2 b \exp[-(x - Lb\cos\phi_L - ab\cos\phi_a)^2/2\sigma^2 - a^2/2\rho^2]da}{\sum_{b\in\{-1,1\}} \int_{a=0}^{\infty} a \exp[-(x - Lb\cos\phi_L - ab\cos\phi_a)^2/2\sigma^2 - a^2/2\rho^2]da} \qquad (4.50) \\
&= \frac{1}{w} \cdot \frac{\sum_{b\in\{-1,1\}} b\left[(2w+v^2)\sqrt{\pi/w}\exp(s+v^2/4w)\mathrm{erfc}(-v/2\sqrt{w}) + 2v\right]}{\sum_{b\in\{-1,1\}} v\sqrt{\pi/w}\exp(s+v^2/4w)\mathrm{erfc}(-v/2\sqrt{w}) + 2} \qquad (4.51)
\end{aligned}$$

where $s = (Lyb\cos\phi_L)/\sigma^2$, $w = (\cos^2\phi_a)/2\sigma^2 + 1/2\rho^2$, and $v = (yb\cos\phi_a)/\sigma^2$. Similarly, the estimate of b given y may be evaluated using the above expression, with ab rather than a^2b in the numerator in (4.50). This gives the following expression:

$$G_R(x) = \frac{\sum_{b\in\{-1,1\}} b\left[v\sqrt{\pi/w}\exp(s+v^2/4w)\mathrm{erfc}(-v/2\sqrt{w}) + 2\right]}{\sum_{b\in\{-1,1\}} v\sqrt{\pi/w}\exp(s+v^2/4w)\mathrm{erfc}(-v/2\sqrt{w}) + 2} \qquad (4.52)$$

with s, w, and v as defined above. It is straightforward to see that as $\alpha \to \infty$ (i.e., the amplitude becomes constant and known), the expression in (4.52) approaches $\tanh(|L|x/\sigma^2)$. The E- and M-steps may then be combined as follows:

$$u_i = y_i - \sum_{k\neq i} R_i k \left[L_k G_R(u_k) + \Phi_R(u_k)\right] \qquad (4.53)$$

and the algorithm is formed by iteratively calculating u_i in sequence for all values of i.

4.3 Incoherent Multiuser Detectors

In addressing this problem, we shall use a method first introduced in Chapter 2, which was an innovations-based MLSE approach to data sequence detection in fading channels [26]. Little attention in the literature has been given to the problem of incoherent multiuser detection. In this section, we will conduct some initial analysis on this problem to illustrate techniques for future research. Owing to the inherent complexity of this problem, only brief sketches of receiver designs will be given. Since it has not yet had any application in multiuser detection, we will first discuss its use in deriving incoherent versions of already existing receiver types, which will be used as benchmarks for comparison. We will subsequently present EM-based receivers based on this method for incoherent multiuser detection.

4.3.1 Basic multiuser detector types

In this subsection we will give the equivalents of the conventional and decorrelating detectors based on Yu's MLSD algorithm. The model for the received signal in the incoherent case is given by (2.30), although the Cartesian representation, rather than the polar representation, will be used to express the complex fading terms. Let us express the received signal $\mathbf{y}$ with minor changes in notation to conform more closely to [26]:

$$\mathbf{y} = \mathbf{RB}(\mathbf{a}_i + j\mathbf{a}_q) + \mathbf{z} \tag{4.54}$$

where B is the diagonal matrix of transmitted symbols, $\mathbf{a}_i$ is the vector of in-phase fading components, and $\mathbf{a}_q$ is the vector of quadrature fading components. We know that $\mathbf{a}_i$ and $\mathbf{a}_q$ are jointly Gaussian with zero mean and diagonal covariance matrix with elements ρ_i^2. Let us first consider the incoherent equivalent of the conventional receiver. In non-fading channels, the conventional receiver assumes that the interference is equivalent to Gaussian noise. Extending this concept to the case of incoherent detection, we may define the conventional incoherent detector as the detector which takes the elements of $\mathbf{y}$ from (4.54) as input and uses them, without any interference cancellation or processing, in sequence detection. Unlike the phase-coherent case, the MLSD algorithm from [26] requires knowledge of the variance

of the additive noise in the signal, so the variance of the interference must be taken into account. The variance of each interfering term is given by:

$$E\left[|R_{ik}(x_i + jy_i)b_i|^2\right] = 2R_{ik}^2\rho_i^2 \tag{4.55}$$

We are concerned with the variance of the real and imaginary components individually, which are given by $R_{ik}^2\rho_i^2$. Thus, for the kth user, assuming that the thermal noise is circularly symmetric, the equivalent noise variance in each component for the conventional incoherent receiver is given by

$$\sigma_C^2 = \sum_{i \neq k} R_{ik}^2\rho_i^2 + \sigma^2 \tag{4.56}$$

In the standard conventional receiver, a Gaussian assumption is made on the interference by appealing to the Central Limit Theorem. It is interesting to note that the model definition in (4.54) implies that the real and imaginary components of the vector $\mathbf{y}$ have exactly the Gaussian distribution, since they are merely sums of Gaussian random variables.

We may also consider the incoherent equivalent of the decorrelating receiver. In the coherent case [2], this involves multiplying the received vector on the left by $\mathbf{R}^{-1}$, which linearly removes all MAI, following which each detection is performed individually on each processed observation. To accomplish the equivalent incoherent task, we first multiply on the left in (4.54) by $\mathbf{R}^{-1}$:

$$\mathbf{d} = \mathbf{R}^{-1}\mathbf{y} = \mathbf{B}(\mathbf{a}_i + j\mathbf{a}_q) + \mathbf{w} \tag{4.57}$$

Following this operation, the MAI has been removed from each processed observation d_i, and thus we may apply the MLSD algorithm to each stream of observations individually. Recall from previous work that the equivalent noise variance for the kth user in the decorrelator is given by $\sigma^2[\mathbf{R}^{-1}]_{kk}$, which is always greater than σ^2.

In the sequence detection algorithms proposed above, if equiprobable, antipodal symbols are used, it can be shown that the path metrics for some sequence $\mathbf{s}$ and the sequence $-\mathbf{s}$ are the same. Therefore, when implementing these algorithms (and all subsequent algorithms

of this type), we shall use differential encoding to eliminate the need for absolute knowledge of the sign of the received sequence.

4.3.2 More complex multiuser detector types

Certain modifications to Yu's algorithm are required in order to adapt it to more complex methods of multiuser detection. In particular, modifications must be made to cope with cases where multiple signals are received, in each of which numerous symbols of interest are observed, which are present in multiuser detection as the set of sufficient statistics for the channel. In order to perform any sort of sequence detection with the Viterbi algorithm, we require that the underlying process of random parameters to be detected be modelled as a Markov process, and that the observations be conditionally independent with knowledge of the sequence of parameters [41]. For each observation, we are thus required to calculate a branch metric of the following form [42]:

$$H_{t,a,b} = \log f_y(y_t \mid [\Psi_a,\ \Psi_b]_t) \tag{4.58}$$

where y_t is the observation at time t, and $[\Psi_a,\ \Psi_b]_t$ represents the transition from current state Ψ_a to future state Ψ_b at time t. Permitted state transitions at time t are assumed to be equiprobable.

As we have already discussed, consecutive observations of a fading signal envelope are not independent, and in order to achieve the conditions leading up to (4.58), a whitening operation is required. However, now we must deal with the complication of a vector of simultaneous, dependent observations at each sampling instant. Considering once again the multiuser channel fading model that was given previously in (4.54), we observe that the observed vector $\mathbf{y}$ is multivariate Gaussian. We may calculate the cross-correlation matrix $\mathbf{S}(t_1, t_2)$ between any two realizations of this vector (as distinct from the cross-correlation matrix $\mathbf{R}$ of the user chip sequences):

$$\mathbf{S}(t_1, t_2) = E[\mathbf{y}_1 \mathbf{y}_2^H] = \mathbf{R}\mathbf{B}_1\mathbf{D}(\tau)\mathbf{B}_2\mathbf{R} + 2\sigma^2\mathbf{R}\delta_\tau \tag{4.59}$$

where σ^2 is the noise variance in each of the real and imaginary components of the signal,

$\tau = t_1 - t_2$, $\mathbf{B}_k$ is the diagonal matrix of user's symbol transmissions at time k, δ_τ is the delta function equal to 1 when $\tau = 0$ and zero elsewhere, $\mathbf{D}(\tau)$ is the cross-correlation matrix of the fading envelopes and the superscript H denotes Hermitian transposition. Note that when $t_1 = t_2$, assuming that the fading envelopes are independent, $\mathbf{D}(0)$ becomes a diagonal matrix with elements equal to the variance of each fading envelope. In this case, assuming that the symbols are antipodal, we know that $\mathbf{B}_1\mathbf{D}(\tau)\mathbf{B}_2 = \mathbf{D}(0)$, so the expression in (4.59) simplifies to $\mathbf{S} = \mathbf{R}\mathbf{D}(0)\mathbf{R} + \sigma^2\mathbf{R}$.

To perform the task of whitening this signal, we are faced with the task of finding a set of linear filters where the set of filter outputs $\mathbf{w}_n$ obey $E[\mathbf{w}_u\mathbf{w}_v^*] = \mathbf{0}$ for all $u \neq v$, where $\mathbf{0}$ is the zero matrix. This is a multiple-input, multiple-output filtering problem, for which the solution is assumed to be of the form

$$\mathbf{w}_t = \mathbf{y}_t + \mathbf{P}_1\mathbf{y}_{t-1} + \cdots + \mathbf{P}_T\mathbf{y}_{t-T} \tag{4.60}$$

where $\mathbf{P}_k$ are $n \times n$ filter coefficient matrices. For simplicity, in this case we will use only a first-order filter, i.e., $\mathbf{P}_k = \mathbf{0}$ for $k > 1$. This first-order assumption for Rayleigh channels is shown in [45] to be sufficient for most applications. By an argument similar to the Yule-Walker method [27], we find that the filter tap at time t, $\mathbf{P}_{1,t}$, conditioned on knowledge of the symbol vector, is given by

$$\mathbf{P}_{1,t} = -\left[\mathbf{S}^{-1}(t,t)\mathbf{S}(t,t-1)\right]^H \tag{4.61}$$

Furthermore, the covariance matrix of $\mathbf{w}_t$ is given by

$$\mathbf{C}_t = \mathbf{S}(t,t) - \mathbf{S}(t,t-1)\mathbf{S}^{-1}(t,t)\mathbf{S}(t,t-1) \tag{4.62}$$

Note that $\mathbf{C}_t$ is not necessarily diagonal.

We may now turn our attention to calculating the branch metrics from (4.58). Using our first-order assumption, there are 2^n states and 2^{2n} branch metrics corresponding to every possible pair of current and future states. This is also the number of members of the set of possible symbol vectors at time t and $t-1$. Let $s = [\Psi_a, \Psi_b]_t$ be a member of the set of possible state transitions at time t. The branch metric conditioned on b is given by

$$H_{t,s} = \frac{1}{2}|\mathbf{w}_{t,s}^H \mathbf{C}_{t,s}^{-1} \mathbf{w}_{t,s}|^2 + \Delta(\mathbf{C}_{t,s}) \tag{4.63}$$

where $\Delta(\cdot)$ corresponds to the determinant of a matrix, and where the subscript s denotes that the quantity was calculated conditioned on the state transition set member s. These branch metrics are applied to a trellis diagram of transitions between states, and the Viterbi algorithm [41, 42] is used to determine the maximum likelihood estimate of the sequence of symbols for every user. This receiver design is depicted in Figure 4.7.

