Adaptive multicoding and robust linear-quadratic receivers for uncertain CDMA frequency-selective fading channels

by

Jianjun Ni

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Elecrtrical Engineering (Communications and Signal Processing)

Program of Study Committee
Jennifer Davidson, Major Professor
Richard Barton
Julie Dickerson
Manimaran Govindarasu
Dan Ashlock

Iowa State University

Ames, Iowa

2002

Graduate College
Iowa State University

This is to certify that the doctoral dissertation of

Jianjun Ni

has met the dissertation requirements of Iowa State University

Signature was redacted for privacy.

Committee Member

Signature was redacted for privacy.

Committee Member

Signature was redacted for privacy.

Committee Member

Signature was redacted for privacy.

Committee Member

Signature was redacted for privacy.

Major Professor

Signature was redacted for privacy.

For the Major Program

Table of Contents

List of Figures

Acknowledgements

The successful completion of any doctoral research effort usually requires the support and encouragement of family, friends and colleagues. It is my pleasure to thank the following individuals who have contributed to this dissertation.

First, I would like to thank Dr. Richard Barton for his support throughout this research effort. I own a debt of gratitude to him for his persistent help and guidance during my graduate study. I'm especially grateful that his rigorous academic spirit has imprinted in my life.

I am also grateful to Dr. Jennifer Davidson, Dr. Julie Dickerson, Dr. Manimaran Govindarasu and Dr. Dan Ashlock who served on my committee, and who have taught me much through their courses. Thanks to my colleagues and friends Leilei Lock, Xiangming Kong, Lili Chen, Fei Wan, Vidula Udas and Manoj Daniel for their valuable discussion.

I thank my parents and my sister, who have given so much yet asked for nothing in return. Their unconditional support throughout my life has instilled the confidence in me to pursue my dream.

Most of all, I thank my wife Bo Xu. Whenever I ran into difficulty, I knew where I could get the help and encouragement. She picked up the most tedious work to help me type those messy equations with great patience and carefulness. Her love made this dissertation complete.

Chapter 1. Introduction

Reliability, affordability, timeliness and location are four major concerns in many practical communication systems. Today, the development trends in telecommunications are driven by the high market demands for advanced wireless communications, which include access to a diverse range of services for anyone, anywhere, anytime and at the lowest possible cost. In late 1999, the International Telecommunications Union (ITU) approved five IMT-2000 (International Mobile Telecommunications for the 21st century) terrestrial radio interfaces which indicates the arrival of the third generation (3G) of wireless technology. The mobile communication systems are now migrating from the second generation (GSM, IS-54, IS-95, etc) to the IMT-2000 vision (i.e. UMTS). Wideband Code Division Multiple Access (WCDMA) is one of the major new third generation mobile communication systems being developed within the IMT-2000 framework. Such systems operate in low-power environments as well as time-varying, multiuser environments in which the exact structure of the channel is difficult to determine precisely. We refer to this environment generically as the *mobile wireless channel*, and it serves as a primary motivation for this research.

In this introductory chapter, we first describe in an intuitive way several of the principal types of interference that corrupt WCDMA channels. These include multiple-access interference (MAI), multipath propagation, intersymbol interference (ISI), narrowband interference, and additive wideband channel noise. In addition, the problem of channel uncertainty is particularly addressed. Once these problems are described, the dissertation objective and approach to solving these problems are then discussed.

1.1 Problem Description

In this investigation, our research effort is focused on the design and analysis of low-complexity, adaptive wireless receivers and signal constellations that exhibit good performance characteristics in the presence of multiple sources of complex structured interference as well as significant uncertainty regarding the exact structure of that interference. We refer to receivers that offer superior performance in the presence of uncertainty regarding the exact structure of the channel as *robust receivers*. We consider problems primarily related to the code-division-multiple-access (CDMA) environment. In particular, we study robust adaptive receivers for CDMA channels corrupted by the cumulative effects of multiple-access interference (MAI), multipath propagation, intersymbol interference (ISI), narrowband interference (NBI), and additive wideband channel noise. The work is motivated by the utility and popularity of CDMA as a means for implementation of IMT-2000/UMTS and by the ubiquitous nature of all of these sources of interference in such systems. In addition, because many tactical and commercial wireless networks operate in a mobile environment in which channel parameters vary rapidly over time, it is often difficult or impossible to estimate channel characteristics accurately. Hence, it is often necessary to implement a receiver in a heavily cluttered environment characterized by a high level of uncertainty.

Inherent to every communication system is a channel that links the transmitter and receiver. As we demonstrate later in this dissertation, we can describe the uncertain CDMA fading channel analytically with a simple discrete-time model:

$$r_m = \sum_{l=0}^{L} \hat{\alpha}_l s_{m-l} + \sum_{l=0}^{L} \varepsilon_l s_{m-l} + \gamma_m$$

where r_m represents the received signal for the user of interest (user 0), $\{s_m\}$ represents the sequence of transmitted symbols for user 0, the sequence $\{\hat{\alpha}_l\}_{l=0}^{L}$ represents *known* estimated values of the discrete-time channel impulse-response sequence for user 0, the sequence $\{\varepsilon_l\}_{l=0}^{L}$ represents the uncertainty (*unknown* errors) in the estimates of the impulse-response sequence, and γ_m represents an additive noise vector including the effects of MAI, NBI, and additive white Gaussian noise(AWGN). Hence, we now have a model in which the effects of MAI, NBI and wideband noise are modeled by the additive interference term $\{\gamma_m\}$; the ISI and multipath are modeled together by the known multiplicative interference terms $\{\hat{\alpha}_l\}_{l=0}^{L}$; and the channel uncertainty (including timing and waveform mismatches in the underlying chip-matched filter) is modeled by the separate random multiplicative interference terms $\{\varepsilon_l\}_{l=0}^{L}$.

The research work will concentrate on baseband discrete-time (i.e., digital) receivers. Furthermore, since we are primarily interested in low-complexity decision strategies that can be implemented in small, low-power receivers, we will restrict our discussion to one-shot, single-user receivers. That is, we assume that we are interested in demodulating only the symbol sequence transmitted by a single user and that symbols will be demodulated one at a time.

In contrast to previous studies, which have concentrated primarily on receiver design for a single type of interference under known channel assumptions, *this research effort seeks to identify and analyze receivers that are not only nearly optimal in the presence of multiple sources of interference but also robust in the presence of channel uncertainty.* The following problems are considered.

1. Receivers are designed based on the probabilistic channel model described above that explicitly incorporates multiple sources of additive interference as well as a stochastic structure for the channel uncertainty. We consider a simple, intuitively appealing cost

function that can be maximized to find linear-quadratic (LQ) detectors that are optimal in a certain sense. An adaptive algorithm is derived for LQ detectors that can be viewed as a generalization of the minimum-output-energy algorithm for the MMSE linear detector.

2. Adaptive multicoding – an approach to adaptive modulation for time-varying channels based on the proposed LQ cost function is also studied.

3. The Chernoff bound is derived for the performance analysis of LQ receivers.

4. The problem of binary signaling is considered first, and an extension from binary signals to M-ary signal constellations is developed in a multi-dimensional setting in order to achieve higher data rate transmission.

1.2 Dissertation Organization

The organization of this dissertation is as follows.

Historical Review Chapter 2 provides a brief literature review of the problems considered in this dissertation. Optimal receivers and suboptimal receivers and the trade-off between them are discussed. The uncertainty of fading channels has been recognized and addressed in several pertinent papers. Similarly adaptive modulation techniques have been studied previously for fading channels by several investigators.

Communication System Model Chapter 3 defines a generic CDMA communication system model for fading channels, and an equivalent discrete-time model that explicitly incorporates multiple sources of additive interference and channel uncertainty is rigorously developed. For simplicity of analysis, we restate the model using vector notation and make some necessary assumptions.

Adaptive LQ Receivers Chapter 4 motivates and derives the proposed LQ receivers. A simple, intuitively appealing cost function, the *modified deflection ratio,* is proposed. We

discuss the properties of the proposed LQ cost function and derive a related adaptive algorithm to find the optimal LQ receivers.

Adaptive Multicoding Chapter 5 first introduces the J-divergence as a relevant cost function for receiver and constellation design. Then a novel adaptive modulation scheme based on the J-divergence is developed to find the best signal constellation and receiver pair. Simulation work for binary signals is described in this chapter, and preliminary performance evaluation is also conducted here.

Performance Analysis of LQ Receivers Chapter 6 derives the Chernoff bound for LQ receivers to estimate the probability of bit error. To validate the simulations, the simulated BER of LQ receivers is compared with the Chernoff bound and other bounds.

M-ary Signal Constellations Chapter 7 extends the results of Chapter 4 and Chapter 5 to M-ary signal constellations. Simulation work is presented and results are analyzed. In addition, the information theory perspective on signal dimension expansion is also addressed in this chapter.

Summary and Conclusions Chapter 8 summarizes the primary results of our theoretical analysis and simulation work. We provide a synopsis of the unique contributions of this dissertation and discuss some future investigations based on the results presented here.

Chapter 2. Historical Review

Literature review in the relevant fields is critical for any sizeable research effort. By performing such a historical survey, we can understand the techniques employed by other investigators in similar fields as well as the applications driving similar research activities.

In this chapter, we provide a brief review of selected references in the major area of this research – interference suppression, channel uncertainty and adaptive modulation techniques.

2.1 Interference Suppression

Interference suppression techniques have been studied individually by many different investigators over decades. Various methods have been developed in order to solve the problems caused by MAI, ISI, NBI, and multipath propagation on CDMA channels.

MAI is perhaps the most extensively studied type of interference. The first optimum multiuser detector was developed by S. Verdu in 1983 [1]. After that, the analysis and derivation of optimum multiuser detectors was carried out in [2-4]. For example, the structure of the optimal maximum-likelihood sequence detector for an asynchronous Gaussian multiple-access CDMA channel was studied in [4]. Generally, the optimum receiver is defined as the receiver that selects the most probable sequence of bits given the received signal observed over the time interval. Among other virtues, as long as the set of signature waveform is linearly independent, the optimal detector solves the critical *near-far problem* for multi-user detection, in which the signals of distant or otherwise weaker users are overwhelmed by the signals of stronger users, even when the system is synchronized and

the signals of the users are nearly mutually orthogonal. Unfortunately, the optimal detectors not only requires knowledge of the waveforms of all users but also involves an order of computational complexity that is exponential in the number of users.

In fact, Verdu's optimum receiver uses the Viterbi algorithm for ML sequence estimation. As explained in [5], optimum multiuser demodulation is equivalent to a shortest-path problem in a layered directed graph. Therefore, a suboptimum version of the forward dynamic programming algorithm is adopted in practice, whereby each decision is based on the path corresponding to the cost-to-arrive function computed a fixed number of steps ahead. Thus, a significant reduction in computational complexity is obtained with respect to the block size parameter, but the exponential dependence on the number of users cannot be reduced. It is obvious that its application in practice is limited to communication systems where there only allow a small number of users.

Due to the complexity of the optimal multiuser detector, it is impractical in many real-time situations. As alternatives, suboptimal linear detectors for the same channel have been extensively studied by many investigators [6-9]. For example, linear multiuser detectors for synchronous CDMA channels are studied in [7]. Even though these linear detectors do exhibit higher probability of error than the optimal detector, they display the same *near–far resistance* as the optimal detector [10]. This implies that linear detectors still offer a solution to the near-far problem.

Among the class of linear multi-user detectors, perhaps the most interested one is the minimum-mean-squared-error (MMSE) detector [6, 11]. This detector is not only near-far resistant, but also can be implemented adaptively in a straightforward manner without recourse to training sequences [9, 11] in some cases. Furthermore, the structure of the MMSE detector depends only on the crosscovariance structure between the true bit sequence of the desired user and the received symbol sequence as well as the autocovariance structure of the

received symbol sequence. Therefore, no matter whatever complex structured interference is actually present, only the crosscovariance and autocovariance structure need to be adaptively estimated from the data to implement a detector that coherently compensates for that inference. On this perspective, the adaptive MMSE detector is itself a robust detector.

There are also several suboptimal nonlinear multi-user detectors that have been studied in last decade. Within this category, there are decision feedback detectors [13-15], Multistage detectors [16], successive interference cancellation [17], and detectors based on neural-network architectures [18-20]. These detectors are also near-far resistant and offer performance advantages over linear detectors in some situations. But usually they are more difficult to implement than linear detectors. However, they are still far less complex than the optimal multi-user detectors in many cases.

Similarly, the NBI suppression techniques on CDMA channels have been extensively studied. The traditional solution to this problem of NBI suppression is provided in [22-24]. First, it filters the incoming symbol sequence with a linear transversal filter. This filter estimates the narrowband process and subtracts the estimate from the received symbol sequence. Then the output from the filter is correlated with the spreading sequence of the user. At the end, compare the result to a threshold to make a bit decision. A refinement of this technique [25-27] uses a nonlinear transversal filter. The narrowband component is estimated by applying a linear transversal filter to the received signal minus a nonlinear estimate of the transmitted symbol sequence.

In general, the problems of multipath interference and ISI on CDMA channels are studied separately. That is, multipath interference generally causes both signal fading and ISI, but most research has focused on one or the other. In particular, a great deal of work has been done on designing multiuser receivers that exploit signal diversity to combat fading. Many different investigators have studied variations of the so-called RAKE receiver, which

exploits the frequency diversity present in CDMA signals [28-33]. In addition, array-processing techniques have been applied to the problem of multiuser detection on multipath CDMA channels to exploit the spatial diversity available at the receiver [34-38]. For channels with fast-fading characteristics such as mobile CDMA channels or spread-signature CDMA channels, some very interesting work has been done on employing time-frequency techniques to design receivers that exploit both time and frequency diversity [39-41].

With respect to ISI, it is well known that as long as the bit sequence of all users can be treated as mutually independent streams of independent and equally likely random variables, ISI can be modeled as equivalent to additional MAI. Hence, the optimal detector for ISI on a multiuser channel is fundamentally similar to the optimum multiuser detector without ISI. In the same way, linear multiuser detectors and nonlinear techniques such as decision feedback, multistage detection, and successive interference cancellation can be modified to mitigate both ISI and MAI simultaneously at higher complexity. However, in practical applications in which low complexity is crucial, suboptimum receivers remain an open problem. It is of interest to note that a Bayesian approach similar to the one proposed in this dissertation as a means of dealing with channel uncertainty has also been considered in the context of adaptive channel deconvolution for ISI channels [42, 43]. The success of Bayesian techniques for problems so closely related to the one of interest here is another indication of the potential benefits of this work.

2.2 Channel Uncertainty

In many of the aforementioned references, the analysis assumed exact knowledge of channel parameters. However, in many communication situations, the transmitter and the receiver must be designed without complete knowledge of the channel over which

transmission takes place. In mobile wireless communications, the time-variation of the channel caused by the varying location of the mobile transmitter and receiver with respect to scatterers leads to an uncertain channel.

Information-theoretic research efforts on uncertain channels have produced classes of models to describe many situations arising in mobile wireless communications. Many investigators have studied reliable communications under channel uncertainty from the information-theoretic perspective. For example, the extent to which dimension and geometry of the signal constellation can be exploited to mitigate the loss in capacity caused by channel uncertainty on wideband spread-spectrum fading channels has been studied in [44]. Universal decoding has been studied in [46] for Gaussian channels with a deterministic but unknown parameteric interference. In [47], the mismatch problem with minimum Euclidean distance decoding has been studied. For multiple-access channels, universal decoding has been studied in [48, 49], and mismatched decoding has been studied in [50,51].

In comparison to the amount of work that has been done on interference suppression, the issue of sensitivity of receiver performance to errors in the estimates of relevant channel parameters in multi-user systems has been studied relatively little. Problems associated with tracking errors, such as phase or timing mismatches, have been studied in [52-55] and sensitivity to estimates of fading parameters have been studied in [56-60]. For example, in [59], the variance of the channel measurement error at receiver impacts the channel capacity. Similarly, in [60], the effect of channel estimation error on M-QAM systems in Rayleigh fading channels is studied. More recently there has been some interest in the application of classical techniques from the field of robust statistics to the design and analysis of multiuser receivers for CDMA systems. For example, in [61], censoring of data in the frequency domain was used to reduce the sensitivity of detector performance to unknown narrowband interference. Also, in [62, 63], McKellips and Verdu study the characterization and impact of

uncertain noise distribution and worst-case additive noise distribution with respect to maximum probability of error under power and divergence constraints.

2.3 Adaptive Modulation Techniques

Communications systems are designed to deliver information as fast as possible, consume power as low as possible and cause error as few as possible. Typically, the goal of an adaptive modulation scheme on a communication channel is related to one of the following:

1. Minimize the transmitted power subject to constraints on the throughput and BER.

2. Minimize the BER subject to constraints on the throughput and transmitted power.

3. Maximize the throughput on the channel subject to constraints on the transmitted power and the BER.

Adaptive modulation techniques that address the trade-off of channel throughput, transmitted power and the bit error rate (BER) have been studied by many different investigators. Of particular interest here are techniques designed to combat the effects of channel fading, such as those studied in [64-70]. For example, in [70], the signal constellation is chosen from a discrete set of possibilities with fixed average power in order to maintain a constant BER with a high data rate (large constellation size) when the channel is favorable (shallow fade) and a low data rate (small constellation size) when the channel is unfavorable (deep fade). Similarly, in [65], both the constellation size and the transmitted power are varied to maintain constant BER and constant average transmitted power but maximize the data rate for any particular channel state. In [69], the throughput gain is achieved by combining adaptive modulation and power control for variable rate communications in multiuser environment.

To enhance the performance of the LQ receivers, we investigate adaptive modulation design in parallel with adaptive receiver design. In this dissertation, we use the term *adaptive modulation* to refer to techniques that alter the size or shape of the signal constellation in response to characteristics of the transmission environment. In particular, we adopt M-ary quadrature amplitude modulation (MQAM) as the basic signaling format and consider techniques to determine adaptively the number M of symbols in the constellation and the coordinates of each symbol with respect to a known set of basis signals. Although we propose it in this investigation to provide robustness against channel uncertainty, the notion of *multicoding* or assigning multiple spreading codes to each user has been proposed previously for use on multirate CDMA systems [71-74], where it provides a simple mechanism for accommodating users with different data rates. For example, in [73], the appropriate number of spreading codes is dynamically assigned to each user in order to meet its throughput requirement for multiuser multimedia services via mobile radio channel. Note that, in principle, our definition of adaptive modulation encompasses adaptive power control as well since the average transmitted power of the signaling scheme is one of the parameters determined by the coordinates of the signal constellation. Also note that we are primarily concerned with uncoded modulation schemes in this investigation.

Chapter 3. Communication System Model

Having heuristically described the problem and briefly reviewed the relevant literature, our focus now shifts to a rigorous formulation of the problem of interest.

In this chapter, we first introduce a typical CDMA communication system model in a mobile wireless channel. Next, an equivalent discrete-time model is defined and validated for our particular problem. Finally, in order to make the model tractable, we enumerate some reasonable research assumptions.

3.1 CDMA System Model

It is shown in many different communication texts (see, for example [28]) that a fading channel such as the mobile wireless channel can be modeled mathematically as a time-varying linear filter. Such a filter is characterized by a time-varying channel impulse response $c(\tau;t)$, where $c(\tau;t)$ is the response of the channel at time t due to an impulse applied at time $t-\tau$, where τ represents the time delay variable. Therefore, a mobile wireless channel with additive interference can be illustrated as in Figure 3.1.

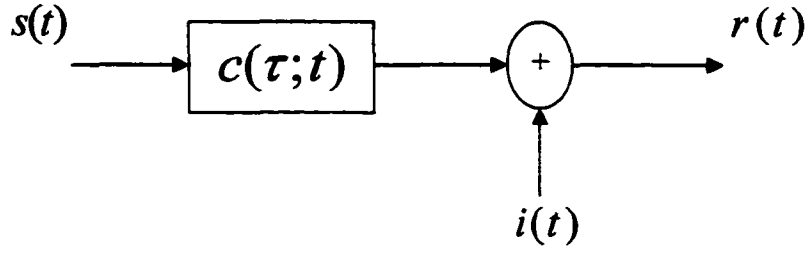

Figure 3.1 Mobile Wireless Channel with Additive Interference

For an input signal $s(t)$, the channel output signal is

$$\begin{aligned} r(t) &= s(t) * c(\tau;t) + i(t) \\ &= \int_{-\infty}^{\infty} c(\tau;t)s(t-\tau)d\tau + i(t) \end{aligned} \tag{3.1}$$

Where $i(t)$ represents all additive interference including MAI, NBI and AWGN.

If the channel is a multipath channel, the time-varying impulse response has a special form

$$c(\tau;t) = \sum_{l=1}^{L} \alpha_l(t)\delta(\tau - \tau_l) \tag{3.2}$$

Where the $\{\alpha_l(t)\}$ represents the attenuation factors for the L propagation paths and $\{\tau_l\}$ are the corresponding time delays. If (3.2) is substituted into (3.1), the received signal has the form

$$r(t) = \sum_{l=1}^{L} \alpha_l(t)s(t-\tau_l) + i(t) \tag{3.3}$$

Therefore, the received signal consists of L multipath components, where each component is attenuated by $\{\alpha_l(t)\}$ and delayed by $\{\tau_l\}$.

Due to bandwidth constraints, the mathematical model described above generally characterizes the mobile wireless channel. In order to design and analyze a robust receiver for this channel, in the next section, we will rigorously develop an equivalent discrete-time model that explicitly incorporates channel uncertainty.

3.2 Discrete-Time Model

The principal goal of this aspect of the research effort is the design and analysis of CDMA receivers that perform well in the presence of significant interference as well as substantial uncertainty regarding the structure of the channel. We propose a novel Bayesian approach to this problem in which receivers are designed based on a probabilistic channel model that explicitly incorporates a stochastic structure for the channel uncertainty. Although the ideas developed in this proposal can be extended quite easily to other types of CDMA modulation, particularly differentially encoded CDMA waveforms, we will confine our discussion to direct-sequence CDMA systems. We assume initially that information is transmitted using binary (but not necessarily antipodal) direct-sequence CDMA modulation. Since we will be concerned primarily with noncoherent systems, we adopt a complex-valued baseband model for the signal. That is, we assume that the transmitted baseband signal for the user of interest (user 0) takes the form

$$s(t) = \sqrt{2P} \sum_{k=-\infty}^{\infty} \sum_{i=0}^{N-1} \left[b_k c_{1,i} + (1-b_k) c_{0,i} \right] \psi(t - \tau - kT - iT_c) e^{i\theta} \quad , \tag{3.4}$$

where P is the average power of user 0, N is the length of the chip sequence, $T_c = T/N$ is the length of the chip interval, $b_k \in \{0,1\}$ represents the transmitted value of bit k for user 0, $\psi(t)$ is a chip waveform of duration T_c, τ is the relative delay for user 0, θ is the relative phase of user 0, and $c_{j,i}$ is the ith component of the user's chip sequence $\mathbf{c}_j$ corresponding to a transmitted bit value of $j \in \{0,1\}$. Without loss of generality, we assume that $\theta = 0$ and $\tau = 0$. We also assume that the chip sequences $\mathbf{c}_0$ and $\mathbf{c}_1$ are known to the receiver. Note that the choice of $\mathbf{c}_0 = -\mathbf{c}_1$ corresponds to an antipodal binary signaling structure, while choosing $\mathbf{c}_0$ and $\mathbf{c}_1$ to be orthogonal spreading sequences corresponds to an orthogonal signaling structure.

To highlight the most salient features of the proposed research, we will confine our discussion to a simple discrete-time model for the channel. This is equivalent, for example, to the assumption that the receiver demodulates the received baseband signal $r(t)$ into a sequence of symbols $\{r_m\}_{-\infty}^{\infty}$ using a noncoherent chip-matched filter. The received symbol vector corresponding to a single detection interval can then be modeled as [99]:

$$r_m = \sum_{k=0}^{K} \sum_{l=0}^{L} \alpha_k(l,m) s_{k,m-l} + \nu_m + \eta_m, \tag{3.5}$$

where $\{s_{k,m}\}$ represents the sequence of transmitted symbols for user k, where $\{\eta_m\}$ is a stationary, zero-mean broadband noise sequence; $\{\nu_m\}$ is a zero-mean, stationary narrowband process; K represents the number of interfering users; and the coefficients $\{\alpha_k(0,m), \alpha_k(1,m), ..., \alpha_k(L,m)\}$ represent the time-varying impulse-response sequence of the discrete-time channel at time m for the transmission of user k. Note that we have made the assumption that the number of interfering symbols for each user is bounded by L, which is assumed to be a fixed upper bound on the length of the discrete-time impulse response of the channel for all users.

To account for the uncertainty in the values of the channel parameters relevant to user 0, we rewrite the sequence $\{r_m\}$ as follows:

$$r_m = \sum_{l=0}^{L} \hat{\alpha}_{0,l} s_{0,m-l} + \sum_{l=0}^{L} \varepsilon_{0,l} s_{0,m-l} + \zeta_m + \nu_m + \eta_m, \tag{3.6}$$

where the sequence $\{\zeta_m\}$ represents the cumulative effects of all of the multiple-access interference, the sequence $\{\hat{\alpha}_{0,l}\}_{l=0}^{L}$ represents the estimated values of the discrete-time channel impulse-response sequence for user 0, and the sequence $\{\varepsilon_{0,l}\}_{l=0}^{L}$ represents the uncertainty (errors) in the estimates of the impulse-response sequence. Note that we have

suppressed the dependence on observation time m for the sequences $\{\hat{\alpha}_{0,l}\}_{l=0}^{L}$ and $\{\varepsilon_{0,l}\}_{l=0}^{L}$. This is equivalent to the assumption that the true structure of the channel is essentially constant over periods of time at least as long as a single detection interval, and the estimates of the channel impulse response are provided as side information updated no more frequently than once per detection interval.

Hence, we now have a model in which the effects of MAI, NBI, and broadband noise are modeled by the additive interference terms $\{\zeta_m\}$, $\{\nu_m\}$, and $\{\eta_m\}$, respectively; the ISI and multipath are modeled together by the known multiplicative interference terms $\{\hat{\alpha}_{0,l}\}_{l=0}^{L}$; and the channel uncertainty (including timing and waveform mismatches in the underlying chip-matched filter) is modeled by the separate random multiplicative interference terms $\{\varepsilon_{0,l}\}_{l=0}^{L}$. We will assume that the various random interference sequences in this model all have mean zero and are independent of each other as well as the data sequence.

Now, because we are interested in one-shot detectors, we consider an observation vector $\mathbf{x}(n)$ of length $N+L$ (for example) with components $x_i(n)=r_{nN+i}$, for $0 \le i \le N+L-1$. Reverting entirely to vector notation and dropping the explicit dependence on the bit interval n, we see that the observation vector $\mathbf{x}$ can be written in the form

$$\mathbf{x} = \mathbf{S}\hat{\boldsymbol{\alpha}} + \mathbf{S}\boldsymbol{\varepsilon} + \boldsymbol{\iota} + \boldsymbol{\zeta} + \boldsymbol{\nu} + \boldsymbol{\eta}, \tag{3.7}$$

where $\mathbf{S}$ is an $(N+L)\times(L+1)$ matrix, $\hat{\boldsymbol{\alpha}}$ and $\boldsymbol{\varepsilon}$ are vectors of length $L+1$, and $\boldsymbol{\iota}$, $\boldsymbol{\zeta}$, $\boldsymbol{\nu}$, and $\boldsymbol{\eta}$ are vectors of length $N+L$. Vectors $\hat{\boldsymbol{\alpha}}$, $\boldsymbol{\varepsilon}$, $\boldsymbol{\iota}$, $\boldsymbol{\zeta}$, $\boldsymbol{\nu}$, and $\boldsymbol{\eta}$ represent estimated channel impulse response, channel estimation errors, intersymbol interference (ISI), multiple access interference(MAI), narrowband interference(NBI) and additive wideband noise respectively. Notice that the distribution of the random vector $\mathbf{x}$ is influenced by the data bits from the user of interest only through the matrix $\mathbf{S}$. Furthermore, the matrix $\mathbf{S}$ is completely

determined by the known signature sequences $\mathbf{c}_0$, $\mathbf{c}_1$ and the current transmitted bit value for user 0. In particular, assuming only a single bit of ISI, the matrix $\mathbf{S}$ can be written as $\mathbf{S} = b_{-1}\mathbf{C}_{-1} + b_0\mathbf{C}_0 + b_{+1}\mathbf{C}_{+1}$, where b_{-1}, b_0, b_{+1} represent the previous bit, current bit, and next bit, respectively and $\mathbf{C}_{-1}, \mathbf{C}_0, \mathbf{C}_{+1}$ are given by

$$\mathbf{C}_{-1} = \begin{bmatrix} c_{0,N-1} & c_{0,N-2} & \cdots & c_{0,N-L} \\ 0 & c_{0,N-1} & \cdots & \vdots \\ 0 & 0 & \cdots & c_{0,N-2} \\ 0 & 0 & \cdots & c_{0,N-1} \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}, \quad \mathbf{C}_0 = \begin{bmatrix} c_{0,0} & 0 & \cdots & 0 \\ c_{0,1} & c_{0,0} & \cdots & \vdots \\ \vdots & c_{0,1} & \cdots & 0 \\ c_{0,N-1} & \vdots & \cdots & 0 \\ 0 & c_{0,N-1} & \cdots & c_{0,0} \\ 0 & 0 & \cdots & c_{0,1} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & c_{0,N-1} \end{bmatrix}, \quad \mathbf{C}_{+1} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 \\ c_{0,0} & 0 & \cdots & 0 \\ \vdots & c_{0,0} & \cdots & 0 \\ c_{0,L-2} & \vdots & \vdots & 0 \\ c_{0,L-1} & c_{0,L-2} & \cdots & c_{0,0} \end{bmatrix}.$$

To simplify the model a bit further, we can combine all of the additive interference into one vector $\boldsymbol{\gamma}$. That is, let

$$\boldsymbol{\gamma} = \boldsymbol{\zeta} + \boldsymbol{\nu} + \boldsymbol{\eta} \tag{3.8}$$

represent the cumulative effects of all of the additive interference – *i.e.*, multiple-access interference (MAI), narrowband interference (NBI), and additive white Gaussian noise (AWGN). Then (3.7) becomes

$$\mathbf{x} = \mathbf{S}\hat{\boldsymbol{\alpha}} + \mathbf{S}\boldsymbol{\varepsilon} + \boldsymbol{\iota} + \boldsymbol{\gamma} \tag{3.9}$$

To complete the channel model, we make the simplifying (and generally reasonable) assumptions that $\hat{\boldsymbol{\alpha}}$, $\boldsymbol{\varepsilon}$, $\boldsymbol{\iota}$, $\boldsymbol{\gamma}$ are mutually uncorrelated, wide-sense stationary random vectors, and that $\boldsymbol{\varepsilon}$ has mean zero.

