CHANNEL CAPACITY, SIGNALING DESIGN, AND STATISTICAL DETECTION FOR MULTIPLE-ANTENNA CHANNELS

by

Hong Wan

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Electrical and Computer Engineering

The University of Utah

May 2012

The University of Utah Graduate School

STATEMENT OF THESIS APPROVAL

This dissertation of **Hong Wan**

has been approved by the following supervisory committee members:

Rong-Rong Chen, Chair		**12/02/2011** Date Approved
Behrouz Farhang-Boroujeny, Member		**12/02/2011** Date Approved
Cynthia Furse, Member		**12/02/2011** Date Approved
Neal Patwari, Member		**12/02/2011** Date Approved
G. Lawrence Zeng, Member		**12/02/2011** Date Approved

and by **Gianluca Lazzi**, Chair of the Department of **Electrical and Computer Engineering**

and by Charles A. Wight, Dean of the Graduate School.

ABSTRACT

This dissertation addresses several key challenges in multiple-antenna communications, including information-theoretical analysis of channel capacity, capacity-achieving signaling design, and practical statistical detection algorithms.

The first part of the thesis studies the capacity limits of multiple-input multiple-output (MIMO) multiple access channel (MAC) via virtual representation (VR) model. The VR model captures the physical scattering environment via channel gains in the angular domain, and hence is a realistic MIMO channel model that includes many existing channel models as special cases. This study provides analytical characterization of the optimal input distribution that achieves the sum-capacity of MAC-VR. It also investigates the optimality of beamforming, which is a simple scalar coding strategy desirable in practice. For temporally correlated channels, beamforming codebook designs are proposed that can efficiently exploit channel correlation.

The second part of the thesis focuses on statistical detection for time-varying frequency-selective channels. The proposed statistical detectors are developed based on Markov Chain Monte Carlo (MCMC) techniques. The complexity of such detectors grows linearly in system dimensions, which renders them applicable to inter-symbol-interference (ISI) channels with long delay spread, for which the traditional trellis-based detectors fail due to prohibitive complexity. The proposed MCMC detectors provide substantial gain over the de facto turbo minimum-mean square-error (MMSE) detector for both synthetic channel and underwater acoustic (UWA) channels. The effectiveness of the proposed MCMC detectors is successfully validated through experimental data collected from naval at-sea experiments.

To my parents.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincerest gratitude to my advisor, Professor Rong-Rong Chen, for giving me the chance to work with her, and sharing with me her keen insights and deep knowledge in many research areas. Without her invaluable guidance and kind inspiration, I would never have been able to finish this dissertation.

I would like to thank Professor Behrouz Farhang-Boroujeny for his constructive advice in many aspects of my research. I would also like to thank Professor Cynthia Furse, Professor Neal Patwari, Professor Larry Zeng, and Professor Gil Shamir, for taking part in my PhD thesis committee and providing valuable suggestions during my qualifying exam, thesis proposal and final defense. I am grateful to Professor Andrew Singer, Dr. Jun Choi Won, Dr. K. Kim from University of Illinois at Urbana-Champaign, and Dr. James Presig from Woods Hole Oceanographic Institution, for their constant support and creative suggestions during our collaborations.

I am indebted to all the colleagues I have worked with at University of Utah. Thanks to Ronghui Peng, Huajun She, Yang Zhao, Xuehong Mao, Pooyan Amini, Harsha Rao, and many others who supported my research in one way or another.

I am heartily thankful to Yingxian Zhu and his family, who helped me in many ways from the first day I entered the United States. Thanks to Jiarong for her support and encouragement. Thanks to the friends who made my life at Utah memorable, especially Guohong Li, Xuejun Yang and the friends I met in the volleyball club. I treasured all precious moments we shared in and after the games.

Finally I would like to thank my parents, my uncle, and my whole family for their love and support throughout my life.

CHAPTER 1

INTRODUCTION

In recent years, multiple-antenna communications have been identified as one of the most practical methods to increase the channel capacity, and also to improve the reliability of wireless communications. Wireless channels that utilize multiple antennas are referred to as multiple-input multiple-output (MIMO) channels. For the next generation cellular and wireless local area networks (LAN), MIMO technology is envisioned to be the core technology to achieve higher data rates. Currently, IEEE is proposing the 802.11n (MIMO) wireless standards, which promise to deliver a data rate of 500 Mbs (about 10 times faster than todays wireless LANs) for wireless transmissions of HDTV and other multimedia devices for streaming video and audio. Successful employment of MIMO technology imposes new challenges in the fields of wireless communication, signal processing, and communication networks.

In this dissertation, we investigate several key challenges in multiple-antenna communications. The main contribution of this dissertation is two-fold. First, we conduct an in-depth study of the information-theoretical capacity of wireless multiple-access channels that employs multiple antennas, and develop practical signaling strategies to achieve the channel capacity. Our study originates from a recently developed channel model via virtual representation (VR). The VR model captures the physical scattering environment via channel gains in the angular domain, and hence is a realistic MIMO channel model that includes many existing channel models as special cases. Our study yields new information-theoretical results that are different from those obtained under other idealized channel models and provides signaling designs that have direct engineering impact. Second, we develop a class of novel statistical detection methods based on Markov Chain Monte Carlo methods for time-varying, frequency-selective channels. We have successfully demonstrated that the proposed

statistical detectors achieve excellent performance for underwater acoustic channels, which are considered as one of the most challenging communication channels due to severe inter-symbol-interference and fast time-variation. Using real data collected from at-sea experiments, the proposed detectors achieve a bit-error- rate that is orders of magnitude less than that of the other state-of-the-art deterministic detectors.

1.1 Capacity and Signaling Design for MIMO MAC

MIMO technology provides powerful means to improve the reliability and capacity of wireless channels. A significant amount of work has been done to study the optimal input distribution and channel capacity of both MIMO single user and multiuser channels [3–9]. Several models have been adopted to capture the statistical correlation of elements of the channel matrices including the i.i.d. model [3], the Kronecker model [10–13], the virtual representation (VR) model [14, 15], and the unitary independent-unitary (UIU) model [7]. The first two models apply only to limited wireless environments where scattering is rich or at least locally rich at either the transmitter or the receiver. The VR and UIU models are more general channel models, and both transform the MIMO channel to a domain such that the channel gains can be justified to be approximately independent.

In the first part of the dissertation [16], we generalize the study of single-user MIMO channel based on VR [14] to the MIMO multiple access channel (MAC) based on VR. We first characterize the optimal input distribution that achieves the sum-capacity. We study the optimality of beamforming, which is a simple scalar coding strategy desirable in practice. We first strengthen the conditions for the optimality of beamforming for the single-user VR model in [14] by proving that there exists a signal-to-noise ratio (SNR) threshold below which beamforming is optimal and above which beamforming is strictly suboptimal. For multiuser case, we show that the capacity-achieving beamforming angle (c.b.a.) of a given user may vary with SNR and beamforming angles of other users. This is in contrast to the single-user case in which the c.b.a. is independent of SNR. We also derive explicit conditions to determine possible c.b.a. for certain MAC-VR channels. For systems with K users, we show that as K goes to infinity, the sum-rates achieved by a large class of power

allocation schemes are within a constant of the sum-capacity, and they grow in the order of $n_r \log K$, where n_r is the number of receive antennas. Furthermore, we obtain conditions under which beamforming is asymptotically capacity-achieving.

We also study the issue of beamforming design for communication systems under limited feedback. The goal of this study is to develop low-complexity signaling design methods for practical systems in which only limited information about the channel state information is available to the transmitter. For such systems, we first design a finite set of beamforming codebooks that are known a *priori* to both the transmitter and receiver. The receiver selects a codebook based on the instantaneous channel state information and feeds back the codeword index to the transmitter for transmission. Compared to previous work [17–19] that consider the idealized independent and identically distributed (i.i.d.) Rayleigh fading channels or spatially correlated channel, the goal of this work is to design beamforming codebooks that can effectively utilize channel temporal correlation, hence improving system performance.

1.2 Markov Chain Monte Carlo (MCMC) Statistical Detection for Time-Varying Channels

A main contribution of this dissertation is the development of statistical detectors for time-varying channels with inter-symbol-interference (ISI). The proposed statistical detectors are developed based on Markov Chain Monte Carlo (MCMC) techniques. The MCMC detectors are stochastic in nature, which makes them fundamentally different from other state-of-the-art detectors that are largely deterministic. This work demonstrates that the proposed MCMC detectors are high performance, low-complexity detectors that can significantly outperform other existing detectors under even the most challenging operating environment.

Earlier versions of the MCMC detectors have been applied to wireless communications and signal processing [20, 21]. These detectors, developed based on the bit-counting approach, typically require a large number of samples (thus high complexity) to achieve satisfactory performance. The MCMC detectors developed in this work are built upon a class of recently proposed MCMC detectors [22–24]. This new class of MCMC detectors utilizes Monte Carlo integration in soft detection, which

yields substantial complexity reduction when compared to earlier MCMC designs [25] It is shown in [22, 26] that for MIMO channels, the MCMC detectors outperform the much studied sphere decoding detectors with a complexity that is order of magnitude less. In [27], it is shown that MCMC detectors outperform the state-of-the-art turbo minimum-mean square-error (MMSE) detectors for various ISI channels considered. These work establish the MCMC detectors as a promising detector of choice for transceiver design.

In this work, we address a more challenging problem of MCMC detection for time-varying ISI channels with long channel memory. Such channels require the design of an entirely new class of MCMC detectors beyond those of [27], together with adaptive channel estimation algorithms, to facilitate joint channel tracking and data detection. The MCMC detector proposed here is based on list channel estimate (MCMC-LCE). In MCMC-LCE, the Gibbs sampler (core part of the MCMC detector) is designed to generate a list of likely pairs of data samples and matching estimates of the channel impulse response (CIR). This is important to overcome the uncertainly in both data and CIR under the difficult channels considered in this work. In contrast, the Gibbs sampler of [27] is for stationary channels, and thus generates only data samples assuming a perfectly known CIR. The idea of performing data detection using a list of channel estimates, rather than a single channel estimate that could be erroneous, is a key technical contribution that makes the proposed receiver highly effective for time-varying channels. Moreover, the choice of a particular form of the VSLMS algorithm and its deployment along within an iterative channel refinement in the context of turbo equalization has led to a significant improvement in the receiver performance.

In order to validate the performance of the proposed MCMC detector, we have tested our receiver design using data collected from actual at sea experiments. The underwater acoustic (UWA) channel has been considered as one of the most challenging channels in use today due to long multipath spread and rapid time variability. Our detector achieves bit-error-rates that are orders of magnitude less than those of the existing detectors developed by other leading UWA communication research groups.

1.3 Dissertation Structure

This dissertation is organized as follows.

Chapter 2 presents the information theoretical analysis on MIMO MAC channels based on the virtual representation model. We first derive necessary and sufficient conditions for capacity-achieving inputs in Section 2.2. An iterative algorithm for computing the optimal input distribution is developed in Section 2.3, followed by a low-complexity version based on a deterministic form of the capacity expression. Section 2.4 provides the main results on the optimality of beamforming. The single-user case and multiuser case are treated separately. In Section 2.4.1, we prove that beamforming can be optimal at low SNR and the beamforming angle is uniquely determined by the channel variance matrix. We also prove the threshold behavior which indicates that there exists a threshold below which the beamforming is strictly optimal. In Section 2.4.2, we investigate the multiuser case and demonstrate by examples that previous results for the Kronecker model may not hold for the virtual representation model. The asymptotic behavior of MIMO MAC capacity is studied in Section 2.5.

Chapter 3 presents a practical beamforming codebook design for temporally correlated multiple-input single-out (MISO) channels. The channel model and system setup are introduced in Section 3.1. In Section 3.2, we present a rotation-based codebook design which adapts the codebook according to the instantaneous CSI. In Section 3.3, several criteria are proposed to construct the root codebook. Numerical results for both coded and uncoded systems are provided in Section 3.4.

Chapter 4 describes the transceiver design for time-varying frequency selective channels. The proposed single-carrier block transmission (SCBT) system is described in Section 4.1. A particular form of adaptive least mean square channel estimation method with variable step-size (VSLMS) is proposed to perform channel tracking as presented in Section 4.2. In Section 4.3, we present the proposed MCMC detector based on list channel estimate. The effectiveness of the proposed design is demonstrated for both synthetic channels and UWA channels.

Chapter 5 summarizes this dissertation and addresses future research directions.

CHAPTER 2

OPTIMALITY OF BEAMFORMING FOR MIMO MULTIPLE ACCESS CHANNELS VIA VIRTUAL REPRESENTATION

The multiple-input multiple-output (MIMO) techniques provide powerful means to improve reliability and capacity of wireless channels. A significant amount of work has been done to study optimal input distributions and the channel capacity of single-user and multi-user MIMO channels (see, e.g., [1, 5, 9, 10, 14, 28, 29]). Several models have been adopted to capture the spatial correlation between the channel gains corresponding to different transmit-receive antenna pairs. These models include the i.i.d. model [28], the Kronecker model [10–12,30], the virtual representation (VR) model [14,31], and the unitary-independent-unitary (UIU) model [1]. The i.i.d. model assumes that the channel gains are independent and identically distributed (i.i.d.), and the Kronecker model assumes that the correlation between the channel gains can be written in terms of the product of the transmit correlation and the receive correlation. These two models apply only to wireless environments with rich or locally rich scattering at either the transmitter or the receiver. The VR and UIU models are more general, and both transform the MIMO channel to a domain such that the channel gains can be justified to be approximately independent. In this work, we adopt the VR model [31], which represents the MIMO channel in a virtual *angular* domain with each channel gain corresponding to one virtual transmit and receive angle pair. In the angular domain, the channel gains can be justified to be approximately independent of each other, although not necessarily identically distributed, because they include different signal paths (corresponding to different transmit and receive angle pairs) with independent random phases.

The single-user MIMO channel based on VR was studied in [14]. In this work [16, 32], we generalize this study to the MIMO multiple access channel (MAC) based on VR, denoted by MAC-VR. We first characterize the optimal input distribution that achieves the sum-capacity. Then we study the optimality of beamforming, which is a simple scalar coding strategy desirable in practice. We first strengthen the conditions for the optimality of beamforming for the single-user VR model in [14] by proving that there exists an signal-to-noise ratio (SNR) threshold below which beamforming is optimal and above which beamforming is strictly suboptimal. This result was illustrated in [14] only numerically. We then study the MAC-VR, for which we present an example to show that the capacity-achieving beamforming angle of a given user may vary with SNR and beamforming angles of other users. This is in contrast to the single-user case in which the capacity-achieving beamforming angle is independent of SNR. We also derive explicit conditions to determine possible capacity-achieving beamforming angles for certain MAC-VR channels. For systems with K users, we show that as K goes to infinity, the sum-rates achieved by a large class of input signaling schemes are within a constant of the sum-capacity, and they grow in the order of $n_r \log K$, where n_r is the number of receive antennas. Furthermore, we obtain conditions under which beamforming is asymptotically capacity-achieving.