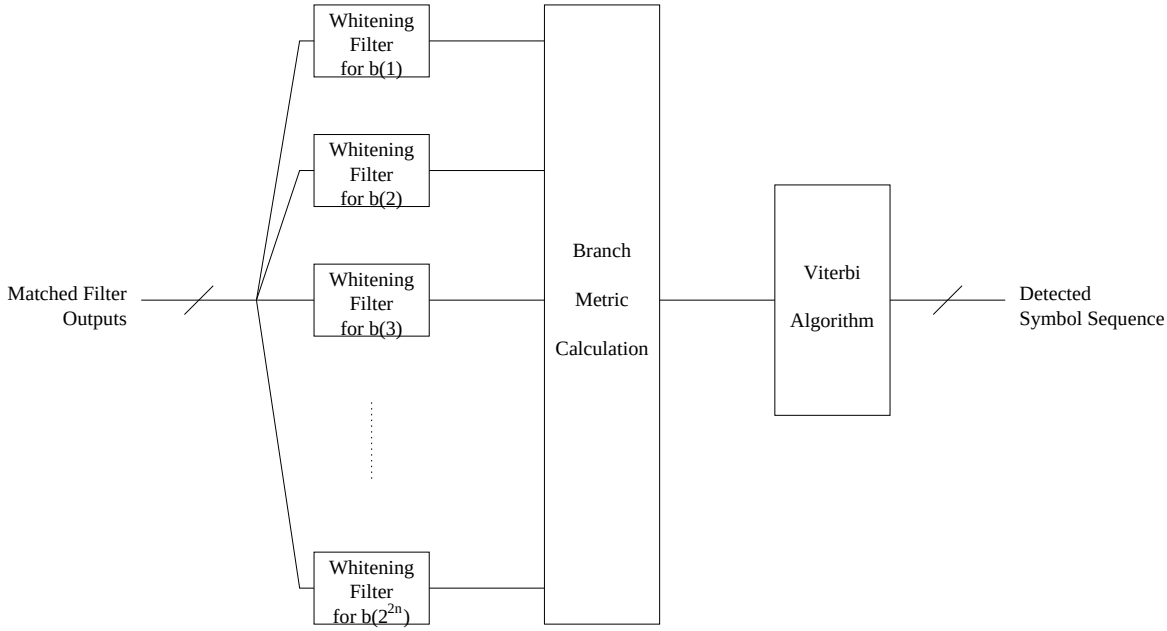

Figure 4.7: Representation of MLSD incoherent multiuser detector

We remark that this receiver structure suffers from the same disadvantage as the optimal multiuser detector, since from the above work the complexity of this receiver is exponential in the number of users. We may therefore consider some alternative receiver structures to mitigate this complexity. In particular, if every user's transmitted symbol is known, aside from that of the user of interest, we need not calculate branch metrics or assign states for those known symbols. Thus, the number of states and the number of branch metrics are reduced to 2 and 4, respectively. Using the EM algorithm, for instance, we may define all symbols other than those of the user of interest to be the "hidden data", to be estimated in

the E-step:

$$Q(\mathbf{b}_i;\, \bar{\mathbf{b}}_i) = E\left[\log f_{\mathbf{y}}(\mathbf{y} \mid \mathbf{b}) \mid \mathbf{y}, \mathbf{b}_i = \bar{\mathbf{b}}_i\right] \tag{4.64}$$

while sequence detection is performed on the user of interest's symbol in the M-step. Georghiades and Han [16] have provided a framework for applying the EM algorithm to MLSD problems. Unfortunately, this problem suffers from complexity problems owing to the calculation in the E-step being dependent on the entire sequence of observations $\mathbf{y}$.

We shall consider a slightly different and simpler case in this work, based on the SAGE algorithm [12]. Assuming we wish to detect every user simultaneously, we may consider the sequence of each user's transmitted symbols as individual parameter spaces and perform MLSD on each one in turn, while using previous estimates for all other users as estimates for the values of those symbols. That is, we will assume that the estimates are equivalent to the "known" values of symbols of other users. This is a concept similar to that expressed by (2.17) taken from [8], which is a simple coordinate-ascent algorithm. We may specify this algorithm concisely as follows:

1. (Initialization step) Perform detection on the received sequence $\mathbf{y}_t$ using the decorrelating receiver to obtain an initial sequence of symbol estimates $\bar{\mathbf{b}}_t$.

2. Let $k = 0$.

3. Perform MLSD on the sequence $\mathbf{y}_t$ to detect symbols from user k, assuming all other users' symbols are equal to the values given by $\bar{\mathbf{b}}_t$.

4. Let the members of $\bar{b}_t$ associated with user k equal the result of step 3.

5. Let $k = (k+1) \bmod a$.

6. If further iterations are desired, go to 3., else stop.

Some results for this algorithm are given in Section 4.4. An APP implementation of this algorithm is possible along lines outlined in Chapter 3, which would require the use of the

Soft Output Viterbi Algorithm (SOVA) [44]. The complexity of this algorithm is such that we will leave its implementation to future work.

It is straightforward to modify this algorithm for the Rician fading case. As was noted in Chapter 2, Rician fading implies that the in-phase and quadrature fading envelopes are each modelled as Gaussian random variables with nonzero mean, indicating both a line-of-sight component and specular components from the transmitter to the receiver. Let us assume that these mean values are known and that they remain constant throughout the detection procedure. The received signal is then given by

$$\mathbf{y} = \mathbf{RB}(\mathbf{x} + \mathbf{a}) + \mathbf{z} \tag{4.65}$$

where $\mathbf{x}$ and $\mathbf{a}$ are complex vectors representing the vector of constant mean amplitudes for every user and the vector of random amplitudes for each user, respectively. We are interested in whitening the random components only, so this has no effect on the whitening procedures outlined above, because the cross-correlation calculation in (4.59) is made with the mean values subtracted. Since $\mathbf{x}$ is assumed known, and since the path metrics are calculated conditioned on knowledge of the symbols, we need merely make an adjustment to the branch metric calculation (4.63) as follows:

$$H_{t,s} = \frac{1}{2}|(\mathbf{w}_{t,s} - \mu_{t,s})^H \mathbf{C}_{t,s}^{-1}(\mathbf{w}_{t,s} - \mu_{t,s})|^2 + \Delta(\mathbf{C}_{t,s}) \tag{4.66}$$

where $\mu_{t,s}$ is the vector of mean values present in the whitened filter output, given by

$$\mu_{t,s} = \mathbf{RB}_t\mathbf{x} + \mathbf{PRB}_{t-1}\mathbf{x} \tag{4.67}$$

Certain similarities exist between this method and the MLSD method for multiuser channels suffering from ISI [5]. Some results are given using this method in a Rician channel with the SAGE algorithm discussed above in section 4.4.

4.4 Experimental Results

In this section, experimental results for the above receivers obtained through numerical simulation will be presented. In most of the following cases, the optimal detector is far too complex to accurately simulate, making it useless as a benchmark for performance comparison. Since the decorrelating detector is the optimal detector in the absence of any knowledge of the received amplitudes, and since it is a robust, simple, and well-known multiuser detector in channels with random amplitudes, our objective in these simulations shall be to improve significantly on the decorrelator's performance. Furthermore, following work earlier in this chapter, we are also interested in how closely the proposed algorithms approach the single-user bound. Above we showed that the BER difference between the single user bound and the decorrelator could be expressed as a constant ratio at high SNR. Our figure of merit for these tests shall therefore be the ratio between the decorrelator BER and the BER for the algorithm to be tested.

4.4.1 Phase-coherent detectors: Rayleigh channels

In this section, results related to the phase-coherent multiuser detectors derived in section 4.2 will be presented for the Rayleigh fading case. Of primary interest in most signal processing algorithms is the relationship between performance and SNR. The EM Algorithm method and the modified EM algorithm method were tested, along with the conventional and decorrelating detectors for 2 users, while the modified EM algorithm alone was tested against the conventional and decorrelating detectors for 4 users. (Using the formulation given above, the EM algorithm detector is essentially intractable for more than two users.) As in [37], a cross-correlation matrix of the form

$$\mathbf{R} = \begin{bmatrix} 1 & r & \cdots & r \\ r & 1 & \cdots & r \\ \vdots & \vdots & \ddots & \vdots \\ r & r & \cdots & 1 \end{bmatrix} \tag{4.68}$$

was used. We used $r = 0.454$ and $r = 0.339$ for 2 and 4 users, respectively. These were

chosen because they led to values of $[\mathbf{R}^{-1}]_{ii} = 1.26 \simeq 1\,\text{dB}$, recalling that the ratio between the performance of the decorrelator and the single-user lower bound was a factor of $[\mathbf{R}^{-1}]_{ii}$ - hence the value chosen was selected because it represented a noticeable difference. For this set of simulations, the average SNR of every user, $S = 2\rho^2/\sigma^2$, was held equal, and plots were generated of bit error rate versus average SNR, as shown in Figure 4.8 for the modified EM algorithm detector. A plot comparing the performance of the numerically implemented EM algorithm detector and the modified EM algorithm detector for 2 users is given in Figure 4.9. This and all subsequent simulations were conducted in MATLAB using $5 \cdot 10^4$ samples, except where otherwise specified. From Figure 4.8, it is clear that our modified EM algorithm method is consistently better than the decorrelating detector, maintaining a ratio of just over 1.1 throughout the range of tested average SNR for two users, and slightly less for four users. Although this is a noticeable improvement, it remains far from the single user bound ratio of 1.26. Furthermore, from Figure 4.9, we observe that the performance of the EM algorithm detector and the modified EM algorithm detector are essentially equivalent over this range of SNR. Due to this result, and owing to the complexity of the numerical calculations required to implement it, in the sequel we shall exclude the EM algorithm detector from further experimentation.

We may also consider the performance of this algorithm for the more general case where the average SNRs of the users are different. It is known that most multiuser detection algorithms tend to have better performance in channels suffering from the near-far effect [2]. In fact, it is intuitively obvious that in channels suffering severely from the near-far effect, with one very powerful user and several much weaker users, the more powerful user should experience nearly single-user performance as the ratio of signal power to interference power approaches infinity. To ensure that our algorithm is fair, with no user suffering excessive interference to improve the performance of another, we require that every user's performance be at least as good as that user's performance under the decorrelator, which has performance independent of interference power. Furthermore, we remark that the extremely simple conventional receiver also approaches single-user performance in these circumstances, so to be successful our algorithm will be required to approach single-user performance more quickly than the conventional receiver.