3.3 Research Assumptions

In order to facilitate design and analysis of Bayesian receivers, we will consider the following set of nominal conditions and some reasonable assumptions:

1. The error vector $\boldsymbol{\varepsilon}$ can be treated as a zero-mean Gaussian random vector with covariance matrix Σ_{ε}. In many cases, if the channel impulse-response vector $\hat{\boldsymbol{\alpha}}$ is estimated adaptively, this is a standard assumption for the error vector $\boldsymbol{\varepsilon}$. As an alternative, $\boldsymbol{\varepsilon}$ can be regarded as a Rayleigh fading component, and $\hat{\boldsymbol{\alpha}}$ as the mean value of the multipath interference on a fading channel. This model for the channel estimation errors has been employed in prior work [56, 57, 75] using Kalman filters.

2. The vector $\boldsymbol{\eta}$ is a zero-mean additive white Gaussian noise (AWGN) vector with covariance matrix $\sigma^2\mathbf{I}$. This is also a standard assumption that is representative of a wide range of natural phenomena. The broadband background noise is often treated as AWGN.

3. The NBI vector $\boldsymbol{\nu}$ is a zero-mean Gaussian random vector with covariance matrix Σ_{ν}. This implies that the AWGN and the NBI can be combined into a single Gaussian random vector with mean zero and covariance matrix $\Sigma_{\gamma} + \sigma^2\mathbf{I}$. Even though NBI has some non-Gaussian behavior, in general, when attempting to design improved NBI suppression techniques [25, 27, 61], it is the non-Gaussian structure of the CDMA signals themselves rather than the non-Gaussian behavior of the NBI that is modeled.

4. The MAI is caused by a number of weaker users that can be accurately modeled as a zero-mean Gaussian process with covariance matrix Σ_{ζ}, together with a few dominant users with arbitrary power. In this scenario, the aggregate additive noise vector $\boldsymbol{\gamma}$ (MAI, NBI, and AWGN) is properly characterized as a uniform mixture of independent Gaussian distributions with different means and possibly different covariance matrices.

First, consider the case when there are several interfering users with independent but identical behavior (including multipath effects) at the receiver. In this situation, assuming

that the data bits for all users are mutually independent, the central limit theorem can be employed to imply that the MAI vector itself should be roughly Gaussian. Based on this implication, a Gaussian model for a component of the MAI is a reasonable nominal assumption and is frequently adopted [76-78]. Moreover, it is straightforward to show that on a CDMA channel with antipodal modulation where the MAI and AWGN are the only sources of interference, treating the MAI as Gaussian with the appropriate covariance matrix leads directly to an optimal detector that is equivalent to the conventional MMSE linear detector. Since the MMSE detector is known to be a robust detector in the presence of heavy MAI, it can be argued that the assumption of Gaussian MAI with the appropriate covariance structure is a good starting point for designing robust detectors. Similarly, it has been shown in [21] in most cases that the maximum divergence between an appropriate Gaussian distribution and the MAI-plus-noise component at the output of the linear stage of an MMSE detector is quite small.

On the other hand, if there are only a few interfering users on a channel with fixed multipath structure, the distribution of the sum of MAI, NBI, and AWGN, given any particular realization of data bits for the interfering users, is conditionally Gaussian. If we assume again that the data bits for all users are mutually independent, the distribution of the aggregate noise will then be a uniform mixture of Gaussian distributions with different means. As a matter of fact, ignoring all interference except MAI and AWGN, and accounting for all users in this manner leads to the true optimal one-shot detector [20]; however, this detector is again exponentially complex in the number of users.

Combining these two arguments, we can conclude that the additive noise component on a slowly fading channel where the MAI is generated by a relatively large number of homogeneous weak interfering users together with a few dominant interfering users, can be modeled as a uniform mixture of Gaussian distributions with different means but the same

covariance matrix. In a similar way, a uniform mixture of Gaussian distributions with different means and different covariance matrices turns to be the model for fast-fading channels. This latter case is equivalent to modeling the channel impulse response for each strong interfering user as an estimated component plus a zero-mean error component, just as we did explicitly for user 0. Hence, using a model that allows different covariance matrices for each component of the mixture distribution corresponding to the additive noise is equivalent to explicitly modeling the uncertainty in the multipath characteristics for each of the dominant interfering users.

Chapter 4 Adaptive LQ Receivers

The discrete-time model defined in Chapter 3 sets the foundation of the communication architecture under investigation. The structure of this channel model incorporates explicitly all significant types of interference as well as substantial channel uncertainty. In addition, some reasonable assumptions have been made to facilitate further design and analysis.

In this chapter, we design robust receivers based on the channel model defined in Chapter 3. First, the statistics for a binary hypothesis testing problem are derived and discussed. Then, adaptive LQ receivers are proposed in order to exploit both the known CSI and the structure of the channel uncertainty. Next, we proposed a simple, intuitively appealing cost function that can be maximized to find LQ receivers that are optimal in a certain sense. At the end of this chapter, we discuss the properties of the proposed LQ cost function and derive a related adaptive algorithm.

4.1 LQ Receivers

Our task in this investigation is to design the robust one-shot, single-user detector based on the channel model developed in Chapter 3. This kind of detector can be implemented as a generalized radial-basis-function (RBF) neural network with Gaussian processing nodes. With only moderate increase in complexity, such detectors can be expected to offer increased robustness and performance improvements comparing to linear detectors. In [19], Mitra and Poor have studied adaptive multi-user detectors of this type that demonstrated some robustness properties. These detectors will be less complex to implement

than the true optimal multiuser or single-user detectors. Such detectors for use on ISI channels have also been studied by Barton, Shaw and Reichart in [80]. They investigated the sensitivity of different processing architectures to mismatches in detector complexity. As results, they found that if implementation considerations require that detector complexity be severely constrained or the required complexity of the optimal detector is severely underestimated, superior performance could be achieved by a nonoptimal detector architecture. On the other hand, as complexity constraints are relaxed or optimal detector complexity is more accurately estimated, the performance of the suboptimal architecture improves very little but the performance of the optimal architecture improves rapidly. If we recall the essential equivalence between ISI and MAI, we see that these results have implications for multi-user receivers as well.

One approach to reducing the complexity of these receivers while still retaining some of the performance advantages is to constrain the number of processing nodes in the detector as much as possible. In the extreme, this results in a detector with only two Gaussian processing nodes. This is equivalent to assuming that the aggregate additive interference is Gaussian and corresponds to the following simple binary hypothesis testing problem [100]:

$$\begin{aligned} & H_0 : \mathbf{x} \sim \mathbb{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0), \\ \text{Versus} \quad & \\ & H_1 : \mathbf{x} \sim \mathbb{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1). \end{aligned} \tag{a}$$

where

$$\begin{aligned} \boldsymbol{\mu}_0 &= \mathbf{C}_0 \hat{\boldsymbol{\alpha}} \\ \boldsymbol{\mu}_1 &= \mathbf{C}_1 \hat{\boldsymbol{\alpha}} \\ \boldsymbol{\Sigma}_0 &= \mathbf{C}_0 \boldsymbol{\Sigma}_\varepsilon \mathbf{C}_0^* + \boldsymbol{\Sigma}_\iota + \boldsymbol{\Sigma}_\gamma \\ \boldsymbol{\Sigma}_1 &= \mathbf{C}_1 \boldsymbol{\Sigma}_\varepsilon \mathbf{C}_1^* + \boldsymbol{\Sigma}_\iota + \boldsymbol{\Sigma}_\gamma \end{aligned}$$

It is straightforward to show that the difference in the covariance matrices under the two hypotheses results entirely from the inclusion of the modeled estimation error term $\boldsymbol{\varepsilon}$. That is, in the absence of any modeled error term, Problem (a) reduces to

$$\begin{aligned} & H_0 : \mathbf{x} \sim \mathbb{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}), \\ \text{versus} \quad & \\ & H_1 : \mathbf{x} \sim \mathbb{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}). \end{aligned} \tag{b}$$

Furthermore, if we assume antipodal CDMA modulation ($\mathbf{C}_0 = -\mathbf{C}_1$), then problem (b) reduces to

$$\begin{aligned} & H_0 : \mathbf{x} \sim \mathbb{N}(-\boldsymbol{\mu}, \boldsymbol{\Sigma}), \\ \text{versus} \quad & \\ & H_1 : \mathbf{x} \sim \mathbb{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}). \end{aligned} \tag{b'}$$

where

$$\boldsymbol{\mu} = \mathbf{C}_1 \hat{\boldsymbol{\alpha}}$$

$$\boldsymbol{\Sigma} = \mathbf{C}_0 \boldsymbol{\Sigma}_\varepsilon \mathbf{C}_0^* + \boldsymbol{\Sigma}_\iota + \boldsymbol{\Sigma}_\gamma$$

It is well known that the optimal detector for Problem (b') is equivalent to the MMSE linear receiver for the single-user detection problem on antipodally modulated, frequency-selective CDMA fading channels. On such channels, the MMSE detector is known to be a robust suboptimal solution to the detection problem that can be implemented in an adaptive fashion provided that the channel is slowly time varying. This suggests that it is worthwhile to investigate optimal solutions to Problem (a) as robust adaptive suboptimal detectors for rapidly time-varying channels or other situations in which good estimates of the channel parameters are not available.

Towards this end, we note that if the channel is stationary and ergodic (at least over reasonably long time intervals), Problem (a) can be readily transformed in a blind adaptive fashion into the slightly more tractable form

$$\begin{aligned} & H_0 : \mathbf{y} \sim \mathbb{N}(-\boldsymbol{\mu}, \boldsymbol{\Sigma}_0), \\ \text{versus} \quad & \\ & H_1 : \mathbf{y} \sim \mathbb{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_1). \end{aligned} \tag{a'}$$

where $\boldsymbol{\mu}$ is known. Since the observation vector $\mathbf{x}$ and the estimated CSI vector $\hat{\boldsymbol{\alpha}}$ are both known to the receiver, it is a simple matter to compute estimates $\boldsymbol{\mu}_{\mathbf{x}}$ and $\boldsymbol{\mu}_{\hat{\alpha}}$ of the corresponding mean vectors in an adaptive fashion. Assuming these estimates are unbiased, we can transform the observation vector $\mathbf{x}$ into a zero-mean observation vector $\mathbf{y}$ using the transformation

$$\mathbf{y} = \mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}} - \tfrac{1}{2}(\mathbf{C}_0 + \mathbf{C}_1)(\hat{\boldsymbol{\alpha}} - \boldsymbol{\mu}_{\hat{\alpha}}), \tag{4.1}$$

where the $(N+L)\times(L+1)$ matrices $\mathbf{C}_0$ and $\mathbf{C}_1$ are defined straightforwardly in terms of the known spreading sequences $\mathbf{c}_0$ and $\mathbf{c}_1$. The new observation vector then takes the form $\mathbf{y} = \boldsymbol{\mu} b + \boldsymbol{\xi}$, where $b \in \{-1, +1\}$ now represents the transmitted bit value, $\boldsymbol{\mu} = \tfrac{1}{2}(\mathbf{C}_1 - \mathbf{C}_0)\hat{\boldsymbol{\alpha}}$ represents a constant transmitted baseband "signal", and

$$\boldsymbol{\xi} = \tfrac{1}{2} b(\mathbf{C}_1 - \mathbf{C}_0)\boldsymbol{\varepsilon} + \tfrac{1}{2}(\mathbf{C}_1 + \mathbf{C}_0)\boldsymbol{\varepsilon} + \hat{\boldsymbol{\imath}} + \hat{\boldsymbol{\gamma}}, \tag{4.2}$$

represents zero-mean additive channel noise. Note, that the ISI and additive interference vectors $\boldsymbol{\imath}$ and $\boldsymbol{\gamma}$ have been replaced with their zero-mean counterparts $\tilde{\boldsymbol{\imath}}$ and $\tilde{\boldsymbol{\gamma}}$. Conceptually then, we have transformed the original detection problem into a more conventional antipodal binary detection problem, which is convenient for expository

purposes. Furthermore, given the transmitted bit value b and the estimated CSI vector $\hat{\boldsymbol{\alpha}}$, the observation vector $\mathbf{y}$ has conditional mean vector

$$\boldsymbol{\mu}_{\mathbf{y}|b,\hat{\boldsymbol{\alpha}}} = \boldsymbol{\mu} b = \tfrac{1}{2} b(\mathbf{C}_1 - \mathbf{C}_0)\hat{\boldsymbol{\alpha}}, \tag{4.3}$$

and conditional covariance matrix

$$\boldsymbol{\Sigma}_{\mathbf{y}|b,\hat{\boldsymbol{\alpha}}} = \begin{cases} \boldsymbol{\Sigma}_0 = \mathbf{C}_0 \boldsymbol{\Sigma}_\varepsilon \mathbf{C}_0^* + \boldsymbol{\Sigma}_\iota + \boldsymbol{\Sigma}_\gamma, & b = -1, \\ \boldsymbol{\Sigma}_1 = \mathbf{C}_1 \boldsymbol{\Sigma}_\varepsilon \mathbf{C}_1^* + \boldsymbol{\Sigma}_\iota + \boldsymbol{\Sigma}_\gamma, & b = +1, \end{cases} \tag{4.4}$$

where $\boldsymbol{\Sigma}_\varepsilon$ represents the covariance matrix of the channel uncertainty vector $\boldsymbol{\varepsilon}$, $\boldsymbol{\Sigma}_\gamma$ represents the covariance matrix of the additive interference vector $\boldsymbol{\gamma}$, and $\boldsymbol{\Sigma}_\iota$ represents the covariance matrix of the ISI vector $\boldsymbol{\iota}$. Note that the matrix $\boldsymbol{\Sigma}_\iota$ also has a known form that is completely determined by $\mathbf{c}_0$, $\mathbf{c}_1$, $\boldsymbol{\Sigma}_\varepsilon$, and $\boldsymbol{\Sigma}_{\hat{\alpha}}$, the covariance matrix of the observed estimated CSI vector $\hat{\boldsymbol{\alpha}}$. (see Appendix C).

Equations (4.3) and (4.4) reveal several interesting properties related to the second-order statistics of the observation vector $\mathbf{y}$. First, in the absence of instantaneous CSI estimates ($\hat{\boldsymbol{\alpha}} = \mathbf{0}$, *i.e.*, completely noncoherent detection), the conditional mean vectors under the binary hypotheses $H_0 : b = -1$ and $H_1 : b = +1$ are identical. Hence, the only second-order information that can be exploited by the receiver to discriminate between the two hypotheses is the difference in the conditional covariance matrices. On the other hand, if some CSI is available ($\hat{\boldsymbol{\alpha}} \neq \mathbf{0}$, *i.e.*, partially or completely coherent detection), then the conditional mean vectors will differ as long as the signal structure under the two hypotheses is not identical ($\mathbf{c}_1 \neq \mathbf{c}_0$). The availability of CSI can obviously be exploited at the receiver by including a linear component in the detector, and the processing gain associated with the linear detector component will be maximized if the signals are antipodal ($\mathbf{c}_1 = -\mathbf{c}_0$).

Similarly, the conditional covariance matrices of $\mathbf{y}$ under the two hypotheses H_0 and H_1 differ only due to the influence of the terms $\mathbf{C}_j\boldsymbol{\Sigma}_\varepsilon\mathbf{C}_j^*$, $j=0,1$. This implies that if the CSI is known precisely ($\boldsymbol{\varepsilon}=\mathbf{0}$, *i.e.*, completely coherent detection), then the covariance matrices under the two hypotheses will be identical and the only second-order information that can be exploited by the receiver to discriminate between the two hypotheses is the difference in the mean vectors. However, if the CSI is not known precisely ($\boldsymbol{\varepsilon}\neq\mathbf{0}$, *i.e.*, partially coherent or noncoherent detection) the difference in the covariance matrices under the two hypotheses will depend on the interaction between the signal structure ($\mathbf{c}_0$ and $\mathbf{c}_1$) and the covariance structure of the channel uncertainty vector $\boldsymbol{\varepsilon}$. Hence, the structure of the channel uncertainty can actually be exploited by the receiver by including a quadratic component in the detector, but only if the signal structure is chosen appropriately. In particular, if an antipodal signal structure is chosen, the structure of the channel uncertainty cannot be exploited by the receiver (at least based on second-order statistics) and will serve only to degrade system performance.

In order to exploit both the available CSI and the structure of the channel uncertainty in the receiver, we propose to develop adaptive LQ receivers in which the detector takes the form

$$\delta(\mathbf{y})=\begin{cases}1, & T(\mathbf{y})\geq\tau,\\ 0, & T(\mathbf{y})<\tau,\end{cases}\quad T(\mathbf{y})=\mathbf{y}^*\boldsymbol{\Phi}\mathbf{y}+2\,\mathrm{Re}\left(\mathbf{h}^*\mathbf{y}\right), \tag{4.5}$$

where $\boldsymbol{\Phi}$ is a Hermitian (conjugate symmetric) matrix that determines the quadratic component of the detector, $\mathbf{h}$ is a complex-valued vector that determines the linear component of the detector, and τ is a real-valued detection threshold.

4.2 Adaptive Algorithm Based on Modified Deflection Ratio

The first question that must be answered in order to implement such a detector is how to choose the components $(\mathbf{\Phi}, \mathbf{h}, \tau)$. One possible approach to this problem is to choose $(\mathbf{\Phi}, \mathbf{h}, \tau)$ to minimize the quantity

$$E\left\{\left[b-\left(T(\mathbf{y})-\tau\right)\right]^2\right\}. \tag{4.6}$$

This leads to a minimum-mean-squared-error (MMSE) LQ receiver. Note that if an antipodal signal structure is employed, the resulting detector will be equivalent to the more familiar MMSE linear detector [6, 9]. Unfortunately, in general, identifying the MMSE LQ detector requires knowledge of the third- and fourth-order moment structure under both hypotheses in addition to the first- and second-order information that is required to identify the linear MMSE detector. Hence, this approach to LQ receiver design is somewhat impractical, particularly if the statistics of the problem are time-varying and must be acquired adaptively. As an alternative to the MMSE approach, we consider the simple expedient of identifying the optimal (necessarily LQ) detector corresponding to the hypotheses H_0 and H_1 under the additional assumption that the additive noise vector γ is Gaussian. While this is clearly not an accurate assumption in most cases, it is nevertheless an assumption that often leads to suboptimal receivers with excellent performance characteristics. In particular, if an antipodal signal structure is employed, this approach will again lead to a detector that is equivalent to the MMSE linear detector. Since the MMSE receiver is known to be a robust receiver in the presence of heavy MAI, it can be argued that the assumption of Gaussian additive interference is a good starting point for designing robust detectors even on multiple-access channels. Similarly, it has been shown in [21] that the maximum divergence between an appropriate Gaussian distribution and the MAI-plus-noise component at the output of the

linear stage of an MMSE detector is quite small in most cases. Assuming the matrices Σ_0 and Σ_1 are invertible, this approach leads to an LQ detector with components given by

$$\begin{aligned} \mathbf{h} &= \left(\Sigma_0^{-1} + \Sigma_1^{-1}\right)\boldsymbol{\mu}, \\ \Phi &= \Sigma_0^{-1} - \Sigma_1^{-1}, \\ \tau &= \ln\left(|\Sigma_1|/|\Sigma_0|\right) - \boldsymbol{\mu}^* \Phi \boldsymbol{\mu}. \end{aligned} \tag{4.7}$$

While Equations (4.7) provide a possible solution to the problem of identifying the components of the LQ receiver, it is not a completely satisfying solution for two reasons. First, this solution requires the inversion of the matrices Σ_0 and Σ_1. Since the system is time-varying, the matrices Σ_0 and Σ_1 must be tracked adaptively, and the inverses must be recomputed accordingly. Since matrix inversion is a computationally intensive operation, frequent recomputation of these inverses is undesirable. While it is possible to track both the matrices and their inverses directly using subspace decomposition techniques (as discussed, for example in [81,82]), it is of interest to identify adaptive solutions for $(\Phi, \mathbf{h}, \tau)$ that do not require inversion of Σ_0 and Σ_1. Second, the solution given by Equations (4.7) requires prior knowledge of the signal structure for the problem and does not provide any insight or methodology for choosing a signal structure adaptively in order to optimize receiver performance.

To address both of these deficiencies of solution in (4.7), we consider an alternative approach to adaptive LQ receiver design, which allows us to identify simultaneously both a signal structure $(\mathbf{c}_0, \mathbf{c}_1)$ and a detector structure $(\Phi, \mathbf{h}, \tau)$ that are jointly optimal in a certain sense. Toward this end, we assume for the moment that the matrices Σ_ε, and Σ_γ are known to the receiver, along with the current estimated CSI vector $\hat{\boldsymbol{\alpha}}$ (from which we can also estimate $\Sigma_{\hat{\alpha}}$). In this case, both the mean vector μ and the two covariance matrices Σ_0 and

$\boldsymbol{\Sigma}_1$ are explicit functions of $\mathbf{c}_0$ and $\mathbf{c}_1$ as given in Equations (4.3) and (4.4) above. To simultaneously adapt both the signal structure and the detector structure, we seek a solution $\left(\tilde{\mathbf{c}}_0, \tilde{\mathbf{c}}_1, \tilde{\mathbf{h}}, \tilde{\boldsymbol{\Phi}}\right)$ to the constrained maximization problem

$$\max_{(\mathbf{c}_0, \mathbf{c}_1, \mathbf{h}, \boldsymbol{\Phi})} D\left(\mathbf{c}_0, \mathbf{c}_1, \mathbf{h}, \boldsymbol{\Phi}\right) \text{ subject to: } \left\|\mathbf{c}_0\right\|^2 = \left\|\mathbf{c}_1\right\|^2 \leq E_b, \tag{4.8}$$

where $\left\|\mathbf{c}_j\right\|^2 = \sum_{i=0}^{N-1} c_{ji}^2$, E_b represents the energy transmitted per bit, and

$$D\left(\mathbf{c}_0, \mathbf{c}_1, \mathbf{h}, \boldsymbol{\Phi}\right) = \frac{\left|\boldsymbol{\mu}^* \left(\mathbf{h}_0 + \mathbf{h}_1\right) + \frac{1}{4} \mathrm{Tr}\left[\left(\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_0\right)\boldsymbol{\Phi}\right]\right|^2}{\mathbf{h}_0^* \boldsymbol{\Sigma}_0 \mathbf{h}_0 + \mathbf{h}_1^* \boldsymbol{\Sigma}_1 \mathbf{h}_1 + \frac{1}{4} \mathrm{Tr}\left[\boldsymbol{\Phi}\boldsymbol{\Sigma}_0 \boldsymbol{\Phi}\boldsymbol{\Sigma}_1\right]}, \quad \mathbf{h} = \mathbf{h}_0 + \mathbf{h}_1. \tag{4.9}$$

The optimal detector structure is completed by defining the threshold $\tilde{\tau}$ as before; that is,

$$\tilde{\tau} = \ln\left(\left|\tilde{\boldsymbol{\Sigma}}_1\right| / \left|\tilde{\boldsymbol{\Sigma}}_0\right|\right) - \tilde{\boldsymbol{\mu}}^* \tilde{\boldsymbol{\Phi}} \tilde{\boldsymbol{\mu}}. \tag{4.10}$$

Perhaps the most intriguing questions regarding this approach to receiver design are the proper choices for the cost function and the uncertainty class. To keep the problem tractable, it would be desirable to choose a cost function and an uncertainty class in such a way that the resulting problem could be decomposed into a set of tractable, independent problems for each of the possible components of the hypothesized mixture distribution. On the other hand, it is also desirable that the uncertainty class reflects the actual structure of the uncertainty in the problem and that the cost function is indicative of the relative probability of error between two candidate distributions. Unfortunately, these two goals are generally in conflict, and we will be forced to settle for a suitable compromise. We refer to the cost function $D\left(\mathbf{c}_0, \mathbf{c}_1, \mathbf{h}, \boldsymbol{\Phi}\right)$ as the *modified deflection ratio*. A more restricted version of this cost function has been studied previously in [83], where it was shown to have some desirable properties. For the

particular problem of detecting a zero-mean Gaussian signal in the presence of zero-mean Gaussian noise, this cost function is equivalent to the modified deflection ratio discussed in [84, 85].

To see that this solution is indeed optimal in some sense, we note that, for a fixed signal structure, the maximum value of the modified deflection ratio is given by

$$D^{*}(\mathbf{c}_0,\mathbf{c}_1)=\max_{(\mathbf{h}_0,\mathbf{h}_1,\boldsymbol{\Phi})} D(\mathbf{c}_0,\mathbf{c}_1,\mathbf{h},\boldsymbol{\Phi})=\boldsymbol{\mu}^{*}\left(\boldsymbol{\Sigma}_0^{-1}+\boldsymbol{\Sigma}_1^{-1}\right)\boldsymbol{\mu}+\tfrac{1}{4}\mathrm{Tr}\left(\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}_1^{-1}+\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_0^{-1}-2\mathbf{I}\right), \tag{4.11}$$

which is attained (except for an arbitrary scaling factor) if and only if $\boldsymbol{\Phi}=\boldsymbol{\Sigma}_0^{-1}-\boldsymbol{\Sigma}_1^{-1}$, $\mathbf{h}_0=\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}$, and $\mathbf{h}_1=\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}$. Hence, for a fixed signal structure, maximizing the modified deflection ratio leads to a detector structure that is equivalent to the optimal Gaussian detector.

This approach also leads to efficient adaptive algorithms. To see this, we consider first the problem of maximizing $D(\mathbf{c}_0,\mathbf{c}_1,\mathbf{h},\boldsymbol{\Phi})$ for a fixed signal pair $(\mathbf{c}_0,\mathbf{c}_1)$. While an adaptive algorithm could be derived by attempting to maximize the modified deflection ratio directly over all possible sets $(\mathbf{h}_0,\mathbf{h}_1,\boldsymbol{\Phi})$, an easier and more stable approach is based on the fact that the set $\left(\hat{\mathbf{h}}_0,\hat{\mathbf{h}}_1,\hat{\boldsymbol{\Phi}}\right)$ maximizes the modified deflection ratio if and only if (modulo a scaling factor)

$$\hat{\mathbf{h}}_0=\frac{\tilde{\mathbf{h}}_0}{\tilde{\mathbf{h}}_0^{*}\boldsymbol{\Sigma}_0\tilde{\mathbf{h}}_0},\quad \hat{\mathbf{h}}_1=\frac{\tilde{\mathbf{h}}_1}{\tilde{\mathbf{h}}_1^{*}\boldsymbol{\Sigma}_1\tilde{\mathbf{h}}_1},\quad \hat{\boldsymbol{\Phi}}=\frac{\tilde{\boldsymbol{\Phi}}}{\mathrm{Tr}\left[\tilde{\boldsymbol{\Phi}}\boldsymbol{\Sigma}_0\tilde{\boldsymbol{\Phi}}\boldsymbol{\Sigma}_1\right]},$$

where $\left(\tilde{\mathbf{h}}_0,\tilde{\mathbf{h}}_1,\tilde{\boldsymbol{\Phi}}\right)$ is chosen to solve the dual minimization problem:

$$\begin{aligned}
&\min_{\mathbf{h}_0} \mathbf{h}_0^{*}\boldsymbol{\Sigma}_0\mathbf{h}_0 \text{ subject to } \boldsymbol{\mu}^{*}\mathbf{h}_0 = 1, \\
&\min_{\mathbf{h}_1} \mathbf{h}_1^{*}\boldsymbol{\Sigma}_1\mathbf{h}_1 \text{ subject to } \boldsymbol{\mu}^{*}\mathbf{h}_1 = 1, \\
&\min_{\boldsymbol{\Phi}} \mathrm{Tr}\left[\boldsymbol{\Phi}\boldsymbol{\Sigma}_0\boldsymbol{\Phi}\boldsymbol{\Sigma}_1\right] \text{ subject to } \mathrm{Tr}\left[\left(\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_0\right)\boldsymbol{\Phi}\right] = 1.
\end{aligned} \tag{4.12}$$

An adaptive solution to this set of constrained minimization problems, which does not involve matrix inversion, can be derived straightforwardly using either the least-mean-square (LMS) algorithm or the recursive least squares (RLS) algorithm. The LMS solution, for example, leads to an adaptive algorithm analogous to the well-known minimum-output-energy (MOE) implementation of the linear MMSE detector proposed in [11].

Chapter 5 Adaptive Multicoding

Adaptive LQ receivers has been developed in Chapter 4 in order to exploit both the known CSI and the structure of the channel uncertainty. A related adaptive algorithm has also been derived based on the properties of the proposed cost function the modified deflection ratio.

In this chapter, we design the adaptive modulation scheme to enhance the performance of the proposed LQ receivers. First, we discuss some important distance criteria for signal design. Second, the adaptive multicoding technique is introduced to maximize the J-divergence for searching the optimal signal constellation. A gradient search algorithm is also developed for system simulation. At the end, we conduct some simulation work to evaluate the robustness of LQ receivers and the efficacy of adaptive multicoding.

5.1 Distance Criteria

In the area of signal design for communication systems, the optimal signals are defined as those that minimize the probability of error. However, the optimization of a statistical distance measure between competing hypotheses has become an alternative approach to signal design for two reasons. First, in many cases, direct minimization of the probability of error in order to determine an optimum signal set is not possible. This may be because an explicit analytical expression for the error probability is too difficult to find, or even if it can be found, the expression may be too complicated for analytical or numerical minimization. Therefore, it is useful to search for signal selection criteria that may be weaker than error probability but are easier to evaluate and manipulate. Second, the classical design

strategies (such as Bayes, minimax, and Neyman Pearson) for optimal signal detection and other decision problems require a complete statistical description of the data in order to specify the optimum decision rule structure. However, it has been demonstrated that procedures designed around a particular model may perform poorly when actual data statistics differ from those assumed. Thus, since there is frequently some uncertainty concerning the statistical structure of the data, it is of interest to find decision procedures that are robust, that is, which perform well despite small variations from the assumed statistical model. It is shown in [86] for the general case that robustness in terms of risk implies robustness in terms of distance, a fact which, together with the added tractability of the distance measure, enhances the desirability of using the later criterion.

In the search for suitable criteria, we often consider a simple binary hypothesis-testing problem in which we assume that there are two possible hypotheses, H_0 and H_1, corresponding to two possible probability distributions P_0 and P_1, respectively. We may write this problem as

$$\begin{aligned} & H_0 : \mathbf{x} \sim P_0 \\ \text{versus} & \\ & H_1 : \mathbf{x} \sim P_1 \end{aligned} \tag{5.1}$$

where the notation "$\mathbf{x} \sim P$" denotes the condition "$\mathbf{x}$ has distribution P". It is well known that the optimal detector in both a Bayesian and Neyman-Pearson sense is the likelihood ratio test

$$L(\mathbf{x}) = \frac{p_1(\mathbf{x})}{p_0(\mathbf{x})} \underset{H_0}{\overset{H_1}{\gtrless}} \tau \tag{5.2}$$

where $p_0(\mathbf{x})$ and $p_1(\mathbf{x})$ are probability density functions for the two hypotheses, and τ is some threshold. Note that the error probability depends on the total distribution of the

likelihood ratio. The likelihood ratio $L(\mathbf{x})$ will play a prominent part in the following discussions.