Our study for the single-user case generalizes that in [10, 29] for the Kronecker model, and is different from [33] for the double-scattering model [34]. Our study for the MAC-VR also differs from [9] which assumes perfect channel state information at the transmitter, and from [35], which assumes finite feedback. We also note that the results we derive for the MAC-VR are applicable to the MIMO-MAC Kronecker (MAC-Kr) model in [12]. However, certain results valid for the MAC-Kr may not hold for the MAC-VR as demonstrated in later sections.

2.1 Channel Model and Virtual Representation

We consider the K-user MIMO MAC, in which K users transmit to one base station (BS) with each user equipped with n_t antennas and the BS equipped with n_r antennas. The channel between each user k and the BS is assumed to be the frequency-flat, slow fading MIMO channel. The received signal at the BS is an n_r-

dimensional vector $Y \in \mathcal{C}^{n_r}$ and is given by

$$Y = \sum_{k=1}^{K} \sqrt{\frac{p^k}{n_t}} H^k X^k + W, \tag{2.1}$$

where $X^k \in \mathcal{C}^{n_t}$ is the input vector of user k that satisfies the power constraint $E[X^{k\dagger} X^k] \leq n_t$, $(\cdot)^\dagger$ denotes the Hermitian operator, p^k represents the effective SNR of user k at each receive antenna, $W \in \mathcal{C}^{n_r}$ is a proper complex Gaussian noise vector that consists of i.i.d. entries with zero-mean and unit-variance, and $H^k \in \mathcal{C}^{n_r \times n_t}$ is the channel matrix of user k. The entries of H^k are identically distributed with unit variance, i.e., $E[|H_{m,j}|^2] = 1$ for all $m = 1, \ldots, n_r$ and $j = 1, \ldots, n_t$. As we show in Figure 2.1, when there are finite number of scatters, in general, these entries are correlated because each channel gain in the antenna domain (corresponding to a transmit and receive antenna pair) captures all of the signal paths. For each H^k, we follow [14] to consider its virtual representation $H^k = A_r \tilde{H}^k A_t^\dagger$, where A_r and A_t are two-dimensional spatial Fourier matrices. The matrix $\tilde{H}^k$ is referred to as a virtual representation of H^k. Each element of $\tilde{H}^k$, referred to as the virtual coefficient, represents the channel gain corresponding to one transmit and receive virtual angle pair. The virtual coefficients are independent and not identically distributed (i.n.d.) random variables and each is assumed to be a zero-mean proper complex random variable with a symmetric distribution around the origin. The independency among virtual coefficients can be justified because they capture different sets of signal paths with independent random phases, each corresponding to a different pair of transmit and receive angle. The correlation of the channel gains in the antenna domain is implicitly determined by the i.n.d. channel gains in virtual domain and the Fourier transform between the two domains. [1]

A virtual representation of the MIMO MAC channel (2.1) is given by

$$\tilde{Y} = \sum_{k=1}^{K} \sqrt{\frac{p^k}{n_t}} \tilde{H}^k \tilde{X}^k + \tilde{W}, \tag{2.2}$$

[1] We note that a channel matrix with arbitrary correlation in the antenna domain may not necessarily have a meaningful virtual representation with an i.n.d. channel matrix in the angular domain [31].

where $\tilde{X}^k = A_t^\dagger X^k$, $\tilde{Y} = A_r^\dagger Y$, and $\tilde{W} = A_r^\dagger W$. Due to the unitarity of A_t, the input power constraint in the virtual domain does not change, i.e., $E(\tilde{X}^{k\dagger}\tilde{X}^k) \leq n_t$. Given $\tilde{H}^k$, we define the (m,j)-th element of the variance matrix V^k as $V^k_{m,j} = \mathrm{Var}(\tilde{H}^k_{m,j})$, for $1 \leq m \leq n_r$ and $1 \leq j \leq n_t$, which characterizes the second order statistics of $\tilde{H}^k$.

In [12], the MIMO MAC channel (2.1) is considered assuming that each user's channel follows the kronecker channel model

$$H^k = \Phi^{1/2} H^k_w (\Sigma^k)^{1/2} \quad k = 1, \cdots, K \tag{2.3}$$

where elements of H^k_w are i.i.d, zero-mean, unit-variance complex Gaussian $\mathcal{CN}(0,1)$ random variables, $\Sigma^k \in \mathcal{C}^{n_t \times n_t}$ is the deterministic transmit correlation matrix of user k, $\Phi \in C^{n_r \times n_r}$ is the deterministic receive correlation matrix which is assumed to be the same for all users in the MAC channel. We will show that this MAC channel model reduces to a special case of (2.2). To see this, we perform eigenvalue decomposition of $\Phi^{1/2}$ and $(\Sigma^k)^{1/2}$ to obtain $\Phi^{1/2} = R\Lambda R^\dagger$ and $(\Sigma^k)^{1/2} = T^k \Sigma^k T^{k\dagger}$, where R and T^k are unitary matrices; $\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_{n_r})$ and $\Sigma^k = \mathrm{diag}(\gamma^k_1, \cdots, \gamma^k_{n_t})$ are diagonal matrices containing the eigenvalues of $\Phi^{1/2}$ and $(\Sigma^k)^{1/2}$, respectively. We have

$$H^k = R\left(\Lambda R^\dagger H^k_w T^k \Gamma^k\right) T^{k\dagger} = R\tilde{H}^k T^{k\dagger}, \tag{2.4}$$

where $\tilde{H}^k = \Lambda R^\dagger H^k_w T^k \Gamma^k$. Since the elements of H^k_w are i.i.d. $\mathcal{CN}(0,1)$ distributed, elements of $R^\dagger H^k_w T^k$ are also i.i.d. $\mathcal{CN}(0,1)$ distributed. It follows that the elements of $\tilde{H}^k$ are independent, zero-mean complex-Gaussian random variables, and the variance of the (m,n)-th element $\tilde{H}^k_{m,n}$, denoted by $V^k_{m,n}$ equals $\lambda_m \cdot \gamma^k_n$. Hence, using (2.4), the MIMO kronecker MAC reduces to (2.2) once we let $\tilde{X}^k = T^{k\dagger} X^k$, $\tilde{Y} = R^\dagger Y$, and $\tilde{W} = R^\dagger W$. Hence, the results presented in this paper for the MIMO MAC via virtual representation (2.2) are also applicable to the MIMO kronecker MAC studied in [12].

For the Kronecker model, V^k takes on a special product-form: $V^k = \Lambda \cdot \Gamma_k$ and therefore $V^k_{m,n} = \lambda_m \cdot \gamma^k_n$. This product relation in general does not hold for the virtual representation model with elements in V^k can take arbitrary nonnegative values. The latter model is also more general because each virtual coefficient $\tilde{H}^k_{m,n}$ is only required to be a zero-mean, proper-complex random variable with a symmetric distribution

around the origin. The Gaussian assumption is not imposed. For the Kronecker model, however, $\tilde{H}^k_{m,n}$ is assumed to be Gaussian distributed.

2.2 Optimal Input Distribution

By adopting the virtual representation for the MIMO MAC, the sum capacity is given by

$$\begin{aligned} C &= \max_{\tilde{Q}^k, k=1,\cdots,K} \mathcal{I}\left(\tilde{Q}^1, \cdots \tilde{Q}^K\right) \\ &= \max_{\tilde{Q}^k, k=1,\cdots,K} E\left[\log\det\left(I + \sum_{k=1}^{K} \frac{p^k}{n_t} \tilde{H}^k \tilde{Q}^k \tilde{H}^{k\dagger}\right)\right] \\ s.t. \quad & \mathrm{Tr}(\tilde{Q}^k) \leq n_t \quad \text{for} \quad k = 1, \cdots, K \end{aligned} \tag{2.5}$$

where $\tilde{Q}^k = E(\tilde{X}^k \tilde{X}^{k\dagger})$ is the covariance matrix of the input vector of user k in the virtual domain.

We need to be cautious when interpreting this capacity equation. Throughout this work, we consider ergodic capacity, assuming that perfect channel state information is available at the receiver via training and that only channel statistics are known at the transmitter. The latter is a realistic assumption because the channel statistics change over much larger time scales than that of the channel gains. Accordingly, capacity is based on a quasi-static analysis where the channel state varies randomly from burst to burst. Within a burst the channel is assumed to be unchanged and it is also assumed that sufficient bits are transmitted for the standard infinite time horizon of information theory to be meaningful. It is also assumed that the channel is memoryless, i.e., each burst draws an independent channel realization. Thus, the capacity measurement can be performed over sufficient number of bursts to achieve average performance, which justifies the study of ergodic capacity.

We would also like to point out that, in this work, we focus on small-scale fading, with the understanding that the large scale fading such as path loss and reflection loss, can be treated at a much larger time scale through system-level design such as power control. The results presented in this work apply to a time duration over which the large scale fading is roughly unchanged and only small-scale fading contributes to the channel variation.

It is shown in [14] that for the single user case in which $K = 1$, the optimal input covariance matrix under the virtual channel representation is diagonal. In the following Theorem, we extend this result in [14] to MIMO-MAC-VR in which $K > 1$, and show that the optimal input covariance matrix remains diagonal for each user.

An important consequence of the optimal input covariance matrices being diagonal is as follows. Since $X^k = A_t \tilde{X}^k$, for MIMO-MAC-VR, the optimal signaling scheme is for each user to transmit independent data streams to the n_t column vectors, corresponding to n_t virtual transmitting angles of the Fourier matrix A_t. For the MIMO MAC kronecker model, since $X^k = T^k \tilde{X}^k$, the optimal signaling scheme, therefore, is for each user to transmit independent data streams to the n_t column vectors of T^k, which are the eigenvectors of the transmitter covariance matrix (Σ^k). Hence, Here we generalize [12] Theorem 1 for the MIMO MAC kronecker model to the MIMO MAC VR model. The proof presented above follows the techniques of [14], which does not reply on the Gaussian assumption, whereas the proof of [12] Theorem 1 follows a different approach that relies on the Gaussian assumption.

Next, we derive a necessary and sufficient condition of optimality for the input covariance matrices.

Theorem 1 *The diagonal covariance matrices* $\{\tilde{Q}^k, k = 1, \cdots, K\}$ *achieve the sum-capacity if and only if for every* $1 \le k \le K$ *and* $1 \le j \le n_t$*, we have*

$$E\ Tr\left[A^{-1}\left(\tilde{h}_j^k \tilde{h}_j^{k\dagger} - \frac{1}{n_t}\tilde{H}^k \tilde{Q}^k \tilde{H}^{k\dagger}\right)\right] \begin{cases} = 0, & \text{if } \lambda_j^k > 0, \\ \le 0, & \text{if } \lambda_j^k = 0, \end{cases} \tag{2.6}$$

where $\tilde{Q}^k = diag(\lambda_1^k, \cdots, \lambda_{n_t}^k)$*,* $A = I_{n_r} + \sum_{l=1}^{K} \frac{p^l}{n_t} \tilde{H}^l \tilde{Q}^l \tilde{H}^{l\dagger}$*, and* $\tilde{h}_j^k$ *denotes the* j*-th column of* $\tilde{H}^k$*.*

Proof. See 2.8.1.

For the single user case, by letting $A = I_{n_t}$, one can show that (2.6) is equivalent to (5) of [36]. We can rewrite the left-hand side of (2.6) as $E\Big[\tilde{h}_j^{k\dagger} A^{-1} \tilde{h}_j^k - \frac{1}{n_t}\sum_{l=1}^{n_t} \lambda_l^k \tilde{h}_l^{k\dagger} A^{-1} \tilde{h}_j^l\Big]$ It implies that if the optimal power allocation $\lambda_j^k > 0$, then it must satisfy

$$\lambda_j^k = \frac{\lambda_j^k E[\tilde{h}_j^{k\dagger} A^{-1} \tilde{h}_j^k]}{\frac{1}{n_t}\sum_{l=1}^{n_t} \lambda_l^k E[\tilde{h}_l^{k\dagger} A^{-1} \tilde{h}_l^k]} = \frac{\lambda_j^k \frac{\Sigma^k}{n_t} E[\tilde{h}_j^{k\dagger} A^{-1} \tilde{h}_j^k]}{\frac{1}{n_t}\sum_{l=1}^{n_t} \lambda_l^k \frac{\Sigma^k}{n_t} E[\tilde{h}_l^{k\dagger} A^{-1} \tilde{h}_l^k]} \tag{2.7}$$

As discussed in [36], conditioned upon the channel realization, the term $\tilde{h}_j^{k\dagger} A^{-1} \tilde{h}_j^k$ relates to the instantaneous minimum mean square error (MMSE) on the linear estimation of $\tilde{X}_j^k$, which is the transmitted symbol from the j-th virtual angle of user k, as follows:

$$\mathrm{MMSE}_j^k = 1 - \lambda_j^k \frac{p^k}{n_t} \tilde{h}_j^{k*} A^{-1} \tilde{h}_j^k \tag{2.8}$$

The averaged MMSE, denoted by MMSE_j^k, is given by $\overline{\mathrm{MMSE}}_i^k = 1 - \lambda_j^k \frac{p^k}{n_t} E[\tilde{h}_j^{k*} A^{-1} \tilde{h}_j^k]$ It immediately follows from (2.7) that

$$\lambda_j^k = \frac{1 - \overline{\mathrm{MMSE}}_j^k}{\frac{1}{n_t} \sum_{l=1}^{n_t} (1 - \overline{\mathrm{MMSE}}_l^k)} \tag{2.9}$$

Based on the above derivations, we can construct an iterative algorithm to compute the optimal covariance matrices of all users that satisfy condition (2.6). Our algorithm extends the results in [36] into the generalized multi-user scenario.

By taking turns to refine the input of one user while fixing the rest, the problem is reduced to the single user scenario, except with known interference from other users. If the channel is reduced to the Kronecker model, the proposed algorithm is equivalent to the one proposed in [30], and readers are suggested to refer to [30] for the convergence analysis.

In information theory for MIMO communications, it is well known that to obtain close-form analytic formulas for the mean capacity or outage capacity of MIMO channels is an open problem when transmit/receive antennas are modeled to be spatially correlated. Once consequence is that the computation of capacity-achieving input covariance is of very high complexity since the calculation of the MIMO capacity requires integrating over the probability density distribution of random channel matrices. The solution is only available for the special case of i.i.d. Rayleigh distributed MIMO channels, where the ergodic capacity can be obtained [3] by using the eigenvalue distribution of the Wishart matrix in an integral form. Such results can not be applied to the virtual representation model where the channel elements are not identically Gaussian distributed. Therefore, although the proposed algorithm provides the optimal solution that satisfies all conditions in (2.6), the computational complexity is high since it is based on the capacity formula in the expectation form.