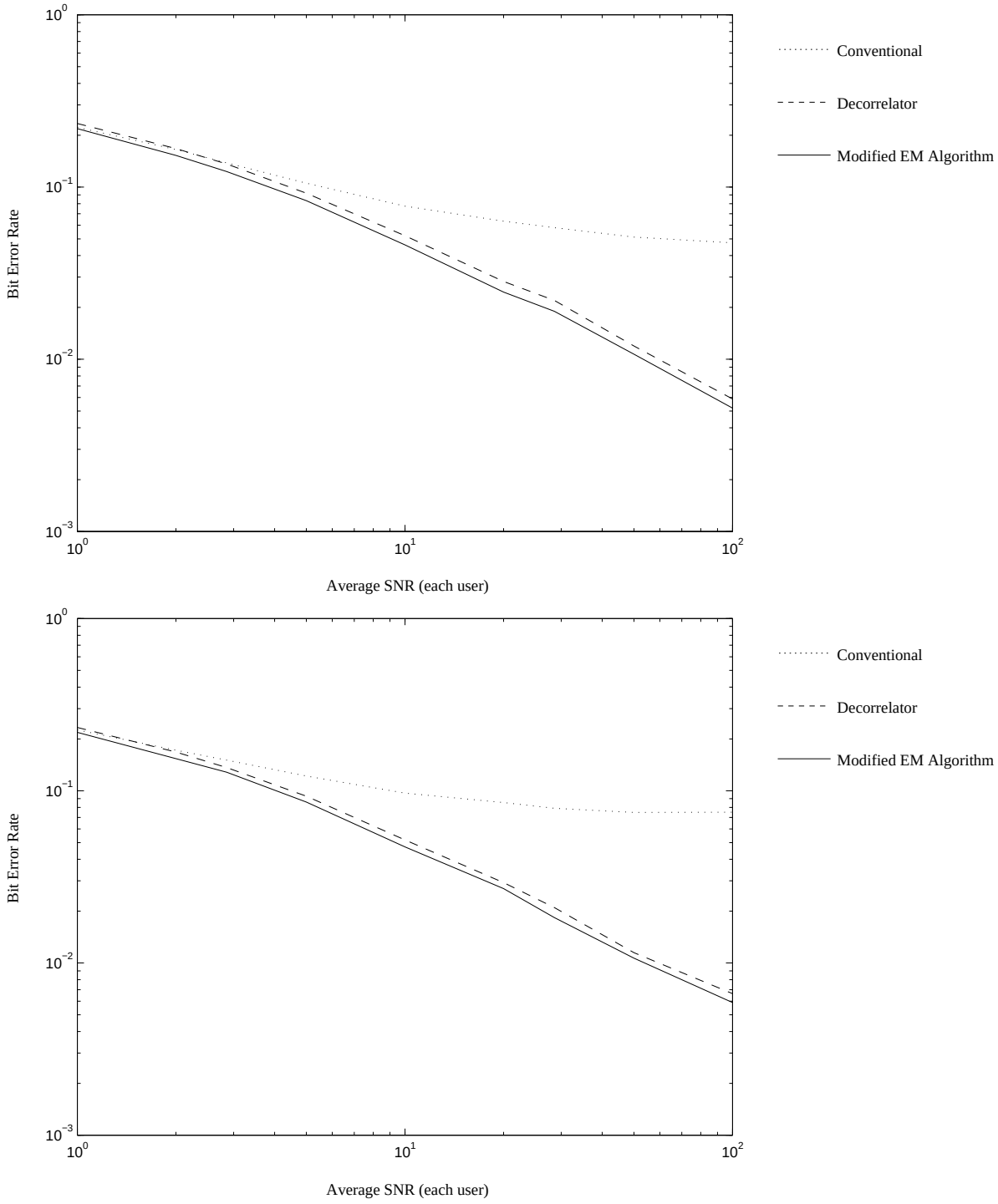

Figure 4.8: Plot of bit error rate versus average SNR for two users (top) and four users (bottom)

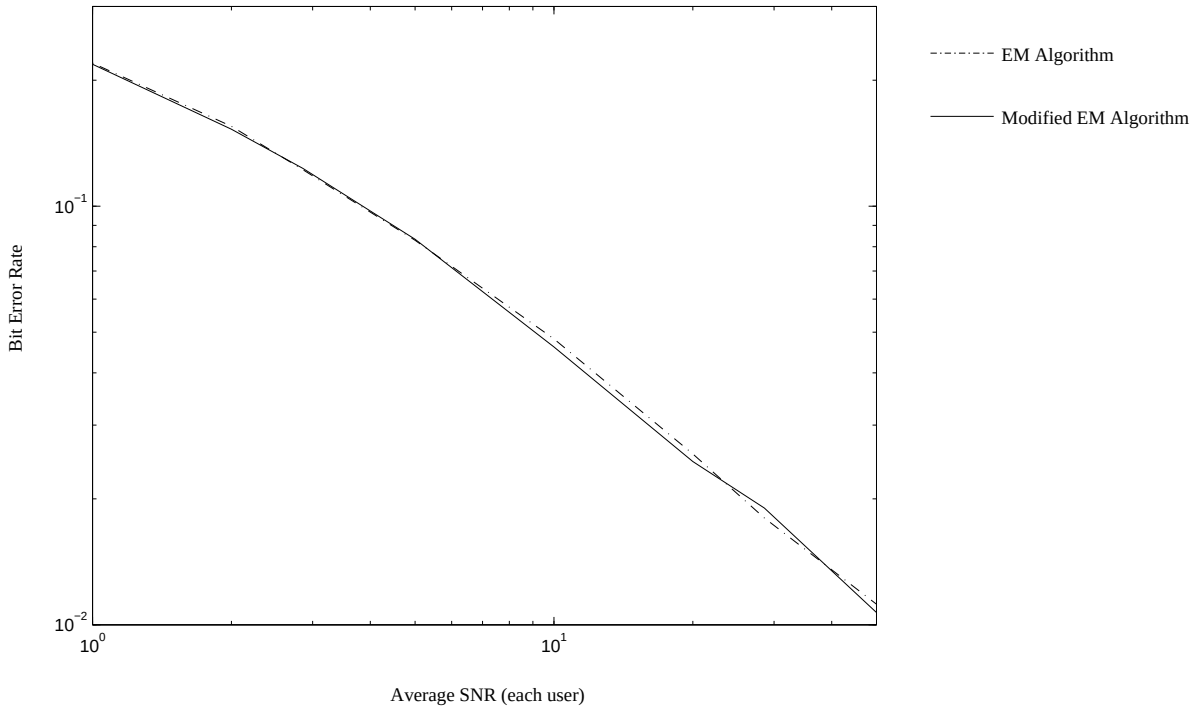

Figure 4.9: Comparison of EM algorithm and modified EM algorithm in changing average SNR for two users

Simulation results for the near-far case are shown in Figure 4.10 for two and four users, using the same definitions of $\mathbf{R}$ as the previous case. In each simulation, the value of the user of interest's average SNR was held constant at 7 dB, while that of the interfering user or users was varied. (In cases of more than one interfering users, every interfering user had the same average SNR). The results indicate that the modified EM algorithm receiver exceeds the BER performance of the decorrelator by nearly a factor of 1.1 everywhere. As predicted, the performance of both the conventional receiver and the given algorithm slowly approach the single-user bound as average interference SNR drops, although the modified EM algorithm's performance begins to approach single-user performance at interference SNRs of roughly 0 dB. Still, the results indicate good performance of this algorithm in near-far channels.

The relationship between the performance of this algorithm and the cross-correlation coefficients is also of interest. Previously, we chose values of $\mathbf{R}$ specifically to produce a significant separation between the performance of the decorrelating detector and the single

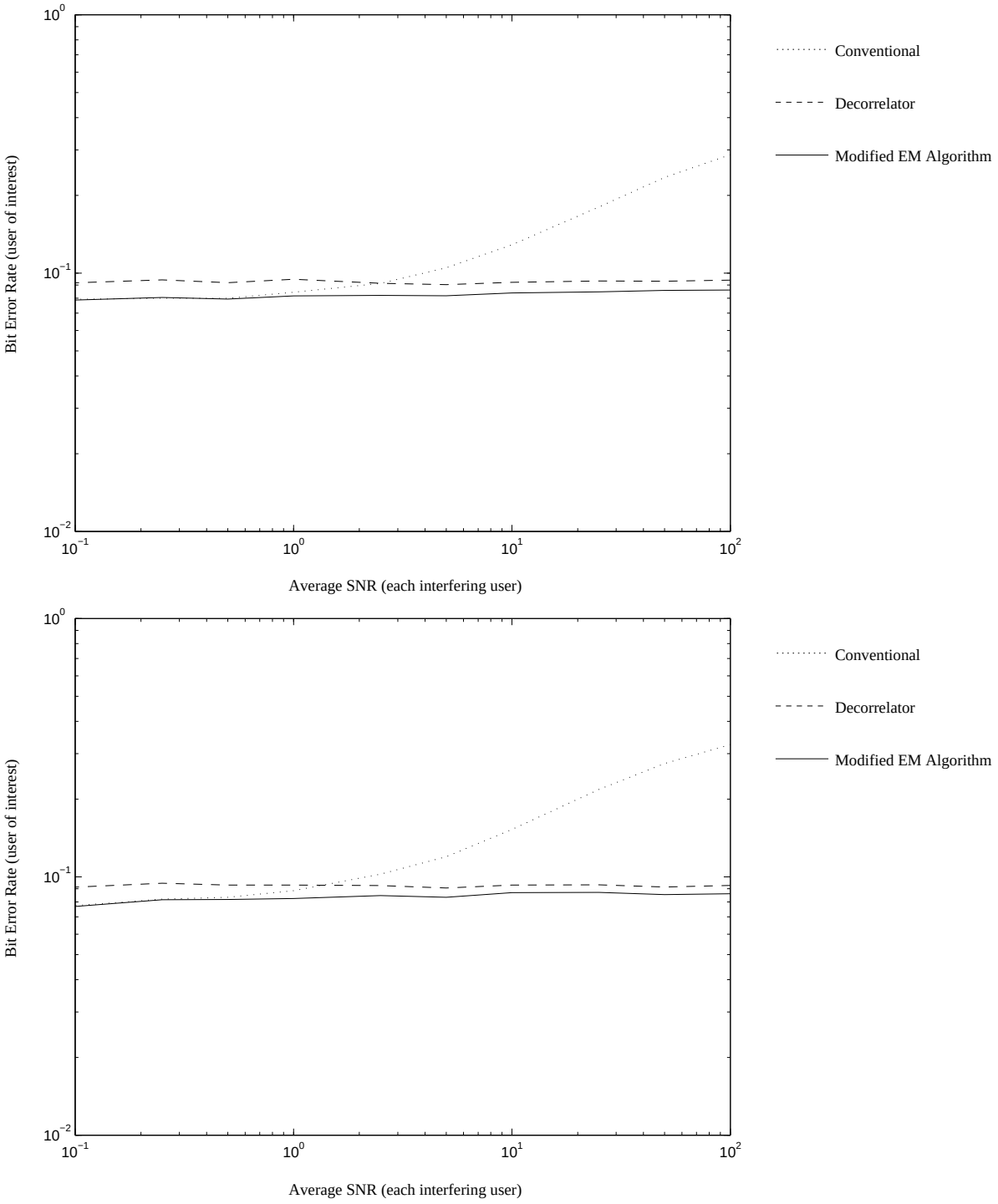

Figure 4.10: Plot of bit error rate for the user of interest versus average interfering user SNR for two users (top) and four users (bottom)