The notion of a *distance* between two probability distributions is quite useful. The further apart we can make these distributions, hopefully the smaller will be the probability of mistaking one for the other. Therefore, various distance measures have been studied as simple substitutes for the error probability. Among them, there is a general class of measures of discrimination between probability measures p_0 and p_1 known as f-divergence [87] or Ai-Silvey distances [88]. Mathematically, these distance measures are given by

$$d(p_0, p_1) = h\{E_0[C(L(\mathbf{x}))]\} \tag{5.3}$$

where E_0 indicates that the expectation is taken with respect to p_0 and where $C(.)$ is a continuous, convex real function and $h(.)$ is an increasing real function of a real variable. Many well-known measures of discrimination including Kullback-Leibler (KL) distance, J-divergence, Bhattacharyya distance and Kolmogorov variational distance, as well as other commonly used measures are members of this class. In the following, we will briefly discuss KL distance, J-divergence, and Bhattacharyya distance. Then we will explain why we choose J-divergence as our distance criterion for adaptive multicoding.

KL Distance Kullback Leiber (KL) distance (also called relative entropy in [89]) between two probability density functions $p_0(\mathbf{x})$ and $p_1(\mathbf{x})$ is defined in [90] as

$$\begin{aligned} KL(p_0, p_1) &= E_1\{\ln(\frac{p_1(\mathbf{x})}{p_0(\mathbf{x})})\} \\ &= \int p_1(\mathbf{x}) \ln(\frac{p_1(\mathbf{x})}{p_0(\mathbf{x})} d\mathbf{x} \\ &= E_1\{\ln(L(\mathbf{x})\} \end{aligned} \tag{5.4}$$

and

$$KL(p_1, p_0) = E_0\{\ln(\frac{p_0(\mathbf{x})}{p_1(\mathbf{x})})\}$$
$$= \int p_0(\mathbf{x}) \ln(\frac{p_0(\mathbf{x})}{p_1(\mathbf{x})}) d\mathbf{x} \quad (5.5)$$
$$= -E_0\{\ln(L(\mathbf{x})\}$$

In statistics, it represents the expected logarithm of the likelihood ratio. It can be shown that KL distance is always non-negative and is zero if and only if $p_0(\mathbf{x}) = p_1(\mathbf{x})$ [89]. However, it is not a true distance between distributions since it is not symmetric ($KL(p_0, p_1) \neq KL(p_1, p_0)$ in general) and does not satisfy the triangle inequality. Nonetheless, it is often useful to think of KL Distance as a "distance" between distributions. Surprisingly, when the amplitude distribution of the observation vector is symmetric, the KL distance becomes a symmetric function. In [91], it is shown that the dependence of the detector's performance on signal characteristics can be related to the KL distance when the noise has statistically independent, identically distributed (i.i.d.) components. Its properties determine the impact of signal set selection as well as noise amplitude distribution on performance. However, KL distance cannot provide numeric estimates of error rates, even though it does express the dependence of the error probabilities on the choice of signals. Rigorously speaking, the KL distance only determines asymptotic performance.

J-Divergence J-divergence was first introduced by H. Jeffreys in [92]. It is defined as the difference in the expectations of the log-likelihood ratio under two hypotheses.

$$J = E_1\{\ln(L(\mathbf{x})\} - E_0\{\ln(L(\mathbf{x})\} \quad (5.6)$$

It is a symmetric form of the KL distance, that is

$$J = KL(p_0, p_1) + KL(p_1, p_0) \quad (5.7)$$

The J-divergence satisfies all the properties for a distance (metric) except the triangle inequality. Several other properties of the J-divergence and various applications to classification and hypothesis testing are studied by Kullback in [93]. It is also noticed that the J-divergence depends only on the first moments (mean values) of the likelihood ratio while the probability of error requires the total distribution of the likelihood ratio. It can be shown that the J-divergence is a convex function of the likelihood ratio. Some applications of J-divergence to signal selection are studied in [94]. For Gaussian processes with unequal mean and same covariance, the J-divergence signal selection criterion yields signals that are in fact also optimal on an error probability basis. This result is unfortunately not universal. For Gaussian processes with different means and covariances, the J-divergence yields results at low SNR that are well correlated with probability of error, but not well correlated at high SNR.

Bhattacharyya Distance Bhattacharyya distance is defined as

$$B = -\ln \rho \tag{5.8}$$

where ρ is the Bhattacharyya coefficient defined as

$$\begin{aligned} \rho &= \int \sqrt{p_0(\mathbf{x}) p_1(\mathbf{x})}\, d\mathbf{x} \\ &= \int \sqrt{\frac{p_1(\mathbf{x})}{p_0(\mathbf{x})}}\, p_0(\mathbf{x})\, d\mathbf{x} \\ &= \mathbf{E}_0\{\sqrt{L(\mathbf{x})}\} \end{aligned} \tag{5.9}$$

In fact, it is a special case of the *cumulant generating function* of the test statistic under H_0 which is related to the well-known Chernoff bound [79]. The Bhattacharyya distance is also a convex function of the likelihood ratio. It has many interesting properties. One property discovered by Kailin and Bradt [95] is described in the following theorem.

Theorem: If for two sets of parameters α and β, we have $B(\alpha) > B(\beta)$ or $\rho(\alpha) < \rho(\beta)$, there exists a set $\pi = (\pi_0, \pi_1)$ of prior probabilities for which $P_e(\alpha, \pi) < P_e(\beta, \pi)$.

The J-divergence has this property as well, but it is hard to assert something more than just existence about the proper set of prior probabilities. For the relationship between J-divergence and Bhattacharyya distance, the inequality

$$J \geq 8B \tag{5.10}$$

can be established. For some applications, the Bhattacharyya distance is claimed to be superior to the J-divergence for Guassian processes [96].

For the problem of interest, it is straightforward to show that the quantity $D^*(\mathbf{c}_0, \mathbf{c}_1)$ in (4.11) is equivalent to the J-divergence between the two Gaussian distributions corresponding to the signal pair $(\mathbf{c}_0, \mathbf{c}_1)$ (see Appendix A). Hence, solving Problem (4.8) leads to the optimal Gaussian detector for the signal pair that gives maximum divergence distance between the two corresponding Gaussian hypotheses. Since maximizing the J-divergence is known to be a useful criterion for signal selection in Gaussian detection problems [91, 93, 97], this is a desirable result. In addition, for the problem of interest, the derivatives of the J-Divergence with respect to the coordinates of the signals are much easier to derive than those for the Bhattacharyya distance. Therefore, we will choose the J-divergence as distance criteria to find the optimal signal constellations.

5.2 Adaptive Multicoding

To solve simultaneously for the optimal signal structure and the corresponding optimal detector, we note that the value of $D^*(\mathbf{c}_0,\mathbf{c}_1)$ can be rewritten in the form

$$D^*(\mathbf{c}_0,\mathbf{c}_1)=\boldsymbol{\mu}^*\mathbf{h}+\tfrac{1}{4}\mathrm{Tr}(\boldsymbol{\Sigma}_0\boldsymbol{\Phi}\boldsymbol{\Sigma}_1\boldsymbol{\Phi}), \tag{5.11}$$

where $(\mathbf{h},\boldsymbol{\Phi})$ represents the optimal detector for $(\mathbf{c}_0,\mathbf{c}_1)$. Further, assuming that the set $(\mathbf{h},\boldsymbol{\Phi})$ is given, the formula for the gradient of $D^*(\mathbf{c}_0,\mathbf{c}_1)$ is straightforward (see Appendix B). Hence, to solve simultaneously for the signal structure and the detector structure, we can use a recursive gradient descent procedure that alternates between updates of $(\mathbf{h},\boldsymbol{\Phi})$ to solve (5.11) and updates of $(\mathbf{c}_0,\mathbf{c}_1)$ to solve (4.7).

The gradient descent algorithm works as follows. First, the information bit sequence and the signature sequence matrix are generated. Then, we form the channel tap covariance matrix and the background noise vector. At the beginning, we set the channel uncertainty coefficient equal to zero, which represent the channel is completely coherent. Later we will increase the channel uncertainty to different levels in order to evaluate the impact on receivers' performance. Based on the given channel uncertainty, the estimated channel tap vector and the channel estimate error vector are independently generated. Then, we set some initial values including a random start point, a threshold, and a step size. Now we can compute the mean vector and the covariance matrices of the transformed observation given the start point $(\mathbf{c}_0,\mathbf{c}_1)$. By solving (4.7) and (5.11) we can find the associated detector $(\mathbf{h},\boldsymbol{\Phi})$ and J-divergence with respect to the start point. Then, we can find the gradient descent direction by using the derived gradient expression of J-divergence. Adding a small step size along the gradient direction from the start point, we reach a new start point $(\mathbf{c}_0,\mathbf{c}_1)$. Find the new detector $(\mathbf{h},\boldsymbol{\Phi})$ corresponding to the new start point by solving (4.7). Then plug the new

detector into (5.11) to find the corresponding new J-divergence. At this point, compare the changing rate of J-divergence with the threshold. If the changing rate is greater than the threshold, adapt the step size and add further along the current gradient direction, then repeat the above procedure until the changing rate of J-divergence is less than the given threshold, which indicts the maximum value of J-divergence is approximately achieved for the current pair of signal structure $(\mathbf{c}_0,\mathbf{c}_1)$ and receiver structure $(\mathbf{h},\Phi)$. Hence, we can compute the transformed observation based on the current data, and apply LQ detector to detect the transmitted information bit, count the detection error by comparing with the original information data. By following the above procedure, we are able to evaluate the system performance under various uncertainty levels by adjusting the channel uncertainty coefficient in the range from zero to one. Figure 5.1 illustrates the algorithm in a flow chart.

To actually implement the procedure described above, we need estimates of Σ_ε and Σ_γ, and we must restrict the search space for the signals $(\mathbf{c}_0,\mathbf{c}_1)$ to a reasonable parametric set. In fact, since the desired signal structure will be computed adaptively at the receiver, the parameters describing the new signals must be communicated to the transmitter whenever the signal structure is updated. As a result, it is desirable to keep the dimensionality of the parameter space as small as possible. One possibility is to assign two linearly independent "basis" signature sequences $(\mathbf{s}_0,\mathbf{s}_1)$ to each user and adapt the signal structure by searching over the two-dimensional signal subspace spanned by the known basis signatures. We define $\mathbf{c}_0$ and $\mathbf{c}_1$ as

$$\begin{aligned} \mathbf{c}_0 &= \gamma_0\mathbf{s}_0 + \gamma_1\mathbf{s}_1, \quad \text{where } \gamma_0^2+\gamma_1^2 \le E_b \\ \mathbf{c}_1 &= \gamma_2\mathbf{s}_0 + \gamma_3\mathbf{s}_1, \quad \text{where } \gamma_2^2+\gamma_3^2 \le E_b \end{aligned} \tag{5.12}$$

then choose γ_0, γ_1, γ_2 and γ_3 to maximize $D^*(\mathbf{c}_0,\mathbf{c}_1)$ in (5.11). We refer to the above technique as *adaptive multicoding*.

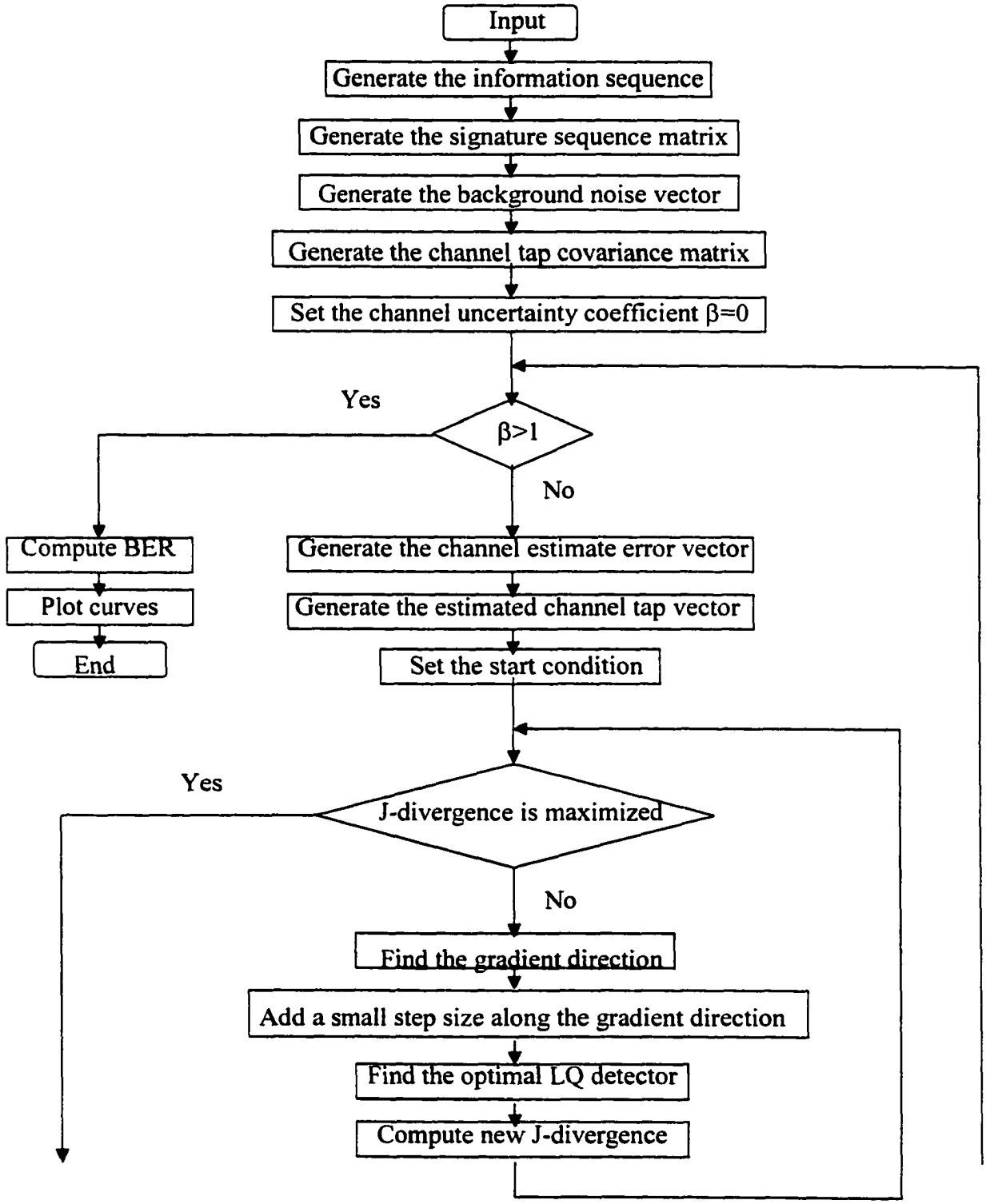

Figure 5.1 Flow Chart of Gradient Descent Algorithm

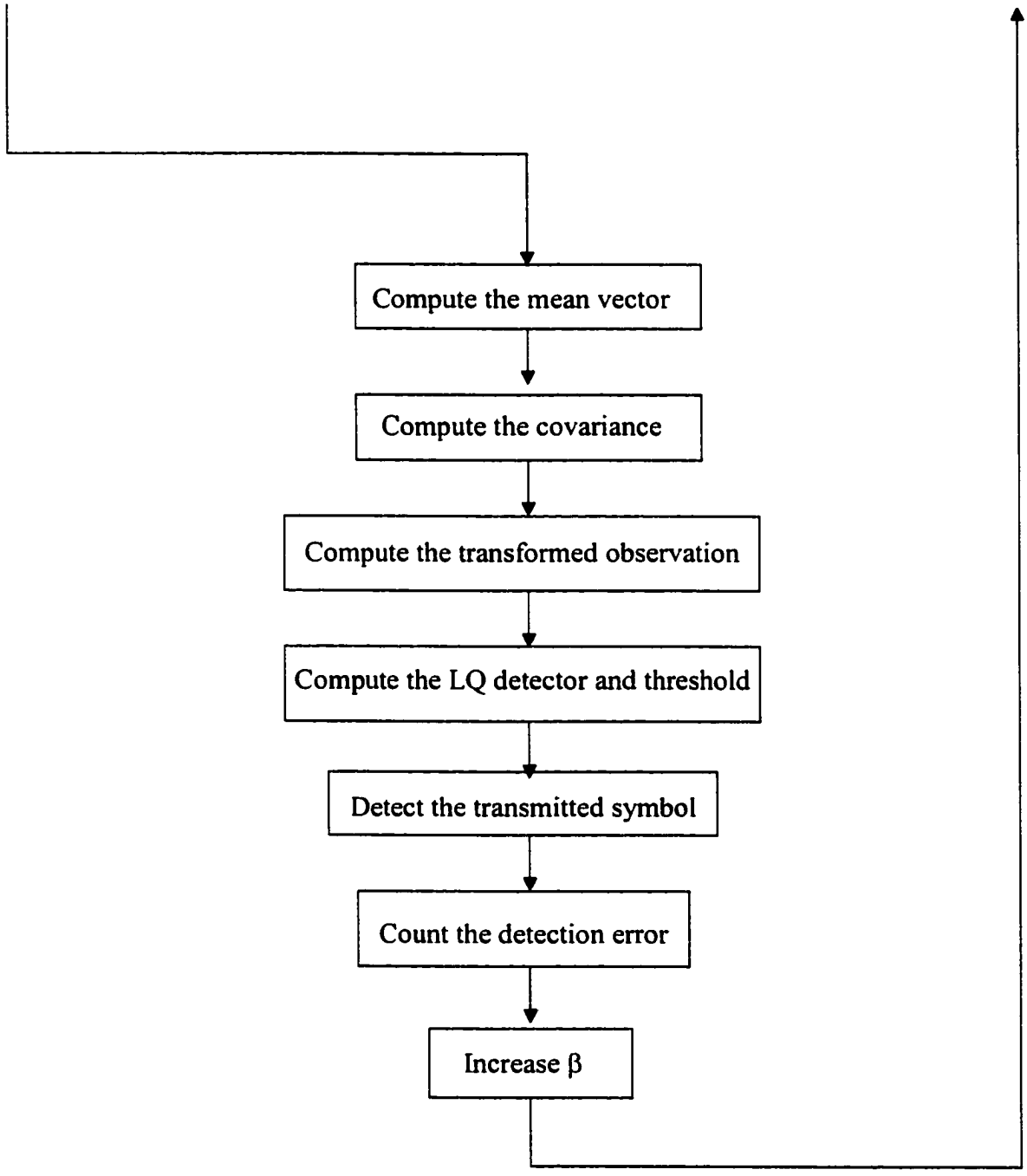

Figure 5.1 (Continued)

This approach has an additional advantage related to estimating Σ_ε and Σ_γ. That is, an initial training period can be adopted prior to signal adaptation, during which we set $\mathbf{c}_0 = \mathbf{s}_0$, $\mathbf{c}_1 = -\mathbf{s}_0$ for the first half of the training period and $\mathbf{c}_0 = \mathbf{s}_1$, $\mathbf{c}_1 = -\mathbf{s}_1$ for the second half of the training period. Σ_ε and Σ_γ can then be estimated from the received data (*i.e.*, blindly) during the training period in a straightforward manner. Once the training period is ended, signal adaptation can begin, and the estimates of Σ_ε and Σ_γ can be updated using a decision directed approach.

5.3 Simulation Work

In order to evaluate the efficacy of adaptive multicoding and the robustness of LQ detectors, we have completed a simulation of the proposed adaptive signal/receiver scheme in the binary case. The performance of LQ detectors and linear MMSE detectors are compared for both fixed signaling and adaptive signaling. The procedure of this simulation work is presented below. Throughout this section, we make the following assumptions [101]:

1. Both signals have equal energy, and the spreading gain is $N = 31$ ($\|\mathbf{c}_0\|^2 = \|\mathbf{c}_1\|^2 = 31$).
2. The additive noise on the channel consists of ISI and AWGN only.
3. The ISI is generated by frequency-selective Rayleigh fading with $L+1 = 16$ resolvable paths that are independent but not identically distributed. The fading parameters are assumed constant over blocks of 100 bits and independent from block to block. The total power of the fading process (*i.e.*, the total variance for all paths) is normalized to one.
4. The statistics of the channel are known *a priori*; in particular, the covariance matrix is diagonal and:

$$\begin{aligned}
\boldsymbol{\Sigma}_{\alpha} &= \boldsymbol{\Sigma}_{\varepsilon} + \boldsymbol{\Sigma}_{\hat{\alpha}}, \\
\boldsymbol{\Sigma}_{\varepsilon} &= \beta \boldsymbol{\Sigma}_{\alpha}, \\
\boldsymbol{\Sigma}_{\hat{\alpha}} &= (1-\beta)\boldsymbol{\Sigma}_{\alpha}, \\
\boldsymbol{\Sigma}_{\gamma} &= \sigma^2 \mathbf{I},
\end{aligned} \tag{5.13}$$

where $\boldsymbol{\Sigma}_{\alpha}$ represents the covariance matrix of the channel, $0 \le \beta \le 1$ represents the *uncertainty coefficient* of the channel and $\sigma^2 > 0$ represents the variance of the additive channel noise (excluding ISI). A completely coherent channel corresponds to $\beta = 0$, and a completely noncoherent channel corresponds to $\beta = 1$. The channel SNR is given by $E_b/N_0 = 3\,1/\sigma^2$.

First, let us compute the conditional statistics. From the observation vector given in (4.1) and (4.2), for known CSI $\hat{\boldsymbol{\alpha}}$, and under the assumption that $\boldsymbol{\varepsilon}$, $\hat{\mathbf{i}}$ and $\hat{\boldsymbol{\gamma}}$ are uncorrelated, the conditional statistics can be obtained as follows.

Under hypothesis H_0,

$$\begin{aligned}
\mathbf{y}_0 &= \tfrac{1}{2}(\mathbf{C}_{00} - \mathbf{C}_{01})\hat{\boldsymbol{\alpha}} + \mathbf{C}_{00}\boldsymbol{\varepsilon} + \hat{\mathbf{i}} + \hat{\boldsymbol{\gamma}} \\
\boldsymbol{\mu}_0 &= \tfrac{1}{2}(\mathbf{C}_{00} - \mathbf{C}_{01})\hat{\boldsymbol{\alpha}} \\
\boldsymbol{\Sigma}_0 &= \mathbf{C}_{00}\boldsymbol{\Sigma}_{\varepsilon}\mathbf{C}_{00}^{*} + \boldsymbol{\Sigma}_{\mathbf{i}} + \boldsymbol{\Sigma}_{\gamma}
\end{aligned} \tag{5.14}$$

Under hypothesis H_1,

$$\begin{aligned}
\mathbf{y}_1 &= \tfrac{1}{2}(\mathbf{C}_{01} - \mathbf{C}_{00})\hat{\boldsymbol{\alpha}} + \mathbf{C}_{01}\boldsymbol{\varepsilon} + \hat{\mathbf{i}} + \hat{\boldsymbol{\gamma}} \\
\boldsymbol{\mu}_1 &= \tfrac{1}{2}(\mathbf{C}_{01} - \mathbf{C}_{00})\hat{\boldsymbol{\alpha}} \\
\boldsymbol{\Sigma}_1 &= \mathbf{C}_{01}\boldsymbol{\Sigma}_{\varepsilon}\mathbf{C}_{01}^{*} + \boldsymbol{\Sigma}_{\mathbf{i}} + \boldsymbol{\Sigma}_{\gamma}
\end{aligned} \tag{5.15}$$

where $\mathbf{C}_{00}$, $\mathbf{C}_{01}$ represent respectively the signature sequence matrices for the cases of 0 and 1 transmitted as the current bit. The noise vector $\hat{\boldsymbol{\gamma}}$ can be generated from its covariance matrix $\boldsymbol{\Sigma}_{\gamma}$ with some known variance σ^2. Now we know everything about the observations except ISI. Let $b \in \{0,1\}$ represent the transmitted bit value, and assume the CSI vector $\boldsymbol{\alpha}$ is

stationary with zero mean. Then $\hat{\mathbf{i}}$ and $\boldsymbol{\Sigma}_{\mathbf{i}}$ can be derived for the problem of interest (see Appendix C) as

$$\begin{aligned}\hat{\mathbf{i}} &= (b_{-1}-\tfrac{1}{2})(C_{-11}-C_{-10})\boldsymbol{\alpha}_{-1}+(b_{+1}-\tfrac{1}{2})(C_{+11}-C_{+10})\boldsymbol{\alpha}_{+1}\\ \boldsymbol{\Sigma}_{\mathbf{i}} &= \tfrac{1}{4}((C_{-11}-C_{-10})\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}(C_{-11}-C_{-10})^{*}+(C_{+11}-C_{+10})\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}(C_{+11}-C_{+10})^{*})\end{aligned} \tag{5.16}$$

where b_{-1} and b_{+1} represent the transmitted bit values for previous bit and next bit, respectively, $\mathbf{C}_{-10}$, $\mathbf{C}_{-11}$ represent the signature sequence matrices for 0 and 1 transmitted in the previous bit, and $\mathbf{C}_{+10}$, $\mathbf{C}_{+11}$ represent the signature sequence matrices for 0 and 1 transmitted in the subsequent.

As for linear detectors that exploit antipodal signaling ($\mathbf{C}_{\bullet 0}=-\mathbf{C}_{\bullet 1}$), the conditional statistics have a simpler expression.

Under hypothesis H_0,

$$\begin{aligned}\mathbf{y}_0 &= \mathbf{C}_{00}\hat{\boldsymbol{\alpha}}+\mathbf{C}_{00}\boldsymbol{\varepsilon}+\hat{\mathbf{i}}+\hat{\boldsymbol{\gamma}}\\ \boldsymbol{\mu}_0 &= \mathbf{C}_{00}\hat{\boldsymbol{\alpha}}\\ \boldsymbol{\Sigma}_0 &= \mathbf{C}_{00}\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}\mathbf{C}_{00}^{*}+\boldsymbol{\Sigma}_{\mathbf{i}}+\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}\end{aligned} \tag{5.17}$$

Under hypothesis H_1,

$$\begin{aligned}\mathbf{y}_1 &= -\mathbf{C}_{00}\hat{\boldsymbol{\alpha}}-\mathbf{C}_{00}\boldsymbol{\varepsilon}+\hat{\mathbf{i}}+\hat{\boldsymbol{\gamma}}\\ \boldsymbol{\mu}_1 &= -\mathbf{C}_{00}\hat{\boldsymbol{\alpha}}\\ \boldsymbol{\Sigma}_1 &= \mathbf{C}_{00}\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}\mathbf{C}_{00}^{*}+\boldsymbol{\Sigma}_{\mathbf{i}}+\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}\end{aligned} \tag{5.18}$$

ISI vector and its covariance matrix have the following forms:

$$\begin{aligned}\hat{\mathbf{i}} &= (1-2b_{-1})C_{-10}\boldsymbol{\alpha}_{-1}+(1-2b_{+1})C_{+10}\boldsymbol{\alpha}_{+1}\\ \boldsymbol{\Sigma}_{\mathbf{i}} &= C_{-10}\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}C_{-10}^{*}+C_{+10}\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}C_{+10}^{*}\end{aligned} \tag{5.19}$$

Having computed all of the necessary statistics, we can use Matlab to simulate the system. Documentation of simulation code is presented in details below.

- **Generate the information sequence**: A {0, 1} sequence is randomly generated with equal probability by using functions "rand" and "fix".
- **Generate the signature sequence matrix**: We choose two orthogonal Gold sequences as *basis* vectors. One sequence is used for transmitting "0" and the other is used for transmitting "1". The corresponding signature sequence matrices are generated by permutation of basis vectors.
- **Generate the background noise vector**: Given the variance of the background noise, an AWGN vector is generated. First, use function "randn" to generate a standard AWGN vector with zero mean and variance one. Then use function "chol" to do the Cholosky decomposition of the covariance matrix with the given variance on its diagonal. Multiply the result with the standard AWGN vector to get the desired background noise vector.
- **Form the channel tap covariance matrix**: We assume the channel has 16 resolvable fading paths that are uncorrelated but not identically distributed. A diagonal matrix is formed with all the variances of the channel taps on the diagonal by descending order. The total power of the fading process (*i.e.*, the total variance for all paths) is normalized to one.
- **Evaluate under various uncertainty level**: We evaluate the system performance under different uncertainty level by using a loop to vary the channel uncertainty coefficient from 0 to 1.
- **Generate the channel estimate error vector**: Given the channel uncertainty level, use (5.13) to compute the covariance matrix of estimate error. Then use

functions "randn" and "chol" to generate the channel estimate error vector by assuming it is Guassian with zero mean.

- **Generate the estimated channel tap vector**: Use the same approach for generating the channel estimate error vector.
- **Linear Detectors**: (1). *Fixed signaling*: use antipodal signals for linear detectors. Choose one signature sequence matrix $\mathbf{C}_0$ for transmitting "0", and use $\mathbf{C}_1 = -\mathbf{C}_0$ for transmitting "1". Use (5.17), (5.18) and (5.19) to compute the transformed observation, the mean vector and the covariance matrix. Find the linear detector and compare the test statistics to the threshold to determine the information bit. Count an error if it is not identical with the transmitted bit. (2). *Adaptive signaling*: still use antipodal signals but rotate the fixed signal structure on the plane spanned by the two basis vectors. Use function "fminbnd" to maximize the J-divergence in order to find the optimal signal structure. Once the optimal signals are found, follow the procedure for fixed signaling to determine the performance.
- **LQ Detectors**: (1) *Fixed signaling*: use orthogonal signals for LQ detectors. Since the two basis vectors are orthogonal, we simply choose them. Use (5.14), (5.15) and (5.16) to compute the transformed observation, the mean vector and the covariance matrix. Find the LQ detector and compare the test statistics to the threshold to determine the information bit. Count an error if it is not identical with the transmitted bit. (2) *Adaptive signaling*: no restriction on signal structure, just search the plane spanned by the two basis vectors to find the optimal signal structure that maximizes the J-divergence. Use the gradient descent algorithm developed in last section to determine the optimal

pair of the receiver structure and signal structure. Detect the information bits and count the errors.

- **Compute BER and plot curves**: The bit error probabilities for all scenarios are computed. The curves demonstrating BER vs. Uncertainty are plotted.

The performance comparison between linear detectors and LQ detectors for both fixed signals and adaptive signals is presented in next section.