Algorithm 1: Algorithm to compute the optimal covariance matrices

For a K-user MIMO-MAC-VR system, each user having n_r receive antennas and n_t transmit antennas:
1) given SNR $p^1, \cdots, p^K$ and variance matrices $V^1, \cdots, V^K$, initialize $\tilde{Q}^1, \cdots, \tilde{Q}^K$ to be identity matrices;
2) For each user $k = 1, \cdots, K$:
I.Fix $\tilde{Q}^{k'}$ for all other users $k' \neq k$;
II.For iteration index $n = 1, 2, \cdots$
a. update $\overline{\text{MMSE}}_j^{k(n)}$ as

$$\overline{\text{MMSE}}_j^{k(n)} = E\left[1 - \lambda_j^{k(n-1)} \frac{p^k}{n_t} \tilde{h}_j^{k*} \left(I + \sum_{l \neq k} \frac{p^l}{n_t} \tilde{H}^l \tilde{Q}^l \tilde{H}^{l\dagger} + \frac{p^k}{n_t} \tilde{H}^k \tilde{Q}^{k(n-1)} \tilde{H}^{k\dagger}\right)^{-1} \tilde{h}_j^k\right]$$

b. update $\{\lambda_1^{k(n)}, \cdots, \lambda_{n_t}^{k(n)}\}$ as

$$\begin{cases} \lambda_j^{k(n)} = 0 & p^k E\left[\tilde{h}_j^{k+} A^{-1} \tilde{h}_j^k\right] \leq \sum_{l=1}^{n_t} \left(1 - \overline{\text{MMSE}}_j^{k(n)}\right) \\ \lambda_j^{k(n)} = \frac{1 - \overline{\text{MMSE}}_j^{k(n)}}{\frac{1}{n_t} \sum_{l=1}^{n_t} (1 - \overline{\text{MMSE}}_l^{k(n)})} & otherwise \end{cases}$$

c. if $||\tilde{Q}^{k(n)} - \tilde{Q}^{k(n-1)}||^2 < \varepsilon_1$, return $\tilde{Q}^k = \tilde{Q}^{k(n)}$ and step forward to the next user;
III. if $\max_{j,k} ||\nabla_{\lambda_j^k} \mathcal{I}(\tilde{Q}^1, \cdots, \tilde{Q}^K)||^2 < \varepsilon_2$, return $\{\tilde{Q}^1, \cdots, \tilde{Q}^K\}$; or else repeat step II.

In Section 2.3, we propose a low-complexity algorithm that does not require the computation of the expectation.

2.3 Low Complexity Signalling Design

It is well known that the distribution of eigenvalues of a large class of random matrix ensembles converges to deterministic limiting distribution as the matrix dimension goes to infinity. A random matrix shows fewer random fluctuations as its dimension increases. According to a central limit theorem in the random matrix theory [37], the distribution of random determinants converges to Gaussian, which indicates that with a certain limiting ratio between the numbers of transmit and receive antennas, the MIMO random capacity is asymptotically Gaussian as the number of antennas goes to infinity. For a MIMO channel with finite number of

antennas, the asymptotic results still serves as a good approximation. Specifically, Girko's random matrix results can be applied to the virtual representation channel model to compute the asymptotic normalized capacity. The asymptotic capacity can be directly presented in the form of a two-dimensional spatial scattering channel function. The advantage is, following this approach, the computation of the asymptotic capacity requires only the second-order statistics, i.e., the channel variance matrix, while the computation of nonasymptotic capacity requires the detailed probability density functions about the marginals of all virtual channel elements. In [14], it has been shown that the asymptotics are accurate even for moderate numbers of transmit and receive antennas. A direct indication is that the virtual channel variance matrix is all we need to accurately characterizing the capacity even for nonGaussian channels. This result can be easily leveraged to the multi-user scenarios when the channel hardens with multiple users. As an efficient low-complexity approach leading to a near-optimum signaling solution, it provides important insights to how the second order statistics of channel elements affects the MIMO capacity.

For single-user channels with independent and nonidentically distributed (IND) entries, a similar approach was presented in [1], where the channel takes a simpler form. In [1], it is shown that the optimal capacity for the single-user IND channels can be approximated by

$$C \approx \sum_{j=1}^{n_t} \log_2\left(\frac{1+p\beta_j^*}{e^{\alpha_j^*\beta_j^*}}\right) + \sum_{i=1}^{n_r} \log_2\left(1+\frac{1}{n_t}\sum_{j=1}^{n_t}\lambda_j^* V_{i,j}\alpha_j^*\right) \tag{2.10}$$

where $\{\lambda_1^*, \cdots, \lambda_{n_t}^*\}$ denotes the optimal power allocation, and $\{\alpha_j^*, \beta_j^*, j = 1, \cdots, n_t\}$ are the solutions to the following pair of equations

$$\alpha_j^* = \frac{p}{1+p\beta_j^*} \tag{2.11}$$

$$\beta_j^* = \sum_{i=1}^{n_r} \frac{\lambda_j^* V_{i,j}}{n_t + \sum\limits_{j=1}^{n_t} \lambda_j^* V_{i,j}\alpha_j^*} \tag{2.12}$$

It is shown in [1] that asymptotically, the signal-to-interference-plus-noise ratio (SINR) exhibited by the signal radiated from the j-th virtual angle at the output of a linear MMSE receiver is approximated by $p\beta_j^*$, while the corresponding MMSE is approximated by α_j^*/p.

Considering a K-user system with each user having n_t transmit antennas and n_r receive antennas, the capacity expression (2.5) has the same form as that of a single-user system with $K * n_t$ transmit antennas and n_r receive antennas, except for the power constraints being imposed to each user independently. Whereas the asymptotic limits in [1] were derived independent of input power profile, so we can generalize the asymptotic capacity expression (2.10) for MAC channels by replacing $(\alpha_j^*, \beta_j^*, V_{i,j})$ with $(\alpha_j^{l^*}, \beta_j^{l^*}, V_{i,j}^l)$, $\sum_{j=1}^{n_t}$ with $\sum_{l=1}^{K}\sum_{j=1}^{n_t}$, and $\frac{1}{n_t}\lambda_j^*$ with $\frac{1}{n_t}\lambda_j^{l^*}$, where $l = 1, \cdots, K$ is user index. We write down the expression as follows:

$$C \approx \sum_{l=1}^{K}\sum_{j=1}^{n_t} \log_2\left(\frac{1+p^l\beta_j^{l^*}}{e^{\alpha_j^{l^*}\beta_j^{l^*}}}\right) + \sum_{i=1}^{n_r} \log_2\left(1+\frac{1}{n_t}\sum_{l=1}^{K}\sum_{j=1}^{n_t}\lambda_j^{l^*}V_{i,j}^l\alpha_j^{l^*}\right) \tag{2.13}$$

where $\{\lambda_j^{l^*}, j = 1, \cdots, n_t,\ l = 1, \cdots, K\}$ denotes the optimum power allocation, and $(\alpha_j^{l^*}, \beta_j^{l^*})$ satisfies

$$\begin{aligned} \alpha_j^{l^*} &= \frac{p^l}{1+p^l\beta_j^{l^*}} \\ \beta_j^{l^*} &= \sum_{i=1}^{n_r} \frac{\lambda_j^{l^*}V_{i,j}^l}{n_t + \sum_{l=1}^{K}\sum_{j=1}^{n_t}\lambda_j^{l^*}V_{i,j}^l\alpha_j^{l^*}} \end{aligned} \tag{2.14}$$

Let $\boldsymbol{\lambda} = \{\lambda_j^l\}, \boldsymbol{\alpha} = \{\lambda_j^l\}, \boldsymbol{\beta} = \{\beta_j^l\}$, where $j = 1, \cdots, n_t, l = 1, \cdots, K$. Now we show that the optimal solution to

$$I(\boldsymbol{\lambda}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{l=1}^{K}\sum_{j=1}^{n_t} \log_2\left(\frac{1+p^l\beta_j^l}{e^{\alpha_j^l\beta_j^l}}\right) + \sum_{i=1}^{n_r} \log_2\left(1+\frac{1}{n_t}\sum_{l=1}^{K}\sum_{j=1}^{n_t}\lambda_j^lV_{i,j}^l\alpha_j^l\right) \tag{2.15}$$

must satisfy

$$\lambda_j^l = \frac{\alpha_j^l\beta_j^l}{\frac{1}{n_t}\sum_{k=1}^{n_t}\alpha_k^l\beta_k^l.} \tag{2.16}$$

To maximize (2.15), it is equivalent to maximize

$$\tilde{\mathcal{I}}(\boldsymbol{\lambda}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{l=1}^{K}\sum_{j=1}^{n_t} \ln(1+p^l\beta_j^l) + \sum_{i=1}^{n_r} \ln\left(1+\frac{1}{n_t}\sum_{l=1}^{K}\sum_{j=1}^{n_t}\lambda_j^lV_{i,j}^l\alpha_j^l\right) - \sum_{l=1}^{K}\sum_{j=1}^{n_t}\alpha_j^l\beta_j^l \tag{2.17}$$

Accordingly, we can introduce the Lagrangian multiplier and generate a set of KKT conditions for user m as

$$\begin{cases} \frac{\partial \tilde{\mathcal{I}}}{\partial \lambda_k^m} = u^m & \text{for} \quad \lambda_k^m > 0 \\ \frac{\partial \tilde{\mathcal{I}}}{\partial \lambda_k^m} < u^m & \text{for} \quad \lambda_k^m = 0 \\ u^m \left(\sum_{k=1}^{n_t} \lambda_k^m - n_t \right) = 0 \end{cases} \tag{2.18}$$

To solve this optimization problem, we need to deal with the first order derivative of (2.17) to all the power coefficients $\lambda_j^l, l = 1, \cdots, K, j = 1, \cdots, n_t$. We will derive the derivative of the three terms in (2.17) to an arbitrary λ_k^m respectively. For the first term, the derivative is

$$\frac{\partial \sum_{l=1}^{K} \sum_{j=1}^{n_t} \ln(1 + p^l \beta_j^l)}{\partial \lambda_k^m} = \sum_{l=1}^{K} \sum_{j=1}^{n_t} \frac{p^l \partial \beta_j^l / \partial \lambda_k^m}{1 + p \beta_j^l} = \sum_{l=1}^{K} \sum_{j=1}^{n_t} \alpha_j^l \frac{\partial \beta_j^l}{\partial \lambda_k^m} \tag{2.19}$$

For the second term, the derivative is

$$\begin{aligned} \frac{\partial \sum_{i=1}^{n_r} (1 + \frac{1}{n_t} \sum_{l=1}^{K} \sum_{j=1}^{n_t} \lambda_j^l V_{i,j}^l \alpha_j^l)}{\partial \lambda_k^m} &= \sum_{i=1}^{n_r} \frac{\sum_{l=1}^{K} \sum_{j=1}^{n_t} \lambda_j^l V_{i,j}^l \frac{\partial \alpha_j^l}{\partial \lambda_k^m} + V_{i,k}^m \alpha_k^m}{n_t + \sum_{l=1}^{K} \sum_{j=1}^{n_t} \lambda_j^l V_{i,j}^l \alpha_j^l} \\ &= \sum_{l=1}^{K} \sum_{j=1}^{n_t} \beta_j^l \frac{\partial \alpha_j^l}{\partial \lambda_k^m} + \sum_{i=1}^{n_r} \frac{V_{i,k}^m \alpha_k^m}{n_t + \sum_{l=1}^{K} \sum_{j=1}^{n_t} \lambda_j^l V_{i,j}^l \alpha_j^l} \end{aligned} \tag{2.20}$$

For the third term, the derivative is

$$\frac{\partial \left(- \sum_{l=1}^{K} \sum_{j=1}^{n_t} \alpha_j^l \beta_j^l \right)}{\partial \lambda_k^m} = - \sum_{l=1}^{K} \sum_{j=1}^{n_t} \left(\beta_j^l \frac{\partial \alpha_j^l}{\partial \lambda_k^m} + \alpha_j^l \frac{\partial \beta_j^l}{\partial \lambda_k^m} \right) \tag{2.21}$$

Summing up (2.19), (2.20), and (2.21), we obtain

$$\frac{\partial \tilde{\mathcal{I}}}{\partial \lambda_k^m} = \sum_{i=1}^{n_r} \frac{V_{i,k}^m \alpha_k^m}{n_t + \sum_{l=1}^{K} \sum_{j=1}^{n_t} \lambda_j^l V_{i,j}^l \alpha_j^l} = \frac{\alpha_k^m \beta_k^m}{\lambda_k^m} \tag{2.22}$$

and therefore we have $\lambda_k^m \frac{\partial \mathcal{I}}{\partial \lambda_k^m} = \alpha_k^m \beta_k^m$.

From the first two lines of KKT condition, we have $\lambda_k^m \frac{\partial \mathcal{I}}{\partial \lambda_k^m} = \lambda_k^m u^m$ for arbitrary k. Thus,

$$u^m = u^m \frac{1}{n_t}\Big(\sum_{k=1}^{n_t} \lambda_k^m\Big) = \frac{1}{n_t}\Big(\sum_{k=1}^{n_t} u^m \lambda_k^m\Big) = \frac{1}{n_t}\Big(\sum_{k=1}^{n_t} \lambda_k^m \frac{\partial \tilde{\mathcal{I}}}{\partial \lambda_k^m}\Big) = \frac{1}{n_t}\Big(\sum_{k=1}^{n_t} \alpha_k^m \beta_k^m\Big) \tag{2.23}$$

Then we can update λ_k^m as

$$\lambda_k^m = \frac{\lambda_k^m \frac{\partial \mathcal{I}}{\partial \lambda_k^m}}{u^m} = \frac{\alpha_k^m \beta_k^m}{\frac{1}{n_t}\sum\limits_{j=1}^{n_t} \alpha_j^m \beta_j^m} \tag{2.24}$$

Based on the above derivations, we can develop an iterative algorithm to find $\{\lambda_j^{l^*}, \alpha_j^{l^*}, \beta_j^{l^*}\}$ as follows:

Algorithm 2: The low-complexity algorithm to compute the optimal covariance matrices

---

For a K-user MIMO-MAC-VR system, initiate $\tilde{Q}^1, \cdots, \tilde{Q}^K$ to be identity matrices, then take the following steps:

1) For each user $k = 1, \cdots, K$,

a. Fix $\{\alpha_j^k\}$ and $\{\beta_j^k\}$ for all other users $k' \neq k$;

b. Update $\{\alpha_j^k\}$ and $\{\beta_j^k\}$, where $j = 1, \cdots, n_t$ according to (2.14) till the values converge;

c. Update $\{\lambda_j^{k(n)}, j = 1, \cdots, n_t\}$ according to (2.16);

d. if $||\tilde{Q}^{k(n)} - \tilde{Q}^{k(n-1)}||^2 < \varepsilon_1$, return $\tilde{Q}^k = \tilde{Q}^{k(n)}$ and step forward to the next user;

2) check the convergence condition, if it is satisfied, stop iteration; or else repeat steps 1 and 2.

Since $\alpha_j^l \approx p^l \mathrm{MMSE}_j^l$ and $\beta_j^l \approx \frac{\mathrm{SINR}_j^l}{p^l}$, we have

$$\alpha_j^l \cdot \beta_j^l \approx p^l \mathrm{MMSE}_j^l \cdot \frac{\mathrm{SINR}_j^l}{p^l} = \Big(p^l \mathrm{MMSE}_j^l\Big)\frac{1 - \mathrm{MMSE}_j^l}{p^l \mathrm{MMSE}_j^l} = 1 - \mathrm{MMSE}_j^l$$

When replacing MMSE_j^l by $\overline{\mathrm{MMSE}}_j^l$, the power updating step in (2.16) resembles that of (2.9).