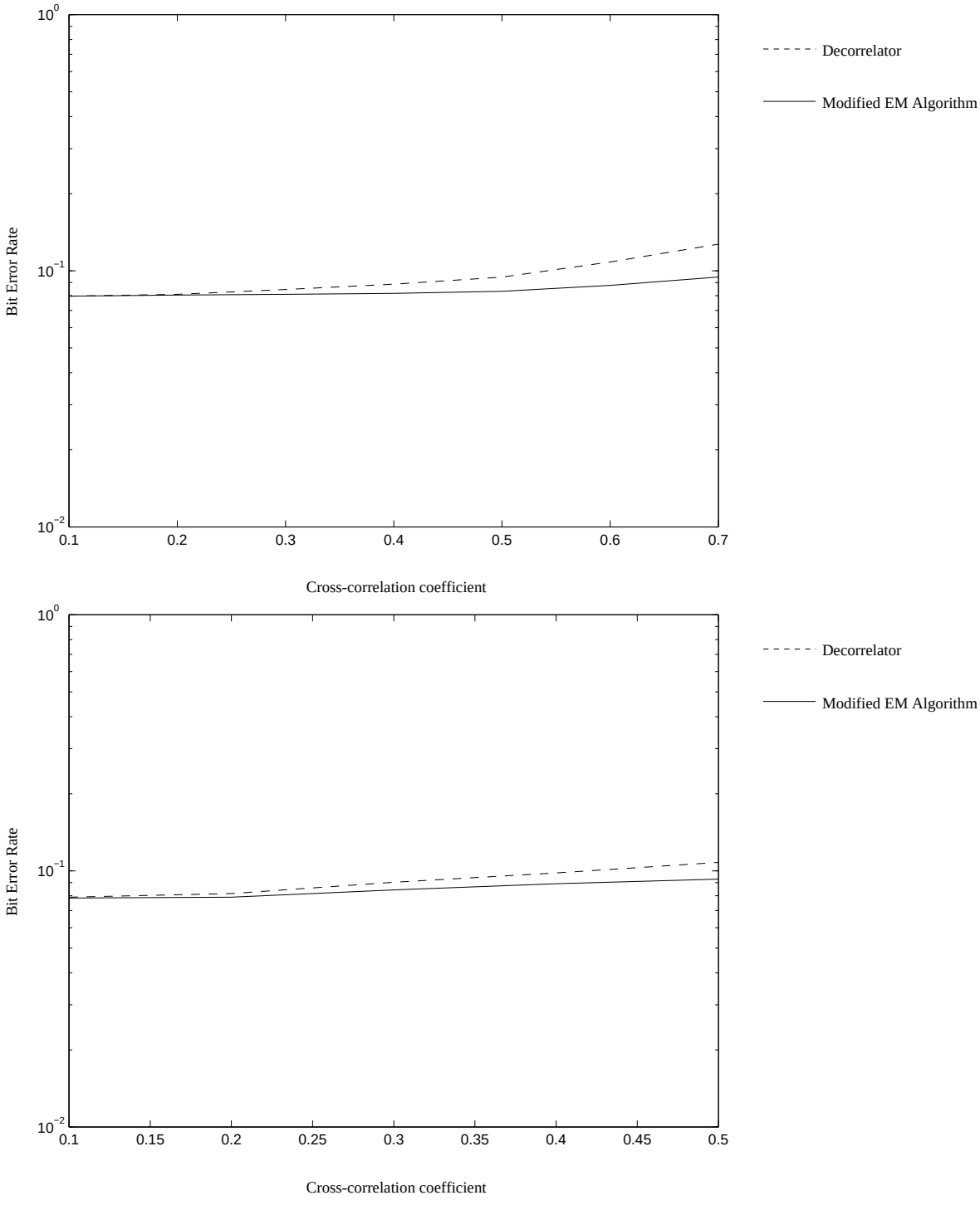

Figure 4.11: Plot of bit error rate versus cross-correlation coefficient r for two users (top) and four users (bottom)

user bound. A practical system would be likely to have smaller cross-correlation than the one we have chosen. Simulations were performed for the two and four user cases, where each user's average SNR was held at 7 dB, with a wide range of values of r. Results are shown in Figure 4.11. In those plots, in the two user case it is observed that for small values of r (i.e. $r \leq 0.1$), the difference between the modified EM algorithm and the decorrelator is essentially nil. As cross-correlation increases (roughly $r \geq 0.3$), the ratio of the decorrelator BER and the modified EM algorithm BER approaches slightly less than half the difference between the decorrelator and the single user bound (on a log scale). In the four user case, slightly less distinction is observed between the performance of the decorrelating detector and the modified EM algorithm. Here, the log scale difference between the two detectors is consistently in the vicinity of 1/3 the distance between the decorrelator and the single user bound.

4.4.2 Phase-coherent detectors: Rician channels

Owing to the similarity of the modified EM algorithm detectors for both the Rayleigh and Rician cases, this subsection will largely follow the previous one. However, we must deal with the complication of an additional parameter: α, the ratio of the line-of-sight signal component power to the scattered component power. We remark that if $\alpha = 0$, the channel reduces to the Rayleigh channel. Our preliminary observations have indicated that small values of α, i.e. $\alpha < 0.5$, are virtually undistinguished from the Rayleigh channel in terms of performance. We have thus concentrated our study on higher values of α, between 1 and 10. In principle, channels with even higher values of α may be examined, but the scattering component of such channels would be so small that the receiver for fixed amplitudes would likely suffice. In the sequel, in accordance with the assumptions made in derivation, we assume that both the amplitude and absolute phase of the fading component is known to the receiver.

For the sake of accurate comparison, values of $\mathbf{R}$ are the same in this section as in the previous section. Again, for this set of simulations, the average SNR of every user, for which we used the value of $S = (|L|^2 + 2\rho^2)/\sigma^2$ (from [34]), was held equal, and plots were

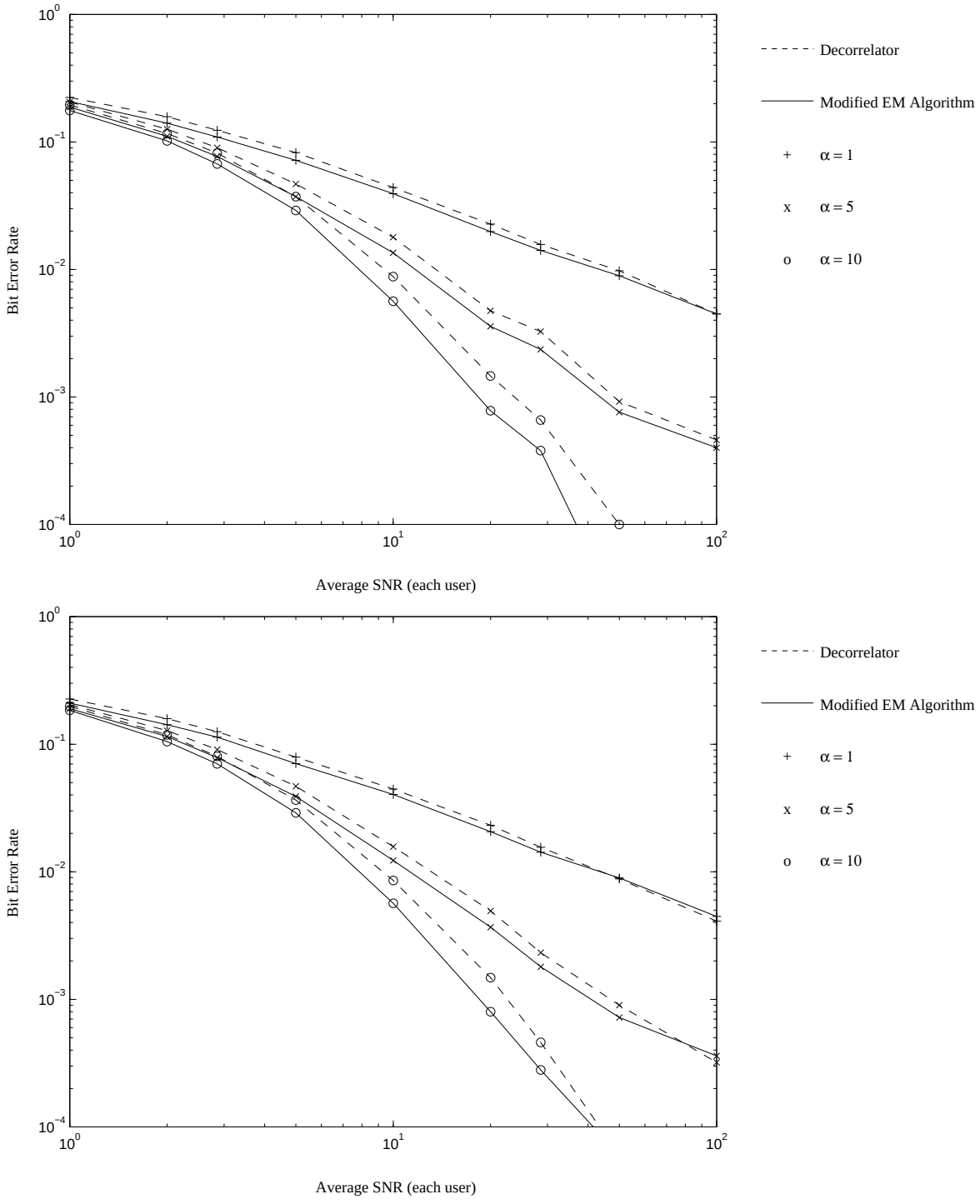

Figure 4.12: Plot of bit error rate versus average SNR for two users (top) and four users (bottom)

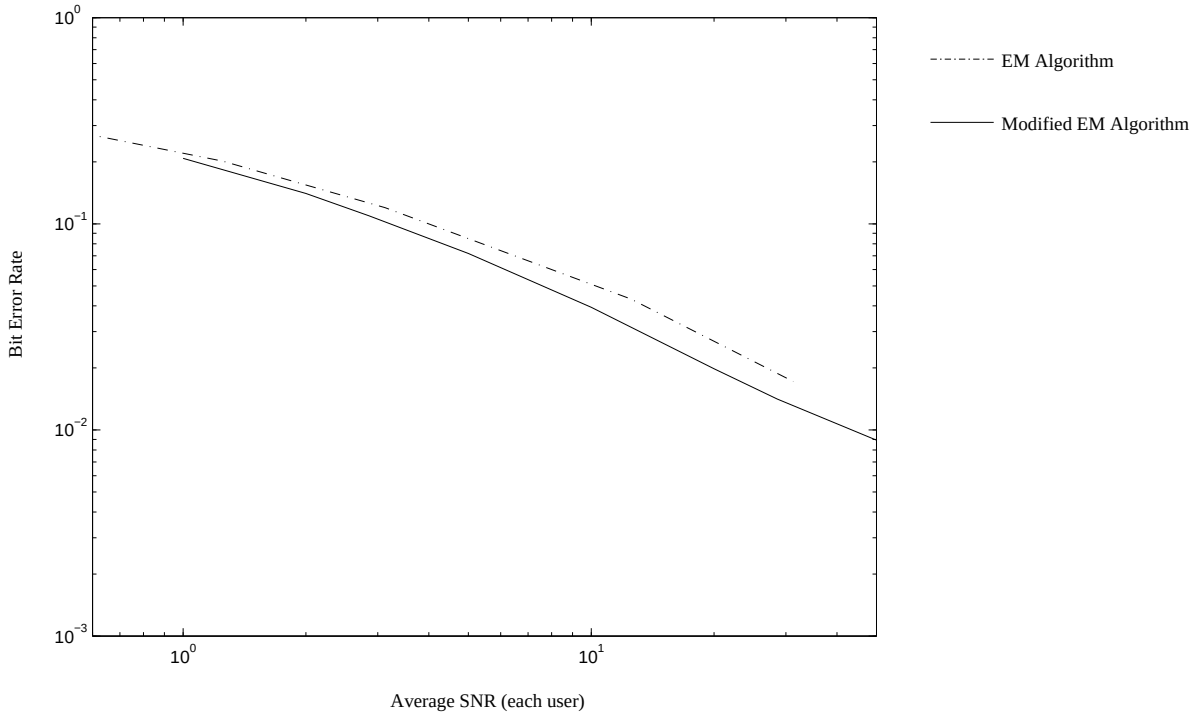

Figure 4.13: Comparison of EM-based detector and modified EM detector for $\alpha = 1$

generated of bit error rate versus average SNR, as shown in Figure 4.12. Results for the conventional detector are not presented in order to simplify the plots. The observed results are somewhat similar to the Rayleigh case. As anticipated, not only did increasing the value of α improve the overall performance of the system, it also increased the separation between the decorrelating detector's performance and that of the modified EM algorithm. It is also noticed that the performance of the modified EM algorithm approaches, and even begins to trail, the performance of the decorrelating detector at high SNR. This is likely related to the phenomena observed in Figure 4.1, where at very high SNR the difference between the single-user bound and the decorrelator performance are diminished. Nonetheless, Figure 4.12 clearly demonstrates that the modified EM algorithm performs well over a broad range of low to moderate SNR. Additionally, a comparison of the modified EM algorithm detector and the numerically implemented EM-based receiver is shown in Figure 4.13, for $\alpha = 1$. The EM-based receiver was simulated using $1 \cdot 10^4$ samples. The performance of the EM-based detector is significantly worse than the modified EM algorithm detector in this case, most

likely due to the approximations that were required in order to numerically calculate several integrals. For this reason, we shall focus the remainder of our analysis on the performance of the modified EM algorithm receiver.