5.4 Performance Evaluation

The results of the simulations are illustrated in this section in Figures 5.2-5.7.

In Figure 5.2, the performance of an LQ receiver employing orthogonal modulation is compared with the performance of a linear MMSE receiver employing antipodal modulation. Note that for any given SNR, the LQ receiver outperforms the linear MMSE receiver for high uncertainty, but the linear MMSE receiver outperforms the LQ receiver for low uncertainty. As the SNR increases, the cross point of the two performance curve moves toward left side of the graph. This is consistent with the fact that, as the SNR on the channel increases, the dominant error mechanism on the channel becomes the mismatch between the detector and the channel that results from inaccurate CSI estimates.

If adaptive modulation is employed, is there any improvement in the performance? The signal sets for both the linear detector and the LQ detector were chosen adaptively to give the best performance in each case relative to the instantaneous CSI and the known channel statistics. The results of simulations for these cases are illustrated in figure 5.3 and5.4.

In Figure 5.3, which illustrates the case of adapting a signal set for the linear detector, only $\mathbf{c}_0$ was chosen adaptively, since antipodal modulation ($\mathbf{c}_1 = -\mathbf{c}_0$) was assumed. Since the

signal structure is fixed, we only rotate the antipodal signal pair to maximize the divergence distance that depends on the level of channel uncertainty. Note that only a slight gain is obtained by adaptive modulation for the linear detector.

In Figure 5.4, which illustrates the case of adapting a signal set for the LQ detector, both $\mathbf{c}_0$ and $\mathbf{c}_1$ were chosen adaptively subject to the equal-energy constraint. Thus, we choose the signal pair to maximize the divergence distance with more freedom. There is a significant performance improvement for the LQ detector with low SNR; however, with increasing SNR, adaptive modulation can actually perform worse than orthogonal signaling for high uncertainty. This indicates that the divergence distance is not a good approximation for the probability of bit error for high SNR.

In Figure 5.5, the LQ detector and linear detector with adaptive modulation are compared. The performance of the adaptive LQ detector is always at least as good as the performance of the linear detector. This is the desired result and indicates not only that the modified deflection ratio is a good criterion for signal selection for this problem, but also that the recursive LMS algorithm discussed above converges to a nearly optimal solution. Furthermore, as one would expect, the LQ detector becomes strictly better than the linear detector as the uncertainty on the channel increases at a fixed SNR. Finally, as the SNR on the channel increases, the level of uncertainty at which the LQ detector begins to significantly outperform the linear detector converges to zero.

The results of the same sequence of simulations are summarized and plotted as a function of SNR in Figure 5.6. Here the increasing advantage of the LQ receiver relative to a desired bit error rate is clearly displayed. In particular, for values of $\beta > 0.2$, the performance gain associated with the LQ detector in the operating range of interest (say $P_e < 10^{-3}$) is dramatic.

At the end of this section, we present the optimal binary signal constellation under different levels of channel uncertainty. Figure 5.7 illustrates how the signal pair varies between antipodal signals and orthogonal signals. Antipodal signals are chosen for coherent channels while orthogonal signals are chosen for noncoherent channels.

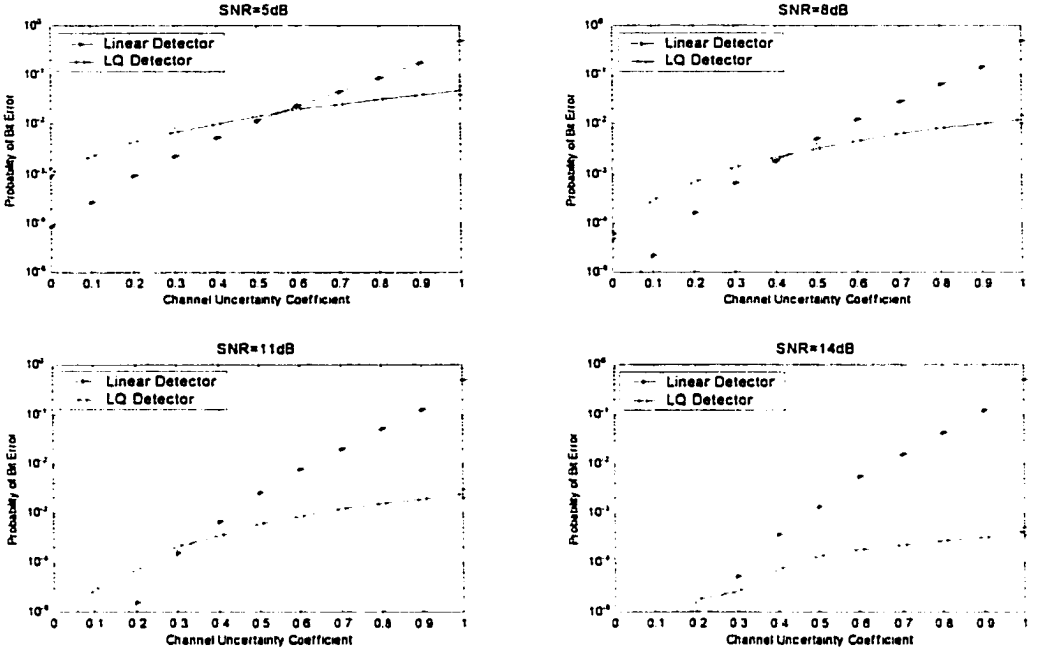

Figure 5.2 Linear vs. LQ Detectors for Fixed Signaling

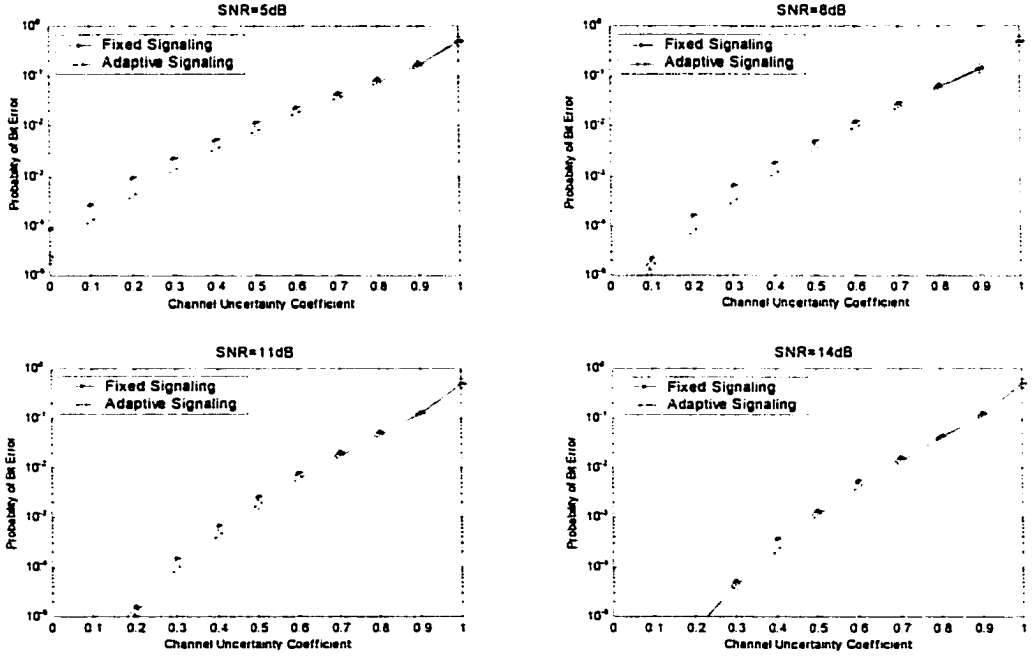

Figure 5.3 Fixed vs. Adaptive Signaling for Linear Detector

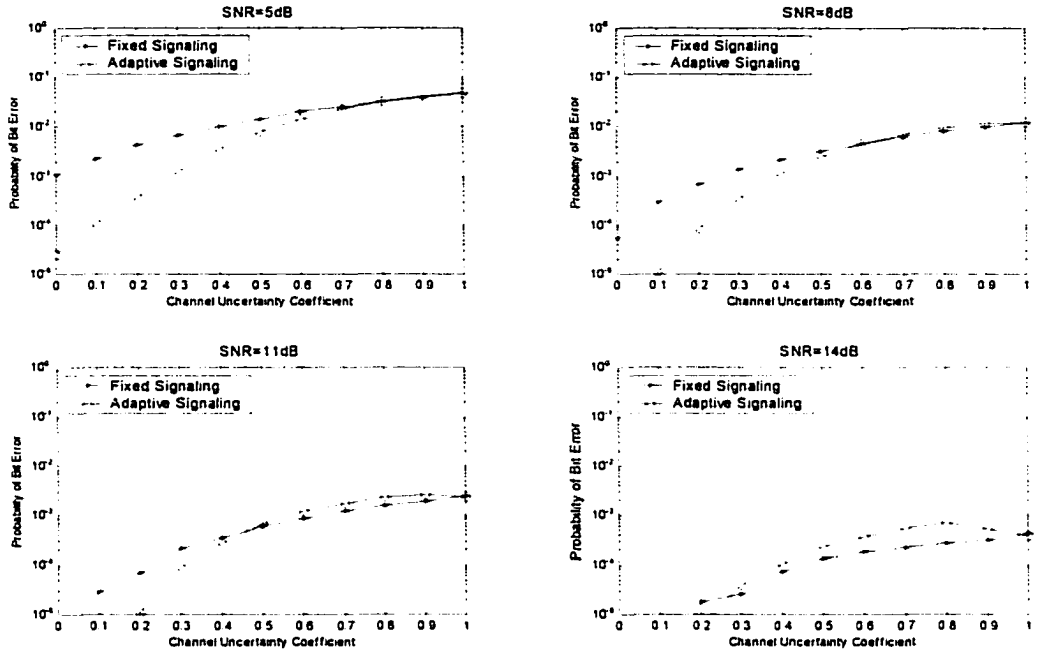

Figure 5.4 Fixed vs. Adaptive Signaling for LQ Detector

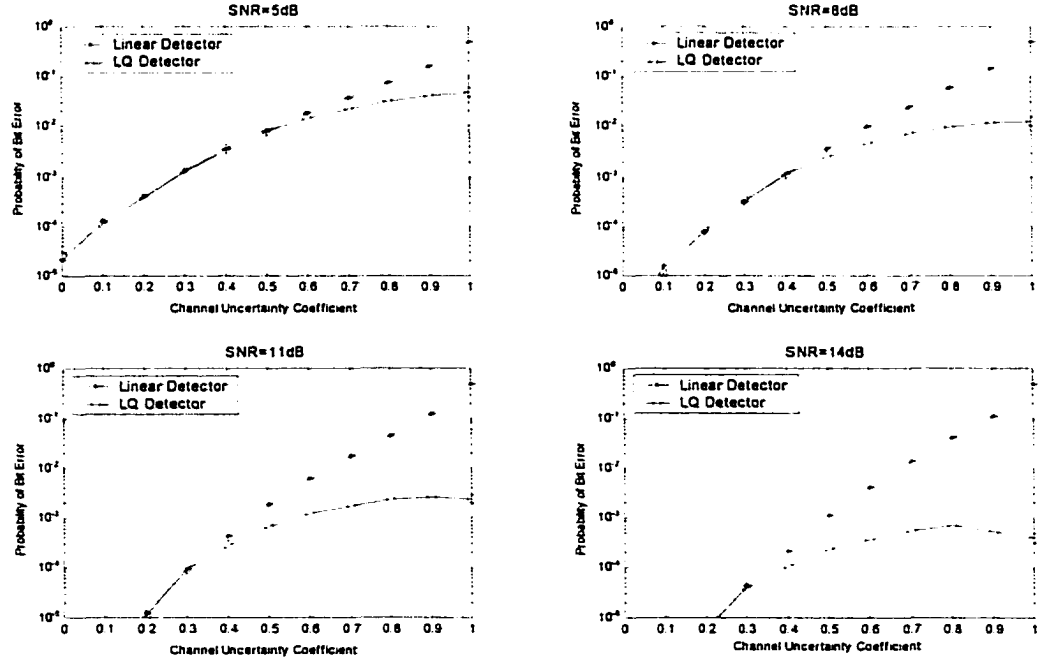

Figure 5.5 Linear vs. LQ Detectors for Adaptive Signaling

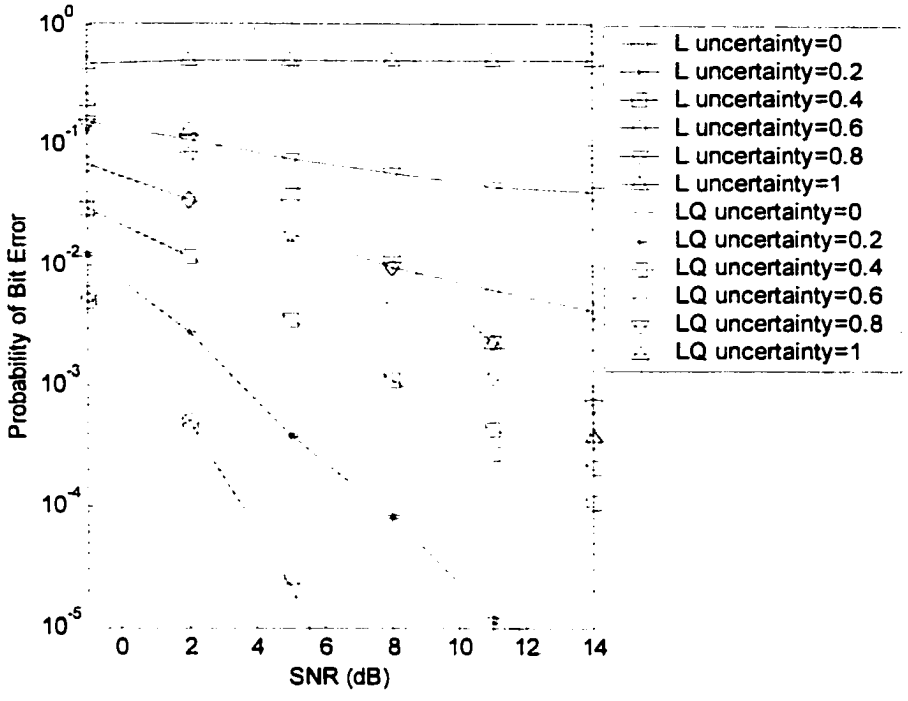

Figure 5.6 Performance of Linear vs. LQ Receiver – P_e vs. SNR

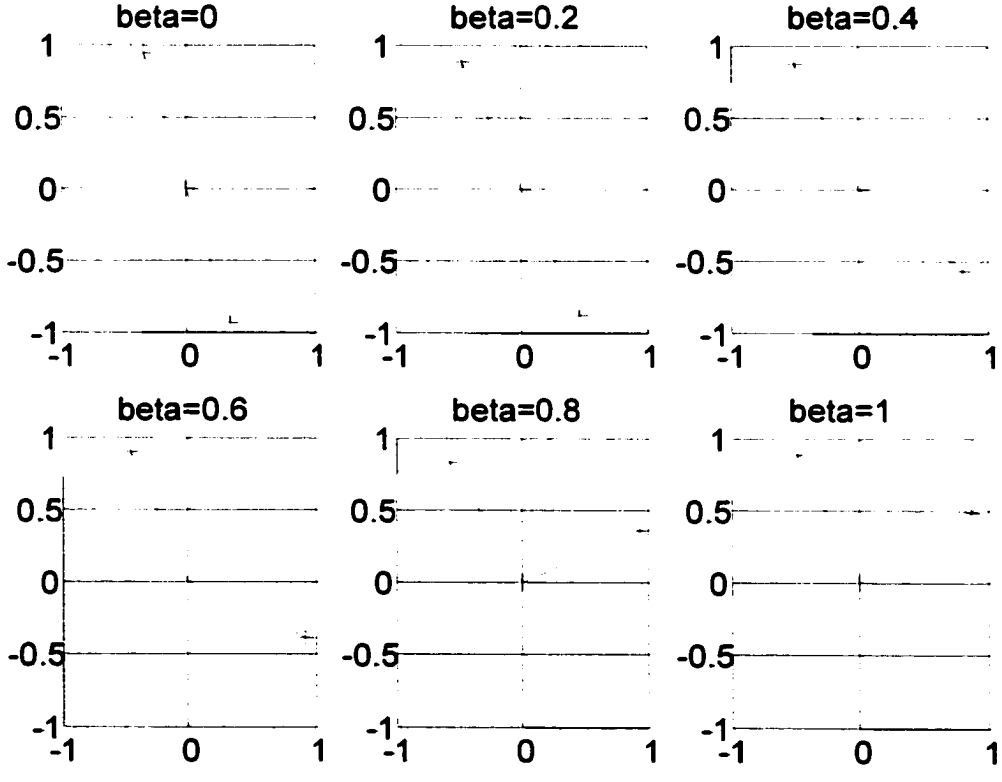

Figure 5.7 Optimal Binary Signal Constellations under Uncertainties

Chapter 6 Performance Analysis of LQ Receivers

The most desirable cost function is the probability of error that is the most common performance evaluation criterion for communication reliability. Unfortunately, it is intractable in many cases. While J-divergence is useful for signal selection because of its feasibility, it is not tightly correlated with probability of error at high SNR.

In this chapter, we study some potential cost functions. Although the probability of error for LQ receivers has no analytical closed form, we can often resort to Chernoff bound in many practical applications. Particularly, we study some bounds related to Bhattachayya distance and J-divergence. At the end of this chapter, we explain why we prefer J-divergence rather than others.

6.1 Intractability of BER

For the general binary hypothesis testing problem described in (a), the probability of bit error is defined as follows [79],

$$P_e = P\{\frac{p_1(\mathbf{x})}{p_0(\mathbf{x})} \geq \tau \mid H_0\} + P\{\frac{p_1(\mathbf{x})}{p_0(\mathbf{x})} < \tau \mid H_1\} \tag{6.1}$$

where

$$\tau = \frac{\pi_0}{\pi_1}$$

and π_0 and π_1 are priori probabilities for two hypotheses. Thus, we can rewrite (6.1) as

$$P_e = \pi_0 P(L \geq \tau) + \pi_1 P(L < \tau) \tag{6.2}$$

where

$$L = \frac{p_1(\mathbf{x})}{p_0(\mathbf{x})}$$

is the likelihood ratio. By definition,

$$P(L \geq \tau) = E\{I_{[\tau,\infty)}(L)\} \tag{6.3}$$

where $I_{[\tau,\infty)}(L)$ is the indicator function of the set $[\tau,\infty)$ defined by

$$I_{[\tau,\infty)}(x) = \begin{cases} 1 & if\ x \geq \tau \\ 0 & if\ x < \tau \end{cases} \tag{6.4}$$

Therefore, if we can compute $E\{I(L)\}$, then we can compute the probability of error. Unfortunately, $E\{I(L)\}$ is intractable in many cases. However, in the real world, it is usually sufficient to obtain some good bounds on the error probably. An alternative solution is to find a function $f \geq I$ for which $E\{f(L)\}$ is easy to compute, so that $E\{f(L)\}$ can bound $E\{I(L)\}$. Various bounds discussed later in this chapter are based on those functions. For example, $E\{L^s\}$ is used for deriving Chernoff bound; and $E\{(L-1)\log L\}$ is applied for a lower bound related to J-divergence.

In order to analyze the performance of the LQ detector in (4.5), we must compute the probabilities $P_j(T(\mathbf{y}) \geq \tau)$ for $j = 0$, 1. For the general binary hypothesis testing problem in (a), it can be discussed more easily if we first transform the observations to some new vectors whose components are all independent. We call this procedure *whitening transformation*. It is

well known [98] that if $\mathbf{\Sigma}_0$ is positive definite and $\mathbf{\Sigma}_1$ is Hermitian, then there exists a nonsingular matrix $\mathbf{W}$ such that

$$\begin{aligned} \mathbf{W}\mathbf{\Sigma}_0\mathbf{W}^T &= \mathbf{I} \\ \mathbf{W}\mathbf{\Sigma}_1\mathbf{W}^T &= \mathbf{\Lambda} \end{aligned}, \tag{6.5}$$

where $\mathbf{I}$ is identity and $\mathbf{\Lambda}$ is diagonal. Without loss of generality, we assume that covariance matrices $\mathbf{\Sigma}_0$ and $\mathbf{\Sigma}_1$ satisfy the above assumption. Thus, $\mathbf{\Sigma}_0$ and $\mathbf{\Sigma}_1$ can be simultaneously diagonalized by the whitening matrix $\mathbf{W}$. $\mathbf{W}$ can be found by the following procedure[98]:

1. Cholesky Decomposition:

Since $\mathbf{\Sigma}_0$ is positive definite, there exists a nonsingular matrix $\mathbf{A}$ such that

$$\mathbf{\Sigma}_0 = \mathbf{A}^*\mathbf{A}, \tag{6.6}$$

that is

$$(\mathbf{A}^*)^{-1}\mathbf{\Sigma}_0\mathbf{A}^{-1} = \mathbf{I}. \tag{6.7}$$

2. Takagi Decomposition:

Let $\mathbf{B} = (\mathbf{A}^*)^{-1}\mathbf{\Sigma}_1\mathbf{A}^{-1}$, $\mathbf{\Sigma}_1$ is Hermitian, so is $\mathbf{B}$. Then there exists a unitary matrix $\mathbf{U}$ such that

$$\mathbf{U}^*\mathbf{B}\mathbf{U} = \mathbf{\Lambda} \tag{6.8}$$

where $\mathbf{\Lambda} = diag(\lambda_1, \lambda_2, ..., \lambda_n)$, and $\{\lambda_i\}$ are eigenvalues of $\mathbf{B}$.

3. Let $\mathbf{W} = \mathbf{A}^{-1}\mathbf{U}$.

After the whitening transformation, we have a new binary hypothesis testing problem

$$\begin{aligned} & H_0 : \mathbf{z} \sim \mathbb{N}(\mathbf{m}_0, \mathbf{I}), \\ \text{versus} & \\ & H_1 : \mathbf{z} \sim \mathbb{N}(\mathbf{m}_1, \mathbf{\Lambda}). \end{aligned} \tag{c}$$

where $\mathbf{z} = \mathbf{W}\mathbf{y}, \mathbf{m}_0 = \mathbf{W}\boldsymbol{\mu}_0, \mathbf{m}_1 = \mathbf{W}\boldsymbol{\mu}_1$. We can compute the new test statistics

$$\begin{aligned} T(\mathbf{z}) &= \tfrac{1}{2}\mathbf{z}^T(\mathbf{I} - \mathbf{\Lambda}^{-1})\mathbf{z} + (\mathbf{m}_1^T\mathbf{\Lambda}^{-1} - \mathbf{m}_0^T\mathbf{I})\mathbf{z} + C \\ &= \sum_{k=1}^{n}(a_k z_k^2 + b_k z_k) + C \\ &= \sum_{k=1}^{n} T_k(z_k) + C \end{aligned} \tag{6.9}$$

where $T_k(z_k) = a_k z_k^2 + b_k z_k$,

$a_k = \tfrac{1}{2}(1 - \lambda_k^{-1})$,

$b_k = m_{1k}\lambda_k^{-1} - m_{0k}$,

$C = \tfrac{1}{2}(\log(|\,\mathbf{\Sigma}_0\,| / |\,\mathbf{\Sigma}_1\,|) + \boldsymbol{\mu}_0^T\mathbf{\Sigma}_0^{-1}\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1^T\mathbf{\Sigma}_1^{-1}\boldsymbol{\mu}_1)$.

It is straightforward to show that $z_1, z_2, ..., z_n$ are independent Gaussian random variables under distributions:

$$\begin{aligned} & H_0 : \mathbf{z}_k \sim \mathbb{N}(\mathbf{m}_{0k}, 1), \\ \text{versus} & \\ & H_1 : \mathbf{z}_k \sim \mathbb{N}(\mathbf{m}_{1k}, \lambda_k). \end{aligned} \tag{c'}$$

Thus, the probability density function of $T_k(z_k)$ under H_j can be determined by a method using characteristic functions [12]. It follows that the characteristic function

$$\Phi_{T_k}(\omega) = \int_{-\infty}^{\infty} e^{j\omega T_k} p_{T_k}(T_k) dT_k \tag{6.10}$$

of the random variable $T_k(z_k) = a_k z_k^2 + b_k z_k$ equals

$$\Phi_{T_k}(\omega) = E(e^{j\omega T_k(Z_k)}) = \int_{-\infty}^{\infty} e^{j\omega T_k(Z_k)} p_{Z_k}(Z_k) dZ_k \tag{6.11}$$

If, therefore, the above integral can be written in the form

$$\int_{-\infty}^{\infty} e^{j\omega T_k} h(T_k) dT_k \tag{6.12}$$

it will follow that

$$p_{T_k}(T_k) = h(T_k) \tag{6.13}$$

In our case,

$$\Phi_{T_k}(\omega) = \int_{-\infty}^{\infty} e^{j\omega T_k} P_{T_k}(T_k) dT_k$$

$$= \int_{-\infty}^{\infty} e^{j\omega(a_k Z_k^2 + b_k Z_k)} P_{Z_k}(Z_k) dZ_k$$

$$= \int_{-\infty}^{\infty} e^{j\omega(a_k Z_k^2 + b_k Z_k)} \frac{1}{\sqrt{2\pi}\sigma_{jk}} e^{\frac{(Z_k - m_{jk})^2}{2\sigma_{jk}^2}} dZ_k$$

$$= \int_{-\infty}^{0} e^{j\omega T_k} \frac{1}{\sqrt{2\pi}\sigma_{jk}} e^{-\frac{(\frac{-b_k-(b_k^2+4a_kT_k)^{\frac{1}{2}}}{2a_k}-m_{jk})^2}{2\sigma_{jk}^2}} (-2(b_k{}^2+4a_kT_k)^{-\frac{1}{2}}) dT_k$$

$$+ \int_{0}^{\infty} e^{j\omega T_k} \frac{1}{\sqrt{2\pi}\sigma_{jk}} e^{-\frac{(\frac{-b_k+(b_k^2+4a_kT_k)^{\frac{1}{2}}}{2a_k}-m_{jk})^2}{2\sigma_{jk}^2}} 2(b_k{}^2+4a_kT_k)^{-\frac{1}{2}} dT_k \tag{6.14}$$

Hence, the above yields,

$$p_{T_k}(T_k) = \frac{2}{(2\pi(b_k^2+4a_kT_k)\sigma_{jk}^2)^{\frac{1}{2}}}(e^{-c_k}u(T_k) - e^{-d_k}u(-T_k)) \tag{6.15}$$

where $c_k = \frac{(\frac{-b_k-(b_k^2+4a_kT_k)^{\frac{1}{2}}}{2a_k}-m_{jk})^2}{2\sigma_{jk}^2}$,

$$d_k = \frac{(\frac{-b_k+(b_k^2+4a_kT_k)^{\frac{1}{2}}}{2a_k}-m_{jk})^2}{2\sigma_{jk}^2},$$

$\sigma_{jk}^2 = 1$, for $j = 0$,

$\sigma_{jk}^2 = \lambda_k$, for $j = 1$.

and $u(T_k)$ is step function. This result can be reduced to the gamma density in [79] under zero-mean Guassian distribution. The probability density, p_T, of $T(\mathbf{z}) = \sum_{k=1}^{n} T_k(z_k)$ is the n-fold convolution

$$p_T = p_{T_1}(T_1) * p_{T_2}(T_2) * \ldots * p_{T_n}(T_n). \tag{6.16}$$

Unfortunately, no general closed form is known.

6.2 Chernoff Bound

We can use the result after the whitening transformation to obtain the Chernoff bound of the LQ detector. The Chernoff bound is defined in [79] as

$$P_e \le \pi_0^{1-s}\pi_1^{s} e^{u_{T,0}(s)}, \quad 0 \le s \le 1 \tag{6.17}$$

where $u_{T,0}(s)$ is the cumulant generating function (cgf) of $T(\mathbf{z})$ under H_0 . In the problem of interest,

$$u_{T,0}(s) = \log(E\{e^{sT(\mathbf{z})} \mid H_0\}) \tag{6.18}$$

$$= \log(E\{e^{s\sum_{k=1}^{n}(a_k z_k^2 + b_k z_k)} e^{sC} \mid H_0\})$$

$$= sC + \log(E\{\prod_{k=1}^{n} e^{s(a_k z_k^2 + b_k z_k)} \mid H_0\})$$

$$= sC + \log(\prod_{k=1}^{n} E\{e^{s(a_k z_k^2 + b_k z_k)} \mid H_0\})$$

$$= sC + \sum_{k=1}^{n} \log(E\{e^{s(a_k z_k^2 + b_k z_k)} \mid H_0\})$$

and

$$E\{e^{s(a_k z_k^2 + b_k z_k)} \mid H_0\}$$

$$= \int_{-\infty}^{\infty} e^{s(a_k z_k^2 + b_k z_k)} \frac{1}{\sqrt{2\pi}} e^{\frac{(z_k - m_{0k})^2}{2}} dz_k$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}z_k^2 + sa_k z_k^2 + m_{0k} z_k + sb_k z_k - \frac{1}{2}m_{0k}^2} dz_k$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}[(1-2sa_k)z_k^2 - 2(m_{0k}+sb_k)z_k + m_{0k}^2]} dz_k$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{z_k^2 - 2\frac{m_{0k}+sb_k}{1-2sa_k}z_k + \frac{m_{0k}^2}{1-2sa_k}}{2(1-2sa_k)^{-1}}} dz_k$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{(z_k - \frac{m_{0k}+sb_k}{1-2sa_k})^2 - (\frac{m_{0k}+sb_k}{1-2sa_k})^2 + \frac{m_{0k}^2}{1-2sa_k}}{2[(1-2sa_k)^{-\frac{1}{2}}]^2}} dz_k$$

$$= \frac{1(1-2sa_k)^{-\frac{1}{2}}}{\sqrt{2\pi}(1-2sa_k)^{-\frac{1}{2}}} \int_{-\infty}^{\infty} \frac{(z_k - \frac{m_{0k}+sb_k}{1-2sa_k})^2}{2[(1-2sa_k)^{-\frac{1}{2}}]^2} dz_k e^{\frac{(1-2sa)m_{0k}^2 - (m_{0k}+sb_k)^2}{2(1-2sa_k)}}$$

$$= (1-2sa_k)^{-\frac{1}{2}} e^{\frac{2sa_k m_{0k}^2 + 2sb_k m_{0k} + s^2 b_k^2}{2(1-2sa_k)}} \tag{6.19}$$

if $\pi_0 = \pi_1 = \frac{1}{2}$, then

$$p_e \le \frac{1}{2} e^{\mu_{T,0}(s)}$$

$$p_e \le \frac{1}{2} e^{\{sC + \sum_{k=1}^{n} \log[(1-2sa_k)^{-\frac{1}{2}} e^{\frac{2sa_k m_{0k}^2 + 2sb_k m_{0k} + s^2 b_k^2}{2(1-2sa_k)}}]\}}$$

$$P_e \le \frac{1}{2} e^{\{Cs + \sum_{k=1}^{n} [-\frac{1}{2}\log(1-2a_k s) + \frac{2m_{0k}(a_k m_{0k} + b_k)s + b_k^2 s^2}{2(1-2a_k s)}]\}}, \quad 0 \le s \le 1. \tag{6.20}$$

The bound is minimized by the value s_0 solving

$$\frac{\partial u_{T,0}(s)}{\partial s} = 0$$

$$= C + \sum_{k=1}^{n} [\frac{a_k}{1-2sa_k} + \frac{(2a_k m_{ok}^2 + 2b_k m_{0k} + 2sb_k^2)2(1-2sa_k)}{4(1-2sa_k)^2} - \frac{(2sa_k m_{0k}^2 + 2sb_k m_{0k} + s^2 b_k^2)(-4a_k)}{4(1-2sa_k)^2}]$$

$$= C + \sum_{k=1}^{n} [\frac{a_k}{1-2sa_k} + \frac{(a_k m_{0k}^2 + b_k m_{0k} + sb_k^2)(1-2sa_k) + (2sa_k m_{0k}^2 + 2sb_k m_{0k} + s^2 b_k^2)4a_k}{(1-2sa_k)^2}]$$

$$= C + \sum_{k=1}^{n} \frac{(a_k + b_k m_{0k} + a_k m_{0k}^2) + (b_k^2 - 2a_k^2)s - a_k b_k^2 s^2}{(1-2a_k s)^2} \tag{6.21}$$

Therefore, the optimum s_0 can be solved numerically. For the i.i.d. case,

$$\begin{aligned} &a_1 = a_2 = \ldots = a_n = a, \\ &b_1 = b_2 = \ldots = b_n = b, \\ &m_{01} = m_{02} = \ldots = m_{0n} = m_0 \end{aligned} \tag{6.22}$$

the optimal s_0 can be determined analytically as follows.

$$\begin{aligned} 0 &= C(1-2as)^2 + n[(a + bm_o + am_0^2) + (b^2 - 2a^2)s - ab^2 s^2] \\ &= c(1 - 4as + 4a^2 s^2) + n(a + bm_0 + am_0^2) + n(b^2 - 2a^2)s - nab^2 s^2 \\ &= (4a^2 C - nab^2)s^2 - [4aC - n(b^2 - 2a^2)]s + [C + n(a + bm_0 + am_0^2)] \end{aligned} \tag{6.23}$$

Therefore,

$$s_0 = \frac{[4aC - n(b^2 - 2a^2)] \pm \{[4aC - n(b^2 - 2a^2)]^2 - 4(4a^2 C - nab^2)[C + n(a + bm_0 + am_0^2)]\}^{\frac{1}{2}}}{2(4a^2 C - nab^2)}, \quad 0 \le s_0 \le 1. \tag{6.24}$$

For other cases, we resort to numerical computation.