2.4 Optimality of Beamforming

One practical issue in realistic propagation environments is the channel rank deficiency caused by double scattering or keyhole effects. Rank deficiency may decrease the channel spatial multiplexing gain and thereby severely degrade the MIMO

capacity. Based on physical measurements, A keyhole channel [38] has been proposed which indicates the occurrence of a rank-deficient channel where the channel has only one single degree of freedom even though the channel fading is uncorrelated. In this model, the channel elements are in the product form of two independent complex Gaussian variables, instead of being the complex Gaussian random variables as assumed in the i.i.d. Gaussian channels. Similarly, a double scattering MIMO channel model that includes both the fading correlation and rank deficiency was proposed [34], and it is pointed out that for this type of channels, there is no spatial correlation among the transmit and receive antennas but the channel still shows a poor rank property. The keyhole channel model can actually be viewed as a special case of double scattering channel model. Both types of channel models exhibit significant degradation on the achievable capacity, and require different signaling strategies other than the equal power solution for the i.i.d. channels. Especially for the keyhole channels, the capacity-achieving covariance should be rank one.

In previous sections, we considered the optimality conditions of the capacity-achieving inputs and proposed a low-complexity approach to obtain near-optimum solutions. However, those designs might still be too complicated to implement in realistic communication systems. An appealing alternative scheme is beamforming, which is of even lower complexity. In such scenario, only one diagonal element, say λ_j^k, in the input covariance matrix $\tilde{Q}^k$ is nonzero. Hence, all transmission power is allocated to the j-th (transmit) virtual angle, which we refer to as the beamforming angle of user k. One advantage of beamforming is that the scalar codec techniques can be used, which greatly simplifies the transmission, so that beamforming is quite desirable in practise. The beamforming scheme does not guarantee to achieve sum capacity. Thus, the research on optimality region of beamforming becomes meaningful.

In the literature, most work was focused on the Kronecker model. In the single-user case when the transmitter knows only the channel covariance matrix and the receiver knows perfect CSI, it was shown that beamforming is optimal under certain conditions for single-sided correlation environment [10] or double-sided correlation environment [29]. In [29], the conditions for beamforming to be optimal is shown to depend on SNR and the eigenvalues of channel covariance matrix, and it is revealed

the capacity-achieving beamforming should be along the direction of strongest eigenvalue of the channel covariance matrix. For the Kronecker model, it can be easily shown that the eigenvectors of channel covariance matrix are the same the as the eigenvectors of the transmit correlation matrix. For the virtual representation model, however, there are few results available on the issue of optimality of beamforming. In [14], it is shown that beamforming along one transmit virtual angle is asymptotically optimal at low SNR regime, for single-user MIMO channels via virtual representation. General conditions for the optimality of beamforming has not been derived yet for the virtual representation model, in either single-user or multi-use case.

Therefore, here we study the conditions under which beamforming achieves the sum capacity (2.5) under MIMO-MAC-VR. Given a K-user system. Assume that all users perform beamforming. Let $\tilde{h}_{\mathrm{B}}^{k}$ denote the column of $\tilde{H}^{k}$ that corresponds to the beamforming angle of user k. Since each user k performs beamforming, we have $\tilde{H}^{k}\tilde{Q}^{k}\tilde{H}^{k\dagger} = n_t\tilde{h}_{\mathrm{B}}^{k}\tilde{h}_{\mathrm{B}}^{k\dagger}$. Hence,

$$I + \sum_{l=1}^{K} \frac{p^l}{n_t}\tilde{H}^{l}\tilde{Q}^{l}\tilde{H}^{l\dagger} = I + \sum_{l=1}^{K} p^k\tilde{h}_{\mathrm{B}}^{k}\tilde{h}_{\mathrm{B}}^{k\dagger}.$$

. By substituting it into (2.6), we obtain a different form of the optimality conditions as: beamforming is optimal, in the sense of achieving the sum capacity (2.5), if and only if for every $j = 1, \cdots, n_t$, $k = 1, \cdots, K$, the condition

$$f_j^k(p^k) = E\left[\tilde{h}_j^{k\dagger}A^{-1}\tilde{h}_j^{k} - \tilde{h}_{\mathrm{B}}^{k\dagger}A^{-1}\tilde{h}_{\mathrm{B}}^{k}\right] \leq 0 \tag{2.25}$$

is satisfied, where

$$A = I + \sum_{l=1}^{K} p^l\tilde{h}_{\mathrm{B}}^{l}\tilde{h}_{\mathrm{B}}^{l\dagger} \tag{2.26}$$

Since each user k performs beamforming, we have $\tilde{H}^{k}\tilde{Q}^{k}\tilde{H}^{k\dagger} = n_t\tilde{h}_{\mathrm{B}}^{k}\tilde{h}_{\mathrm{B}}^{k\dagger}$. Hence,

$$I + \sum_{l=1}^{K} \frac{p^l}{n_t}\tilde{H}^{l}\tilde{Q}^{l}\tilde{H}^{l\dagger} = I + \sum_{l=1}^{K} p^k\tilde{h}_{\mathrm{B}}^{k}\tilde{h}_{\mathrm{B}}^{k\dagger}.$$

Based on the above derivations, we further investigate properties of optimal beamforming angles and how SNR affects the optimality of beamforming. We consider the single user case and the multi user case separately.

2.4.1 Single-User Case

It is observed in [14] numerically that there exists an SNR threshold below which beamforming is optimal and above which beamforming is suboptimal. Here we provide a mathematical proof of this threshold behavior. We first define the *sum-variance* of the i-th virtual angle as $\sum_{j=1}^{n_r} V_{j,i}$.

Theorem 2 *For a single-user VR channel, beamforming to the i-th virtual angle is optimal (capacity-achieving) if and only if*

(a) The i-th virtual angle has a sum-variance that is strictly larger than the sum-variance of any other virtual angles. Thus, the capacity-achieving beamforming angle is unique.

(b) The SNR is below a threshold $p < p_s$, where p_s is a fixed constant.

Proof. see Section 2.8.2.

As stated in Theorem 2 (a), the i-th angle is the unique capacity-achieving beamforming angle if and only if it has a sum-variance that is strictly larger than the sum-variance of any other virtual angles. This is a generalization of previous results on the optimality of beamforming in a single user case. In the special case when $\tilde{H}^k$ is an i.i.d complex Gaussian matrix, since all the virtual angles have the same sum-variance, it follows from Theorem 2 (a) that none of the beamforming angle is capacity-achieving. This is consistent with the well known result that the capacity-achieving input distribution for this $\tilde{H}^k$ is the proper complex Gaussian distribution with equal power allocation over each virtual angle. Hence, for $\tilde{H}^k$ beamforming is not optimal. In the case of Kronecker model, the sum-variance of each transmit angle is proportional to the eigenvalue of the transmit correlation matrix. Therefore, the capacity-achieving beamforming strategy is to transmit in the eigenmode along the spatial direction corresponding to the strongest eigenvalue. This is also consistent with the results published in previous papers.

2.4.2 Multi User Case

As opposed to the single-user case where the capacity-achieving beamforming angle (c.b.a.) of a given user is unique, we first present an example to show that

for a multi user system, the c.b.a. of a particular user may vary with the SNRs and beamforming angles of other users in the system.

Example 1 *Consider a two-user MAC-VR with $n_t = n_r = 2$. The variance matrices are $V^1 = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}$, $V^2 = \begin{pmatrix} 2.5 & 0 \\ 0 & 1.5 \end{pmatrix}$. From condition (2.25) we find that, if $(p^1, p^2) = (-15\ dB, -10\ dB)$, the first virtual angle is the c.b.a. for both users. If $(p^1, p^2) = (-3\ dB, -10\ dB)$, however, the second virtual angle becomes the c.b.a. for user 2, while the first angle is still the c.b.a. for user 1.*

In general, any of the n_t virtual angles could be the c.b.a. at certain SNR. However, as shown in Theorem 3, when the variance matrix satisfies certain properties, some of the virtual angles cannot be the c.b.a.

Theorem 3 *Consider user k in a K-user MAC-VR. The i-th virtual angle cannot be the c.b.a. of user k if there exists another virtual angle j such that*

$$V^k_{m,i} \leq V^k_{m,j} \quad \textit{for every } m = 1, \cdots, n_r. \tag{2.27}$$

Proof. see Section 2.8.3.

Example 2 *Assume that $V^k = \begin{pmatrix} 1.5 & 1.6 & 0.3 \\ 0.6 & 0.8 & 1.2 \end{pmatrix}$. Since (2.27) is satisfied for $i = 1$ and $j = 2$, it follows from Theorem 3 that the first virtual angle cannot be the c.b.a for user k. This result holds independent of the SNR and other users' beamforming angles.*

An immediate corollary of Theorem 3 is as follows:

Corollary 1 *If there exists a virtual angle i such that $V^k_{m,i} > V^k_{m,j}$, for every $1 \leq m \leq n_r$ and $j \neq i$, then angle i is the only possible c.b.a.*

Corollary 1 implies that for the MAC-Kr, the c.b.a. of user k is the j-th angle that maximizes $\{\gamma^k_i, i = 1, \cdots, n_t\}$. If there are multiple angles that maximize $\{\gamma^k_i\}$, then beamforming cannot be optimal.

2.5 Power Allocation for Large Systems

We consider the optimality of beamforming for large systems, namely, systems where the number of users is much larger than the number of receive antennas.

Beamforming was shown [39] to be optimal asymptotically in the case that perfect CSI is known to the transmitters. Beamforming is suboptimal in the scenario where transmitters have no CSI [3]. For the Kronecker model, it has been proven [12] that beamforming is asymptotically optimal when each user transmits to the direction of the strongest eigenvector of its transmit correlation matrix. Therefore, an interesting question is, for MIMO-MAC-VR, is that whether the asymptotically optimal transmit strategy simply beamforming or not, with finite number of transmit/receive antennas and let the number of uses goes to infinity.

In this section, we show that under mild conditions, the sum-capacity of a K-user system, denoted by $C(K)$, grows in the order of $n_r \log K$. Furthermore, we present conditions under which the sum-rate achieved by a power allocation scheme $\boldsymbol{\lambda}$ is within a constant of $C(K)$ as K goes to infinity.

Given a power allocation $\boldsymbol{\lambda} = \{\lambda_j^k, k = 1, \cdots, K, j = 1, \cdots, n_t\}$, let $A_K = I_{n_r} + \sum_{k=1}^{K} \frac{p^k}{n_t} \tilde{H}^k \tilde{Q}^k \tilde{H}^{k\dagger} = I_{n_r} + \sum_{k=1}^{K} \sum_{j=1}^{n_t} \frac{p^k}{n_t} \lambda_j^k \tilde{h}_j^k \tilde{h}_j^{k\dagger}$. The sum-rate achieved by $\boldsymbol{\lambda}$ is given by $I(\boldsymbol{\lambda}, K) = E[\log\det A_K]$. We apply Jensen's inequality to obtain an upper bound $\bar{I}(\boldsymbol{\lambda}, K)$ such that

$$\begin{aligned} I(\boldsymbol{\lambda}, K) &= E[\log\det A_K] \leq \log\det E(A_K) \\ &= \sum_{m=1}^{n_r} \log\left(1 + \sum_{k=1}^{K} \sum_{j=1}^{n_t} \frac{p^k}{n_t} \lambda_j^k V_{m,j}^k\right) \\ &= \bar{I}(\boldsymbol{\lambda}, K). \end{aligned} \tag{2.28}$$

Proposition 1 below shows that under mild conditions, $\bar{I}(\boldsymbol{\lambda}, K)$ is asymptotically tight as $K \to \infty$.

Proposition 1 *Assume that* $M_4 = \sup_{m,j,k} (p^k)^2 E(|\tilde{h}_{m,j}^k|^4) < \infty$. *If there exists a constant* $c > 0$ *such that* $\boldsymbol{\lambda}$ *satisfies*

$$\min_{1 \leq m \leq n_r} \liminf_{K \to \infty} \frac{1}{K n_t} \sum_{k=1}^{K} \sum_{j=1}^{n_t} \frac{p^k}{n_t} \lambda_j^k V_{m,j}^k \geq c > 0, \tag{2.29}$$

then we have $I(\boldsymbol{\lambda}, K) = \bar{I}(\boldsymbol{\lambda}, K) + o(1)$, *where* $o(1)$ *converges to zero as* $K \to \infty$.

The proof is presented in Section 2.8.4.

Note that although both proofs apply the SLLN, the proof of Proposition 1 differs from that of [12, Lemma 2] for the MAC-Kr in that we provide a sufficient condition (2.29), which guarantees that (2.63) holds. From Proposition 1 we obtain Corollary 2 below.

Corollary 2 *For any $\boldsymbol{\lambda}$ that satisfies (2.29), we have $I(\boldsymbol{\lambda}, K) = C(K) + O(1) = n_r \log K + O(1)$, where $O(1)$ denotes a bounded quantity as $K \to \infty$. Hence, $I(\boldsymbol{\lambda}, K)$ grows in the order of $n_r \log K$, and asymptotically it differs from the sum-capacity $C(K)$ by only a constant.*

Sketch of proof. It is sufficient to show that there exists constants u_1 and u_c such that

$$\begin{aligned} n_r \log K + u_c + o(1) &\leq I(\boldsymbol{\lambda}, K) \\ &\leq C(K) \\ &\leq \max_{\boldsymbol{\lambda}} \bar{I}(\boldsymbol{\lambda}, K) \\ &\leq n_r \log K + u_1 + o(1). \end{aligned} \tag{2.30}$$

From (2.29) we can find u_c such that $\bar{I}(\boldsymbol{\lambda}, K) \geq n_r \log K + u_c$. This, combined with Proposition 1, leads to the first inequality of (2.30). The last inequality of (2.30) utilizes $M_2 = \sup_{m,j,k} p^k V_{m,j}^k < \infty$. ■

Next, we consider a simple example for which we can characterize the term $O(1)$ in Corollary 2 for various $\boldsymbol{\lambda}$. The accuracy of these computations will be verified in Section 2.6.

Example 3 *Assume that all users have the same $p^k = 2$ and the same variance matrix $V = \begin{bmatrix} 2 & 0.5 \\ 0.5 & 1 \end{bmatrix}$. The virtual elements in $\tilde{H}^k$ are complex Gaussian distributed. Assume that each user adopts the same power allocation $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$ such that $\lambda_1 + \lambda_2 = 2$. For each $\boldsymbol{\lambda}$ that satisfies the assumptions of Proposition 1, we have*

$$\begin{aligned} I(\boldsymbol{\lambda}, K) &= \bar{I}(\boldsymbol{\lambda}, K) + o(1) \\ &= 2\log(K) + \log(2\lambda_1 + 0.5 \cdot \lambda_2) + \\ &\quad \log(0.5 \cdot \lambda_1 + \lambda_2) + o(1). \end{aligned} \tag{2.31}$$

Consider the following three power allocations for which Proposition 1 and Corollary 2 are applicable.