Again, we are also concerned with the performance of our algorithm in fading environments suffering from the near-far problem. Simulation results for the near-far case are shown in Figure 4.14 for two and four users, using the same definitions of $\mathbf{R}$ as the previous case. Here we give results for the cases of $\alpha = 1$ and $\alpha = 5$. As well, as before, in each simulation the value of the user of interest's average SNR was held constant at 7 dB, while that of the interfering user or users was varied. The results are largely consistent with our observations for the Rayleigh fading case. For two users, we observe a nearly constant BER improvement of a factor of 1.15 for $\alpha = 1$ and 1.26 for $\alpha = 5$ over a wide range of interfering user SNR for the two user case (slightly less for each value of α in the four user case), and in both cases as the interference diminishes both approach single user performance slightly more quickly than the conventional receiver. An interesting phenomenon is noticed as interference SNR becomes very large - the performance of the modified EM algorithm seems to decline in performance. Again, this is perhaps due to the behavior noticed in Figure 4.1 for very high SNR signals. Furthermore, it again seems clear from these results that the performance of our algorithms improves as α increases.

Finally, as in the Rayleigh fading case, we wish to determine how the cross-correlation coefficient r affects the performance of the modified EM algorithm. Using an average SNR of 7 dB for each user, experiments were conducted with $0.1 \leq r \leq 0.7$ for 2 users, and $0.1 \leq r \leq 0.5$ for 4 users. Results are shown in Figure 4.15 for $\alpha = 1$, 5, and 10. As expected, improvement is noticed as r increases, and the gap enlarges as α increases. It is interesting to note that there is a discernable improvement in the high-α case even for $r = 0.1$, which seems to improve slightly as the number of users is increased, unlike in the Rayleigh case.

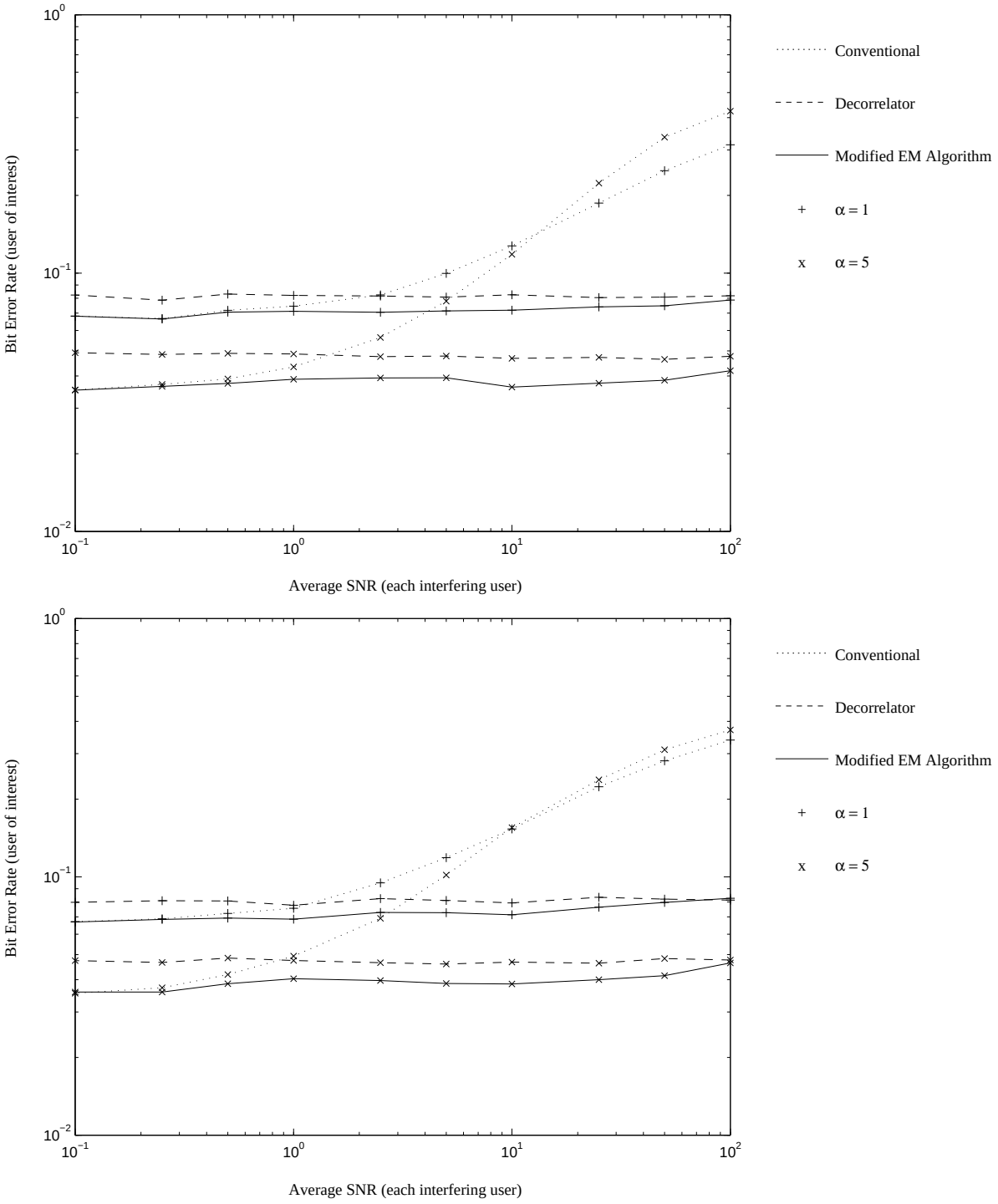

Figure 4.14: Plot of the user of interest's bit error rate versus average interfering user SNR for two users (top) and four users (bottom)

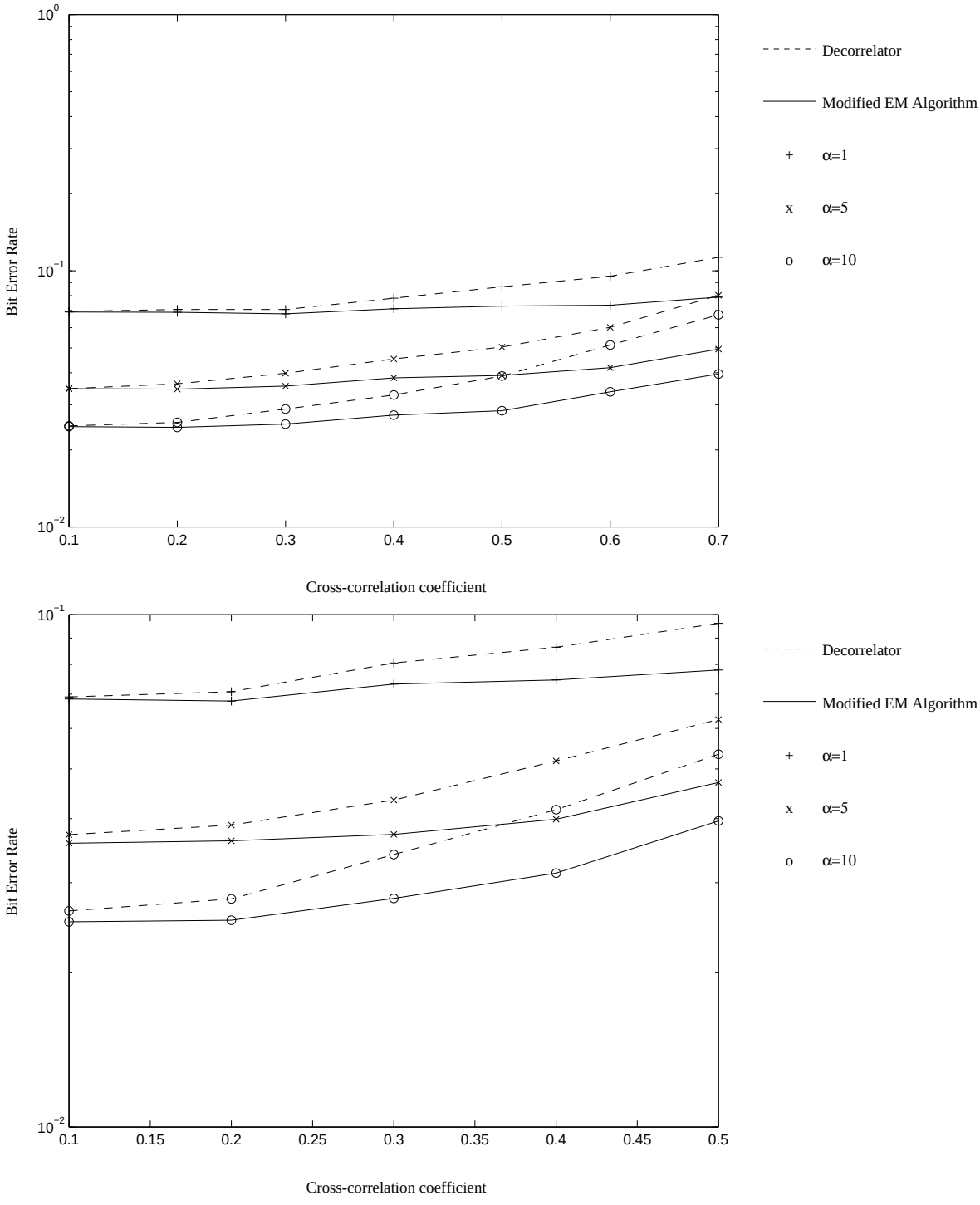

Figure 4.15: Plot of bit error rate versus cross-correlation coefficient r for two users (top) and four users (bottom) in Rician fading

4.4.3 Incoherent detectors

As we have seen, the problem of incoherent detection requires an entirely different approach from that of phase-coherent detection. Additionally, the problem of evaluating incoherent detectors is significantly complicated by the inclusion of the Doppler frequency parameter, which is used to calculate the autocorrelation sequence of the fading signal, as in (2.29). Furthermore, although earlier we stated that it was sufficient to design our detector using the assumption that the fading amplitudes could be modelled as first-order Markov processes, to put our detector to a realistic test requires a more stringent model on the simulated channel. In our case, a model of order 5 was used, which for most values of Doppler frequency expressed the autocorrelation function to at least the first null. We intended our discussion of incoherent multiuser detection to be merely a brief introduction to the topic. Thus, in this analysis we shall consider the effects of Doppler frequency, SNR, and the near-far effect on incoherent reception for the two-user channel only.