For the LQ detector, there are several other bounds that can be found in closed form. Kailath obtained the J-divergence and the Bhattacharyya distance for problem (a) in [94] respectively,

$$\begin{aligned} J &= \int_{\Gamma} (L-1)\log(L) p_0 dx \\ &= \tfrac{1}{2} trace[(\boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}_1)(\boldsymbol{\Sigma}_1^{-1} - \boldsymbol{\Sigma}_0^{-1}) + (\boldsymbol{\Sigma}_0^{-1} + \boldsymbol{\Sigma}_1^{-1})(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1)(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1)^*] \end{aligned} \tag{6.25}$$

and

$$\begin{aligned} B &= -\log\left(\int_{\Gamma}(p_0 p_1)^{\frac{1}{2}} d\mu\right) \\ &= \tfrac{1}{8}(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1)^{*}\left(\tfrac{\Sigma_0+\Sigma_1}{2}\right)^{-1}(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1) + \tfrac{1}{2}\log\left(\frac{\left|\frac{\Sigma_0+\Sigma_1}{2}\right|}{(|\Sigma_0||\Sigma_1|)^{\frac{1}{2}}}\right) \end{aligned} \tag{6.26}$$

It can be shown in general that $J \geq 8B$. Corresponding bounds were introduced by Kobayahsi and Thomas in [45],

$$P_e > \pi_0 \pi_1 e^{-J/2} \tag{6.27}$$

and

$$\pi_0 \pi_1 e^{-2B} \leq Pe \leq (\pi_0 \pi_1)^{\frac{1}{2}} e^{-B} \tag{6.28}$$

It is noticed that the Bhattacharyya upper bound is the particular case of the Chernoff bound with s=1/2.

We compute the above bounds and compare them with the simulated probability of bit error of the LQ detector in Figure 6.1 – 6.4. These figures illustrate that all bounds behave well at lower SNR. The Bhattachayya lower and upper bounds actually bound the simulated probability of bit error. The Chernoff bound bounds from above and J-divergence bound bounds from below. The Bhattacharyya lower bound is tighter than the J-divergence bound, which coincides with the theoretical analysis. With increasing SNR, the upper bounds appear lower than the simulated BER. We conjecture that this is caused by average. That is, we average the signal sets in order to compute those bounds since the channel is time varying. The sample mean is probably greater than the actual mean at high SNR. Despite this slight inconsistency, those bounds can work as good benchmarks to validate the simulation work.

The simulation results also imply that the Chernoff or Bhattacharyya bounds might be better cost functions compare to J-divergence. We need further explain why we use J-

divergence in this investigation. As discussed in Chapter 4, we intended to find the optimal receiver structure and signal set simultaneously by maximizing the modified deflection ratio. It turns out that the J-divergence is the maximum value of the modified deflection ratio that leads to the optimal Gaussian detector for a fixed signal structure. After the optimal detector is found, our adaptive algorithm maximizes the J-divergence to identify the optimal signal set. Furthermore, assuming that the receiver structure is given, the gradient of J-divergence can be straightforwardly obtained. We used a recursive gradient search procedure to alternate between the receiver structure and the signal set to find the pair that is jointly optimal in a certain sense. Hence, J-divergence is employed in this investigation not only for signal selection, but also for receiver design. Plus, J-divergence is well known to be a useful criterion for signal selection. As for Chernoff or Bhattacharyya bound, even though it is tighter than J-divergence bound and its analytical closed form is available, the efficient algorithm to maximize it as a cost function for receiver design and signal selection is still a challenging open problem.

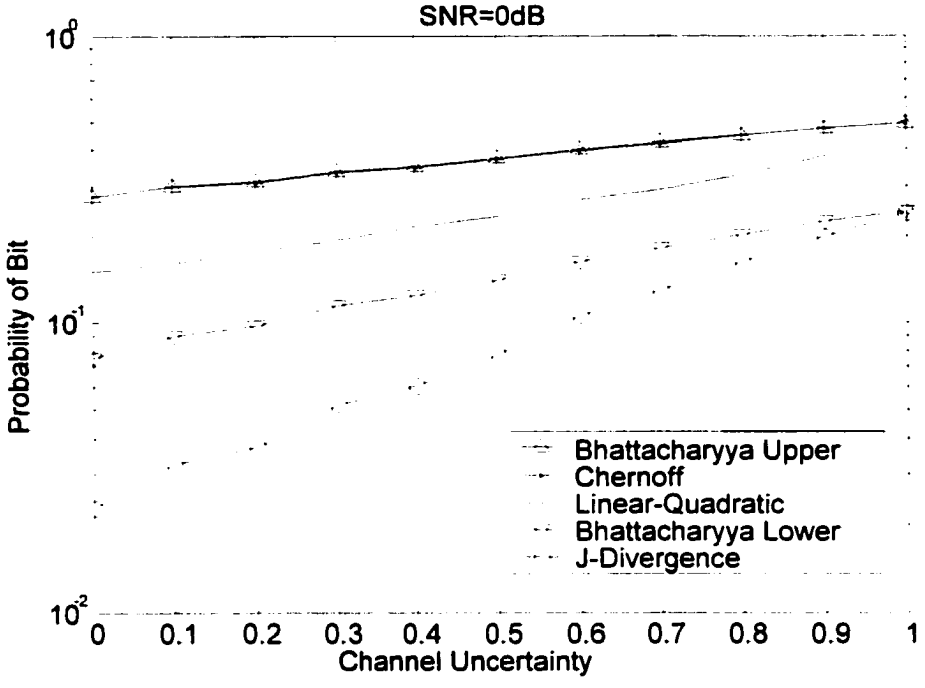

Figure 6.1 Performance Analysis of LQ Detectors (SNR=0dB)

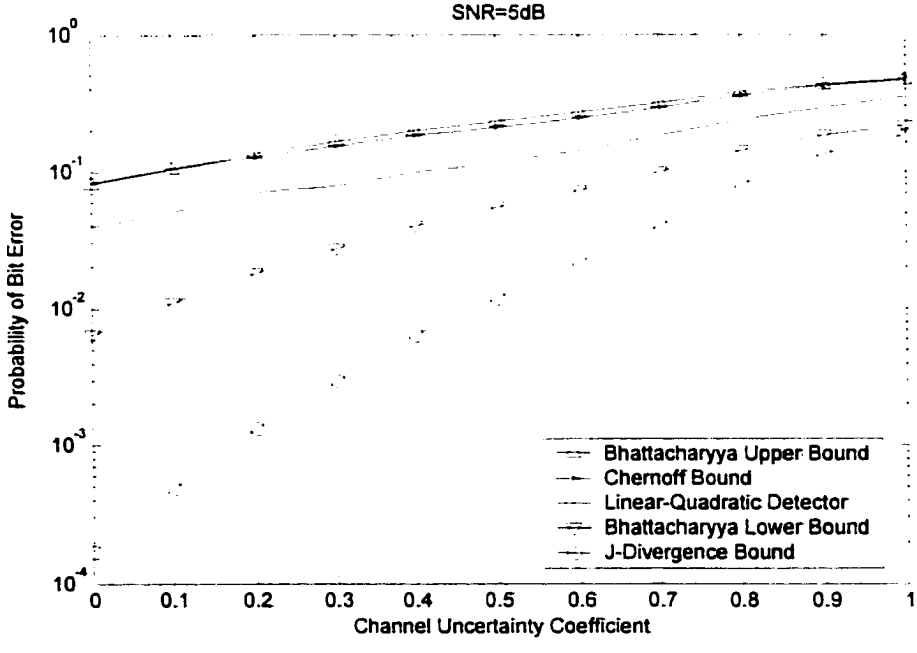

Figure 6.2 Performance Analysis of LQ Detectors (SNR=5dB)

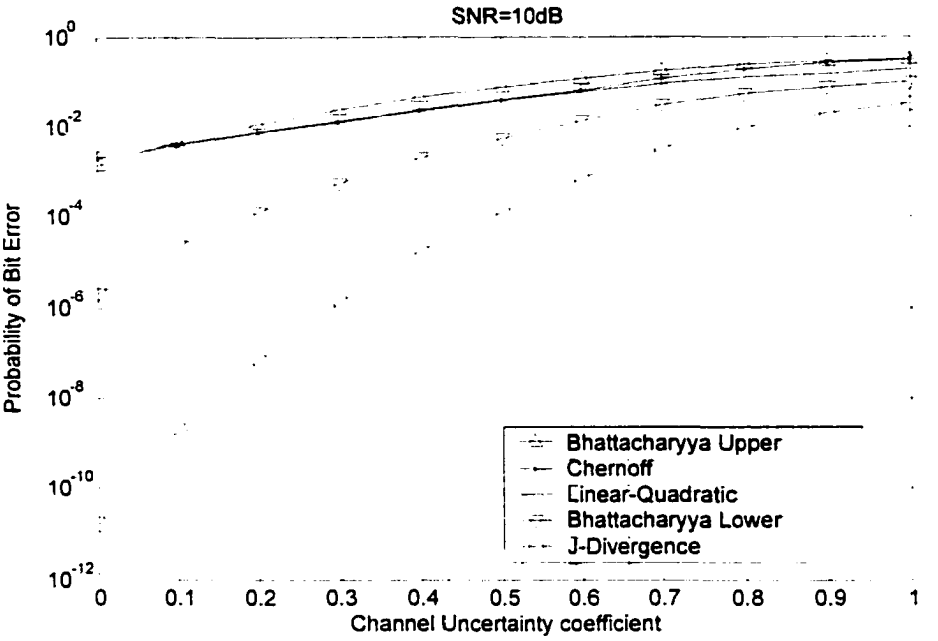

Figure 6.3 Performance Analysis of LQ Detectors (SNR=10dB)

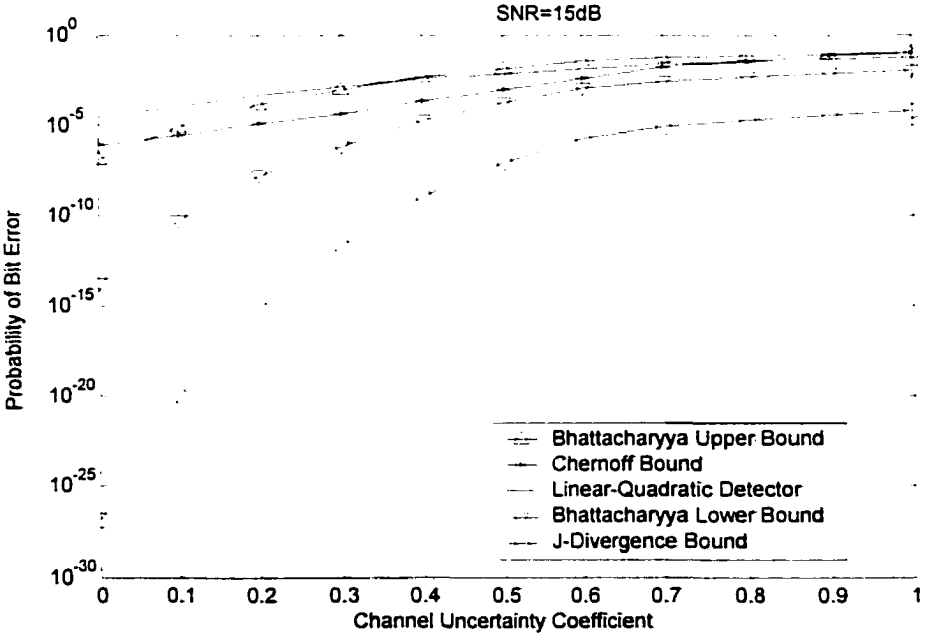

Figure 6.4 Performance Analysis of LQ Detectors (SNR=15dB)

Chapter 7 M-ary Signal Constellations

So far, our investigation has focused on adaptive modulation and receiver design for binary signals. In order to support high date rate communications, the further study of the corresponding design strategy is necessary.

In this chapter, we extend the design schemes in Chapter 4 and Chapter 5 from binary signals to M-ary signal constellations. First, we study adaptive modulation and receiver design for M-ary signals in a two dimensional signal subspace. Later, in order to enhance the system performance, the signal subspace is extended to multi-dimension based on the information theory perspective on signal dimension expansion.

7.1 Adaptive Multicoding in M-ary Signal Constellations

When the channel is in a shallow fade, a transmission with higher data rate can be supported. Hence, we can transmit signals with a larger constellation size. In this section, we extend our approach developed in Chapter 4 and Chapter 5 from binary signals to M-ary signals.

The observation vector at the receiver remains the same form given in (3.9):

$$\mathbf{x} = \mathbf{S}\hat{\boldsymbol{\alpha}} + \mathbf{S}\boldsymbol{\varepsilon} + \boldsymbol{\iota} + \boldsymbol{\gamma} \tag{7.1}$$

We can again transform the observation vector into a zero-mean observation vector as follows:

$$\mathbf{y} = \mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}} - \bar{\mathbf{C}}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\mu}_{\hat{\alpha}}) \tag{7.2}$$

where

$$\bar{\mathbf{C}} = \frac{1}{M}\sum_{i=0}^{M-1}\mathbf{C}_i$$

Under the assumption that the aggregate additive interference is Gaussian, the problem is equivalent to the following M-ary hypothesis testing problem [102]:

$$H_i : \mathbf{y} \sim \mathbb{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), \quad i \in \{0,1,...,M-1\} \tag{d}$$

Given the transmitted bit value i and the estimated CSI vector $\hat{\boldsymbol{\alpha}}$, the observation vector has the conditional mean vector

$$\boldsymbol{\mu}_{\mathbf{y}|i,\hat{\alpha}} = \left(\mathbf{C}_i - \bar{\mathbf{C}}\right)\hat{\boldsymbol{\alpha}}, \quad i \in \{0,1,...,M-1\} \tag{7.3}$$

and conditional covariance matrix

$$\boldsymbol{\Sigma}_{\mathbf{y}|b,\hat{\alpha}} = \mathbf{C}_i\boldsymbol{\Sigma}_{\varepsilon}\mathbf{C}_i^{*} + \boldsymbol{\Sigma}_{\iota} + \boldsymbol{\Sigma}_{\gamma}, \quad i \in \{0,1,...,M-1\}. \tag{7.4}$$

If all symbols are equally likely, the maximum likelihood detector is equivalent to the maximum a post-priori (MAP) detector. We can take hypothesis H_0 as a benchmark, and the MAP detector takes the form:

$$\delta(\mathbf{y}) = \arg\{\max_{j\in\{1,2,...,M-1\}}\{0,\ \log L_{0j}(\mathbf{y})\}\} \tag{7.5}$$

where $L_{0j}(\mathbf{y})$ is the likelihood ratio between H_j and H_0. In fact, the MAP detector consists of $M-1$ LQ detectors. By means of the same adaptive multicoding criteria, the optimal signal structure can be obtained by maximizing the minimum divergence distance between $\binom{M}{2}$ signal pairs in the M-ary constellations; that is,

$$\max_{\{\mathbf{c}_0,\mathbf{c}_1,\ldots,\mathbf{c}_{M-1}\}} \min_{i\neq j} D^*(\mathbf{c}_i,\mathbf{c}_j) = [(\boldsymbol{\mu}_j-\boldsymbol{\mu}_i)^*\left(\boldsymbol{\Sigma}_i^{-1}+\boldsymbol{\Sigma}_j^{-1}\right)(\boldsymbol{\mu}_j-\boldsymbol{\mu}_i)+\mathrm{Tr}\left(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j^{-1}+\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}_i^{-1}-2\mathbf{I}\right)] \qquad (7.6)$$

where $i,j\in\{0,1,\ldots,M-1\}$.

The above adaptive modulation scheme has been simulated for M=4 in a two-dimensional signal subspace and compared with the standard QPSK scheme in Figure 7.2-7.5. Under the power constraint that the average symbol energy is not greater than unity, the adaptive scheme gives almost the same performance as QPSK for lower SNR. As the SNR increases, the two performance curves diverge. The adaptive scheme outperforms QPSK at the low and high uncertainty ends while QPSK outperforms the adaptive scheme in the medium uncertainty range. The simulation results indicate that there is not much benefit for adaptation compared to QPSK in a two-dimensional signal subspace. In addition, the claims in [91, 94] that J-divergence is not a good indicator of probability of error at high SNR are supported here, as the fixed QPSK constellation outperforms the optimal adaptive constellation at intermediate values of the uncertainty parameter.

At the end of this section, we present the optimal quaternary signal constellation under different levels of channel uncertainty. Figure 7.1 illustrates that the optimal constellation has approximately the same structure of QPSK. Recall that for binary signals, the signal constellation was adapted towards antipodal signals for coherent channels and orthogonal signals for noncoherent channels. QPSK signaling on the unit circle is equivalent to two pairs of antipodal signals that are orthogonal to each other. It is the optimal signal

constellation in a Euclidean sense and turns out to be very nearly optimal also for an adaptive scheme at both ends of the uncertainty range. Adaptation in such a scenario has little benefit compared to QPSK.

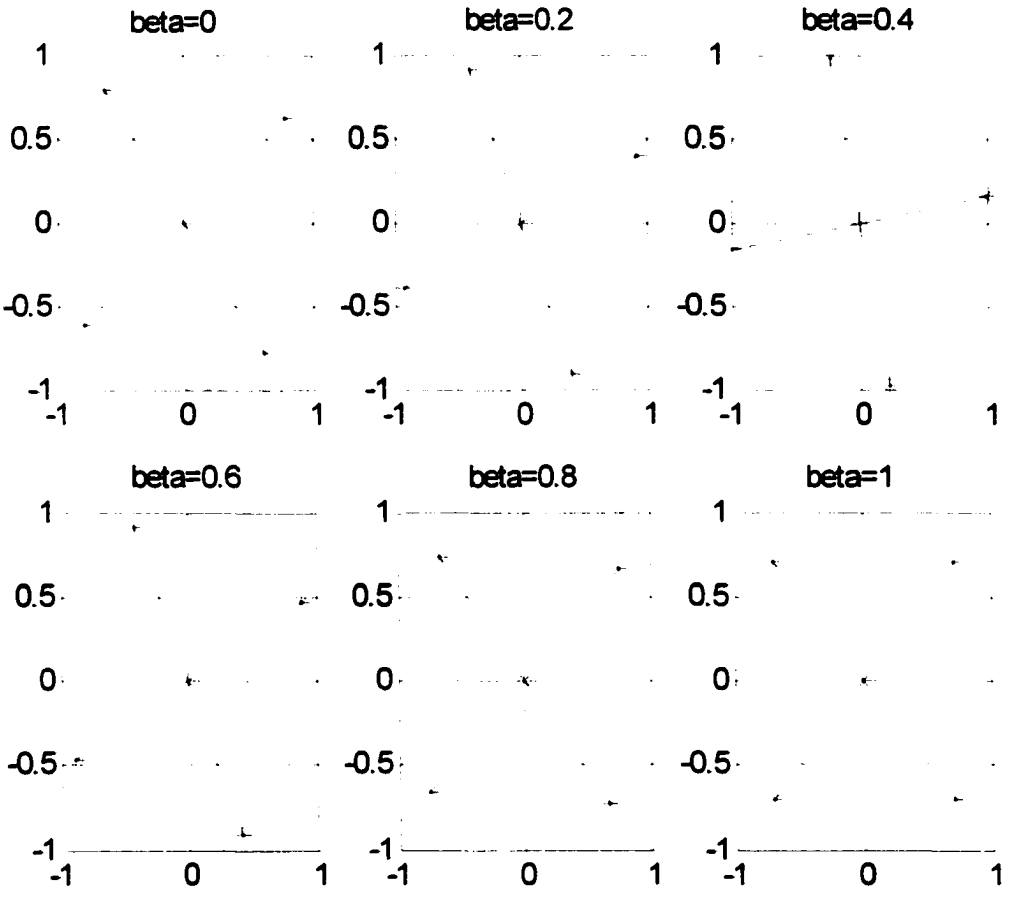

Figure 7.1 Optimal Quaternary Signal Constellations under Uncertainties

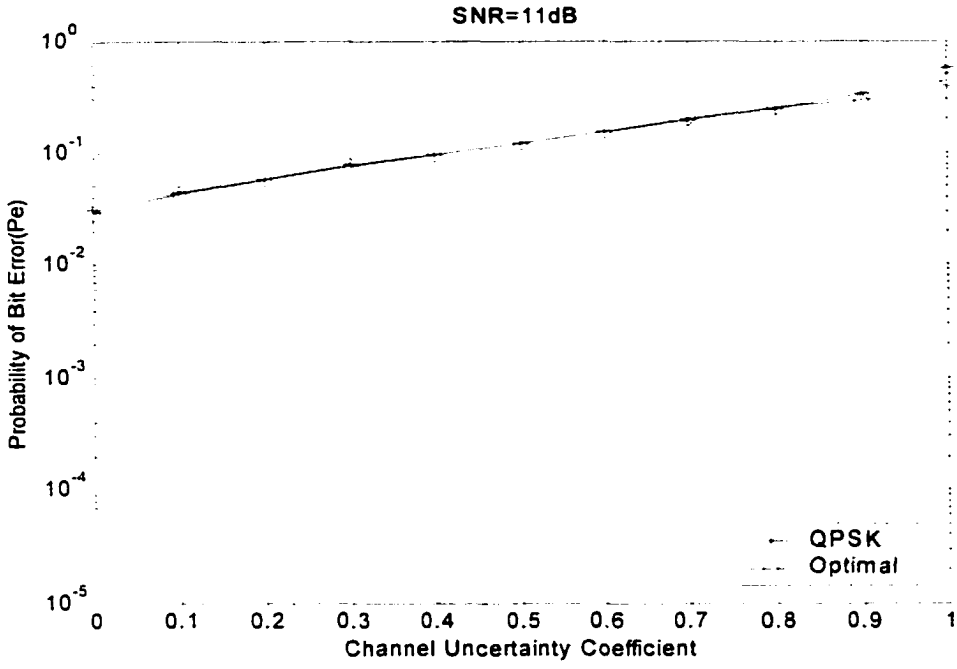

Figure 7.2 OPSK vs. Optimal Four Signal Scheme for SNR=11dB

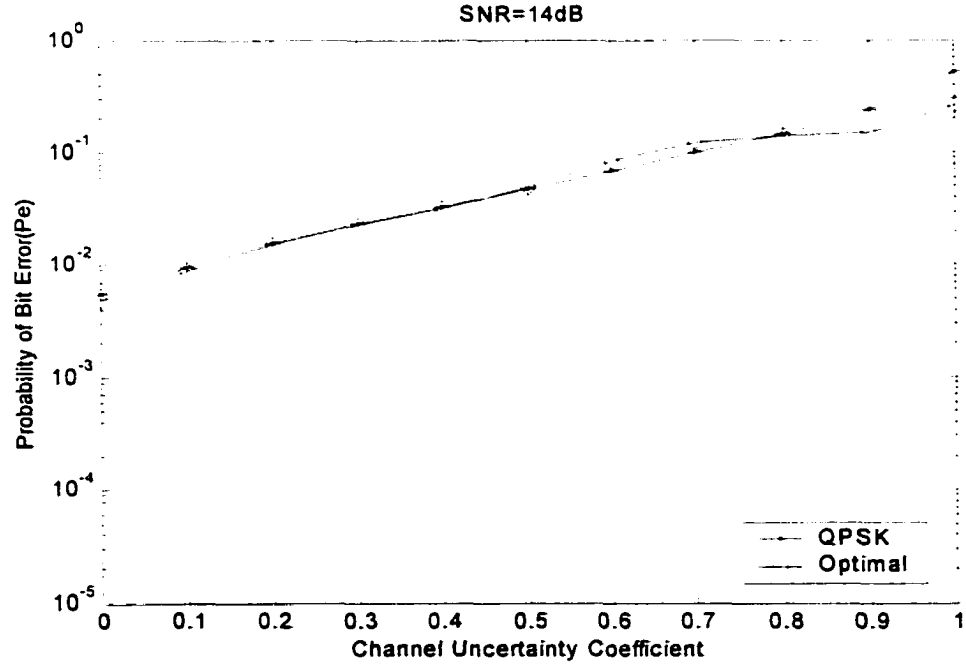

Figure 7.3 QPSK vs. Optimal Four Signal Scheme for SNR=14dB

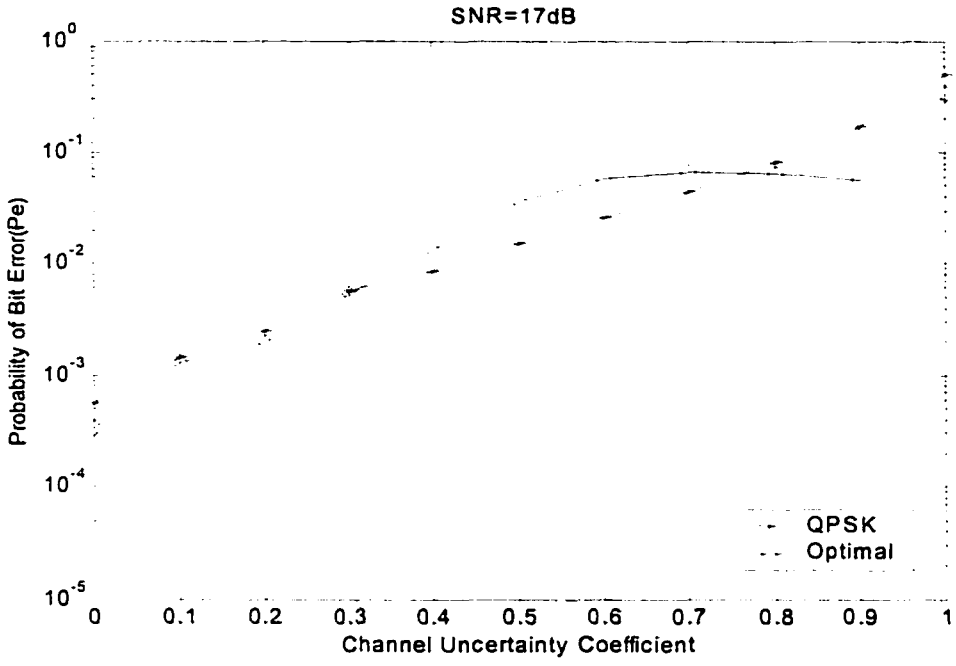

Figure 7.4 QPSK vs. Optimal Four Signal Scheme for SNR=17dB

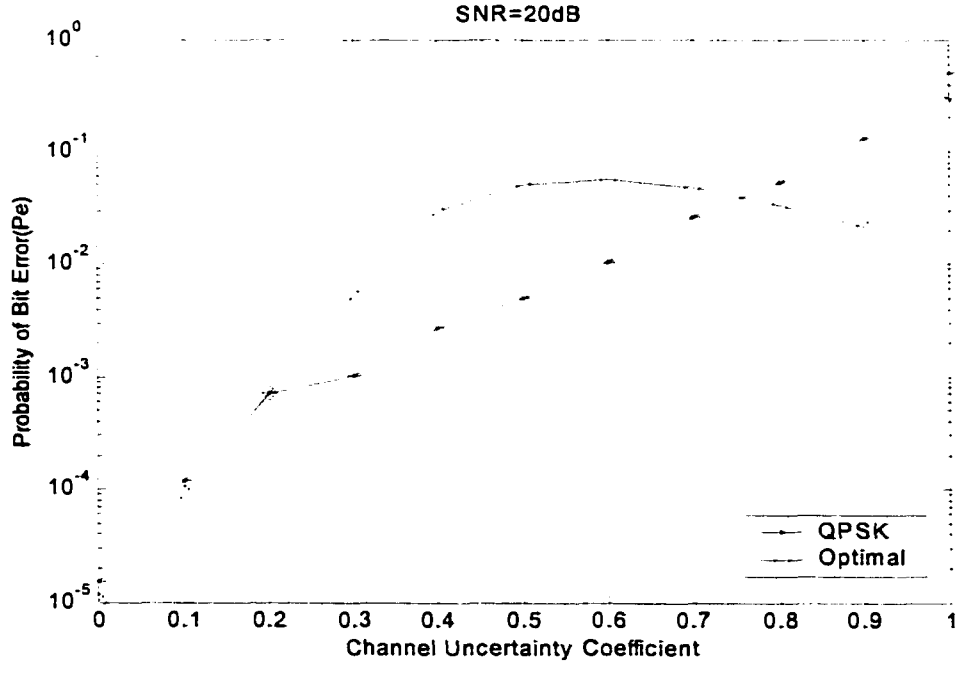

Figure 7.5 QPSK vs. Optimal Four Signal Scheme for SNR=20dB

7.2 Information Theory Perspective on Dimension Expansion

In section 7.1, simulation results demonstrate that adaptive multicoding for quaternary signals in a two-dimension subspace doesn't improve the system performance significantly as it does for binary signals. We conjecture that if we increase the dimensionality of the signal subspace, an adaptive multicoding scheme will once again enhance the system performance. Barton's recent work in [44] provides some theoretical support for this conjecture on information theory perspective.