(1) The beamforming scheme $\boldsymbol{\lambda}_{BF}$ where each user beamforms to the first virtual angle which has the largest sum-variance. Since $\lambda_1 = 2, \lambda_2 = 0$, it follows from (2.31) that

$$\begin{aligned} I(\boldsymbol{\lambda}_{BF}, K) &= \bar{I}(\boldsymbol{\lambda}_{BF}, K) + o(1) \\ &= 2\log K + \log(4) + o(1). \end{aligned} \tag{2.32}$$

(2) The equal power allocation $\boldsymbol{\lambda}_{Eq}$ such that $\lambda_1 = \lambda_2 = 1$. From (2.31) we have

$$\begin{aligned} I(\boldsymbol{\lambda}_{Eq}, K) &= \bar{I}(\boldsymbol{\lambda}_{Eq}, K) + o(1) \\ &= 2\log K + \log(15/4) + o(1) \\ &= 2\log K + 1.9069 + o(1). \end{aligned} \tag{2.33}$$

(3) We can choose (λ_1, λ_2) to maximize the summation of the two constant terms in (2.31). This yields the optimized solution $\boldsymbol{\lambda}^* = (\lambda_1^*, \lambda_2^*) = (5/3, 1/3)$. Hence, from (2.31) we have

$$\begin{aligned} I(\boldsymbol{\lambda}^*, K) &= \bar{I}(\boldsymbol{\lambda}^*, K) + o(1) \\ &= 2\log K + \log(49/12) + o(1) \\ &= 2\log K + 2.0297 + o(1). \end{aligned} \tag{2.34}$$

The constant term in (2.34) is slightly greater than that of $\boldsymbol{\lambda}_{BF}$ in (2.32) and that of $\boldsymbol{\lambda}_{Eq}$ in (2.33). This example demonstrates that beamforming may not be asymptotically optimal for the MAC-VR, even though it is asymptotically optimal for MAC-Kr [12, Theorem 7]. Corollary 3 below provides a sufficient condition under which beamforming is asymptotically optimal for MAC-VR.

Corollary 3 *Beamforming is asymptotically optimal for the MAC-VR: $I(\boldsymbol{\lambda}_{BF}, K) = C(K) + o(1)$, as $K \to \infty$, if each user k beamforms to a virtual angle i_k that satisfies $V_{m,i_k}^k \geq V_{m,j}^k$ for every $1 \leq m \leq n_r$ and $1 \leq j \leq n_t$, and there exists a constant $c > 0$ such that*

$$\min_{1 \leq m \leq n_r} \liminf_{K \to \infty} \frac{1}{K} \sum_{k=1}^{K} p^k V_{m,i_k}^k \geq c > 0. \tag{2.35}$$

Sketch of proof. Given $V^k_{m,i_k} \geq V^k_{m,j}$, one can show that $C(K) \leq \bar{I}(\boldsymbol{\lambda}_{\mathrm{BF}}, K)$. Condition (2.35) ensures that $I(\boldsymbol{\lambda}_{\mathrm{BF}}, K) = \bar{I}(\boldsymbol{\lambda}_{\mathrm{BF}}, K) + o(1)$. Thus Corollary 3 follows.

Considering a special case of Corollary 3 in which we let i_k be the virtual angle that maximizes $\{b^k_j, j = 1, \cdots, n_t\}$, then we obtain the same result as [12, Theorem 7] that beamforming is always asymptotically optimal for MAC-Kr. In comparison, as shown in Example 3, there exists MAC-VR such that beamforming is not asymptotically optimal. This difference, again, is due to the general structure of the variance matrix for MAC-VR.

2.6 Numerical Results

In this section, we present numerical examples to illustrate the theoretical results given in previous sections. Four power allocation schemes are considered: the equal power allocation ($\boldsymbol{\lambda}_{\mathrm{Eq}}$), the beamforming scheme ($\boldsymbol{\lambda}_{\mathrm{BF}}$), the optimal power allocation ($\boldsymbol{\lambda}_{\mathrm{Opt}}$) found by the algorithm of [30], and a low-complexity power allocation algorithm derived based on [40] ($\boldsymbol{\lambda}_{\mathrm{Low}}$). Let $I(\boldsymbol{\lambda})$ denote the sum-rate achieved by $\boldsymbol{\lambda}$. We first consider a single-user system with $n_r = n_t = 5$. The virtual coefficients in $\tilde{H}$ are assumed to be complex Gaussian distributed with the same variance matrix as the one in [14], given by

$$V = \frac{25}{5.7} \begin{bmatrix} 0.1 & 0 & 1 & 0 & 0 \\ 0 & 0.1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0.25 & 0 \\ 0 & 0 & 1 & 0 & 0.25 \end{bmatrix}. \tag{2.36}$$

Such a variance matrix could represent a physical environment with two very small scatterers (corresponding to the first two columns), two bigger scatterers (corresponding to the last two columns), and one large scattering cluster (corresponds to the all one column). The third virtual angle is the beamforming angle because it has the largest sum-variance. Figure 2.2 shows that $I(\boldsymbol{\lambda}_{\mathrm{Low}})$ is very close to $I(\boldsymbol{\lambda}_{\mathrm{Opt}})$ for the entire range of SNRs considered. $I(\boldsymbol{\lambda}_{\mathrm{Eq}})$ is near optimal only at high SNR and $I(\boldsymbol{\lambda}_{\mathrm{BF}})$ is optimal only when SNR is below the threshold of 0.29 dB. This is consistent with the threshold behavior proved in Theorem 2.

It is also interesting to investigate the power allocation schemes. In this example, even if the third virtual angle is much stronger than the rest, $I(\boldsymbol{\lambda}_{\mathrm{Eq}})$ still converges

to the channel capacity at high SNR. In Figure 2.3, we plot the optimal power at all virtual angles $\{\lambda_i, i = 1, \cdots, n_t\}$ versus SNR. While at low SNR it is optimal to allocate all power to the strongest virtual angle, at high SNR it tends to be optimal to spread power equally over all virtual angles, no matter how weak some of these angles are. This complies with the waterfilling strategy when the transmitter knows the perfect CSI, and has been indicated in our low-complexity algorithm 2. In the case of $K = 1$, according to (2.14), we would have $\alpha_j\beta_j \approx 1$ for large values of $\Lambda >> 0$, which forces the optimal solution to equal power allocation as shown in 2.16. In Figure 2.4, we make comparisons between the optimal power allocation and the solution from the low-complexity algorithm 2. Algorithm 2 tends to put more power on the dominant virtual angle but it also appears to converge to $I(\boldsymbol{\lambda}_{\mathrm{Eq}})$ as SNR increases.

In Figure 2.5, we plot the beamforming conditions defined in Section 2.4.1 for the third virtual angle $i = 3$. Since $f_2(p) = f_1(p)$ and $f_5(p) = f_4(p)$, Figure 2.5 plots $f_1(p)$ and $f_4(p)$ and shows that when SNR is below 0.29 dB, both functions are negative and thus beamforming to the third virtual angle is optimal. In this example, it shows that i.i.d. inputs may achieve near-optimal performance when the transmission is over sufficiently rich scattering channels, or when each of virtual transmit angles appears to be symmetric in the variance matrix and shows similar fading strength.

In Figure 2.6, we examine the accuracy of Proposition 1 and Corollary 2 by comparing the asymptotic expressions (2.32)-(2.34) in Example 3 with numerical values of $I(\boldsymbol{\lambda}, K)$ obtained through Monte Carlo integration. Three functions $I(\boldsymbol{\lambda}_{\mathrm{BF}}, K) - 2\log K$, $I(\boldsymbol{\lambda}_{\mathrm{Eq}}, K) - 2\log K$, and $I(\boldsymbol{\lambda}^{*}, K) - 2\log K$ are plotted to confirm that as K increases, they indeed converge to the predicted constants 2, 1.9069, and 2.0297, respectively.

The optimality of beamforming heavily depends on the structure of the variance matrix V. The example we looked at above considers the environment with large transmit antenna correlation, where one or several virtual transmit angles are much stronger than the rest. In that case, it can be optimal to transmit more power on these strong virtual angles. In contrary, in the case where the correlation among transmit antennas is negligible, $I(\boldsymbol{\lambda}_{\mathrm{Eq}})$ can be superior to $I(\boldsymbol{\lambda}_{\mathrm{BF}})$ in most SNR region.

An example is shown in Figure 2.7 by simply taking the transpose of V to be the variance matrix.

In the multi user case, we consider the typical style of variance matrices with dominant angles. For simplicity, we assume the dominant angles have all ones while other angles have only nonzero values at diagonal elements. We randomize all those nonzero elements to be uniformly distributed in $(0, 1)$. Examples of four users with one or two dominant angles are shown as follows. We assume a system with $K = 4$ users, where each user have $n_t = n_r = 5$ number of transmit and receive angles, the variance matrices are listed as above. Assuming user 1 have two dominant angles while other three all have only one dominant angle whose positions are randomly assigned. The results are plotted in Figure 2.8. As shown in this figure, Algorithm 2 still performs close to the optimum; the beamforming solution is optimal at low SNR but at high SNR ($\Gamma = 15dB$), it is about $1dB$ less than optimum; while equal power allocation exhibits $1dB$ degradation in all SNR region.

$$V^1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0.1 & 0 & 0 & 0 \\ 1 & 0 & 0.3 & 0 & 0 \\ 1 & 0 & 0 & 0.5 & 0 \\ 1 & 0 & 0 & 0 & 0.1 \end{bmatrix} \quad V^2 = \begin{bmatrix} 0.25 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0.25 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0.4 \end{bmatrix}$$
$$V^3 = \begin{bmatrix} 0.2 & 0 & 1 & 0 & 0 \\ 0 & 0.1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0.5 & 0 \\ 0 & 0 & 1 & 0 & 0.15 \end{bmatrix} \quad V^4 = \begin{bmatrix} 0.1 & 0 & 0 & 1 & 0 \\ 0 & 0.6 & 0 & 1 & 0 \\ 0 & 0 & 0.3 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0.8 \end{bmatrix} \tag{2.37}$$

Figure 2.9 considers a multi user system in which each V^k is generated independently, taking a form similar to (2.36). The beamforming angle of user k is chosen to be the virtual angle with the largest sum-variance. Because $\{V^k\}$ satisfy conditions of Corollary 3, $\boldsymbol{\lambda}_{\mathrm{BF}}$ is asymptotically optimal. This is confirmed in Figure 2.9. The curve for $I(\boldsymbol{\lambda}_{\mathrm{Opt}}, K)$ is not shown due to high complexity for computing $\boldsymbol{\lambda}_{\mathrm{Opt}}$. Instead, we provide a simple sum-capacity upper bound $C(K) \leq n_r \log(1 + KM_2)$ as a performance benchmark for large K. Hence, the gap between $I(\boldsymbol{\lambda}_{\mathrm{BF}}, K)$ and $C(K)$ is less than the small gap shown in Figure 2.9 between $I(\boldsymbol{\lambda}_{\mathrm{BF}}, K)$ and the upper bound. The gap becomes negligible as K increases, confirming the optimality of beamforming. For small K, $I(\boldsymbol{\lambda}_{\mathrm{Low}}, K)$ closely approximates $I(\boldsymbol{\lambda}_{\mathrm{Opt}}, K)$ (not shown).

For large K, $I(\boldsymbol{\lambda}_{\mathrm{Low}}, K)$ and $I(\boldsymbol{\lambda}_{\mathrm{BF}}, K)$ merge quickly and become indistinguishable after $K \geq 30$. We note that $I(\boldsymbol{\lambda}_{\mathrm{Eq}}, K)$ is inferior to $I(\boldsymbol{\lambda}_{\mathrm{BF}}, K)$ by roughly a constant, even though it achieves the same asymptote of $n_r \log K = 5 \log K$. This is consistent with Corollary 2.

Up to now, we do all the simulations assuming the channel is Gaussian distributed. However, our results should be applicable to arbitrary channel distributions due to the definition of virtual representation model. To address this, we replaced the Gaussian distributed channel with uniformly distributed channel [1] and repeated the previous simulations. Uniformly distributed channels achieve higher capacity than the Gaussian channels but the distribution only affects the scaling of MIMO channel capacity. The results we presented in this work always hold regardless of the channel distribution. In Figure 2.10, we repeat the single-user simulation and the figure shows similar results to the Gaussian distribution case which leads to the same conclusions as above. Similarly, Figure2.11 shows the results of multi user case with uniformly distributed channel, as a comparison to Figure 2.8. Figure 2.12 shows the sum-rates in the large system with Gaussian/Uniform distribution. As confirmed in this figure, the sum capacity converges to be deterministic and is independent of the channel distribution. Both distributions converge to the same capacity level.

2.7 Conclusion

In this work, we study the optimal input distribution and the optimality of beamforming for the MIMO single user channel and the MIMO MAC based on virtual representation. We proposed a low-complexity algorithm to design near-optimum input distribution. We obtain the general condition for the beamforming to be optimal, and demonstrate that in contrast to the single user case, the optimal beamforming angle of a given user to achieve sum capacity for the MIMO MAC depends on the SNRs and beamforming angles of other users. We derive the conditions to disprove a beamforming angle from being the optimal beamforming angle by extracting information from the structure of channel variance matrix. The selection of optimal beamforming angles, however, requires further investigation. In particular, it

will be interesting to study practical algorithms for the joint selection of beamforming angles for the MIMO MAC.

2.8 Appendix

2.8.1 Proof of Theorem 1

Let us rewrite the capacity formulation here:

$$\begin{aligned} C &= \max_{\tilde{Q}^k, k=1,\cdots,K} \mathcal{I}\left(\tilde{Q}^1, \cdots \tilde{Q}^K\right) \\ &= \max_{\tilde{Q}^k, k=1,\cdots,K} E\left[\log\det\left(I + \sum_{k=1}^{K} \frac{p^k}{n_t} \tilde{H}^k \tilde{Q}^k \tilde{H}^{k\dagger}\right)\right] \\ s.t. \quad & \mathrm{Tr}(\tilde{Q}^k) \leq n_t \quad \text{for} \quad k = 1, \cdots, K \end{aligned}$$

We separate the proof of Theorem 1 into two parts. In the first part, we prove that the capacity-achieving covariance matrices $\tilde{Q}^k$ are diagonal matrices for all $k = 1, \cdots, K$. To make the proof, we follow the technique in []; In the second part we derive the set of necessary and sufficient conditions for optimal covariance matrices as shown in the Theorem.