Let us first consider the effect of Doppler frequency. From (2.29), we know that the autocorrelation function is given by a function of $\omega_m t$, where ω_m is the maximum angular Doppler frequency, given by

$$\omega_m = \frac{\omega_c v}{c} \tag{4.69}$$

where v is the velocity of the mobile, $c = 3 \cdot 10^8$ m/s is the speed of light, and $\omega_c = 2\pi f_c$ is the carrier angular frequency. Furthermore, since we are concerned with a discrete-time data transmission system, we may substitute k/R_s for t, where R_s is the symbol rate, and $k \in \{..., -1, 0, 1, ...\}$ is the symbol index. Thus, we now have

$$\omega_m t = \frac{2\pi f_c v}{R_s c} k = A_m k \tag{4.70}$$

where we call A_m the autocorrelation coefficient. Let us consider some arbitrary but reasonable values for the autocorrelation coefficient in (4.70). In a terrestrial example, suppose the carrier frequency f_c is given by 3 GHz, the mobile velocity v is given by 27.8 m/s (i.e. 100 km/hr), and the symbol rate R_s is given by 10^4 baud. This gives a value of $A_m \simeq 0.2$, which results in a value of 0.99 times the central maximum for $k = 1$. In a low-Earth-orbit

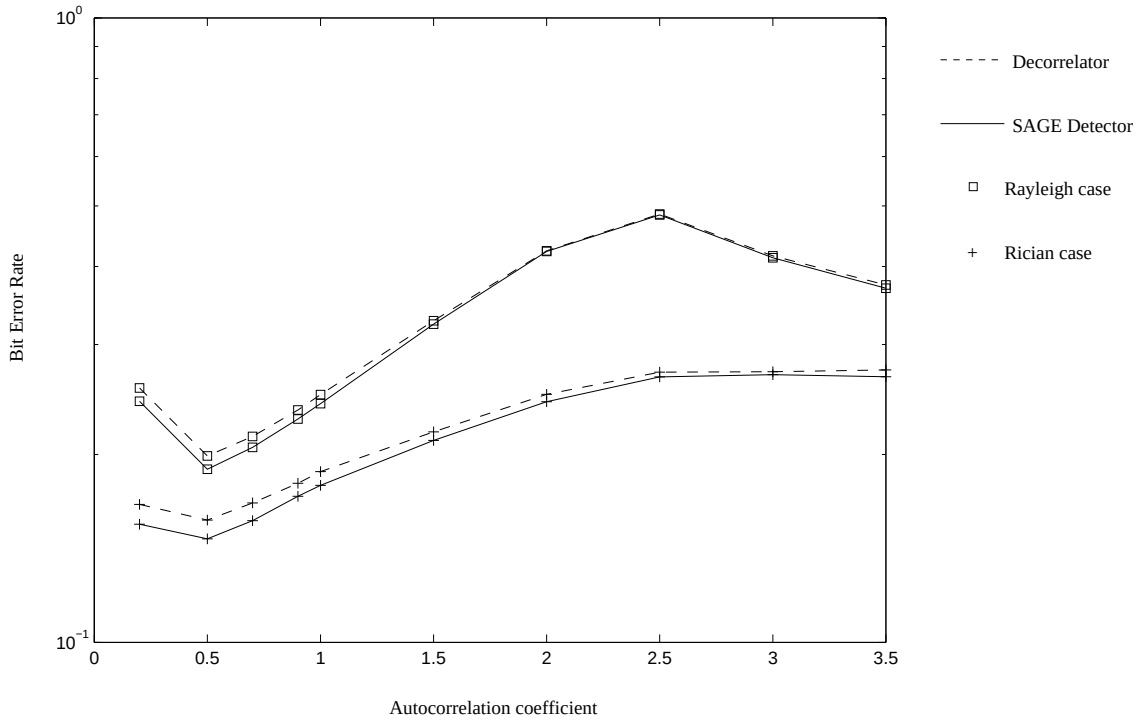

Figure 4.16: Plot of bit error rate versus autocorrelation coefficient

satellite communication example, suppose the carrier frequency f_c is given by 12 GHz, the relative velocity between the satellite and ground station v is given by 278 m/s (i.e. 1000 km/hr), and the symbol rate R_s is given by 10^5 baud. This gives a value of $A_m \simeq 0.7$, which results in a value of 0.88 times the central maximum for $k = 1$.

The effect of changing ω_m was analyzed for the Rayleigh and Rician ($\alpha = 1$) cases, for two users, with average SNR of each user equal to 7 dB. The cross-correlation coefficient between the users was chosen to be $r = 0.454$. Results are shown in Figure 4.16. The behavior of this curve is somewhat unexpected, as we would normally expect the bit error rate to increase monotonically with A_m. It is possible that the observed effects are a result of the first-order assumption used in designing the receiver. However, we also notice that the performance gap between the decorrelating detector and the SAGE detector decreases as A_m increases, diminishing to nearly zero for very large values of A_m. As a result of all these effects, we shall choose a small value of A_m close to the best performing value in Figure 4.16, $A_m = 0.64$, which corresponds to an autocorrelation of almost exactly 0.9 times the

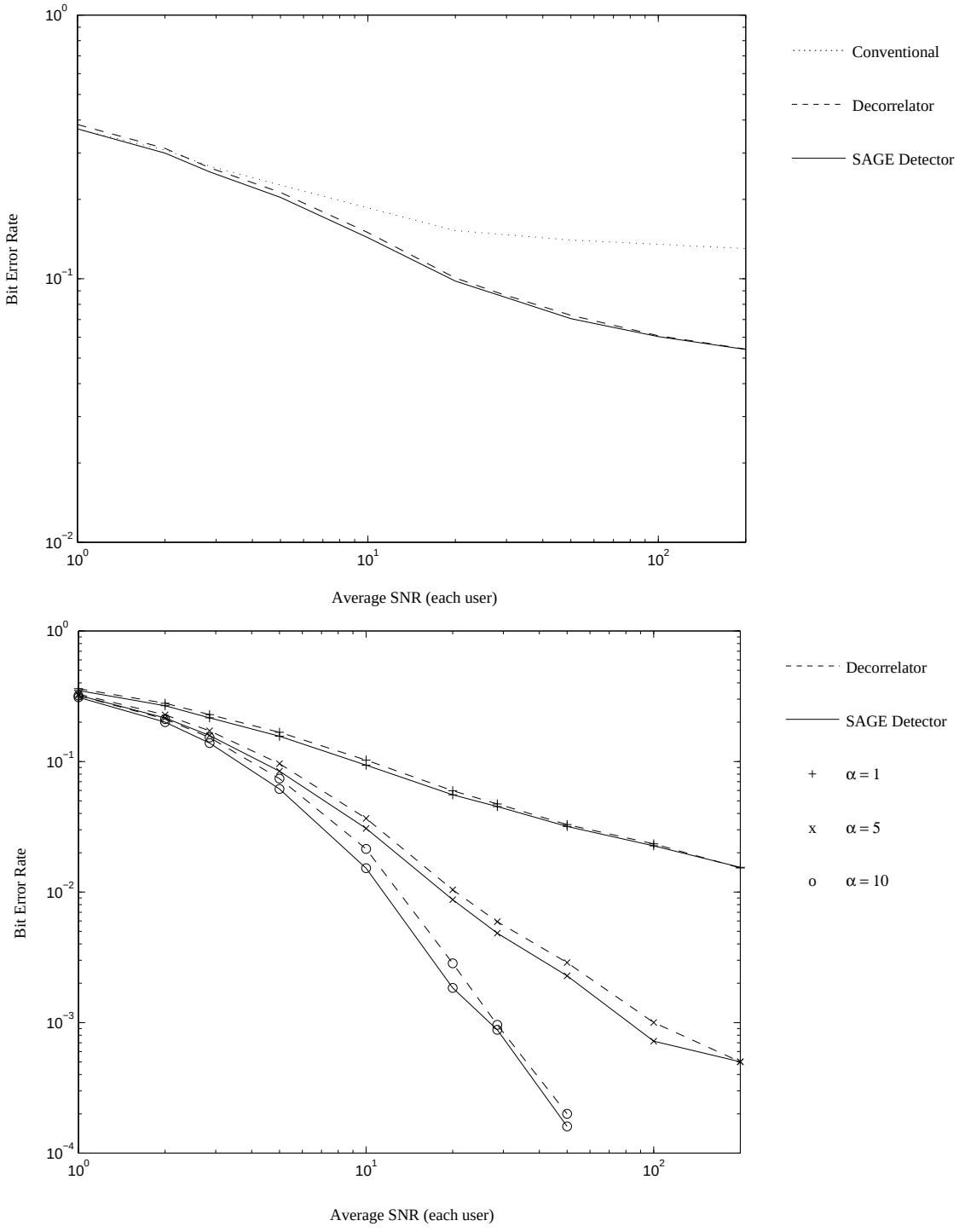

Figure 4.17: Plot of bit error rate versus average SNR for Rayleigh (top) and Rician (bottom) fading models

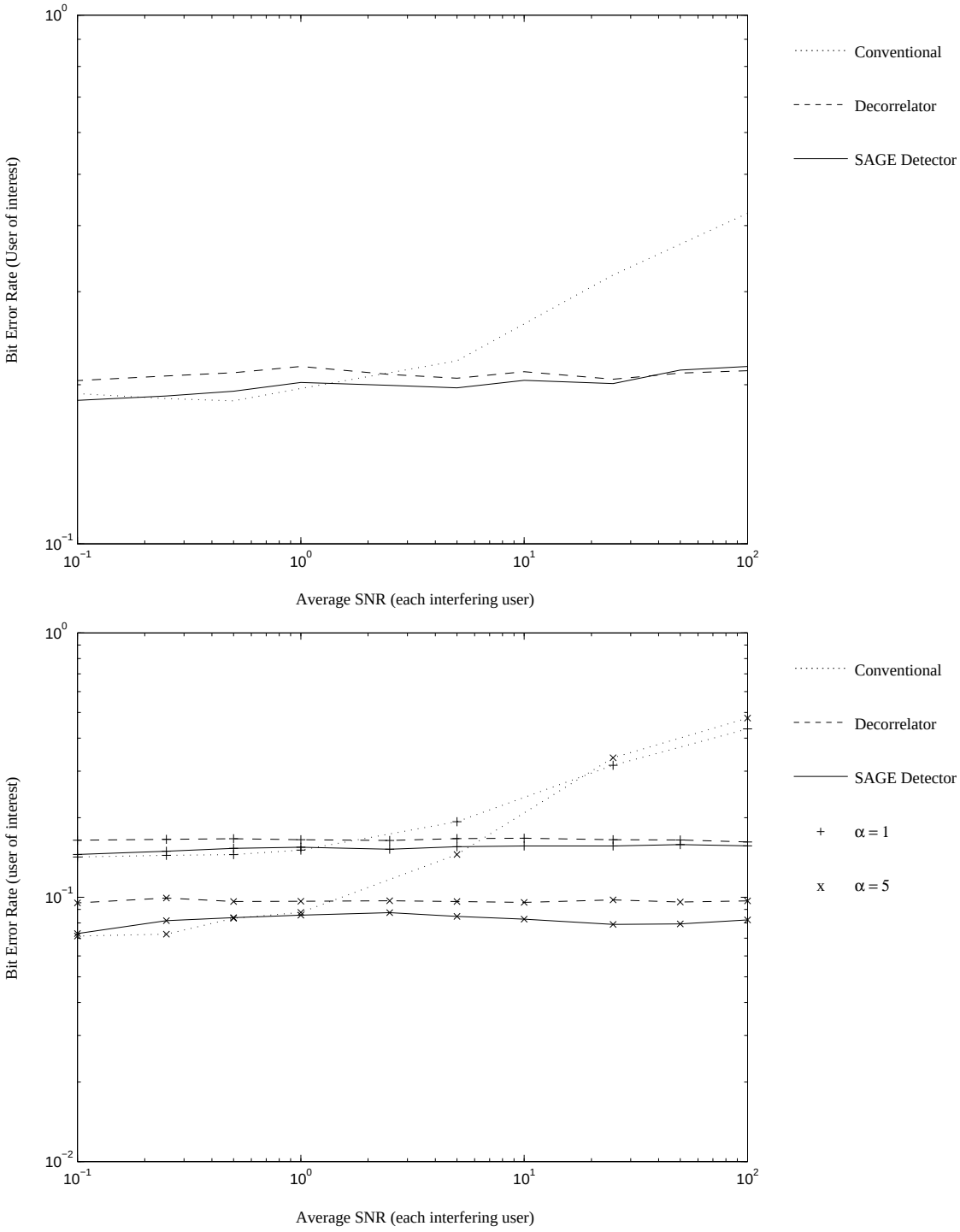

Figure 4.18: Plot of bit error rate of the user of interest versus average interfering user SNR for the Rayleigh (top) and Rician (bottom) fading models

maximum for $k = 1$.