In [44], an asymptotic analysis of the information capacity of wideband CDMA fading channels has been accomplished under a scenario similar to that of this dissertation. In that work, several useful upper and lower bounds on mutual information between channel input and output conditioned on an observed CSI estimate have been developed for channels with large spreading gain. These bounds are stated explicitly in terms of the dimension of the signal constellation, the number of resolvable fading paths on the channel, the current estimate of channel state, and the mean-squared-error of the channel estimate. It has been demonstrated that the dimension of the signal constellation has a significant impact on the mutual information on the channel and that the dimension necessary to achieve the maximum mutual information depends critically on the accuracy of the CSI estimates and the number of resolvable paths on the channel. For a wideband noncoherent channel, if the geometry of the constellation is adapted in a particular manner so that both the dimension and the fourth moment of the constellation increase as the number of resolvable paths increase, the channel capacity can be made arbitrarily close to the capacity for a single-user AWGN channel with the same SNR. Further, it is shown that the maximum conditional mutual information on the channel in the special case of perfect CSI estimates is achieved with a Gaussian constellation, but only by letting the dimension of the constellation increase faster than the received SNR.

The above idea heuristically inspires us to expand the subspace dimensionality for adaptive multicoding for M-ary signals. Intuitively, we know that dimension-expansion can somehow increase the "distance" between signal pairs. For example, the maximin (i.e., the maximum of the minimum pair wise) Euclidean distance for quaternary signals on a unit circle is 1.414 and the corresponding signal constellation is equivalent to QPSK. When the signal dimension expands to three dimensions, the maximin Euclidean distance for quaternary signals on a unit sphere increase to 1.633. However, the maximin Euclidean distance for quaternary signals on a four-dimension or higher dimension unit sphere is still 1.633. Indeed, it is straightforward to show that the maximin Euclidean distance for M-ary signals is achieved on an (M-1) dimension unit sphere where equal pair-wise distance can be obtained. Here, what we are really interested is not Euclidean distance but J-divergence. By expanding the dimension of the signal subspace, we gain more freedom for adaptive multicoding based on J-divergence. In addition, simulation results demonstrate that the maximin J-divergence increases with the increase of the signal dimensionality. Therefore, it is of interest to investigate to what extent the dimension expansion of the signal subspace can enhance the system performance.

7.3 Simulation Results

In this section, we present simulation results to evaluate the adaptive multicoding scheme for M-ary signals described in section 7.1 by expanding the dimensionality of the signal subspace. Due to computational complexity, we only expand the two dimensions to three and four dimensions. Again, we consider a quaternary signal set, that is, M=4. For a three- or four-dimensional signal subspace, we need to search the three- or four-dimensional sphere to find the optimal 12 or 16 coordinates corresponding to the four signals that

maximize the minimum pair-wise J-divergence. In this simulation, uniform channel fading is assumed, which means that the covariance matrix of the channel is an identity matrix. The spreading gain is N=31. We simulate several channels corresponding to different numbers of resolvable paths on the channel including L=1, 5, and 15. It is expected to see some impact on the system performance through the signal subspace expansion. In Figures 7.6-7.8, the performance comparison of dimension expansion for SNR=17dB and L=1, 5, and 15 is illustrated, respectively. The result for lower SNR case (SNR=7dB) is demonstrated in Figure 7.9. For ease of interpretation, the curves are labeled as QPSK-2D (3D, 4D) and Optimal-2D (3D, 4D), which denote the maximin Euclidean distance scheme and the adaptive multicoding approach in designated subspace, respectively.

By inspecting the simulation results, several observations can be made:

1. Regardless of dimension, adaptive multicoding outperforms QPSK for the two extremes – completely coherent channels and completely noncoherent channels.
2. The performance of QPSK is enhanced by expanding dimensionality from two to three. The reason is that the maximum maximin Euclidean distance for four signals is achieved in the three-dimension subspace. No more performance improvement can be made when dimensionality reaches three or higher. Due to the inevitable singularity in simulation, we didn't plot the QPSK-4D curve. But it is expected to be the same as the QPSK-3D curve.
3. For adaptive multicoding based on J-divergence, more dimensionality, more improvement, but an asymptotical limit is expected.
4. Adaptive multicoding in a four-dimensional subspace overall outperforms the same scheme in a two-dimensional subspace. In particular, when channel uncertainty is high, significant performance gain is obtained. This indicates that dimension expansion of the signal subspace can enhance the system performance.

5. It is also observed that at intermediate level of uncertainty, QPSK in a three-dimension subspace consistently outperforms adaptive multicoding in a four-dimensional subspace. Again, this is a negative result probably due to the mismatch between J-divergence and probability of error at high SNR. The results at lower SNR (SNR=7dB) shown in Figure 7.9 validate this conjecture.

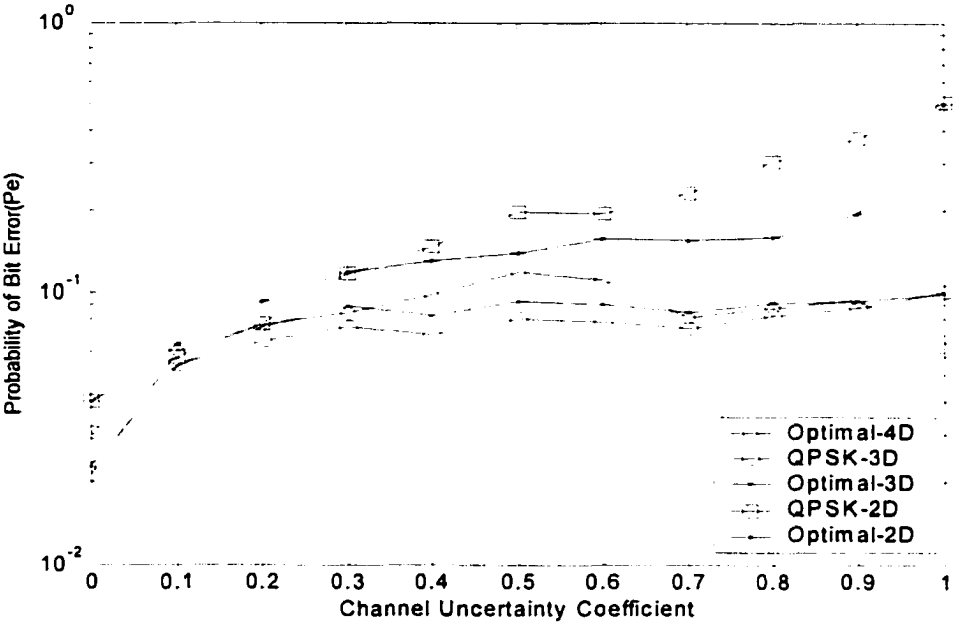

Figure 7.6 Performance Comparison of Dimension Expansion for L=1& SNR=17dB

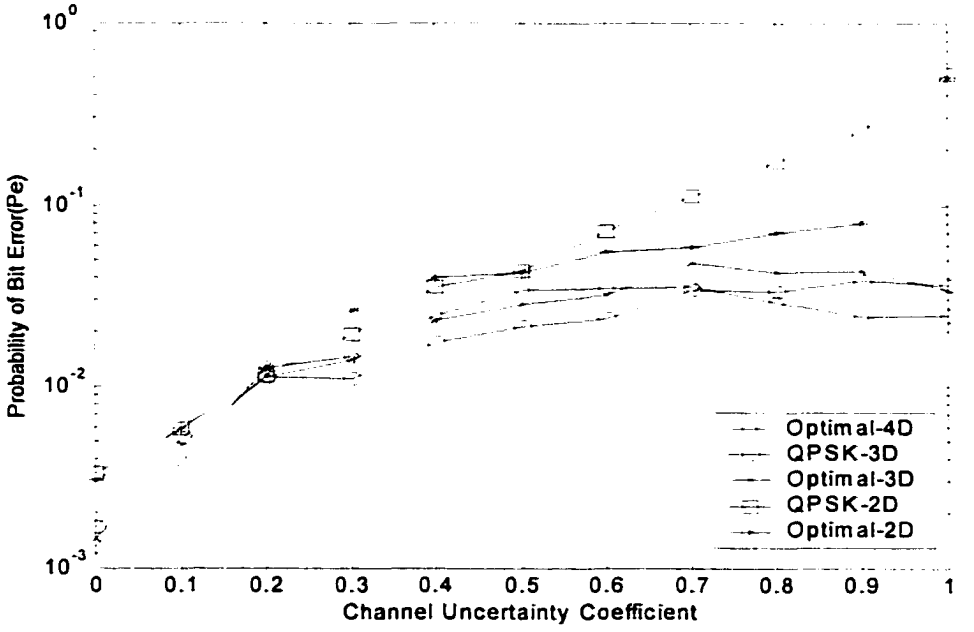

Figure 7.7 Performance Comparison of Dimension Expansion for L=5& SNR=17dB

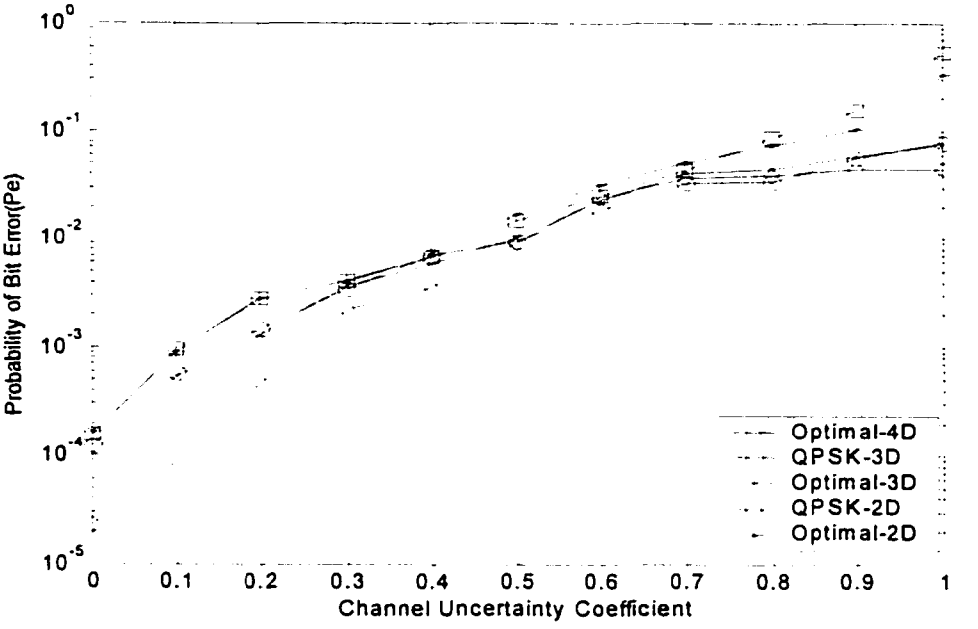

Figure 7.8 Performance Comparison of Dimension Expansion for L=15& SNR=17dB

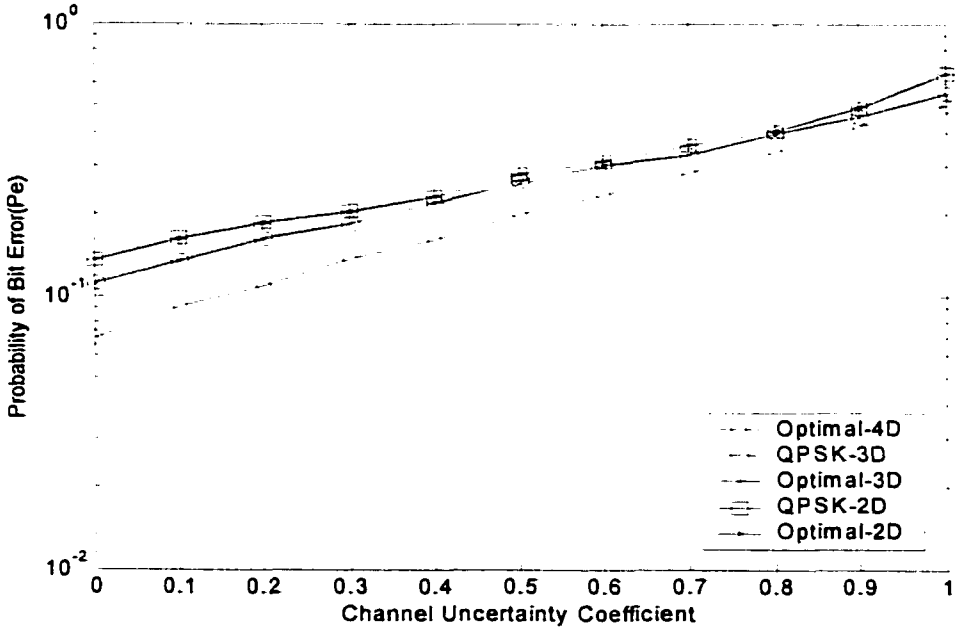

Figure 7.9 Performance Comparison of Dimension Expansion for L=15& SNR=7dB

Chapter 8 Summary and Conclusions

In this dissertation, we have investigated the problem of adaptive modulation and receiver design in the presence of channel uncertainty on a mobile wireless channel. The primary objectives were to identify and analyze robust LQ receivers and a corresponding adaptive multicoding scheme in such an environment.

Based on a detailed review of the relevant literature, there are several unique contributions in this dissertation. In addition, there are several interesting problems that have emerged. In this final chapter, we summarize the unique contributions and recommend new areas of challenging research.

8.1 Summary of Unique Contributions

1. **Adaptive LQ Receivers** In chapter 4, we identified and analyzed the robust receivers based on a channel model that incorporates explicitly all significant sources of interference as well as substantial channel uncertainty. First, the conditional statistics for a binary hypothesis testing problem were derived and discussed. It is evident that the availability of CSI can be exploited at the receiver by including a linear component in the detector, and the processing gain associated with the linear detector component will be maximized if the signals are antipodal. On the other hand, in the presence of channel uncertainty the structure of the uncertainty can actually be exploited by the receiver by including a quadratic component in the detector, but only if the signal structure is chosen appropriately. In order to exploit both CSI and the structure of the channel uncertainty, we proposed and developed adaptive LQ receivers. Next, we proposed a simple, intuitively

appealing cost function that can be maximized to find LQ receivers that are optimal in a certain sense. Finally, we discussed the properties of the proposed LQ cost function and derived an efficient adaptive algorithm by which the optimum LQ receiver and the corresponding optimal signal sets can be found simultaneously. In particular, this approach leads to a detector that is equivalent to the well-known MMSE linear detector if an antipodal signal structure is employed.

2. **Adaptive Multicoding** In Chapter 5, we employed J-divergence as a distance measure criterion to develop a novel adaptive modulation scheme which we refer to as adaptive multicoding. We adapt the signal constellation by searching over the two-dimension signal subspace spanned by two known basis signatures. Performance evaluation based on simulation results for binary signals shows that the adaptive LQ detector outperforms the linear detector. This indicates not only that the J-divergence is a good criterion for signal selection for this problem, but also that the recursive LMS algorithm converges to a nearly optimal solution. Furthermore, the LQ detector becomes strictly better than the linear detector as the uncertainty on the channel increases at a fixed SNR. Finally, as the SNR on the channel increases, the level of uncertainty at which the LQ detector begins to significantly outperform the linear detector converges to zero. In addition, the analysis illustrates that the optimal signal pair varies between antipodal signals and orthogonal signals corresponding to coherent channels and noncoherent channels.

3. **Performance Analysis of LQ Receivers** In Chapter 6, the Chernoff bound for the LQ receiver was derived for performance analysis since no closed form of probability of error can be obtained. The numerically computed Chernoff bound and other associated bounds were compared to the simulated probability of bit error of the LQ receiver. This analysis confirms that J-divergence is a reasonable approximation at low SNR but less appropriate at high SNR.

4. **M-ary Signal Constellations** In Chapter 7, we extended the approach developed in chapter 2 for binary signals to M-ary signals. In this scenario, the optimal detector consists of M-1 LQ detectors. The optimal signal structure can be obtained by maximizing the minimum divergence distance between $\binom{M}{2}$ signal pairs in M-ary constellations. The simulation results for quaternary signals indicate that the adaptive scheme gives almost the same performance as QPSK does in a two-dimensional signal subspace for lower SNR. By further expanding the signal subspace to a higher dimensional subspace, the adaptive multicoding scheme overall outperforms QPSK. This indicates that adaptive multicoding can enhance the system performance through the dimension expansion of the signal subspace.

8.2 Open Areas of Research

During the course of this research, several aspects of adaptive LQ receivers became apparent which are worthy of further study. These are summarized below.

1. **Performance Evaluation of Binary LQ Receivers**. A comprehensive performance evaluation of binary LQ receivers will be performed in order to extend and complete the preliminary results discussed above. Of particular interest will be the effect of MAI on the performance of the receivers and the determination of performance characteristics such as asymptotic multiuser efficiency and near-far resistance for the proposed class of LQ receivers.

2. **Receiver Sensitivity**. The sensitivity of LQ receivers to estimates of channel statistics will be studied. Although it has been demonstrated that the proposed adaptive LQ receivers are much less sensitive to errors in instantaneous CSI estimates than linear MMSE receivers, the performance of LQ detectors in the presence of errors in estimates of channel covariance structure remains to be determined. Of particular interest will be errors in the

estimates of Σ_ϵ and Σ_γ, which will be the most difficult components of channel covariance structure to determine precisely.

3. **Adaptive Algorithms**. The convergence properties of adaptive algorithms for LQ signal and receiver design will be studied. Adaptive algorithms based on RLS and subspace tracking methods will be investigated in addition to the LMS approach currently under study.

4. **Dimension Expansion** Due to computational complexity, we only investigated expansion of the signal subspace to four dimensions with spreading gain 31. If faster adaptive algorithms become available, higher dimensionality and higher spreading gain will be simulated to determine the relationship to the theoretical analysis presented in [44]. In addition, it is of interest to find the analytical proof that the maximin J-divergence increases with the increase of the signal dimensionality.

5. **Distance Criterion** J-divergence is used as the distance criterion in this investigation due to its tractability. However, it is not closely correlated to probability of error at high SNR. Other distance criteria will be studied. Bhattacharyya distance would be a good candidate if a corresponding optimization scheme could be found.

Appendix A. J-divergence between Two Gaussian Distributions

The problem of interest is (a') described in chapter 4.

$$H_0 : \mathbf{y} \sim \mathbb{N}(-\boldsymbol{\mu}, \boldsymbol{\Sigma}_0),$$

verse

$$H_1 : \mathbf{y} \sim \mathbb{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_1).$$

Under H_0:

$$p_0(\mathbf{y}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\boldsymbol{\Sigma}_0|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{y}+\boldsymbol{\mu})^{*} \Sigma_0^{-1}(\mathbf{y}+\boldsymbol{\mu})\right)$$

Under H_1:

$$p_1(\mathbf{y}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\boldsymbol{\Sigma}_0|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^{*} \Sigma_1^{-1}(\mathbf{y}-\boldsymbol{\mu})\right)$$

First, compute Kullback-Leibler distances $KL(p_0, p_1)$ and $KL(p_1, p_0)$

$$KL(p_0, p_1) = E_0\{\log \frac{p_0(\mathbf{y})}{p_1(\mathbf{y})}\}$$

$$KL(p_1, p_0) = E_1\{\log \frac{p_1(\mathbf{y})}{p_0(\mathbf{y})}\}$$

$$= -E_1\{\log \frac{p_0(\mathbf{y})}{p_1(\mathbf{y})}\}$$

$$\frac{p_0(\mathbf{y})}{p_1(\mathbf{y})} = \frac{|\boldsymbol{\Sigma}_1|^{\frac{1}{2}}}{|\boldsymbol{\Sigma}_0|^{\frac{1}{2}}} \exp\left(\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^{*} \Sigma_1^{-1}(\mathbf{y}-\boldsymbol{\mu}) - \frac{1}{2}(\mathbf{y}+\mu)^{*} \Sigma_0^{-1}(\mathbf{y}+\boldsymbol{\mu})\right)$$

$$\log \frac{p_0(\mathbf{y})}{p_1(\mathbf{y})} = \log \frac{|\boldsymbol{\Sigma}_1|^{\frac{1}{2}}}{|\boldsymbol{\Sigma}_0|^{\frac{1}{2}}} + \frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^{*} \Sigma_1^{-1}(\mathbf{y}-\boldsymbol{\mu}) - \frac{1}{2}(\mathbf{y}+\boldsymbol{\mu})^{*} \Sigma_0^{-1}(\mathbf{y}+\boldsymbol{\mu})$$

$$KL(p_0, p_1) = E_0\{\log \frac{p_0(\mathbf{y})}{p_1(\mathbf{y})}\}$$

$$= E_0\{Tr(\log\frac{p_0(\mathbf{y})}{p_1(\mathbf{y})}\}$$

$$= E_0\{Tr(\log\frac{|\boldsymbol{\Sigma}_1|^{\frac{1}{2}}}{|\boldsymbol{\Sigma}_1|^{\frac{1}{2}}} + \frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^*\boldsymbol{\Sigma}_1^{-1}(\mathbf{y}-\boldsymbol{\mu}) - \frac{1}{2}(\mathbf{y}+\boldsymbol{\mu})^*\boldsymbol{\Sigma}_0^{-1}(\mathbf{y}+\boldsymbol{\mu}))\}$$

$$= \log\frac{|\boldsymbol{\Sigma}_1|^{\frac{1}{2}}}{|\boldsymbol{\Sigma}_0|^{\frac{1}{2}}} + \frac{1}{2}E_0\{Tr(\boldsymbol{\Sigma}_1^{-1}(\mathbf{y}-\boldsymbol{\mu})(\mathbf{y}-\boldsymbol{\mu})^* - \boldsymbol{\Sigma}_0^{-1}(\mathbf{y}+\boldsymbol{\mu})(\mathbf{y}+\boldsymbol{\mu})^*)\}$$

$$= \log\frac{|\boldsymbol{\Sigma}_1|^{\frac{1}{2}}}{|\boldsymbol{\Sigma}_0|^{\frac{1}{2}}} + \frac{1}{2}E_0\{Tr(\boldsymbol{\Sigma}_1^{-1}(\mathbf{y}+\boldsymbol{\mu}-2\boldsymbol{\mu})(\mathbf{y}+\boldsymbol{\mu}-2\boldsymbol{\mu})^* - \boldsymbol{\Sigma}_0^{-1}(\mathbf{y}+\boldsymbol{\mu})(\mathbf{y}+\boldsymbol{\mu})^*)\}$$

$$= \log\frac{|\boldsymbol{\Sigma}_1|^{\frac{1}{2}}}{|\boldsymbol{\Sigma}_0|^{\frac{1}{2}}} + \frac{1}{2}E_0\{Tr(\boldsymbol{\Sigma}_1^{-1}(\mathbf{y}+\boldsymbol{\mu})(\mathbf{y}+\boldsymbol{\mu})^* - 2\boldsymbol{\mu}^*(\mathbf{y}+\boldsymbol{\mu}) - 2\boldsymbol{\mu}(\mathbf{y}+\boldsymbol{\mu})^* + 4\boldsymbol{\mu}\boldsymbol{\mu}^*) - \boldsymbol{\Sigma}_0^{-1}(\mathbf{y}+\boldsymbol{\mu})(\mathbf{y}+\boldsymbol{\mu})^*\}$$

$$= \log\frac{|\boldsymbol{\Sigma}_1|^{\frac{1}{2}}}{|\boldsymbol{\Sigma}_0|^{\frac{1}{2}}} + \frac{1}{2}(Tr(\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_0 - 0 - 0 + 4\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^* - \boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_0))$$

$$= \log\frac{|\boldsymbol{\Sigma}_1|^{\frac{1}{2}}}{|\boldsymbol{\Sigma}_0|^{\frac{1}{2}}} + \frac{1}{2}Tr(\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_0 + 4\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^* - \mathbf{I})$$

$$KL(p_1, p_0) = -E_1\{\log\frac{p_0(\mathbf{y})}{p_1(\mathbf{y})}\}$$

$$= -E_1\{Tr(\log\frac{p_0(\mathbf{y})}{p_1(\mathbf{y})}\}$$

$$= -E_1\{Tr(\log\frac{|\boldsymbol{\Sigma}_1|^{\frac{1}{2}}}{|\boldsymbol{\Sigma}_0|^{\frac{1}{2}}} + \frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^*\boldsymbol{\Sigma}_1^{-1}(\mathbf{y}-\boldsymbol{\mu}) - \frac{1}{2}(\mathbf{y}+\boldsymbol{\mu})^*\boldsymbol{\Sigma}_0^{-1}(\mathbf{y}+\boldsymbol{\mu}))\}$$

$$= -\log\frac{|\boldsymbol{\Sigma}_1|^{\frac{1}{2}}}{|\boldsymbol{\Sigma}_0|^{\frac{1}{2}}} - \frac{1}{2}E_1\{Tr(\boldsymbol{\Sigma}_1^{-1}(\mathbf{y}-\boldsymbol{\mu})(\mathbf{y}-\boldsymbol{\mu})^* - \boldsymbol{\Sigma}_0^{-1}(\mathbf{y}+\boldsymbol{\mu})(\mathbf{y}+\boldsymbol{\mu})^*)\}$$

$$= -\log\frac{|\boldsymbol{\Sigma}_1|^{\frac{1}{2}}}{|\boldsymbol{\Sigma}_0|^{\frac{1}{2}}} - \frac{1}{2}E_1\{Tr(\boldsymbol{\Sigma}_1^{-1}(\mathbf{y}-\boldsymbol{\mu})(\mathbf{y}-\boldsymbol{\mu})^* - \boldsymbol{\Sigma}_0^{-1}(\mathbf{y}-\boldsymbol{\mu}+2\boldsymbol{\mu})(\mathbf{y}-\boldsymbol{\mu}+2\boldsymbol{\mu})^*)\}$$

$$= -\log \frac{|\boldsymbol{\Sigma}_1|^{\frac{1}{2}}}{|\boldsymbol{\Sigma}_0|^{\frac{1}{2}}} - \frac{1}{2} E_1\{Tr(\boldsymbol{\Sigma}_1^{-1}(\mathbf{y}-\boldsymbol{\mu})(\mathbf{y}-\boldsymbol{\mu})^* - \boldsymbol{\Sigma}_0^{-1}((\mathbf{y}-\boldsymbol{\mu})(\mathbf{y}-\boldsymbol{\mu})^* + 2\boldsymbol{\mu}^*(\mathbf{y}-\boldsymbol{\mu}) + 2\boldsymbol{\mu}(\mathbf{y}-\boldsymbol{\mu})^* + 4\boldsymbol{\mu}\boldsymbol{\mu}^*))\}$$

$$= -\log \frac{|\boldsymbol{\Sigma}_1|^{\frac{1}{2}}}{|\boldsymbol{\Sigma}_0|^{\frac{1}{2}}} - \frac{1}{2}(Tr(\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_1 - 0 - 0 - 4\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^*))$$

$$= -\log \frac{|\boldsymbol{\Sigma}_1|^{\frac{1}{2}}}{|\boldsymbol{\Sigma}_0|^{\frac{1}{2}}} - \frac{1}{2} Tr(\mathbf{I} - \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_1 - 4\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^*)$$

$$= -\log \frac{|\boldsymbol{\Sigma}_1|^{\frac{1}{2}}}{|\boldsymbol{\Sigma}_0|^{\frac{1}{2}}} + \frac{1}{2} Tr(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_1 + 4\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^* - \mathbf{I})$$

J-divergence is the sum of two KL distances.

$$J = KL(p_0, p_1) + KL(p_1, p_0)$$

$$= \frac{1}{2} Tr(\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_0 + 4\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^* - \mathbf{I} + \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_1 + 4\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}\boldsymbol{\mu}^* - \mathbf{I})$$

$$= \frac{1}{2}(4\boldsymbol{\mu}^*(\boldsymbol{\Sigma}_0^{-1} + \boldsymbol{\Sigma}_1^{-1})\boldsymbol{\mu} + Tr(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_0^{-1} - 2\mathbf{I})$$

$$= 2\boldsymbol{\mu}^*(\boldsymbol{\Sigma}_0^{-1} + \boldsymbol{\Sigma}_1^{-1})\boldsymbol{\mu} + \frac{1}{2} Tr(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_0^{-1} - 2\mathbf{I})$$

This result coincides with $D^*(\mathbf{c}_0, \mathbf{c}_1)$ in (4.11) by a scaling factor 2.