1). To prove the capacity-achieving $\{\tilde{Q}^k, k = 1, \cdots, K\}$ are diagonal matrices, it is sufficient to prove that the optimal $\tilde{Q}^1$ is diagonal. In another word, we fix the values of all other covariance matrices $\tilde{Q}^k, k = 2, \cdots, K$ so that the mutual information can be expressed as $\mathcal{I}\left(\tilde{Q}^1\right)$. We define two sets of matrices as:

$$\Omega := \{\tilde{Q} : \tilde{Q} \text{ is positive semidefinte, and Tr } \{\tilde{Q}\} \leq n_t\}$$

$$\Omega_\Lambda := \{\Lambda : \Lambda \text{ is diagonal, and } \Lambda \in \Omega\}$$

We first consider to optimize the mutual information in (2.5) over the set Ω_Λ. In this scenario, the differentiable function $\mathcal{I}(\tilde{Q}^1)$ is strictly concave over the convex set Ω_Λ. Therefore, there must exist a unique Λ^* that maximizes $\mathcal{I}\left(\tilde{Q}^1\right)$ over Ω_Λ, and Λ^* satisfies the following condition:

$$\delta\mathcal{I}(\Lambda^*; \Lambda - \Lambda^*) \leq 0, \quad \forall \Lambda \in \Omega_\Lambda \tag{2.38}$$

where

$$\delta\mathcal{I}(\Lambda^*; \Lambda - \Lambda^*) := \lim_{\alpha \to 0} [\mathcal{I}(\Lambda^* + \alpha(\Lambda - \Lambda^*)) - \mathcal{I}(\Lambda^*)]$$

The left-hand side of (2.38) can be computed as follows:

$$\begin{aligned}
&\delta\mathcal{I}(\Lambda^*;\Lambda-\Lambda^*)\\
&\quad=\frac{d}{d\alpha}\mathcal{I}(\Lambda^*+\alpha(\Lambda-\Lambda^*))\mid_{x=0}\\
&\quad=\frac{d}{d\alpha}\mathrm{E}\left[\log\det\left(I+\sum_{k=2}^{K}\frac{p^k}{n_t}\tilde{H}^k\tilde{Q}^k\tilde{H}^{k\dagger}+\frac{p^1}{n_t}\tilde{H}^1(\Lambda^*+\alpha(\Lambda-\Lambda^*))\tilde{H}^{1\dagger}\right)\right]\mid_{\alpha=0}
\end{aligned} \tag{2.39}$$

since the matrix

$$\left(I+\sum_{k=2}^{K}\frac{p^k}{n_t}\tilde{H}^k\tilde{Q}^k\tilde{H}^{k\dagger}+\frac{p^1}{n_t}\tilde{H}^1(\Lambda^*+\alpha(\Lambda-\Lambda^*))\tilde{H}^{1\dagger}\right) \tag{2.40}$$

is positive definite, we can further compute the derivative of the above equation to be

$$\delta\mathcal{I}(\Lambda^*;\Lambda-\Lambda^*)=\mathrm{E}\ \mathrm{Tr}\left\{\left(I+\sum_{k=2}^{K}\frac{p^k}{n_t}\tilde{H}^k\tilde{Q}^k\tilde{H}^{k\dagger}+\frac{p^1}{n_t}\tilde{H}^1\Lambda^*\tilde{H}^{1\dagger}\right)^{-1}\frac{p^1}{n_t}\tilde{H}^1(\Lambda-\Lambda^*)\tilde{H}^{1\dagger}\right\}$$

In this way, we convert the condition 2.38 to the following form:

$$\mathrm{E}\ \mathrm{Tr}\left\{\left(I+\sum_{k=2}^{K}\frac{p^k}{n_t}\tilde{H}^k\tilde{Q}^k\tilde{H}^{k\dagger}+\frac{p^1}{n_t}\tilde{H}^1\Lambda^*\tilde{H}^{1\dagger}\right)^{-1}\frac{p^1}{n_t}\tilde{H}^1(\Lambda-\Lambda^*)\tilde{H}^{1\dagger}\right\}\leq 0 \tag{2.41}$$

for $\forall\Lambda\in\Omega_\Lambda$. Next, we will prove that Λ^* maximizes the mutual information $\mathcal{I}\left(\tilde{Q}^1\right)$ when the optimization is performed over the set Ω. It is sufficient to show that

$$\delta\mathcal{I}(\Lambda^*;\tilde{Q}^1-\Lambda^*)\leq 0,\ \forall\tilde{Q}^1\in\Omega \tag{2.42}$$

where

$$\delta\mathcal{I}(\Lambda^*;\tilde{Q}^1-\Lambda^*)=\lim_{\alpha\to 0}\frac{1}{\alpha}\left[\mathcal{I}\left(\Lambda^*+\alpha(\tilde{Q}^1-\Lambda^*)\right)-\mathcal{I}(\Lambda^*)\right]$$

first of all, we express $\tilde{Q}^1=\Lambda^1+\bar{\Lambda}^1$ to separate the diagonal elements from the nondiagonal element, where Λ^1 is a diagonal matrix containing diagonal elements of

$\tilde{Q}^1$, and $\bar{\Lambda}^1$ is composed of all the nondiagonal elements of $\tilde{Q}^1$ while the diagonal components of $\bar{\Lambda}^1$ are all zeroes. Accordingly, we have

$$\begin{aligned}
&\delta\mathcal{I}(\Lambda^*; \tilde{Q}^1 - \Lambda^*) \\
&\quad = \mathrm{E}\ \mathrm{Tr}\left\{ \left(I + \sum_{k=2}^{K} \frac{p^k}{n_t} \tilde{H}^k \tilde{Q}^k \tilde{H}^{k\dagger} + \frac{\Gamma}{n_t} \tilde{H}^1 \Lambda^* \tilde{H}^{1\dagger} \right)^{-1} \frac{p^1}{n_t} \tilde{H}^1 (\tilde{Q} - \Lambda^*) \tilde{H}^{1\dagger} \right\} \\
&\quad = \mathrm{E}\ \mathrm{Tr}\left\{ \left(I + \sum_{k=2}^{K} \frac{p^k}{n_t} \tilde{H}^k \tilde{Q}^k \tilde{H}^{k\dagger} + \frac{\Gamma}{n_t} \tilde{H}^1 \Lambda^* \tilde{H}^{1\dagger} \right)^{-1} \frac{p^1}{n_t} \tilde{H}^1 (\Lambda^1 - \Lambda^*) \tilde{H}^{1\dagger} \right\} \\
&\quad + \mathrm{E}\ \mathrm{Tr}\left\{ \left(I + \sum_{k=2}^{K} \frac{p^k}{n_t} \tilde{H}^k \tilde{Q}^k \tilde{H}^{k\dagger} + \frac{\Gamma}{n_t} \tilde{H}^1 \Lambda^* \tilde{H}^{1\dagger} \right)^{-1} \frac{p^1}{n_t} \tilde{H}^1 \bar{\Lambda}^1 \tilde{H}^{1\dagger} \right\}
\end{aligned} \tag{2.43}$$

Next, we examine the two right-side terms in (2.43). The first term in the preceding equation has been shown to be no greater than zero in (2.41). To evaluate the second term, we denote the columns of matrix $\tilde{H}^1$ by $\tilde{h}_1^1, \tilde{h}_2^1, \cdots, \tilde{h}_{n_t}^1$. Accordingly, we can write the second term into

$$\begin{aligned}
&\mathrm{E}\ \mathrm{Tr}\left\{ \left(I + \sum_{k=2}^{K} \frac{p^k}{n_t} \tilde{H}^k \tilde{Q}^k \tilde{H}^{k\dagger} + \frac{p^1}{n_t} \tilde{H}^1 \Lambda^* \tilde{H}^{1\dagger} \right)^{-1} \frac{p^1}{n_t} \tilde{H}^1 \bar{\Lambda}^1 \tilde{H}^{1\dagger} \right\} \\
&\quad = \sum_{m,l=1, m\neq l}^{n_t} \mathrm{E}\ \mathrm{Tr}\left\{ \left(I + \sum_{k=2}^{K} \frac{p^k}{n_t} \tilde{H}^k \tilde{Q}^k \tilde{H}^{k\dagger} + \frac{p^1}{n_t} \sum_{i=1} (n_t) \lambda_i^* \tilde{h}_i^1 \tilde{h}_i^{1\dagger} \right)^{-1} \frac{p^1}{n_t} \bar{\Lambda}_{m,l}^1 \tilde{h}_k^1 \tilde{h}_k^{1\dagger} \right\}
\end{aligned}$$

where λ_i^* is the ith diagonal entry of Λ^*.

In the above sum, consider a particular term

$$\begin{aligned}
&\mathrm{E}\ \mathrm{Tr}\left\{ \left(I + \sum_{k=2}^{K} \frac{p^k}{n_t} \tilde{H}^k \tilde{Q}^k \tilde{H}^{k\dagger} + \frac{p^1}{n_t} \sum_{i=1} n_t \lambda_i^* \tilde{h}_i^1 \tilde{h}_i^{1\dagger} \right)^{-1} \frac{p^1}{n_t} \bar{\Lambda}_{1,2}^1 \tilde{h}_1^1 \tilde{h}_2^{1\dagger} \right\} \\
&\quad = \mathrm{Tr}\Big\{ \mathrm{E}\Big[\mathrm{E}\big[(I + \sum_{k=2}^{K} \frac{p^k}{n_t} \tilde{H}^k \tilde{Q}^k \tilde{H}^{k\dagger} + \frac{p^1}{n_t} \sum_{i=1}^{n_t} \lambda_i^* \tilde{h}_i^1 \tilde{h}_i^{1\dagger})^{-1} \\
&\qquad \cdot \frac{p^1}{n_t} \bar{\Lambda}_{1,2}^1 \tilde{h}_1 \tilde{h}_2^{1\dagger} \mid \tilde{h}_2^1, \tilde{h}_3^1, \cdots, \tilde{h}_{n_t}^1 \big] \Big] \Big\}
\end{aligned} \tag{2.44}$$

According to the property of the virtual representation channel model, the columns of $\tilde{H}^1$ are independent. Therefore, the distribution of $\tilde{h}_1^1$ does not depend on $\tilde{h}_2^1, \cdots, \tilde{h}_{n_t}^1$. Note that the component inside the above expectation

$$\left(I + \sum_{k=2}^{K} \frac{p^k}{n_t} \tilde{H}^k \tilde{Q}^k \tilde{H}^{k\dagger} + \frac{p^1}{n_t} \sum_{i=1}^{n_t} \lambda_i^* \tilde{h}_i^1 \tilde{h}_i^{1\dagger} \right)^{-1} \frac{p^1}{n_t} \bar{\Lambda}_{1,2}^1 \tilde{h}_{1,2}^1 \tilde{h}_1^1 \tilde{h}_2^{1\dagger}$$

is an odd function of $\tilde{h}_1^1$, which means, if we replace $\tilde{h}_1^1$ by $-\tilde{h}_1^1$, this entry would change to its antisymmetric value. Since each element in $\tilde{H}^1$ follows a symmetric distribution around the origin, we have the inner expectation

$$\mathrm{E}\left[()^{-1}\cdot\frac{p^1}{n_t}\bar{\Lambda}_{1,2}^1\tilde{h}_1^1\tilde{h}_2^1 \mid \tilde{h}_2^1,\tilde{h}_3^1,\cdots,\tilde{h}_{n_t}^1\right]=0 \tag{2.45}$$

Hence, the particular term we considered in the sum on the right-hand side of (31) is zero. Following the same reason, all the terms in the sum are zeroes. Therefore,

$$\delta\mathcal{I}(\Lambda^*;\tilde{Q}^1-\Lambda^*)\leq 0,\ \forall\tilde{Q}^1\in\Omega \tag{2.46}$$

2). Since we have proven that the capacity-achieving $\{\tilde{Q}^k, k=1,\cdots,K\}$ are diagonal, we can define $\tilde{Q}^k=\mathrm{diag}\{\lambda_1^k,\cdots,\lambda_{n_t}^k\}$ for $\forall k\in\{1,2,\cdots,K\}$, where λ_j^k denotes the j-th diagonal element of $\tilde{Q}^k$. We have $\lambda_j^k\geq 0$ for $\forall j,k$ due to the fact that the covariance matrices are Hermitian. The optimal solution of (2.5) is characterized by a set of Karush-Kuhn-Tucker (KKT) conditions, which we derive in the following steps. Consider a specific user k. Letting $\tilde{Q}^k=\mathrm{diag}\{\lambda_1^k,\ldots,\lambda_{n_t}^k\}$ and $A=I+\sum_{l=1}^{K}\frac{p^l}{n_t}\tilde{H}^l\tilde{Q}^l\tilde{H}^{l\dagger}$, the Lagrangian of the objective function (2.5) can be written as

$$\mathcal{L}=E\log\det A-\mu^k\left(\sum_{j=1}^{n_t}\lambda_j^k-n_t\right), \tag{2.47}$$

where μ^k is the Lagrange multiplier corresponding to the power constraint $\sum_{j=1}^{n_t}\lambda_j^k=n_t$.

Note that $A=I+\sum_{j=1}^{n_t}\frac{p^k}{n_t}\lambda_j^k\tilde{h}_j^k\tilde{h}_j^{k\dagger}+\sum_{l\neq k}^{K}\frac{p^l}{n_t}\tilde{H}^l\tilde{Q}^l\tilde{H}^{l\dagger}$, we obtain

$$\frac{\partial\mathcal{L}}{\partial\lambda_j^k}=E\left[\mathrm{tr}\left(A^{-1}\frac{p^k}{n_t}\tilde{h}_j^k\tilde{h}_j^{k\dagger}\right)\right]-\mu^k=0,\quad\text{if }\lambda_j^k>0. \tag{2.48}$$

$$\frac{\partial\mathcal{L}}{\partial\lambda_j^k}=E\left[\mathrm{tr}\left(A^{-1}\frac{p^k}{n_t}\tilde{h}_j^k\tilde{h}_j^{k\dagger}\right)\right]-\mu^k\leq 0,\quad\text{if }\lambda_j^k=0. \tag{2.49}$$

It follows that

$$\mu^k\lambda_j^k=E\left[\mathrm{tr}\left(A^{-1}\frac{p^k}{n_t}\lambda_j^k\tilde{h}_j^k\tilde{h}_j^{k\dagger}\right)\right],$$

and

$$\begin{aligned}\mu^k n_t=\mu^k\left(\sum_{j=1}^{n_t}\lambda_j^k\right)&=\sum_{j=1}^{n_t}E\left[\mathrm{tr}\left(A^{-1}\frac{p^k}{n_t}\lambda_j^k\tilde{h}_j^k\tilde{h}_j^{k\dagger}\right)\right]\\&=E\left[\mathrm{tr}\left(A^{-1}\frac{p^k}{n_t}\tilde{H}^k\tilde{Q}^k\tilde{H}^{k\dagger}\right)\right]\end{aligned}$$

Therefore, we obtain

$$\mu^k = \frac{1}{n_t} E\left[\mathrm{tr}\left(A^{-1}\frac{p^k}{n_t}\tilde{H}^k\tilde{Q}^k\tilde{H}^{k\dagger}\right)\right] \tag{2.50}$$

Substituting (2.50) into (2.48) and (2.49) to yield

$$E\left[\mathrm{tr}\left\{A^{-1}\frac{p^k}{n_t}\tilde{h}_j^k\tilde{h}_j^{k\dagger}\right\}\right] - \mu^k = \frac{p^k}{n_t}E\left[\mathrm{tr}\left\{A^{-1}\left(\tilde{h}_j^k\tilde{h}_j^{k\dagger} - \frac{1}{n_t}\tilde{H}^k\tilde{Q}^k\tilde{H}^{k\dagger}\right)\right\}\right] \tag{2.51}$$

Thus, (2.6) of Theorem 1 follows.