We also consider the effect of SNR on bit error rate for a two-user system. As indicated above, we used $A_m = 0.64$, as well as $r = 0.454$, as in the phase-coherent two user case. The results for Rayleigh and Rician ($\alpha = 1$, 5, and 10) are shown in Figure 4.17. Not surprisingly, similar behavior is observed between these plots and Figures 4.8 and 4.12. Again, the gap between the decorrelating detector and the SAGE algorithm detector increases with α, though it is consistently narrower than for the coherent case - only negligible improvement for the Rayleigh case, a factor of 1.07 for the case of $\alpha = 1$, a factor of 1.15 for the case of $\alpha = 5$, and a factor of 1.26 for the case of $\alpha = 10$, with local variations. In every case the decorrelator and SAGE detector performance become identical for very high average SNR.

As before, we are also interested in the performance of this algorithm under conditions of widely different received powers. The same r and A_m as above were used with two users with the Rayleigh and Rician ($\alpha = 1$ and 5) fading models. Results are shown in Figure 4.18. As expected, the results are somewhat similar to Figures 4.10 and 4.14, as we observe a gradual approach of single-user-bound performance as the interfering average SNR approaches zero. However, for the Rayleigh case, it is interesting to note that the performance drops to that of the decorrelator as interfering SNR increases, which is unlike the phase-coherent case.

4.5 Discussion of Results

In this chapter, we have developed and analyzed several receiver structures for use in fading channels. The experimental results presented in the previous section have indicated some potential applications for EM-based multiuser detection in fading channels. We remark first that, assuming fast algorithms can be found to calculate the various coefficients required, the modified EM algorithm (for phase-coherent detection) and the SAGE algorithm (for incoherent detection) have complexity $O(n^2)$, where n is the number of users, which is the same order as the decorrelator. However, we furthermore remark that these algorithms obviously cannot be implemented with the same or less complexity than the decorrelator, since they all require a decorrelator "front end" to provide initial symbol and amplitude estimates.

As we have described, each algorithm presented is subject to the single-user bound, which severely curtails the improvements that can be made over the decorrelating detector. Nonetheless, such improvements, while generally not very large, have been observed to be nontrivial, and the described algorithms have been shown to cross a significant fraction of the gap - and the entire gap under certain near-far circumstances. Thus, these algorithms would likely find use in applications where sufficient computing power is available, and where a 10-20 % improvement in performance would be considered significant. An example of such a system would be a heavily-used cell in a dense urban area, where a small increase in performance per user would increase the capacity of the system by a few subscribers, hence reducing blocking effects and possibly eliminating the requirement to erect a new cell tower.

Furthermore, from (4.70) and the given examples, it seems clear that the low velocity and (relatively) low carrier frequencies used in terrestrial mobile communications should make accurate phase estimation available to the receivers, implying that phase-coherent multiuser detection may be used. This is good news, since in the terrestrial environment the Rayleigh fading case is frequently encountered, and since the algorithms presented seem to fare much better in phase-coherent environments than incoherent environments for Rayleigh fading channels. We also know that the higher velocities and carrier frequencies used in satellite communication would likely require the use of incoherent detection. Our results show that our methods are useful in these circumstances, since the satellite communication channel is affected by Rician fading [21], and since the presented incoherent detectors did reasonably well under Rician fading.

Chapter 5

Conclusion and Recommendations for Further Study

In this thesis, we have demonstrated important results related to multiuser detection using the EM algortihm. Methods have been indicated for integrating turbo detection into an EM-based multiuser detection framework. Furthermore, we have shown that improvements may be made over the decorrelating detector using amplitude estimation and symbol detection based on the EM algorithm concept in both Rayleigh and Rician channels, with or without accurate knowledge of the signal phase. In this chapter, we will present a summary of the contributions of this work to the field, as well as a brief discussion of opportunities for future research.

5.1 Summary of Contributions

A list of contributions of this work to the field of multiuser detection is as follows:

- A mechanism for including extrinsic information from coding into the interference cancellation of EM-based multiuser detection;
- An approximate approach to uncoded EM-based multiuser detection based on the turbo principle;

- Bounds on performance in two-user Rayleigh multiuser channels assuming knowledge of the channel state;
- Adaptation of the EM algorithm to the problem of amplitude estimation in Rayleigh and Rician channels for multiuser detection;
- Adaptation of Yu's algorithm [26] to Rician fading channels; and
- Use of Yu's algorithm [26] for MLSD in multiuser detection.

From experimentation, it was found that the adaptation of the EM algorithm for amplitude estimation in multiuser detection delivered a noticeable, though small, improvement over the decorrelating detector. It was furthermore noticed that approaching the single-user bound using the present algorithm was only possible for powerful users in extreme near-far channels, although in such cases the present algorithm approached single-user performance somewhat more quickly than the conventional detector. The complexity required in order to achieve these improvements is greater, though on the same order of magnitude, as the decorrelating detector, assuming simple ways can be found to estimate all the required channel parameters.

5.2 Proposals for Future Research

The current work leaves open some significant issues in practicality. In particular, while the objective of this work was to determine the best performance achievable without resorting to coding, the ways in which the performance of the algorithms presented here may be enhanced through coding should be explored. Furthermore, in developing these algorithms we used the unrealistic assumption of chip and symbol synchronization. In order to develop a more practical system, future research should examine the effect of relaxing this requirement. We also assumed the availability of accurate knowledge of channel parameters, such as SNR, Doppler frequency, and so on. The effect of mismatches in these parameters should be examined. Additionally, some relatively complicated calculations were required in the E-steps of the above algorithms. In order for this method to be made more practical, methods should be

investigated for high-speed implementations of these calculations. The incoherent multiuser detection algorithms presented in this paper should be examined in further detail, possibly including APP information from SOVA, as well as to determine the effects of higher-order implementations. Finally, diversity is known to improve performance in fading channels, and the implementation of the above algorithms with diversity should be examined. The present work took a major step towards making multiuser detectors more practical by relaxing the requirement of complete channel state knowledge. Clearly, many more issues need to be examined before multiuser detectors may be deployed in a realistic wireless channel.

Appendix A

Monotonicity of the EM Algorithm

In this appendix, we shall re-present the results from [10] which showed that the parameter estimate obtained from the EM Algorithm increases monotonically in likelihood with each iteration. Consider the conditional PDF of the complete data x conditioned on the incomplete data y and the parameter set $\theta \in \Omega$:

$$f_x(x \mid y, \theta) = \frac{f_x(x \mid \theta)}{f_y(y \mid \theta)} \tag{A.1}$$

Taking the logarithm of both sides and rearranging, we have

$$\log f_x(x \mid \theta) = \log f_x(x \mid y, \theta) + L(\theta) \tag{A.2}$$

where $L(\theta) = \log f_y(y \mid \theta)$ is the log-likelihood function for the parameter θ. Taking the expected value conditioned on y and the assignment $\theta = \bar{\theta}$, we now have

$$Q(\theta;\, \bar{\theta}) = H(\theta;\, \bar{\theta}) + L(\theta) \tag{A.3}$$

where $Q(\theta; \bar{\theta})$ is the expression from the E-step in (1.1), and $H(\theta; \bar{\theta}) = E[\log f_x(x|y, \theta)|y, \theta = \bar{\theta}]$. We now restate Lemma 1 from [10]:

Lemma. *For any pair* $(\theta,\, \bar{\theta})$ *in* $\Omega \times \Omega$, $H(\theta;\, \bar{\theta}) \leq H(\bar{\theta};\, \bar{\theta})$ *with equality if and only if* $f_x(x \mid y, \bar{\theta}) = f_x(x \mid y, \theta)$ *almost everywhere.*

Proof. This is a corrolary of Jensen's inequality [46].

A generalized EM algorithm is defined as any iteration where $\bar{\theta}_{n+1} = M(\bar{\theta}_n)$, where $M(\cdot)$ is some mapping $\Omega \to \Omega$, and where

$$Q(M(\bar{\theta}_n);\, \bar{\theta}_n) \geq Q(\bar{\theta}_n;\, \bar{\theta}_n) \tag{A.4}$$

We may now restate Theorem 1 from [10]:

Theorem. *For every generalized EM algorithm, $L(M(\bar{\theta}_n)) \geq L(\bar{\theta}_n)$, where equality holds if and only if both $Q(M(\bar{\theta}_n);\, \bar{\theta}_n) = Q(\bar{\theta}_n;\, \bar{\theta}_n)$ and $f_x(x \mid y, \bar{\theta}) = f_x(x \mid y, \theta)$ almost everywhere.*

Proof. Consider

$$L(M(\bar{\theta}_n)) - L(\bar{\theta}_n) = [Q(M(\bar{\theta}_n);\, \bar{\theta}_n) - Q(\bar{\theta}_n;\, \bar{\theta}_n)] + [H(\bar{\theta}_n;\, \bar{\theta}_n) - H(M(\bar{\theta}_n);\, \bar{\theta}_n)] \tag{A.5}$$

From (A.4), the difference in Q function is greater than or equal to zero by definition. From the Lemma, the difference in H function is always greater than or equal to zero, with equality under the condition given in the Theorem. Thus the Theorem holds.

The Theorem shows that any mapping $M(\cdot)$ which satisfies (A.4) will lead to a higher likelihood estimate of the parameter θ. As a corrolary to the Theorem, consider a mapping $M(\cdot)$ as follows:

$$M(\bar{\theta}_n) = \underset{\theta \,\in\, \Omega}{\arg\max}\, Q(\theta;\, \bar{\theta}_n) \tag{A.6}$$

that is, take the value of θ which maximizes $Q(\theta; \bar{\theta}_n)$, which forms the EM algorithm M-step. Obviously, condition (A.4) holds, and therefore by the Theorem, the iteration given by the EM algorithm increases monotonically in likelihood with each iteration.

Appendix B

Channel Information and the Decorrelator Solution

In Chapter 3, we asserted that the channel, or intrinsic, information in a symbol observation was given by the decorrelator output for that symbol as well as by the expressions in (3.27) and (3.28). To show this, we note that the element of the whitened vector used to calculate intrinsic information was always the element whose interference was completely cancelled by the whitening operation.

Let $[\cdot]_i$ be the ith row of a matrix. Then the row of $\mathbf{G}^T$ corresponding to the user whose interference is completely cancelled is:

$$[\mathbf{G}^T]_n = [\mathbf{G}^{-1}]_n\mathbf{R} = [0\ 0\ \cdots\ k],\ 0 \leq k \leq 1 \tag{B.1}$$

We also know that:

$$[\mathbf{R}^{-1}]_n\mathbf{R} = [0\ 0\ \cdots\ 1] \tag{B.2}$$

This implies that $[\mathbf{G}^{-1}]_n = k[\mathbf{R}^{-1}]_n$, and from linear algebra we know that this solution is unique. Furthermore, we know that the decorrelating detector implies some noise enhancement. Let e be the noise enhancement, where $e \geq 1$:

$$E\left[([\mathbf{R}^{-1}]_n\mathbf{z})^2\right] = e^2\sigma^2 \tag{B.3}$$

where $\mathbf{z}$ is the multiuser noise vector with cross-correlation matrix $\mathbf{R}$, and σ^2 is the variance of the noise in each observation. We also know that:

$$E\left[\left([\mathbf{G}^{-1}]_n \mathbf{z}\right)^2\right] = \sigma^2 \tag{B.4}$$

From above, obviously $e = k$, so therefore the output of the decorrelator and the whitening filter for the user whose interference is completely cancelled are identical. Therefore, the decorrelator output may be substituted for the whitened filter outputs in (3.27) and (3.28).