Appendix B. The Gradient of $D^*(\mathbf{c}_0, \mathbf{c}_1)$

Assume we have orthogonal matrix C_{0b} and C_{1b} as basis vectors, the define the global optimal vectors as:

$C_{00} = r_0 C_{0b} + r_1 C_{1b}$
where $r_0^2 + r_1^2 \le 1$

$C_{01} = r_2 C_{0b} + r_3 C_{1b}$
where $r_2^2 + r_3^2 \le 1$

or
$r_0^2 + r_1^2 + r_2^2 + r_3^2 \le 2$

Let
$C_{0p} = zero(N+L, L+1);$
$C_{1p} = zeros(N+L, L+1);$
$C_{0p} = (1:L,:) = C_{0b}(N+1:N+L,:)$
$C_{1p}(1:c,:) = C_{1b}(N+1:N+L,:)$
$C_{-10} = r_0 C_{0p} + r_1 C_{1p}$
$C_{-11} = r_2 C_{0p} + r_3 C_{1p}$

$C_{0n} = zeros(N+L, L+1)$
$C_{1n} = zeros(N+L, L+1)$
$C_{0n}(N+L:N+L,:) = C_{0b}(1:L,:)$
$C_{1n}(N+L:N+L,:) = C_{1b}(1:L,:)$
$C_{10} = r_0 C_{0n} + r_1 C_{1n}$
$C_{11} = r_2 C_{0n} + r_3 C_{1n}$

Figure out $\boldsymbol{\mu}, \boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}_1$

$$\boldsymbol{\mu} = \frac{1}{2}(C_{01} - C_{00})\hat{\boldsymbol{\alpha}}$$
$$= \frac{1}{2}[(r_2 - r_0)C_{0b} + (r_3 - r_1)C_{1b}]\hat{\boldsymbol{\alpha}}$$
$$\boldsymbol{\Sigma}_0 = C_{00}\boldsymbol{\Sigma}_\varepsilon C_{00}^{'} + \boldsymbol{\Sigma}_\iota + \sigma^2\mathbf{I}$$
$$\boldsymbol{\Sigma}_1 = C_{01}\boldsymbol{\Sigma}_\varepsilon C_{01}^{'} + \boldsymbol{\Sigma}_\iota + \sigma^2\mathbf{I}$$
$$\boldsymbol{\Sigma}_\varepsilon = \beta\boldsymbol{\Sigma}_\alpha$$

$$\boldsymbol{\Sigma}_\iota = \frac{1}{4}\{[C_{-11} - C_{-10}]\boldsymbol{\Sigma}_\alpha[C_{-11} - C_{-10}]^{'} + [C_{11} - C_{10}]\boldsymbol{\Sigma}_\alpha[C_{11} - C_{10}]^{'}\}$$
$$= \frac{1}{4}\{[(r_2 - r_0)C_{0p} + (r_3 - r_1)C_{1p}]\boldsymbol{\Sigma}_\alpha[(r_2 - r_0)C_{0p} + (r_3 - r_1)C_{1p}]^{'}$$
$$+ [(r_2 - r_0)C_{0n} + (r_3 - r_1)C_{1n}]\boldsymbol{\Sigma}_\alpha[(r_2 - r_0)C_{0n} + (r_3 - r_1)C_{1n}]^{'}\}$$

$$= \frac{1}{4}\{(r_2^2 - 2r_2r_0 + r_0^2)C_{0p}\boldsymbol{\Sigma}_\alpha C_{0p}^{'} + (r_2r_3 - r_2r_1 - r_0r_3 + r_0r_1)C_{0p}\boldsymbol{\Sigma}_\alpha C_{0p}^{'}$$
$$+ (r_2r_3 - r_2r_1 - r_0r_3 + r_0r_1)C_{1p}\boldsymbol{\Sigma}_\alpha C_{0p}^{'} + (r_3^2 - 2r_3r_1 + r_1^2)C_{1p}\boldsymbol{\Sigma}_\alpha C_{1p}^{'}$$
$$+ (r_2^2 - 2r_2r_0 + r_0^2)C_{0n}\boldsymbol{\Sigma}_\alpha C_{0n}^{'} + (r_2r_3 - r_2r_1 - r_0r_3 + r_0r_1)C_{0n}\boldsymbol{\Sigma}_\alpha C_{0n}^{'}$$
$$+ (r_2r_3 - r_2r_1 - r_0r_3 + r_0r_1)C_{1n}\boldsymbol{\Sigma}_\alpha C_{0n}^{'} + (r_3^2 - 2r_3r_1 + r_1^2)C_{1n}\boldsymbol{\Sigma}_\alpha C_{1n}^{'}\}$$

$$= \frac{1}{4}\{(r_2^2 - 2r_2r_0 + r_0^2)[C_{0p}\boldsymbol{\Sigma}_\alpha C_{0p}^{'} + C_{0n}\boldsymbol{\Sigma}_\alpha C_{0n}^{'}]$$
$$+ (r_3^2 - 2r_3r_1 + r_1^2)[C_{1p}\boldsymbol{\Sigma}_\alpha C_{1p}^{'} + C_{1n}\boldsymbol{\Sigma}_\alpha C_{1n}^{'}]$$
$$+ (r_2r_3 - r_2r_1 - r_0r_3 + r_0r_1)[C_{0p}\boldsymbol{\Sigma}_\alpha C_{1p}^{'} + C_{1p}\boldsymbol{\Sigma}_\alpha C_{0p}^{'} + C_{0n}\boldsymbol{\Sigma}_\alpha C_{1n}^{'} + C_{1n}\boldsymbol{\Sigma}_\alpha C_{0n}^{'}]\}$$

$$\boldsymbol{\Sigma}_0 = \beta(r_0C_{0b} + r_1C_{1b}) + \boldsymbol{\Sigma}_\alpha(r_0C_{0b} + r_1C_{1b})^{*} + \boldsymbol{\Sigma}_\iota + \sigma^2\mathbf{I}$$
$$= \beta[r_0^2C_{0b}\boldsymbol{\Sigma}_\alpha C_{0b}^{'} + r_0r_1C_{0b}\boldsymbol{\Sigma}_\alpha C_{1b}^{'} + r_1r_0C_{1b}\boldsymbol{\Sigma}_\alpha C_{0b}^{'} + r_1^2C_{1b}\boldsymbol{\Sigma}_\alpha C_{1b}^{'}] + \boldsymbol{\Sigma}_\iota + \sigma^2\mathbf{I}$$
$$= r_0^2\beta C_{0b}\boldsymbol{\Sigma}_\alpha C_{0b}^{'} + r_0r_1\beta[C_{0b}\boldsymbol{\Sigma}_\alpha C_{1b}^{'} + C_{1b}\boldsymbol{\Sigma}_\alpha C_{0b}^{'}] + r_1^2\beta C_{1b}\boldsymbol{\Sigma}_\alpha C_{1b}^{'} + \boldsymbol{\Sigma}_\iota + \sigma^2\mathbf{I}$$

$$\boldsymbol{\Sigma}_1 = \beta(r_2C_{0b} + r_3C_{1b}) + \boldsymbol{\Sigma}_\alpha(r_2C_{0b} + r_3C_{1b})^{*} + \boldsymbol{\Sigma}_\iota + \sigma^2\mathbf{I}$$
$$= r_2^2\beta C_{0b}\boldsymbol{\Sigma}_\alpha C_{0b}^{'} + r_2r_3\beta[C_{0b}\boldsymbol{\Sigma}_\alpha C_{1b}^{'} + C_{1b}\boldsymbol{\Sigma}_\alpha C_{0b}^{'}] + r_3^2\beta C_{1b}\boldsymbol{\Sigma}_\alpha C_{1b}^{'} + \boldsymbol{\Sigma}_\iota + \sigma^2\mathbf{I}$$

Simplification:

Let

$$E_1 = \beta C_{0b}\boldsymbol{\Sigma}_a C_{0b}^{\cdot}$$

$$E_2 = \beta[C_{0b}\boldsymbol{\Sigma}_a C_{1b}^{\cdot} + C_{1b}\boldsymbol{\Sigma}_a C_{0b}^{\cdot}]$$

$$E_3 = \beta C_{1b}\boldsymbol{\Sigma}_a C_{1b}^{\cdot}$$

$$F_1 = C_{0p}\boldsymbol{\Sigma}_a C_{0b}^{\cdot} + C_{0n}\boldsymbol{\Sigma}_a C_{0n}^{\cdot}$$

$$F_2 = C_{0p}\boldsymbol{\Sigma}_a C_{1p}^{\cdot} + C_{1p}\boldsymbol{\Sigma}_a C_{0p}^{\cdot} + C_{0n}\boldsymbol{\Sigma}_a C_{1n}^{\cdot} + C_{1n}\boldsymbol{\Sigma}_a C_{0n}^{\cdot}$$

$$F_3 = C_{1p}\boldsymbol{\Sigma}_a C_{1p}^{\cdot} + C_{1n}\boldsymbol{\Sigma}_a C_{1n}^{\cdot}$$

then

$$\boldsymbol{\Sigma}_0 = r_0^2 E_1 + r_0 r_1 E_2 + r_1^2 E_3 + \frac{1}{4}[(r_2^2 - 2r_2 r_0 + r_0^2)F_1 + (r_2 r_3 - r_2 r_1 - r_0 r_3 + r_0 r_1)F_2 + (r_3^2 - 2r_3 r_1 + r_1^2)F_3] + \sigma^2\mathbf{I}$$

$$\boldsymbol{\Sigma}_1 = r_2^2 E_1 + r_2 r_3 E_2 + r_3^2 E_3 + \frac{1}{4}[(r_2^2 - 2r_2 r_0 + r_0^2)F_1 + (r_2 r_3 - r_2 r_1 - r_0 r_3 + r_0 r_1)F_2 + (r_3^2 - 2r_3 r_1 + r_1^2)F_3] + \sigma^2\mathbf{I}$$

figure out derivative of $\boldsymbol{\mu}, \boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}_1$

$$\frac{\partial \boldsymbol{\mu}}{\partial r_0} = -\frac{1}{2}C_{0b}\hat{\boldsymbol{\alpha}}$$

$$\frac{\partial \boldsymbol{\mu}}{\partial r_1} = -\frac{1}{2}C_{1b}\hat{\boldsymbol{\alpha}}$$

$$\frac{\partial \boldsymbol{\mu}}{\partial r_2} = \frac{1}{2}C_{0b}\hat{\boldsymbol{\alpha}}$$

$$\frac{\partial \boldsymbol{\mu}}{\partial r_3} = \frac{1}{2}C_{1b}\hat{\boldsymbol{\alpha}}$$

$$\frac{\partial \boldsymbol{\Sigma}_0}{\partial r_0} = 2r_0 E_1 + r_1 E_2 + \frac{1}{4}[-2r_2 F_1 + 2r_0 F_1 - 2r_3 F_2 + 2r_1 F_2]$$

$$\frac{\partial \boldsymbol{\Sigma}_0}{\partial r_1} = r_0 E_1 + 2r_1 E_3 + \frac{1}{4}[-r_2 F_2 + r_0 F_2 - 2r_3 F_3 + 2r_1 F_3]$$

$$\frac{\partial \boldsymbol{\Sigma}_0}{\partial r_2} = \frac{1}{4}[2r_2 F_1 - 2r_0 F_1 + r_3 F_2 - r_1 F_2]$$

$$\frac{\partial \boldsymbol{\Sigma}_0}{\partial r_3} = \frac{1}{4}[r_2 F_2 - r_0 F_2 + 2r_3 F_3 - 2r_1 F_3]$$

$$\frac{\partial \boldsymbol{\Sigma}_1}{\partial r_0} = \frac{1}{4}[-r_2 F_1 + 2r_0 F_2 - r_3 F_2 + r_1 F_2]$$

$$\frac{\partial \boldsymbol{\Sigma}_1}{\partial r_1} = \frac{1}{4}[-r_2 F_2 + r_0 F_2 - 2r_3 F_3 + 2r_1 F_3]$$

$$\frac{\partial \boldsymbol{\Sigma}_1}{\partial r_2} = 2r_2 E_1 + r_3 E_2 + \frac{1}{4}[2r_2 F_1 - 2r_0 F_1 + r_3 F_2 - r_1 F_2]$$

$$\frac{\partial \boldsymbol{\Sigma}_1}{\partial r_3} = r_2 E_2 + 2r_3 E_3 + \frac{1}{4}[r_2 F_2 - r_0 F_2 + 2r_3 F_3 - 2r_1 F_3]$$

Find the derivative of cost function: ($\mathbf{h}, \boldsymbol{\Phi}$ known)

$$D = 4\boldsymbol{\mu}^* \mathbf{h} + Tr[(\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_0)\boldsymbol{\Phi}]$$

$$\frac{\partial D}{\partial r_0} = 4(-\frac{1}{2} C_{0b} \hat{\alpha})^* \mathbf{h} + Tr[(-2r_0 E_1 - r_1 E_2)\Phi]$$

$$\frac{\partial D}{\partial r_1} = 4(-\frac{1}{2} C_{1b} \hat{\alpha})^* \mathbf{h} + Tr[(-r_0 E_2 - 2r_1 E_3)\Phi]$$

$$\frac{\partial D}{\partial r_2} = 4(\frac{1}{2} C_{0b} \hat{\alpha})^* \mathbf{h} + Tr[(2r_2 E_1 + r_3 E_2)\Phi]$$

$$\frac{\partial D}{\partial r_3} = 4(\frac{1}{2} C_{1b} \hat{\alpha})^* \mathbf{h} + Tr[(r_2 E_2 + 2r_3 E_3)\Phi]$$

Let

$$n_0 = 4(-\frac{1}{2} C_{0b} \hat{\boldsymbol{\alpha}})^* h$$

$$n_1 = 4(-\frac{1}{2} C_{1b} \hat{\boldsymbol{\alpha}})^* h$$

$$t_0 = Tr(E_1 \boldsymbol{\Phi})$$

$$t_1 = Tr(E_2 \boldsymbol{\Phi})$$

$$t_4 = Tr(E_3 \boldsymbol{\Phi})$$

Then the gradients of $D^*(\mathbf{c}_0, \mathbf{c}_1)$ are as follows:

$$\frac{\partial D}{\partial r_0} = -2t_0 r_0 - t_1 r_1 + n_0$$
$$\frac{\partial D}{\partial r_1} = -t_1 r_0 - 2t_4 r_1 + n_1$$
$$\frac{\partial D}{\partial r_2} = 2t_0 r_2 + t_1 r_3 - n_0$$
$$\frac{\partial D}{\partial r_3} = t_1 r_2 + 2t_4 r_3 - n_1$$

Appendix C. The Statistics of ISI Vector

a). For Binary Signals:

Assume transmitted signals are equiprobable *a priori*,

$$\mathbf{\iota} = \mathbf{S}_{-1}\boldsymbol{\alpha}_{-1} + \mathbf{S}_{+1}\boldsymbol{\alpha}_{+1}$$
$$= (b_{-1}C_{-11} + (1-b_{-1})C_{-10})\boldsymbol{\alpha}_{-1} + (b_{+1}C_{+11} + (1-b_{+1})C_{+10})\boldsymbol{\alpha}_{+1}$$

$$\mathbf{\iota}_{00} = C_{-10}\boldsymbol{\alpha}_{-1} + C_{+10}\boldsymbol{\alpha}_{+1}$$
$$\mathbf{\iota}_{01} = C_{-10}\boldsymbol{\alpha}_{-1} + C_{+11}\boldsymbol{\alpha}_{+1}$$
$$\mathbf{\iota}_{10} = C_{-11}\boldsymbol{\alpha}_{-1} + C_{+10}\boldsymbol{\alpha}_{+1}$$
$$\mathbf{\iota}_{11} = C_{-11}\boldsymbol{\alpha}_{-1} + C_{+11}\boldsymbol{\alpha}_{+1}$$

$$\boldsymbol{\mu}_{\iota} = \frac{1}{2}(C_{-10}\boldsymbol{\alpha}_{-1} + C_{-11}\boldsymbol{\alpha}_{-1} + C_{+10}\boldsymbol{\alpha}_{+1} + C_{+11}\boldsymbol{\alpha}_{+1})$$
$$= \frac{1}{2}((C_{-10} + C_{-11})\boldsymbol{\alpha}_{-1} + (C_{+10} + C_{+11})\boldsymbol{\alpha}_{+1})$$

$$\hat{\mathbf{\iota}} = \mathbf{\iota} - \boldsymbol{\mu}_{\iota}$$
$$= (b_{-1} - \tfrac{1}{2})(C_{-11} - C_{-10})\boldsymbol{\alpha}_{-1} + (b_{+1} - \tfrac{1}{2})(C_{+11} - C_{+10})\boldsymbol{\alpha}_{+1}$$

$$\hat{\mathbf{\iota}}^{*} = (\mathbf{\iota} - \boldsymbol{\mu}_{\iota})^{*}$$
$$= (b_{-1} - \tfrac{1}{2})\boldsymbol{\alpha}_{-1}^{*}(C_{-11}^{*} - C_{-10}^{*}) + (b_{+1} - \tfrac{1}{2})\boldsymbol{\alpha}_{+1}^{*}(C_{+11}^{*} - C_{+10}^{*})$$

$$E_{\iota} = E\{(\mathbf{\iota} - \boldsymbol{\mu}_{\iota})(\mathbf{\iota} - \boldsymbol{\mu}_{\iota})^{*}\}$$
$$= E\{(b_{-1} - \frac{1}{2})^{2}(C_{-11} - C_{-10})\boldsymbol{\alpha}_{-1}\boldsymbol{\alpha}_{-1}^{*}(C_{+11}^{*} - C_{+10}^{*})$$
$$+ (b_{-1} - \frac{1}{2})(b_{+1} - \frac{1}{2})(C_{-11} - C_{-10})\boldsymbol{\alpha}_{-1}\boldsymbol{\alpha}_{+1}^{*}(C_{+11}^{*} - C_{+10}^{*})$$
$$+ (b_{+1} - \frac{1}{2})(b_{-1} - \frac{1}{2})(C_{+11} - C_{+10})\boldsymbol{\alpha}_{+1}\boldsymbol{\alpha}_{-1}^{*}(C_{-11}^{*} - C_{-10}^{*})$$
$$+ (b_{+1} - \frac{1}{2})^{2}(C_{+11} - C_{+10})\boldsymbol{\alpha}_{+1}\boldsymbol{\alpha}_{+1}^{*}(C_{+11}^{*} - C_{+10}^{*})$$

Case 1: $b_{-1}=0$, $b_{+1}=0$

$$\begin{aligned}\Sigma_{t1} &= \frac{1}{4}(C_{-11}-C_{-10})E\{\alpha_{-1}\alpha_{-1}^{*}\}(C_{+11}^{*}-C_{+10}^{*}) \\ &+\frac{1}{4}(C_{-11}-C_{-10})E\{\alpha_{-1}\alpha_{+1}^{*}\}(C_{+11}^{*}-C_{+10}^{*}) \\ &+\frac{1}{4}(C_{+11}-C_{+10})E\{\alpha_{+1}\alpha_{-1}^{*}\}(C_{-11}^{*}-C_{-10}^{*}) \\ &+\frac{1}{4}(C_{+11}-C_{+10})E\{\alpha_{+1}\alpha_{+1}^{*}\}(C_{+11}^{*}-C_{+10}^{*})\end{aligned}$$

Now, assume α is stationary with zero mean and let $C_{-1}=C_{-11}-C_{-10}$, $C_{+1}=C_{+11}-C_{+10}$, then

Case 1: $b_{-1}=0$, $b_{+1}=0$

$$\Sigma_{t1}=\frac{1}{4}(C_{-1}\Sigma_{\alpha}C_{-1}^{*}+C_{-1}\Sigma_{\alpha}C_{+1}^{*}+C_{+1}\Sigma_{\alpha}C_{-1}^{*}+C_{+1}\Sigma_{\alpha}C_{+1}^{*})$$

Similarly,

Case 2: $b_{-1}=0$, $b_{+1}=1$

$$\Sigma_{t2}=\frac{1}{4}(C_{-1}\Sigma_{\alpha}C_{-1}^{*}-C_{-1}\Sigma_{\alpha}C_{+1}^{*}-C_{+1}\Sigma_{\alpha}C_{-1}^{*}+C_{+1}\Sigma_{\alpha}C_{+1}^{*})$$

Case 3: $b_{-1}=1$, $b_{+1}=0$

$$\Sigma_{t3}=\frac{1}{4}(C_{-1}\Sigma_{\alpha}C_{-1}^{*}-C_{-1}\Sigma_{\alpha}C_{+1}^{*}-C_{+1}\Sigma_{\alpha}C_{-1}^{*}+C_{+1}\Sigma_{\alpha}C_{+1}^{*})$$

Case 4: $b_{-1}=1$, $b_{+1}=1$

$$\Sigma_{t4}=\frac{1}{4}(C_{-1}\Sigma_{\alpha}C_{-1}^{*}+C_{-1}\Sigma_{\alpha}C_{+1}^{*}+C_{+1}\Sigma_{\alpha}C_{-1}^{*}+C_{+1}\Sigma_{\alpha}C_{+1}^{*})$$

$$\begin{aligned}\Sigma_{t} &= \frac{1}{4}(\Sigma_{t1}+\Sigma_{t2}+\Sigma_{t3}+\Sigma_{t4}) \\ &= \frac{1}{4}(\frac{1}{2}(C_{-1}\Sigma_{\alpha}C_{-1}^{*}+C_{-1}\Sigma_{\alpha}C_{+1}^{*}+C_{+1}\Sigma_{\alpha}C_{-1}^{*}+C_{+1}\Sigma_{\alpha}C_{+1}^{*}) \\ &+\frac{1}{2}(C_{-1}\Sigma_{\alpha}C_{-1}^{*}-C_{-1}\Sigma_{\alpha}C_{+1}^{*}-C_{+1}\Sigma_{\alpha}C_{-1}^{*}+C_{+1}\Sigma_{\alpha}C_{+1}^{*})) \\ &= \frac{1}{4}(C_{-1}\Sigma_{\alpha}C_{-1}^{*}+C_{+1}\Sigma_{\alpha}C_{+1}^{*})\end{aligned}$$

$$\boldsymbol{\Sigma}_{\iota} = \tfrac{1}{4}((C_{-11} - C_{-10})\boldsymbol{\Sigma}_{\alpha}(C_{-11} - C_{-10})^{*} + (C_{+11} - C_{+10})\boldsymbol{\Sigma}_{\alpha}(C_{+11} - C_{+10})^{*})$$

Since $\boldsymbol{\alpha} = \hat{\boldsymbol{\alpha}} + \boldsymbol{\varepsilon}$, and $\hat{\boldsymbol{\alpha}}$ and $\boldsymbol{\varepsilon}$ are assumed uncorrelated, then we have $\boldsymbol{\Sigma}_{\alpha} = \boldsymbol{\Sigma}_{\hat{\alpha}} + \boldsymbol{\Sigma}_{\varepsilon}$. Therefore, the covariance matrix of the ISI vector has the form

$$\boldsymbol{\Sigma}_{\iota} = \tfrac{1}{4}((C_{-11} - C_{-10})(\boldsymbol{\Sigma}_{\hat{\alpha}} + \boldsymbol{\Sigma}_{\varepsilon})(C_{-11} - C_{-10})^{*} + (C_{+11} - C_{+10})(\boldsymbol{\Sigma}_{\hat{\alpha}} + \boldsymbol{\Sigma}_{\varepsilon})(C_{+11} - C_{+10})^{*})$$

b). For Quaternary Signals:

Assume transmitted signals are equiprobable *a priori*,

$$\begin{aligned}
\boldsymbol{\iota} &= \mathbf{S}_{-1}\boldsymbol{\alpha}_{-1} + \mathbf{S}_{+1}\boldsymbol{\alpha}_{+1} \\
&= (b_{-1}\mathbf{C}_{-11} + (1-b_{-1})\mathbf{C}_{-10})\boldsymbol{\alpha}_{-1} + (b_{+1}\mathbf{C}_{+11} + (1-b_{+1})\mathbf{C}_{+10})\boldsymbol{\alpha}_{+1}
\end{aligned}$$

$$\mathbf{S}_{-1} = b_{-1}(b_{-1}-1)(b_{-1}-2)(b_{-1}-3)\left(-\frac{\mathbf{C}_{-10}}{6b_{-1}} + \frac{\mathbf{C}_{-11}}{2(b_{-1}-1)} - \frac{\mathbf{C}_{-12}}{2(b_{-1}-2)} + \frac{\mathbf{C}_{-13}}{6(b_{-1}-3)}\right)$$

$$\mathbf{S}_{+1} = b_{+1}(b_{+1}-1)(b_{+1}-2)(b_{+1}-3)\left(-\frac{\mathbf{C}_{+10}}{6b_{+1}} + \frac{\mathbf{C}_{+11}}{2(b_{+1}-1)} - \frac{\mathbf{C}_{+12}}{2(b_{+1}-2)} + \frac{\mathbf{C}_{+13}}{6(b_{+1}-3)}\right)$$

$$\begin{aligned}
\boldsymbol{\iota}_{00} &= \mathbf{C}_{-10}\boldsymbol{\alpha}_{-1} + \mathbf{C}_{+10}\boldsymbol{\alpha}_{+1} \\
\boldsymbol{\iota}_{01} &= \mathbf{C}_{-10}\boldsymbol{\alpha}_{-1} + \mathbf{C}_{+11}\boldsymbol{\alpha}_{+1} \\
\boldsymbol{\iota}_{02} &= \mathbf{C}_{-10}\boldsymbol{\alpha}_{-1} + \mathbf{C}_{+12}\boldsymbol{\alpha}_{+1} \\
\boldsymbol{\iota}_{03} &= \mathbf{C}_{-10}\boldsymbol{\alpha}_{-1} + \mathbf{C}_{+13}\boldsymbol{\alpha}_{+1}
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{\iota}_{10} &= \mathbf{C}_{-11}\boldsymbol{\alpha}_{-1} + \mathbf{C}_{+10}\boldsymbol{\alpha}_{+1} \\
\boldsymbol{\iota}_{11} &= \mathbf{C}_{-11}\boldsymbol{\alpha}_{-1} + \mathbf{C}_{+11}\boldsymbol{\alpha}_{+1} \\
\boldsymbol{\iota}_{12} &= \mathbf{C}_{-11}\boldsymbol{\alpha}_{-1} + \mathbf{C}_{+12}\boldsymbol{\alpha}_{+1} \\
\boldsymbol{\iota}_{13} &= \mathbf{C}_{-11}\boldsymbol{\alpha}_{-1} + \mathbf{C}_{+13}\boldsymbol{\alpha}_{+1}
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{\iota}_{20} &= \mathbf{C}_{-12}\boldsymbol{\alpha}_{-1} + \mathbf{C}_{+10}\boldsymbol{\alpha}_{+1} \\
\boldsymbol{\iota}_{21} &= \mathbf{C}_{-12}\boldsymbol{\alpha}_{-1} + \mathbf{C}_{+11}\boldsymbol{\alpha}_{+1} \\
\boldsymbol{\iota}_{22} &= \mathbf{C}_{-12}\boldsymbol{\alpha}_{-1} + \mathbf{C}_{+12}\boldsymbol{\alpha}_{+1} \\
\boldsymbol{\iota}_{23} &= \mathbf{C}_{-12}\boldsymbol{\alpha}_{-1} + \mathbf{C}_{+13}\boldsymbol{\alpha}_{+1}
\end{aligned}$$

$$\iota_{30} = \mathbf{C}_{-13}\boldsymbol{\alpha}_{-1} + \mathbf{C}_{+10}\boldsymbol{\alpha}_{+1}$$
$$\iota_{31} = \mathbf{C}_{-13}\boldsymbol{\alpha}_{-1} + \mathbf{C}_{+11}\boldsymbol{\alpha}_{+1}$$
$$\iota_{32} = \mathbf{C}_{-13}\boldsymbol{\alpha}_{-1} + \mathbf{C}_{+12}\boldsymbol{\alpha}_{+1}$$
$$\iota_{33} = \mathbf{C}_{-13}\boldsymbol{\alpha}_{-1} + \mathbf{C}_{+13}\boldsymbol{\alpha}_{+1}$$

$$\boldsymbol{\mu}_\iota = \frac{1}{16}(\iota_{00} + \iota_{01} + \iota_{02} + \iota_{03} + \iota_{10} + \iota_{11} + \iota_{12} + \iota_{13} + \iota_{20} + \iota_{21} + \iota_{22} + \iota_{23} + \iota_{30} + \iota_{31} + \iota_{32} + \iota_{33})$$
$$= \frac{1}{4}((\mathbf{C}_{-10} + \mathbf{C}_{-11} + \mathbf{C}_{-12} + \mathbf{C}_{-13})\boldsymbol{\alpha}_{-1} + (\mathbf{C}_{+10} + \mathbf{C}_{+11} + \mathbf{C}_{+12} + \mathbf{C}_{+13})\boldsymbol{\alpha}_{+1})$$
$$= \bar{\mathbf{C}}_{-1}\boldsymbol{\alpha}_{-1} + \bar{\mathbf{C}}_{+1}\boldsymbol{\alpha}_{+1}$$

Where $\bar{\mathbf{C}}_{-1} = \frac{1}{4}\sum_{j=0}^{3}\mathbf{C}_{-1j}$, $\bar{\mathbf{C}}_{+1} = \frac{1}{4}\sum_{j=0}^{3}\mathbf{C}_{+1j}$

$$\tilde{\iota} = \iota - \boldsymbol{\mu}_\iota$$
$$= (\mathbf{S}_{-1} - \bar{\mathbf{C}}_{-1})\boldsymbol{\alpha}_{-1} + (\mathbf{S}_{+1} - \bar{\mathbf{C}}_{+1})\boldsymbol{\alpha}_{+1}$$

$$\tilde{\iota}^* = \boldsymbol{\alpha}_{-1}^*(\mathbf{S}_{-1} - \bar{\mathbf{C}}_{-1}) + \boldsymbol{\alpha}_{+1}^*(\mathbf{S}_{+1} - \bar{\mathbf{C}}_{+1})$$

$$\boldsymbol{\Sigma}_\iota = \mathbf{E}\{\iota\,\iota^*\}$$
$$= \mathbf{E}\{((\mathbf{S}_{-1} - \bar{\mathbf{C}}_{-1})\boldsymbol{\alpha}_{-1} + (\mathbf{S}_{+1} - \bar{\mathbf{C}}_{+1})\boldsymbol{\alpha}_{+1})(\boldsymbol{\alpha}_{-1}^*(\mathbf{S}_{-1} - \bar{\mathbf{C}}_{-1})^* + \boldsymbol{\alpha}_{+1}^*(\mathbf{S}_{+1} - \bar{\mathbf{C}}_{+1})^*)\}$$

Case 1: $b_{-1} = 0,\ b_{+1} = 0$

$$(\mathbf{C}_{-10} - \bar{\mathbf{C}}_{-1})\boldsymbol{\Sigma}_\alpha(\mathbf{C}_{-10} - \bar{\mathbf{C}}_{-1})^* + (\mathbf{C}_{-10} - \bar{\mathbf{C}}_{-1})\boldsymbol{\Sigma}_\alpha(\mathbf{C}_{+10} - \bar{\mathbf{C}}_{+1})^*$$
$$+(\mathbf{C}_{+10} - \bar{\mathbf{C}}_{+1})\boldsymbol{\Sigma}_\alpha(\mathbf{C}_{-10} - \bar{\mathbf{C}}_{-1})^* + (\mathbf{C}_{+10} - \bar{\mathbf{C}}_{+1})\boldsymbol{\Sigma}_\alpha(\mathbf{C}_{+10} - \bar{\mathbf{C}}_{+1})^*$$

Case 2: $b_{-1} = 0,\ b_{+1} = 1$

$$(\mathbf{C}_{-10} - \bar{\mathbf{C}}_{-1})\boldsymbol{\Sigma}_\alpha(\mathbf{C}_{-10} - \bar{\mathbf{C}}_{-1})^* + (\mathbf{C}_{-10} - \bar{\mathbf{C}}_{-1})\boldsymbol{\Sigma}_\alpha(\mathbf{C}_{+11} - \bar{\mathbf{C}}_{+1})^*$$
$$+(\mathbf{C}_{+11} - \bar{\mathbf{C}}_{+1})\boldsymbol{\Sigma}_\alpha(\mathbf{C}_{-10} - \bar{\mathbf{C}}_{-1})^* + (\mathbf{C}_{+11} - \bar{\mathbf{C}}_{+1})\boldsymbol{\Sigma}_\alpha(\mathbf{C}_{+11} - \bar{\mathbf{C}}_{+1})^*$$