2.8.2 Proof of Theorem 2

Following (2.25), we let $f_j(p) = E\Big[\tilde{h}_j^\dagger\Big(I_{n_r} + p\tilde{h}_i\tilde{h}_i^\dagger\Big)^{-1}\tilde{h}_j - \tilde{h}_i^\dagger\Big(I_{n_r} + p\tilde{h}_i\tilde{h}_i^\dagger\Big)^{-1}\tilde{h}_i\Big]$. We first show that

$$f_j'(p) = \frac{\partial f_j(p)}{\partial p} > \frac{1}{p}\left(E\left[\frac{1}{1+p\|\tilde{h}_i\|^2}\right] - 1\right) f_j(p). \tag{2.52}$$

To prove (2.52), we use $A^{-1} = I_{n_r} - \frac{p\tilde{h}_i\tilde{h}_i^\dagger}{1+p\|\tilde{h}_i\|^2}$ to obtain

$$\begin{aligned}
f_j(p) &= E\left[\tilde{h}_j^\dagger A^{-1}\tilde{h}_j - \tilde{h}_i^\dagger A^{-1}\tilde{h}_i\right] \\
&= E\Big[\|\tilde{h}_j\|^2 - \frac{p\|\tilde{h}_j^\dagger\tilde{h}_i\|^2}{1+p\|\tilde{h}_i\|^2}\Big] - E\Big[\|\tilde{h}_i\|^2 - \frac{p\|\tilde{h}_i\|^4}{1+p\|\tilde{h}_i\|^2}\Big] \\
&= E\Big[\frac{\|\tilde{h}_j\|^2(1+p\|\tilde{h}_i\|^2) - \|\tilde{h}_i\|^2(1+p\|\tilde{h}_i\|^2) + p\|\tilde{h}_i\|^4 - p\|\tilde{h}_j^\dagger\tilde{h}_i\|^2}{1+p\|\tilde{h}_i\|^2}\Big] \\
&= E\Big[\frac{\|\tilde{h}_j\|^2 - \|\tilde{h}_i\|^2}{1+p\|\tilde{h}_i\|^2}\Big] + E\Big[\frac{p\|\tilde{h}_j\|^2\|\tilde{h}_i\|^2 - p\|\tilde{h}_j^\dagger\tilde{h}_i\|^2}{1+p\|\tilde{h}_i\|^2}\Big] \\
&= u_j(p) + v_j(p),
\end{aligned} \tag{2.53}$$

where $u_j(p) = E\Big[\frac{\|\tilde{h}_j\|^2 - \|\tilde{h}_i\|^2\|}{1+p\|\tilde{h}_i\|^2}\Big]$ and $v_j(p) = E\Big[\frac{\|\tilde{h}_j\|^2\|\tilde{h}_i\|^2 - \|\tilde{h}_j^\dagger\tilde{h}_i\|^2}{p^{-1}+\|\tilde{h}_i\|^2}\Big]$. Due to the Cauchy-Schwartz inequality, we have $\|\tilde{h}_j\|^2\|\tilde{h}_i\|^2 - \|\tilde{h}_j^\dagger\tilde{h}_i\|^2 > 0$ and thus $v_j(p)$ is a positive, strictly increasing function of p with $v_j'(p) > 0$. It follows that

$$\begin{aligned}\frac{\partial f_j(p)}{\partial p} &= \frac{\partial u_j(p)}{\partial p} + \frac{\partial v_j(p)}{\partial p} \\ &> \frac{\partial u_j(p)}{\partial p} \\ &= E\left[\frac{\|\tilde{h}_i\|^4 - \|\tilde{h}_j\|^2\|\tilde{h}_i\|^2}{(1+p\|\tilde{h}_i\|^2)^2}\right] \\ &= \frac{1}{p}E\left[\frac{\|\tilde{h}_i\|^2 - \|\tilde{h}_j\|^2}{1+p\|\tilde{h}_i\|^2}\right] + \frac{1}{p}E\left[\frac{\|\tilde{h}_j\|^2}{\left(1+p\|\tilde{h}_i\|^2\right)^2}\right] \\ &\quad -\frac{1}{p}E\left[\frac{\|\tilde{h}_i\|^2}{\left(1+p\|\tilde{h}_i\|^2\right)^2}\right]. \end{aligned} \tag{2.54}$$

For the second term of (2.54), we apply the inequality $E(1/x^2) > [E(1/x)]^2$ to obtain

$$\begin{aligned}E\left[\frac{\|\tilde{h}_j\|^2}{\left(1+p\|\tilde{h}_i\|^2\right)^2}\right] &= E\left[\frac{1}{\left(1+p\|\tilde{h}_i\|^2\right)^2}\right]E\left[\|\tilde{h}_j\|^2\right] \\ &> \left(E\left[\frac{1}{1+p\|\tilde{h}_i\|^2}\right]\right)^2 E\left[\|\tilde{h}_j\|^2\right] \\ &= E\left[\frac{1}{1+p\|\tilde{h}_i\|^2}\right]E\left[\frac{\|\tilde{h}_j\|^2}{1+p\|\tilde{h}_i\|^2}\right]. \end{aligned} \tag{2.55}$$

Similarly, for the third term of (2.54), we can show that

$$\begin{aligned}E\left[\frac{\|\tilde{h}_i\|^2}{\left(1+p\|\tilde{h}_i\|^2\right)^2}\right] &= \frac{1}{p}E\left[\frac{1}{1+p\|\tilde{h}_i\|^2} - \frac{1}{\left(1+p\|\tilde{h}_i\|^2\right)^2}\right] \\ &< \frac{1}{p}E\left[\frac{1}{1+p\|\tilde{h}_i\|^2}\right] - E\left[\frac{1}{\left(1+p\|\tilde{h}_i\|^2\right)}\right]^2 \\ &= E\left[\frac{1}{1+p\|\tilde{h}_i\|^2}\right]E\left[\frac{\|\tilde{h}_i\|^2}{1+p\|\tilde{h}_i\|^2}\right]. \end{aligned} \tag{2.56}$$

We then combine (2.54)-(2.56) to obtain (2.52).

Since $\left(E\left[\frac{1}{1+p\|\tilde{h}_i\|^2}\right]-1\right) < 0$, it follows from (2.52) that if $f_j(p) \leq 0$, then we must have $f_j'(p) > 0$. There are two cases:

Case 1: Assume that $f_j(0) < 0$. In this case, if we have $f_j(p) < 0$ for all $p > 0$, then we let $p_{j,s} = \infty$. Otherwise, since $f_j(p)$ is a continuous function, there must exists a zero point $p_0 \in (0, \infty)$ such that $f_j(p_0) = 0$. We claim that p_0 must be the only zero point. Otherwise, let us assume that there is another zero point p_1. Without loss of generality, assume that $p_0 < p_1$. Since $f_j(p)$ is a continuous function, the minimum of $f_j(p)$ over the interval $[p_0, p_1]$ must be negative, because at the right

end point p_1, we have $f_j(p_1) = 0$ and $f_j'(p_1) > 0$. Assume that the minimum is achieved at some interior point $p_m \in (p_0, p_1)$. We must have $f_j'(p_m) = 0$. On the other hand, since $f_j(p_m) < 0$, it follows from (2.52) that $f_j'(p_m) > 0$. Therefore, we reach a contradiction.

Combining that fact that f_j is continuous, $f_j(0) < 0$, p_0 is the only zero point of f_j, and $f_j'(p_0) > 0$, then we have $f_j(p) < 0$ if and only if $p < p_0$. Finally, we let $p_{j,s} = p_0$.

Case 2: Assume that $f_j(0) \geq 0$. In this case, suppose that there exists some $p_0 \in (0, \infty)$ such that $f_j(p_0) \leq 0$. It follows from (2.52) that $f_j'(p_0) > 0$. Similar to the proof of case 1, we consider the minimum of f_j over the interval of $[0, p_0]$ to obtain a contradiction. Hence, when $f_j(0) \geq 0$, we have $f_j(p) > 0$ for all $p > 0$.

The above analysis implies that there exists a unique zero point p_0 such that $f_j(p_0) = 0$. Furthermore, we can show that:

- If $f_j(0) < 0$, then there exists a threshold $p_{j,s} \in (0, \infty]$ such that $f_j(p) < 0$ if and only if $p < p_{j,s}$.
- If $f_j(0) \geq 0$, then we have $f_j(p) > 0$ for all $p > 0$.

Hence, in order for a virtual angle i to be the capacity-achieving beamforming angle, we must have $f_j(0) < 0$, for all $j = 1, \cdots, n_t$ and $j \neq i$. Since $f_j(0) = E(\|\tilde{h}_j\|^2 - \|\tilde{h}_i\|^2)$, angle i must have the largest sum-variance. Theorem 2 then follows by letting $p_s = \min_{j \neq i} p_{j,s}$.

2.8.3 Proof of Theorem 3

We examine condition (2.25) for the virtual angle j that satisfies (2.27). It is sufficient to prove that $f_j^k(p^k) > 0$, thus the beamforming condition is violated, for any j such that $V_{m,i}^k \leq V_{m,j}^k, m = 1, \cdots, n_r$. Here, the user k's beamforming angle is $i_k = i$. To prove this, we apply the matrix inversion lemma which says that if $A = F + EBE^\dagger$, then we have $A^{-1} = F^{-1} - F^{-1}E(B^{-1} + E^\dagger F^{-1} E)^{-1} E^\dagger F^{-1}$. Here we let $A = I_{n_r} + \sum_{l=1}^K p^l \tilde{h}_{i_l}^l \tilde{h}_{i_l}^{l\dagger}$, $F = I_{n_r} + \sum_{l \neq k}^K p^l \tilde{h}_{i_l}^l \tilde{h}_{i_l}^{l\dagger}$, $E = \tilde{h}_i^k$, and $B = p^k$. Letting $D = F^{-1}$, we obtain

$$\begin{aligned} A^{-1} &= D - D\tilde{h}_i^k\big[(p^k)^{-1} + \tilde{h}_i^{k\dagger} D\tilde{h}_i^k\big]^{-1}\tilde{h}_i^{k\dagger} D \\ &= D - p^k \frac{D\tilde{h}_i^k\tilde{h}_i^{k\dagger} D}{1 + p^k\tilde{h}_i^{k\dagger} D\tilde{h}_i^k}. \end{aligned} \tag{2.57}$$

It follows that

$$\begin{aligned} f_j^k(p^k) &= E[\tilde{h}_j^{k\dagger} A^{-1}\tilde{h}_j^k] - [\tilde{h}_i^{k\dagger} A^{-1}\tilde{h}_i^k] \\ &= E[\tilde{h}_j^{k\dagger} D\tilde{h}_j^k] - p^k E\Big[\frac{\big(\tilde{h}_j^{k\dagger} D\tilde{h}_i^k\big)\big(\tilde{h}_i^{k\dagger} D\tilde{h}_j^k\big)}{1 + p^k\tilde{h}_i^{k\dagger} D\tilde{h}_i^k}\Big] - E[\tilde{h}_i^{k\dagger} D\tilde{h}_i^k] + p^k E\Big[\frac{\big(\tilde{h}_i^{k\dagger} D\tilde{h}_i^k\big)\big(\tilde{h}_i^{k\dagger} D\tilde{h}_i^k\big)}{1 + p^k\tilde{h}_i^{k\dagger} D\tilde{h}_i^k}\Big] \\ &= E\left[\frac{\tilde{h}_j^{k\dagger} D^{-1}\tilde{h}_j^k - \tilde{h}_i^{k\dagger} D^{-1}\tilde{h}_i^k}{1 + p^k\tilde{h}_i^{k\dagger} D^{-1}\tilde{h}_i^k}\right] + E\left[p^k \frac{\tilde{h}_j^{k\dagger} D^{-1}\tilde{h}_j^k\tilde{h}_i^{k\dagger} D^{-1}\tilde{h}_i^k - \tilde{h}_j^{k\dagger} D^{-1}\tilde{h}_i^k\tilde{h}_i^{k\dagger} D^{-1}\tilde{h}_j^k}{1 + p^k\tilde{h}_i^{k\dagger} D^{-1}\tilde{h}_i^k}\right]. \end{aligned} \tag{2.58}$$

Due to the Cauchy-Schwartz inequality, the second expectation of (2.58) is always positive. Hence, in order to show that $f_j^k(p^k) > 0$, it suffices to show that conditioned upon any fixed D, the first expectation of (2.58) is positive. Let $c_j = \tilde{h}_j^{k\dagger} D^{-1}\tilde{h}_j^k$ and $c_i = \tilde{h}_i^{k\dagger} D^{-1}\tilde{h}_i^k$. For a fixed D, the two random variables c_j and c_i are independent. Therefore,

$$\begin{aligned} f_j^k(p^k) > E\Big[\frac{c_j - c_i}{1 + p^k c_i}\Big] &= \frac{1}{p^k}\left(E\Big[\frac{1 + p^k c_j}{1 + p^k c_i}\Big] - 1\right) \\ &= \frac{1}{p^k}\left(E[1 + p^k c_j]\cdot E\Big[\frac{1}{1 + p^k c_i}\Big] - 1\right) \\ &> \frac{1}{p^k}\left(\frac{E[1 + p^k c_j]}{E[1 + p^k c_i]} - 1\right) \\ &= \frac{1}{p^k}\left(\frac{1 + p^k E[c_j]}{1 + p^k E[c_i]} - 1\right). \end{aligned} \tag{2.59}$$

Hence, it remains to show that $E[c_j] > E[c_i]$. First, we have

$$\begin{aligned} E[c_j] - E[c_i] &= E[\tilde{h}_j^{k\dagger} D^{-1}\tilde{h}_j^k] - E[\tilde{h}_i^{k\dagger} D^{-1}\tilde{h}_i^k] \\ &= \mathrm{Tr}\left[D^{-1}\cdot E\big(\tilde{h}_j^k\tilde{h}_j^{k\dagger} - \tilde{h}_i^k\tilde{h}_i^{k\dagger}\big)\right]. \end{aligned} \tag{2.60}$$

It follows from (2.27) that $E\big(\tilde{h}_j^k\tilde{h}_j^{k\dagger} - \tilde{h}_i^k\tilde{h}_i^{k\dagger}\big)$ is a diagonal matrix with nonnegative entries. This, combined with the fact that D^{-1} is a positive definite Hermitian matrix with positive diagonal entries, implies that $E[c_j] > E[c_i]$. This proves that $f_j^k(p^k) > 0$. Thus, virtual angle i cannot be the c.b.a. of user k.