Appendix C

Some Important Integral Relations

In this appendix, we will discuss some important integral relations that are used in several expressions in Chapter 4. Specifically, we shall examine integrals of the form

$$\int_{x=0}^{\infty} x^k \exp(-wx^2 + vx)dx \tag{C.1}$$

where k is a positive integer and $w > 0$. Let us first examine the case where $k = 1$. We may expand (C.1) as follows:

$$\frac{v}{2w}\int_0^{\infty} \exp(-wx^2 + vx)dx - \frac{1}{2w}\int_0^{\infty} (v - 2wx)\exp(-wx^2 + vx)dx \tag{C.2}$$

The expression on the right in (C.2) is straightforward to integrate, and the result is $-1/2w$. To solve the expression on the left, we expand further, such that the overall expression becomes:

$$\frac{v}{2w}\exp\left(\frac{v^2}{4w}\right)\int_0^{\infty} \exp\left[-w\left(x - \frac{v}{2w}\right)^2\right]dx + \frac{1}{2w} \tag{C.3}$$

The expression on the left is now straightforward to express in terms of the complementary error function (erfc). The result is:

$$\frac{v}{4w}\sqrt{\frac{\pi}{w}}\exp\left(\frac{v^2}{4w}\right)\operatorname{erfc}\left(-\frac{v}{2\sqrt{w}}\right) + \frac{1}{2w} \tag{C.4}$$

Let us now consider the case where $k = 2$. As before, we may expand the expression:

$$\frac{v}{2w}\int_0^\infty x\exp(-wx^2+vx)dx - \frac{1}{2w}\int_0^\infty x(v-2wx)\exp(-wx^2+vx)dx \tag{C.5}$$

The expression on the left is simply the case solved above, where $k = 1$. On the right, we may integrate by parts, which produces

$$\frac{1}{2w}\int_0^\infty \exp(-wx^2+vx)dx - \frac{1}{2w}x\exp(-wx^2+vx)|_0^\infty + \frac{v}{2w}\int_0^\infty x\exp(-wx^2+vx)dx \tag{C.6}$$

The expression on the left in (C.6) may be solved in a manner similar to the derivation above, whereas the middle expression evaluates to zero. The result becomes:

$$\frac{2w+v^2}{4w^2}\sqrt{\frac{\pi}{w}}\exp\left(\frac{v^2}{4w}\right)\operatorname{erfc}\left(-\frac{v}{2\sqrt{w}}\right) + \frac{v}{4w^2} \tag{C.7}$$

We may now turn our attention to the case where $k \geq 3$. This case is difficult to solve in closed form. However, the following relation is helpful, which is obtained through integration by parts, somewhat analogously to the derivation for $k = 2$:

$$\int_0^\infty x^k\exp(-wx^2+vx)dx = \\ \frac{v}{2w}\int_0^\infty x^{k-1}\exp(-wx^2+vx)dx + \frac{n-1}{2w}\int_0^\infty x^{k-2}\exp(-wx^2+vx)dx \tag{C.8}$$

Applying the expression in (C.8) repeatedly, any case where $k \geq 3$ can be reduced to an expression containing only integrals where $k = 1$ or $k = 2$, which were both solved above.

Bibliography

[1] Sergio Verdu, "Minimum Probability of Error for Asynchronous Gaussian Multiple-Access channels," *IEEE Trans. Info. Theory*, vol. 32, no. 1, January 1986, pp. 85-96.

[2] Ruxandra Lupas and Sergio Verdu, "Linear Multi-User Detectors for Synchronous Code-Division Multiple-Access Channels," *IEEE Trans. Info. Theory*, vol. 35, no. 1, January 1989, pp. 123-36.

[3] Ruxandra Lupas and Sergio Verdu, "Near-Far Resistance of Multiuser Detectors in Asynchronous Channels," *IEEE Trans. Commun.*, vol. 38, no. 4, April 1990, pp. 496-508.

[4] Alexandra Duel-Hallen, "Decorrelating Decision-Feedback Multi-User Detector for Synchronous Code-Division Multiple Access Channels," *IEEE Trans. Commun.*, vol. 41, no. 2, February 1993, pp. 285-90.

[5] Alexandra Duel-Hallen, "A Family of Multiuser Decision-Feedback Detectors for Asynchronous Code-Division Multiple Access Channel," *IEEE Trans. Commun.*, Feb. 1995, pp. 421-434.

[6] Alexandra Duel-Hallen, Jack Holtzman, and Zoran Zvonar, "Multiuser Detection for CDMA Systems," *IEEE Personal Communications*, April 1995, pp. 46-58.

[7] Shimon Moshavi, "Multi-User Detection for DS-CDMA Communications," *IEEE Communications Magazine*, vol. 34, no. 10, October 1996, pp. 124-136.

[8] Laurie B. Nelson and H. Vincent Poor, "Iterative Multiuser Receivers for CDMA Channels: An EM-Based Approach," *IEEE Trans. Commun.*, vol. 44, no. 12, December 1996, pp. 1700-1710.

[9] Laurie B. Nelson and H. Vincent Poor, "Soft-decision Interference Cancellation for AWGN Multi-User Channels," *1994 IEEE Intl. Symp. Info. Theory*, New York, NY, 1994, p. 134.

[10] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *J. Roy. Stat. Soc. B*, vol. 39, 1977, pp. 1-17.

[11] Todd K. Moon, "The Expectation-Maximization Algorithm," *IEEE Signal Processing Magazine*, vol. 13, no. 6, November 1996, pp.47-60.

[12] J. A. Fessler and A. O. Hero, "Space-alternating Generalized Expectation-Maximization Algorithm," *IEEE Trans. Signal Proc.*, vol. 42, no. 10, October 1994, pp. 2664-2677.

[13] Sergio Verdu, *Multiuser Detection*, New York: Cambridge University Press, 1998.

[14] Gene H. Golub and Charles F. Van Loan, *Matrix Computations, third edition*, Baltimore: Johns Hopkins University Press, 1996.

[15] Costas N. Georghiades and Donald L. Snyder, "The Expectation-Maximization Algorithm for Symbol Unsynchronized Sequence Detection," *IEEE Trans. Commun.*, vol. 39, no. 1, January 1991, pp.54-61.

[16] Costas N. Georghiades and Jae Choong Han, "Sequence Estimation in the Presence of Random Parameters Via the EM Algorithm," *IEEE Trans. Commun.*, vol. 45, no. 3, March 1997, pp. 300-308.

[17] H. V. Poor, "On Parameter Estimation in DS/SSMA Formats," *Advances in Communications and Signal Processing*, ed. W. A. Porter and S. C. Kak. London: Springer-Verlag, 1988.

[18] R. Wang and S. D. Blostein, "Maximum Likelihood Multi-User CDMA Receiver for Base-Station Antenna Arrays using EM Algorithm," *19th Biennial Symposium on Communications*, Kingston, ON, 1998, pp. 110-114.

[19] Urs Fawer and Behnaam Aazhang, "A Multiuser Receiver for Code Division Multiple Access Communications over Multipath Channels," *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, February/March/April 1995, pp. 1556-1565.

[20] M. K. Varanasi and B. Aazhang, "Multistage Detection in Asynchronous Code Division Multiple-Access Communications," *IEEE Trans. Commun.*, vol. 38, no. 4, April 1990, pp. 509-519.

[21] A. Chockalingam and Gang Bao, "Probability of Miss Analysis for Packet CDMA Acquisition on Rician-Fading Channels," *IEEE Commun. Letters*, vol. 2, no. 7, July 1998, pp. 177-179.

[22] William C. Jakes, *Microwave Mobile Communications*, New York: John Wiley and Sons, 1974.

[23] Elvino S. Sousa, "Mobile Fading Channels," *ECE 1543 Course Notes*, University of Toronto, 1993.

[24] P. Patel and J. Holtzman, "Analysis of a Simple Successive Interference Cancellation Scheme in DS/CDMA System Using Correlations," *Proc. Globecom '93*, Houston, TX, 1993, pp.76-80.

[25] Xiaoyong Yu and S. Pasupathy, "Innovations-based MLSE for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, February/March/April 1995, pp. 1534-44.

[26] Xiaoyong Yu, *Innovations Based Maximum Likelihood Sequence Estimation for Rayleigh Fading Channels*, Ph. D. Thesis, University of Toronto, 1995.

[27] Athanasios Papoulis, *Probability, Random Variables, and Stochastic Processes, Third Edition*, New York: McGraw-Hill, Inc., 1991.

[28] Desmond P. Taylor, "The Estimate Feedback Equalizer: A Suboptimum Nonlinear Receiver," *IEEE Trans. Commun.*, vol. 21, no. 9, September 1973.

[29] Steven J. Nowlan and Geoffrey E. Hinton, "A Soft Decision-Directed LMS Algorithm for Blind Equalization," *IEEE Trans. Commun.*, vol. 41, no. 2, February 1993, pp. 275-279.

[30] Young-Hoon Kim and Sanyogita Shamsunder, "Adaptive Algorithms for Channel Equalization with Soft Decision Feedback," *IEEE J. Selected Areas in Commun.*, vol. 16, no. 9, December 1998, pp. 1660-1669.

[31] Joachim Hagenauer, Elke Offer, and Lutz Papke, "Iterative Decoding of Binary Block and Convolutional Codes," *IEEE Trans. Info. Theory*, vol. 42, no. 2, March 1996, pp. 429-445.

[32] Louis L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Reading: Addison-Wesley, 1991.

[33] John G. Proakis, *Digital Communications*, New York: McGraw-Hill, Inc., 1995.

[34] William C. Lindsey, "Error Probabilities for Rician Fading Multichannel Reception of Binary and N-ary Signals," *IEEE Trans. Info. Theory*, vol. 10, no. 5, October 1964, pp. 339-350.

[35] J. W. Craig, "A New, Simple, and Exact Result for Calculating the Probability of Error for Two-Dimensional Signal Constellations," *IEEE MILCOM '91 Conf. Record*, Boston, MA, 1991, pp. 25.5.1-25.5.5.

[36] Marvin K. Simon and Dariush Divsalar, "Some New Twists to Problems Involving the Gaussian Probability Integral," *IEEE Trans. Commun.*, vol. 46, no. 2, February 1998, pp. 200-210.

[37] Michael Moher, "An Iterative Multiuser Decoder for Near-Capacity Communications," *IEEE Trans. Commun.*, vol. 46, no. 7, July 1998, pp. 870-880.

[38] Gerhard Bauch and Volker Franz, "Iterative Equalization and Decoding for the GSM-System," *1998 IEEE Veh. Tech. Conf.*, Ottawa, ON, 1998, pp. 2262-2266.

[39] Annie Picart, Pierre Dider, and Alain Glavieux, "Turbo-Detection: A New Approach to Combat Channel Frequency Selectivity," *IEEE ICC '97 Proceedings*, Montreal, PQ, 1997, pp. 1498-1502.

[40] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon Limit Error Correcting Coding and Decoding: Turbo Codes," *IEEE ICC '93 Proceedings*, Geneva, 1993, pp. 1064-1070.

[41] G. David Forney, Jr. "The Viterbi Algorithm," *Proc. IEEE*, vol. 61, no. 3, March 1973, pp. 268-278.

[42] Edward A. Lee and David G. Messerschmitt, *Digital Communication, second edition*, Boston: Kluwer Academic Publishers, 1994.

[43] Jeong Hoon Ko, Jung Suk Joo, and Yong Hoon Lee, "On the Use of Sigmoid Functions for Multistage Detection in Asynchronous CDMA Systems," *IEEE Trans. Veh. Tech.*, vol. 48, no. 2, March 1999, pp. 522-526.

[44] J. Hagenauer and P. Hoeher, "A Viterbi algorithm with soft-decision outputs and its applications," *Proc. Globecom '89*, Dallas, TX, 1989, vol. 3, pp. 1680-1686.

[45] Hong Shen Wang and Pao-Chi Chang, "On Verifying the First-Order Markovian Assumption for a Rayleigh Fading Channel Model," *IEEE Trans. Veh. Tech.*, vol. 45, no. 2, May 1996, pp. 353-357.

[46] Rao, C. R. *Linear Statistical Inference and its Applications.* New York: Wiley, 1965.