Case 3: $b_{-1} = 0,\ b_{+1} = 2$

$$(\mathbf{C}_{-10} - \bar{\mathbf{C}}_{-1})\boldsymbol{\Sigma}_\alpha(\mathbf{C}_{-10} - \bar{\mathbf{C}}_{-1})^* + (\mathbf{C}_{-10} - \bar{\mathbf{C}}_{-1})\boldsymbol{\Sigma}_\alpha(\mathbf{C}_{+12} - \bar{\mathbf{C}}_{+1})^*$$
$$+(\mathbf{C}_{+12} - \bar{\mathbf{C}}_{+1})\boldsymbol{\Sigma}_\alpha(\mathbf{C}_{-10} - \bar{\mathbf{C}}_{-1})^* + (\mathbf{C}_{+12} - \bar{\mathbf{C}}_{+1})\boldsymbol{\Sigma}_\alpha(\mathbf{C}_{+12} - \bar{\mathbf{C}}_{+1})^*$$

Case 4: $b_{-1} = 0,\ b_{+1} = 3$

$$(\mathbf{C}_{-10}-\bar{\mathbf{C}}_{-1})\boldsymbol{\Sigma}_{\alpha}(\mathbf{C}_{-10}-\bar{\mathbf{C}}_{-1})^{\bullet}+(\mathbf{C}_{-10}-\bar{\mathbf{C}}_{-1})\boldsymbol{\Sigma}_{\alpha}(\mathbf{C}_{+13}-\bar{\mathbf{C}}_{+1})^{\bullet}$$
$$+(\mathbf{C}_{+13}-\bar{\mathbf{C}}_{+1})\boldsymbol{\Sigma}_{\alpha}(\mathbf{C}_{-10}-\bar{\mathbf{C}}_{-1})^{\bullet}+(\mathbf{C}_{+13}-\bar{\mathbf{C}}_{+1})\boldsymbol{\Sigma}_{\alpha}(\mathbf{C}_{+13}-\bar{\mathbf{C}}_{+1})^{\bullet}$$

Case 5: $b_{-1}=1,\ b_{+1}=0$

$$(\mathbf{C}_{-11}-\bar{\mathbf{C}}_{-1})\boldsymbol{\Sigma}_{\alpha}(\mathbf{C}_{-11}-\bar{\mathbf{C}}_{-1})^{\bullet}+(\mathbf{C}_{-11}-\bar{\mathbf{C}}_{-1})\boldsymbol{\Sigma}_{\alpha}(\mathbf{C}_{+10}-\bar{\mathbf{C}}_{+1})^{\bullet}$$
$$+(\mathbf{C}_{+10}-\bar{\mathbf{C}}_{+1})\boldsymbol{\Sigma}_{\alpha}(\mathbf{C}_{-11}-\bar{\mathbf{C}}_{-1})^{\bullet}+(\mathbf{C}_{+10}-\bar{\mathbf{C}}_{+1})\boldsymbol{\Sigma}_{\alpha}(\mathbf{C}_{+10}-\bar{\mathbf{C}}_{+1})^{\bullet}$$

Case 6: $b_{-1}=1,\ b_{+1}=1$

$$(\mathbf{C}_{-11}-\bar{\mathbf{C}}_{-1})\boldsymbol{\Sigma}_{\alpha}(\mathbf{C}_{-11}-\bar{\mathbf{C}}_{-1})^{\bullet}+(\mathbf{C}_{-11}-\bar{\mathbf{C}}_{-1})\boldsymbol{\Sigma}_{\alpha}(\mathbf{C}_{+11}-\bar{\mathbf{C}}_{+1})^{\bullet}$$
$$+(\mathbf{C}_{+11}-\bar{\mathbf{C}}_{+1})\boldsymbol{\Sigma}_{\alpha}(\mathbf{C}_{-11}-\bar{\mathbf{C}}_{-1})^{\bullet}+(\mathbf{C}_{+11}-\bar{\mathbf{C}}_{+1})\boldsymbol{\Sigma}_{\alpha}(\mathbf{C}_{+11}-\bar{\mathbf{C}}_{+1})^{\bullet}$$

.....

$$\begin{aligned}\boldsymbol{\Sigma}_{\iota}=\frac{1}{4}(((&\mathbf{C}_{-10}-\bar{\mathbf{C}}_{-1})\boldsymbol{\Sigma}_{\alpha}(\mathbf{C}_{-10}-\bar{\mathbf{C}}_{-1})^{\bullet}+(\mathbf{C}_{-11}-\bar{\mathbf{C}}_{-1})\boldsymbol{\Sigma}_{\alpha}(\mathbf{C}_{-11}-\bar{\mathbf{C}}_{-1})^{\bullet}\\&+(\mathbf{C}_{-12}-\bar{\mathbf{C}}_{-1})\boldsymbol{\Sigma}_{\alpha}(\mathbf{C}_{-12}-\bar{\mathbf{C}}_{-1})^{\bullet}+(\mathbf{C}_{-13}-\bar{\mathbf{C}}_{-1})\boldsymbol{\Sigma}_{\alpha}(\mathbf{C}_{-13}-\bar{\mathbf{C}}_{-1})^{\bullet})\\&+((\mathbf{C}_{+10}-\bar{\mathbf{C}}_{+1})\boldsymbol{\Sigma}_{\alpha}(\mathbf{C}_{+10}-\bar{\mathbf{C}}_{+1})^{\bullet}+(\mathbf{C}_{+11}-\bar{\mathbf{C}}_{+1})\boldsymbol{\Sigma}_{\alpha}(\mathbf{C}_{+11}-\bar{\mathbf{C}}_{+1})^{\bullet}\\&+(\mathbf{C}_{+12}-\bar{\mathbf{C}}_{+1})\boldsymbol{\Sigma}_{\alpha}(\mathbf{C}_{+12}-\bar{\mathbf{C}}_{+1})^{\bullet}+(\mathbf{C}_{+13}-\bar{\mathbf{C}}_{+1})\boldsymbol{\Sigma}_{\alpha}(\mathbf{C}_{+13}-\bar{\mathbf{C}}_{+1})^{\bullet}))\end{aligned}$$

Since $\boldsymbol{\alpha}=\hat{\boldsymbol{\alpha}}+\boldsymbol{\varepsilon}$, and $\hat{\boldsymbol{\alpha}}$ and $\boldsymbol{\varepsilon}$ are assumed uncorrelated, then we have $\boldsymbol{\Sigma}_{\alpha}=\boldsymbol{\Sigma}_{\hat{\alpha}}+\boldsymbol{\Sigma}_{\varepsilon}$. Therefore, the covariance matrix of the ISI vector has the form

$$\begin{aligned}\boldsymbol{\Sigma}_{\iota}=\frac{1}{4}(((&\mathbf{C}_{-10}-\bar{\mathbf{C}}_{-1})(\boldsymbol{\Sigma}_{\hat{\alpha}}+\boldsymbol{\Sigma}_{\varepsilon})(\mathbf{C}_{-10}-\bar{\mathbf{C}}_{-1})^{\bullet}+(\mathbf{C}_{-11}-\bar{\mathbf{C}}_{-1})(\boldsymbol{\Sigma}_{\hat{\alpha}}+\boldsymbol{\Sigma}_{\varepsilon})(\mathbf{C}_{-11}-\bar{\mathbf{C}}_{-1})^{\bullet}\\&+(\mathbf{C}_{-12}-\bar{\mathbf{C}}_{-1})(\boldsymbol{\Sigma}_{\hat{\alpha}}+\boldsymbol{\Sigma}_{\varepsilon})(\mathbf{C}_{-12}-\bar{\mathbf{C}}_{-1})^{\bullet}+(\mathbf{C}_{-13}-\bar{\mathbf{C}}_{-1})(\boldsymbol{\Sigma}_{\hat{\alpha}}+\boldsymbol{\Sigma}_{\varepsilon})(\mathbf{C}_{-13}-\bar{\mathbf{C}}_{-1})^{\bullet})\\&+((\mathbf{C}_{+10}-\bar{\mathbf{C}}_{+1})(\boldsymbol{\Sigma}_{\hat{\alpha}}+\boldsymbol{\Sigma}_{\varepsilon})(\mathbf{C}_{+10}-\bar{\mathbf{C}}_{+1})^{\bullet}+(\mathbf{C}_{+11}-\bar{\mathbf{C}}_{+1})(\boldsymbol{\Sigma}_{\hat{\alpha}}+\boldsymbol{\Sigma}_{\varepsilon})(\mathbf{C}_{+11}-\bar{\mathbf{C}}_{+1})^{\bullet}\\&+(\mathbf{C}_{+12}-\bar{\mathbf{C}}_{+1})(\boldsymbol{\Sigma}_{\hat{\alpha}}+\boldsymbol{\Sigma}_{\varepsilon})(\mathbf{C}_{+12}-\bar{\mathbf{C}}_{+1})^{\bullet}+(\mathbf{C}_{+13}-\bar{\mathbf{C}}_{+1})(\boldsymbol{\Sigma}_{\hat{\alpha}}+\boldsymbol{\Sigma}_{\varepsilon})(\mathbf{C}_{+13}-\bar{\mathbf{C}}_{+1})^{\bullet}))\end{aligned}$$

References

[1] S. Verdu, "Minimum Probability of Error for Asynchronous Multiple Access Communication Systems," *Proc. IEEE Military Commun. Conf.*, 1, Washington, DC, Nov. 1983, pp.213-219.

[2] S. Verdu, "Optimum Sequence Detection of Asynchronous Multiple-Access Communications," Abstracts of Papers: *IEEE Int. Symp. Inform. Theory,* St. Jovite, Canada, September 1983,p.80.

[3] S. Verdu, "Optimum Multi-User Signal Detection," Ph.D. Dissertation, Department of Electrical and Computer Engineering, University of Illinois at Urbara-Champaign. Report T-151 Coordinated Science Laboratory. Urbana, IL, Aug. 1984.

[4] S.Verdu, "Minmum Probability of Error for Asynchronous Gaussian Multiple-Access Channels," *IEEE Transactions on Information Theory*, vol. IT-32, pp. 85-96, 1986.

[5] G.D. Forney, "The Viterbi Algorithm," *Proc. IEEE*, 61(3), pp. 268-278, March 1973.

[6] Z. Xie, R. T. Short, and C. K. Rushforth, "A Family of Suboptimum Detectors for Coherent Multi-user Communications," IEEE *Journal on Selected Areas in Communications*, vol. 8, pp. 683-690, 1990.

[7] R. Lupas and S. Verdu, "Linear Multi-user Detectors for Synchronous Code-Division Multiple Access Channels," IEEE *Transactions on Information Theory*, vol. 35, pp. 123-136, 1989.

[8] R. Lupas and S. Verdu, "Near-Far Resistance of Multi-user Detectors in Asynchronous Channels," IEEE *Transactions on Communications*, vol. 38, pp. 496-508, 1990.

[9] U. Madhow and M. L. Honig, " MMSE Interference Suppression for Direct-Sequence Spread Spectrum CDMA," IEEE *Transactions on Communications*, vol. 42, pp. 3178-3188, 1994.

[10] S. Verdu, "Multi-user Detection," in Advances in Statistical Signal Processing: Signal Detection, vol. 2, H. V. Poor and J. B. Thomas, Eds. Greenwich, CT: JAI Press, 1993.

[11] M. Honig, U. Madhow, and S. Verdu, "Blind Adaptive Multi-user Detection," IEEE *Transactions on Information Theory*, vol. 41, pp. 944-960, July, 1995.

[12] A. Papoulis, *Probability, Random Variables, and Stochastic Process*, Third Edition, WCB McGraw-Hill, 1991.

[13] A. Duel-Hallen, "A Family of Multi-user Decision-Feedback Detectors for Asynchronous Code-Division Multiple-Access Channels," IEEE *Transactions on Communications*, vol. 43, pp. 421-434, 1995.

[14] A. Duel-Hallen, "Equalizers for Multiple Input/Multiple Output Channels and PAM Systems with Cyclostationary Input Sequences," IEEE *Journal on Selected Areas in Communications*, vol. 10, pp. 630-639, 1992.

[15] A. Duel-Hallen, "Decorrelating Decision-Feedback Multiuser Detector for Synchronous Code-Division Multiple-Access Channels," IEEE *Transactions on Communications*, vol. 41, pp. 285-290, 1993.

[16] M. Varanasi and B. Aazhang, "Multistage Detection in Asynchronous CDMA Communication," IEEE *Transactions on Communications*, vol. 38, pp. 509-519, 1990.

[17] P. Patel and J. Holtzman, "Analysis of a Simple Successive Interference Cancellation Scheme in a DS/CDMA System," IEEE *Journal on Selected Areas in Communications – Special Issue on CDMA*, vol. 12, pp. 796-807, 1994.

[18] B. Aazhang, B. P. Paris, and G. Orsak, "Neural Networks for Multi-user Detection in CDMA Communication," IEEE *Transactions on Communications*, vol. 40, pp. 1212-1222, 1992.

[19] U. Mitra and H.V. Poor, "Neural Network Techniques for Adaptive Multi-user Demodulation," IEEE *Journal on Selected Areas in Communications*, vol. 12, pp. 1460-1470, 1994.

[20] H. V. Poor and S. Verdu, "Single-User Detectors for Multi-user Channels," IEEE *Transactions on Communications*, vol. 36, pp. 50-60, 1988.

[21] H. V. Poor and S. Verdu, "Probability of Error in MMSE Multiuser Detection," IEEE *Transactions on Information Theory*, vol. 43, pp. 858-871, 1997.

[22] L. M. Li and L. B. Milstein, "Rejection of Narrow-Band Interference in PN Spread-Spectrum Systems Using Transversal Filters," *IEEE Transactions on Communications*, vol. COM-30, pp. 925-928, 1982.

[23] E. Masry, "Closed-Form Analytical Results for the Rejection of Narrow-Brand Interference in PN Spread-Spectrum Systems- Part I: Linear Prediction Filters," *IEEE Transactions on Communications*, vol. COM-32, pp. 888-896, 1984.

[24] E. Masry, "Closed-Form Analytical Results for the Rejection of Narrow-Brand Interference in PN Spread-Spectrum Systems- Part I: Linear Interpolation Filters," *IEEE Transactions on Communications*, vol. COM-33, pp. 10-19, 1985.

[25] L. A. Rusch and H. V. Poor, "Narrowband Interference Suppression in CDMA Spread Spectrum Communications," *IEEE Transactions on Communications*, vol. 42, pp. 1969-1979, 1994.

[26] H. V. Poor and L. A. Rusch, "Narrowband Interference Suppression in Spread Spectrum CDMA," *IEEE Personal Communications*, pp. 14-27, 1994.

[27] R. Vijayan and H. V. Poor, "Nonlinear Techniques for Interference Suppression in Spread Spectrum Systems," *IEEE Transactions on Communications*, vol. 38, pp. 1060-1065, 1990.

[28] J. G. Proakis, Digital Communications, Third Edition. New York: McGraw Hill, Inc., 1996.

[29] M. Stojanovic and Z. Zvonar, " Linear Multi-user Detection in Time-Varying Multipath Fading Channels," Proceedings of 30th Annual Conference on Information Sciences and Systems, Princeton, NJ, vol. 1, pp. 349-354, 1996.

[30] S. Kondo and L. B. Milstein, "Performance of Multicarrier DS CDMA Systems," *IEEE Transactions on Communications*, vol. 44, pp. 238-246, 1996.

[31] E. A. Sourour and M. Nakagawa, "Performance of Orthogonal Multicarrier CDMA in a Multipath Fading Channel," *IEEE Transactions on Communications*, vol. 44, pp. 356-367, 1996.

[32] Z. Zvonar and D. Brady, "Subooptimal Multi-user Detector for Frequency-Selective Rayleigh Fading Synchronous CDMA Channels," *IEEE Transactions on Communications*, vol. 43, pp. 154-157, 1995.

[33] A. Klein and P. W. Baier, "Linear Unbiased Data Estimation in Mobile Radio Systems Applying CDMA," *IEEE Journal on Selected Areas in Communications*, vol. 11, pp. 1058-1066, Sept., 1993.

[34] M. K. Tsatsanis and G. B. Giannakis, "Blind Estimation of Direct Sequence Spread Spectrum Signals in Multipath," *IEEE Transactions on Signal Processing*, vol. 45, pp. 1241-1252, 1997.

[35] M. K. Tsatsanis and G. B. Giannakis, "Optimal Decorrelating receivers for DS-CDMA Systems: A Signal Processing Framework," *IEEE Transactions on Signal Processing*, vol. 44, pp. 3044-3055, 1996.

[36] A. F. Naguib and A. Paulraj, "A Base-Station Antenna Array Receiver for Cellular DS/CDMA with M-ary Orthogonal Modulation," Proceedings of 28th Asilomar Conference on Signals, Systems & Computers, Pacific Grove, CA, vol. 2, pp. 858-862, 1995.

[37] H, Liu and M.D. Zoltowski, "Blind Equalization in Antenna Array CDMA Systems," *IEEE Transactions on Signal Processing*, vol. 45, pp. 161-172, 1997.

[38] T. F. Wong, T. M. Lok, J. S. Lehnet, and M. D. Zoltowski, "A Linear Receiver for Direct-Sequence Spread-Spectrum Multiple-Access Systems with Antenna Arrays and Blind Adaptation," *IEEE Transactions on Information Theory*, vol. 44, pp. 659-676, 1998.

[39] A. Sayeed and B. Aazhang, "Joint Multipath-Doppler Diversity in Mobile Wireless Communications," *IEEE Transactions on Communications*, vol. 47, no. 1, pp. 123-132, Jan., 1999.

[40] A. Sayeed, A. Sendonaris, and B. Aszhang, "Multi-user Detection in Fast Fading Multipath Environments," *IEEE Journal on Selected Areas in Communications*, vol. 16 no. 9, pp. 1691 –1701, Dec. 1998.

[41] G. W. Wornell, "Spread-Signature CDMA: Efficient Multi-user Communication in the Presence of Fading," *IEEE Transactions on Information Theory*, vol. 41, no. 5, pp. 1418-1438, Sept., 1995.

[42] G. K. Lee, S. B. Gelfand, and M. P. Fitz, "Bayesian Techniques for Known Channel Deconvolution," *IEEE Transactions on Communications*, vol. 44, pp. 448-455, April, 1996.

[43] G. K. Lee, S. B. Gelfand and M. P. Fitz, "Bayesian Techniques for Blind Deconvolution," *IEEE Transactions on Communications*, vol. 44, pp. 826-835, April, 1996.

[44] R. J. Barton, " The Influence of Adaptive Multicoding on the Capacity of CDMA Fading Channels with Patial Side Information, Part I: Theoretic Results," *IEEE Transactions on Information Theory*, 2002, to be submitted.

[45] Kobayashi, H., and J. B. Thomas, "Distance Measures and Related Criteria," Proc. 5th Ann. Allerton Conf. Circuit and System Theory, Monticello, IL, pp. 491-500, 1967.

[46] N. Merhav, "Universal decoding for memoryless Gaussian channels with deterministic interference," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1261–1269, July 1993.

[47] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai (Shitz), "On information rates for mismatched decoders," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1953–1967, Nov. 1994.

[48] M. Feder and A. Lapidoth, "Universal decoding for channels with memory," *IEEE Trans. Inform. Theory*, vol. 44, pp. 1726–1745, Sept.1998.

[49] Y.-S. Liu and B. L. Hughes, "A new universal random coding bound for the multiple-access channel," *IEEE Trans. Inform. Theory*, vol. 42, pp. 376–386, Mar. 1996.

[50] A. Lapidoth, "Mismatched decoding and the multiple-access channel," *IEEE Trans. Inform. Theory*, vol. 42, pp. 1439–1452, Sept. 1996.

[51] A. Lapidoth, "Nearest-neighbor decoding for additive non-Gaussian noise channels," *IEEE Trans. Inform. Theory*, vol. 42, pp. 1520–1529, Sept. 1996.

[52] S. D. Gray, M. Kocic, and D. Brady, "Multi-user Detection in Mismatched Multiple-Access Channels," *IEEE Transactions on Communications*, vol.43, pp. 3080-3089, Dec., 1995.

[53] L.-C. Chu and U. Mitra, "Performance Analysis of an Improved MMSE Multi-User Receive for Mismatched Delay Channels," *IEEE Transactions on Communications*, vol. 46, no.10, pp.1369-1380, Oct., to appear.

[54] L.-C. Chu and U. Mitra, "Improved MMSE-Based Multi-User Detectors for Mismatched Delay Channels," Proceedings of 30th Annual Conference on Information Sciences and Systems, Princeton, NJ, vol. 1, pp. 326-331, March, 1996.

[55] D. G. Strom, "Sensitivity Analysis of Near-Far Resistant DS-CDMA Receivers to Propagation Delay Estimation Errors," Proceedings of Vehicular Technology conference, Sweden, pp. 757-761, 1994.

[56] M. Stojanovic, J. G. Proakis, and J. A. Catipovic, "Analysis of the Impact of Channel Estimation Errors on the Performance of a Decision-Feedback Equalizer in Fading Multipath Channel," *IEEE Transactions on Communications*, vol. 43, pp. 877-886, 1995.

[57] H.-Y. Wu and A. Duel-Hallen, "Multi-User Detection with Differentially Encoded Data for Mismatched Flat Rayleigh Fading DCMA Channels," Proceedings of 39th Annual Conference on Information Sciences and Systems, Princeton, NJ, vol. 1, pp.332-337, March, 1996.

[58] Z. Zvonar and M. Stojanovic, "Performance of Antenna Diversity Multi-user Receivers in CDMA Channels with Imperfect Fading Estimation," *Wireless Personal communications*, vol.3, pp.91-110, 1996.

[59] M. Medard, "The Effect upon Channel Capacity in Wireless Communications of Perfect and Imperfect Knowledge of the Channel," *IEEE Transactions on Information Theory*, vol. 46, no. 3, May 2000.

[60] X. Tang, M. Alouini, and A.J. Goldsmith, "Effect of Channel Estimation Error on M-QA M BER Performance in Rayleigh Fading," *IEEE Transactions on Communications*, vol. 47, no. 12, December 1999.

[61] M. Lops, G.Ricci, and A. M. Tulino, "Robust Multi-user Detection for Synchronous CDMA Systems," Proceedings of 39th Annual Conference on Information Sciences and Systems, Princeton, NJ, vol. 1, pp. 338-342, 1996.

[62] A. L. McKellips and S. Verdu, "Worst Case Additive Noise for Binary-Input Channels and Zero-Threshold Detection Under Constraints of Power and Divergence," *IEEE Transactions on Information Theory*, vol. 43, pp. 1256-1264, July, 1997.

[63] A.L. McKellips an dS.Verdu, "Maximin Performance of Binary-Input Channels with Uncertain Noise Distributions," *IEEE Transactions on Information Theory*, vol. 44, pp.947-972, May, 1998.

[64] M. S. Alouini, X. Tang, and A. Goldsmith, "An Adaptive Modulation Scheme for Simultaneous Voice and Data Transmission over Fading Channels," IEEE *Journal on Selected Areas in Communications*, vol. 17, pp. 837-850, 1999.

[65] A. J. Goldsmith and S. G. Chua, "Variable-Rate Variable-Power MQAM for Fading Channels," IEEE *Transactions on Communications*, vol. 45, no. 10, pp. 1218-1230, Oct., 1997.

[66] A. J. Goldsmith and S. G. Chua, "Adaptive Coded Modulation for Fading Channels," IEEE *Transactions on Communications*, vol. 46, no. 5, pp. 595-602, 1998.

[67] D.L. Goeckel, "Adaptive Coding for Time-Varying Channels Using Outdated Fading Estimates," IEEE *Transactions on Communications*, vol. 47, no. 6, pp. 844-855, June, 1999.

[68] S. Wei and D.L. Goeckel, "Adaptive Signaling Based on Measurements with Statistical Uncertainty," Proceedings of Thirty-Third Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, vol. 1, pp. 27-31, Oct. 24-27, 1999.

[69] X. Qiu and K. Chawla, "On the Performance of Adaptive Modulation in Cellular Systems," IEEE *Transactions on Communications*, vol. 47, no. 6, pp. 884-895, June, 1999.

[70] W. T. Webb, "Variable Rate QAM for Mobile Radio," *IEEE Transactions on Communications*, vol. 43, no. 7,July, 1995.

[71] F. Adachi, K. Ohno, A. Higashi, T. Dohi, and Y. Okumura, "Coherent Multicode DS-CDMA Mobile Radio Access," *IEICE Transactions on Communications*, vol. E79-B, pp. 1316-1325, 1996.

[72] V. K. Bhargava, "High Rate Data Transmission in Mobile and Personal Communications," presented at PIMRC, The Hague, The Netherlands, 1994.

[73] P.-R. Chang and C.-F. Lin, "Design of Spread Spectrum Multicode CDMA Transport Architecture of Multimedia Services," IEEE Journal on Selected Areas in Communications, vol. 18, pp. 99-111, 2000.

[74] K. Takeo, S. Sato, and A. Ogawa, "Multi-Rate Traffic Accommodation Using Multicode Transmission in CDMA Cellular Systems," presented at IEEE International Conference on Communications, 2000.

[75] H.-Y. Wu and A. Duel-Hallen, "Channel Estimation and Multi-user Detection for Frequency-Nonselective Fading Synchronous CDMA Channels," Proceedings of 32nd Allerton Conference on communications, Control and Computing, Allerton, IL, September, 1994.

[76] S. E. Bensley an dB. Aazhang, "Maximum-Likelihood synchronization of a Single User for Code-Division Multiple-Access Communication Systems," *IEEE Transactions on Communications*, vol.46, no.3, pp.392-399, March, 1998.

[77] A. M. Monk, M.Davis, L.B. Milstein, and C.W. Helstrom, "A Noise-Whitening Approach to Multiple Access Noise Rejection – Part I: Theory and Background," *IEEE Journal on Selected Areas in Communications*, vol. 12, pp. 817-827, 1994.

[78] D. Zheng, J. LI, S. L. Miller, and E. G. Strom, "An efficient Code-Timing Estimator for DS-CDMA Signals," *IEEE Transactions on Signal Processing*, vol.45, no. 1, 82-89, Jan., 1997.

[79] H. V. Poor, *An Introduction to Signal Detection and Estimation*, Second Edition, Springer-Verlag, 1994.

[80] R. J. Barton, R. Shaw, and S. P. Reichart, "Performance Analysis of Acoustic Charge Transport Implementation of Nonlinear, Adaptive Neural-Network Equalizers for Ultra-Wideband Communication Channels," Proceedings of 1996 Deal-Use Technologies and Applications Conference, Syracuse NY, June, 1996.

[81] X. Wang and H. V. Poor, "Blind Equalization and Multi-user Detection in Dispersive CDMA Channels," IEEE *Transactions on Communications*, vol. 46, pp. 91-103, Jan., 1998.

[82] X. Wang and H. V. Poor, "Blind Multiuser Detection: A Subspace Approach," IEEE *Transactions on Information Theory*, vol. 44, no. 2, pp. 677-690, March, 1998.

[83] R. J. Barton and H. V. Poor, "On Generalized Signal-to-Noise Ratios in Quadratic Detection," *Mathematics of Control, Signals, and Systems*, vol. 5, pp. 81-91, 1992.

[84] W. A. Gardner, "A Unifying View of Second-Order Measures of Quality for Signal Classification," IEEE *Transactions on Communications*, vol. 28, no. 6, pp. 807-816, June, 1980.

[85] D. Middleton, "Acoustical Signal Detection by Simple Correlators in the Presence of Non-Gaussian Noise," *Journal of the Acoustical Society of America of America*, vol. 34, pp. 1598-1609, 1962.

[86] H. V. Poor, "Robust decision Design Using a Distance Criterion, " *IEE E Transaction on Information Theory*, vol. IT-26, no. 5, pp. 575-587, Sept., 1980.

[87] I. Csiszar, "Information-type measures of difference of probability distributions and indirect observations," *Studia Scient. Math. Hung.*, vol. 2, pp. 299-318, 1967.

[88] S. M. Ali and D. Silvey, "A general class of coefficient of divergence of one distribution from another," *J. Royal Stat. Soc.*, vol. 28, pp. 131-142, 1966.

[89] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc., 1991

[90] S. Kullback and R. A. Leibler, "On information and sufficiency," Ann. Math. Stat., vol. 22, pp. 79-86, 1951.

[91] D. H. Johnson and G. C. Orsak, "Relation of Signal Set Choice to the Performance of Optimal Non-Gaussian Detectors," *IEEE Transactions on Communications*, vol. 41, no. 9, pp. 1319-1328, Sept., 1993.

[92] H. Jeffreys, "An invariant form for the prior probability in estimation problem," *Proc. Roy. Soc. A.*, vol. 186, pp. 453-461, 1946.

[93] S. Kullback, *Information Theory and Statistics*. New York: Wiley, 1959.

[94] T. Kailath, "The Divergence and Bhattacharyya Distance," *IEEE Transactions on Communications*, vol. COM-15, pp. 52-60, 1967.

[95] S. Karlin and R. N. Bradt, "On the design and comparison of dichotomous experiments," Ann. Math. Stat., vol. 27, pp. 390-409,1956.

[96] F. Schweppe, " On the distance between Guassian processes: the state space approach," *Information and Control*, 1967.

[97] T. L. Grettenberg, "Signal Selection in Communication and Radar Systems," *IEEE Transactions on Information Theory*, vol. IT-9, pp. 265-275, Oct., 1963.

[98] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge, 1996.

[99] J. Ni and R. J. Barton, "Adaptive Linear-Quadratic (LQ) Receivers for Time-Varying Frequency-Selective Code-Division-Multiple-Access Channels", Proceedings of Thirty-Third Asilomar Conference on Signals, Systems and Computers, 1999

[100] R. J. Barton and J. Ni, "Adaptive Modulation and Receiver Design for CDMA Fading Channels with Uncertain Channel State Information", Proceedings of Thirty-Eighth Annual Allerton Conference on Communication, Control and Computing, 2000

[101] J. Ni and R. J. Barton, "Multicoding and Adaptive Modulation for Uncertain CDMA Channels", Proceedings of Thirty-Fourth Asilomar Conference on Signals, Systems and Computers, 2000

[102] J. Ni and R. J. Barton, "Adaptive Multicoding and Robust Linear-Quadratic Receivers for Uncertain Frequency-Selective CDMA Fading Channels", *IEEE Transactions on Communications, 2002*, to be submitted