2.8.4 Proof of Proposition 1

First, we write

$$\begin{aligned} I(\boldsymbol{\lambda}, K) - \bar{I}(\boldsymbol{\lambda}, K) &= E \log \frac{\det A_K}{\det E(A_K)} \\ &= E \log \frac{\det A_K}{\prod_{m=1}^{n_r} \left[1 + \sum_{k=1}^{K} \sum_{j=1}^{n_t} \frac{p^k}{n_t} \lambda_j^k V_{m,j}^k \right]} \\ &= E \log \det \tilde{A}_K, \end{aligned} \tag{2.61}$$

where the (m,i)-th element of $\tilde{A}_K$, denoted by $(\tilde{A}_K)_{m,i}$, is given by

$$\begin{aligned} (\tilde{A}_K)_{m,i} &= \frac{\left(1 + \sum_{k=1}^{K} \sum_{j=1}^{n_t} \frac{p^k}{n_t} \lambda_j^k \tilde{h}_{m,j}^k \tilde{h}_{i,j}^{k\dagger} \right)}{\left(1 + \sum_{k=1}^{K} \sum_{j=1}^{n_t} \frac{p^k}{n_t} \lambda_j^k V_{m,j}^k \right)} \\ &= \frac{\left(1 + \sum_{k=1}^{K} \sum_{j=1}^{n_t} \frac{p^k}{n_t} \lambda_j^k \tilde{h}_{m,j}^k \tilde{h}_{i,j}^{k\dagger} \right) / (1 + K n_t)}{\left(1 + \sum_{k=1}^{K} \sum_{j=1}^{n_t} \frac{p^k}{n_t} \lambda_j^k V_{m,j}^k \right) / (1 + K n_t)}. \end{aligned} \tag{2.62}$$

We let $\bar{S}_K$ denote the numerator in the second fraction of (2.62), which equals the average of the summation of independent random variables. It follows from the Strong Law of Large Number (SLLN) and (2.29) that

$$\begin{aligned} 0 &= \lim_{K\to\infty} |\bar{S}_K - E(\bar{S}_K)| \\ &= \limsup_{K\to\infty} \left(E(\bar{S}_K) \left| \frac{\bar{S}_K}{E(\bar{S}_K)} - 1 \right| \right) \\ &\geq \left(\liminf_{K\to\infty} E(\bar{S}_K) \right) \left(\limsup_{K\to\infty} \left| \frac{\bar{S}_K}{E(\bar{S}_K)} - 1 \right| \right) \\ &\geq c \cdot \limsup_{K\to\infty} \left| \frac{\bar{S}_K}{E(\bar{S}_K)} - 1 \right|. \end{aligned}$$

Thus, for $m = i$ we obtain $\lim_{K\to\infty} (\tilde{A}_K)_{i,i} = \lim_{K\to\infty} \bar{S}_K / E(\bar{S}_K) = 1$. For $m \neq i$, since $E(\bar{S}_K) = 0$, it follows from (2.29) and the SLLN that $\lim_{K\to\infty} (\tilde{A}_K)_{m,i} \leq \frac{1}{c} \lim_{K\to\infty} \bar{S}_K = \frac{1}{c} \lim_{K\to\infty} \big(\bar{S}_K - E(\bar{S}_K) \big) = 0$. This proves that $\tilde{A}_K$ converges to the identity matrix and thus $E \lim_{K\to\infty} \big[\log \det \tilde{A}_K \big] = 0$. Assume that the distributions of the random vectors $\{\tilde{h}_j^k\}$ are sufficiently smooth to facilitate exchange of the limit and the expectation operator, we obtain $\lim_{K\to\infty} \big[I(\boldsymbol{\lambda}, K) - \bar{I}(\boldsymbol{\lambda}, K) \big] = \lim_{K\to\infty} \big[E \log \det \tilde{A}_K \big] = E \big[\lim_{K\to\infty} \log \det \tilde{A}_K \big] = 0$.

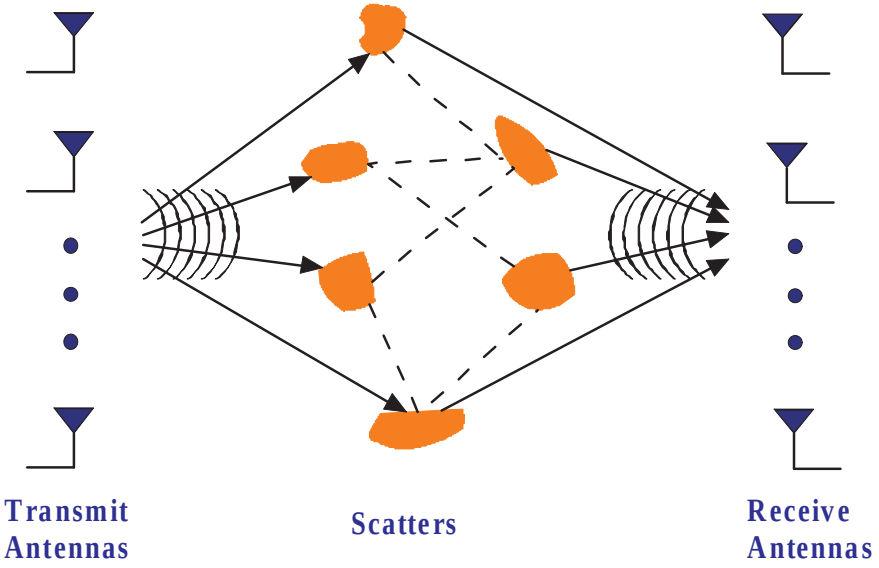


Figure 2.1. The virtual representation model

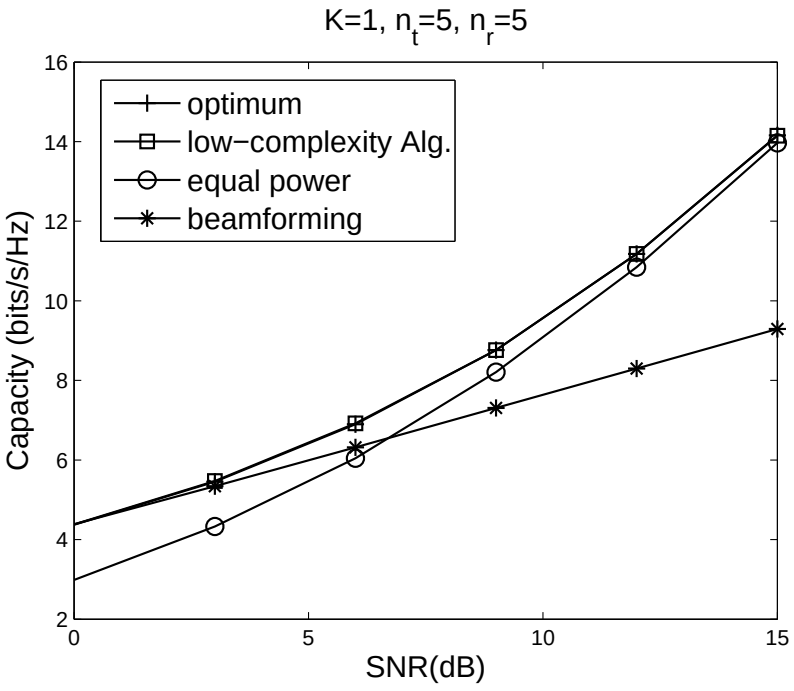


Figure 2.2. Comparisons of capacities achieved by different schemes in the single-user case with $n_t = n_r = 5$

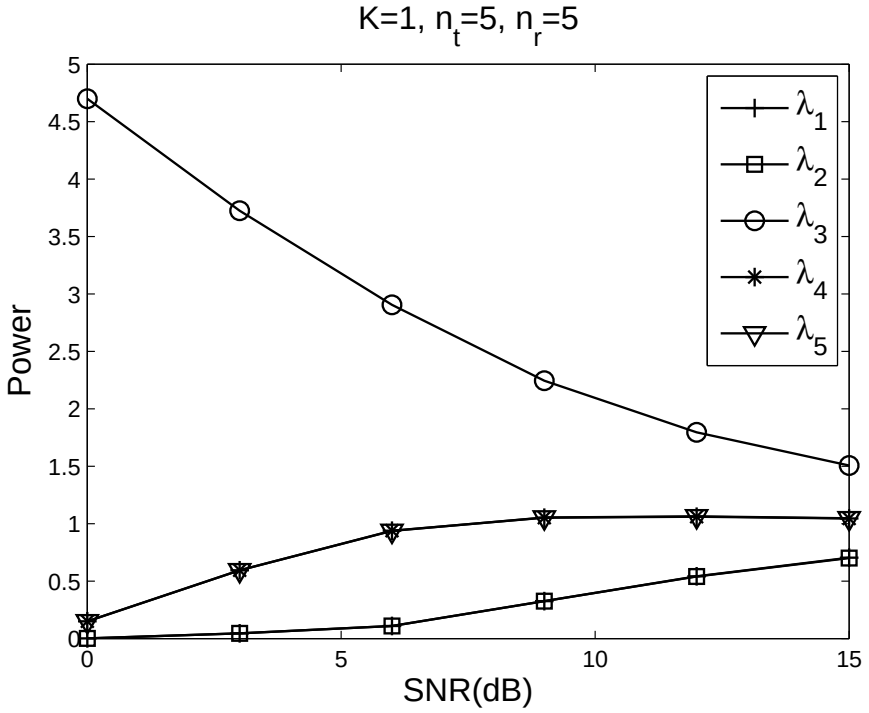


Figure 2.3. The optimal power allocation in the single-user case with $n_t = n_r = 5$

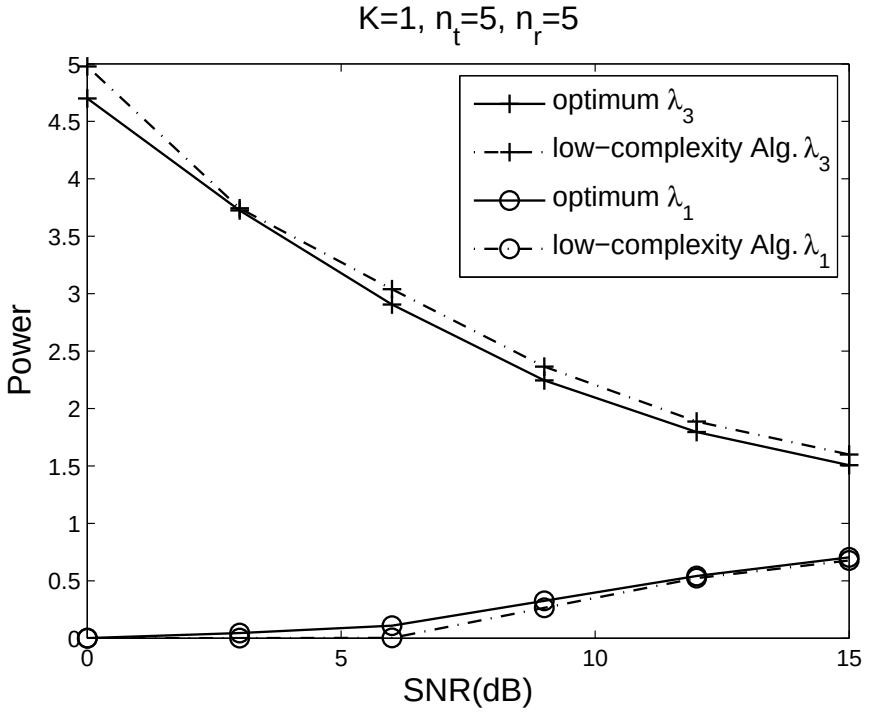


Figure 2.4. Comparison between the optimum and the low-complexity algorithm in the single-user case with $n_t = n_r = 5$

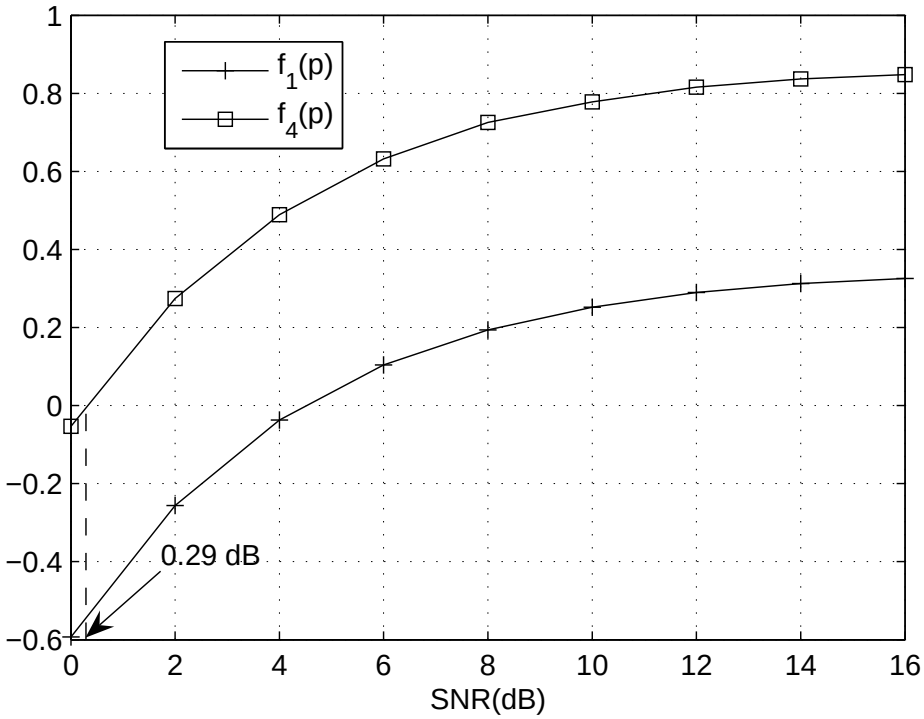


Figure 2.5. The beamforming condition curves, SNR threshold is at 0.29 dB

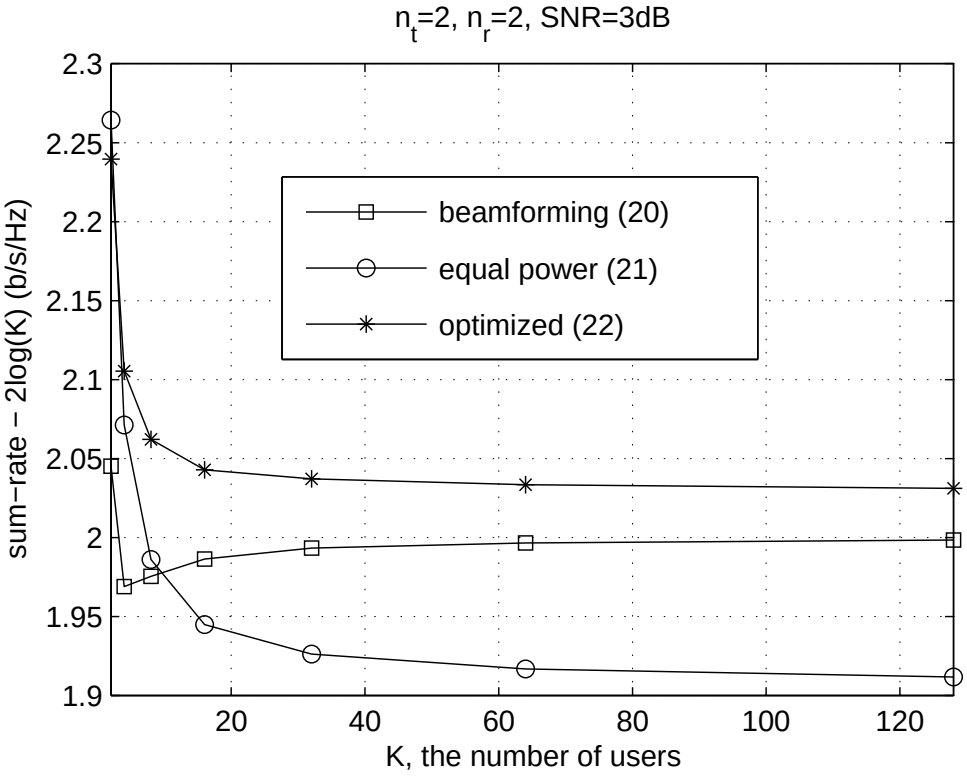


Figure 2.6. Numerical verification of Example 3.

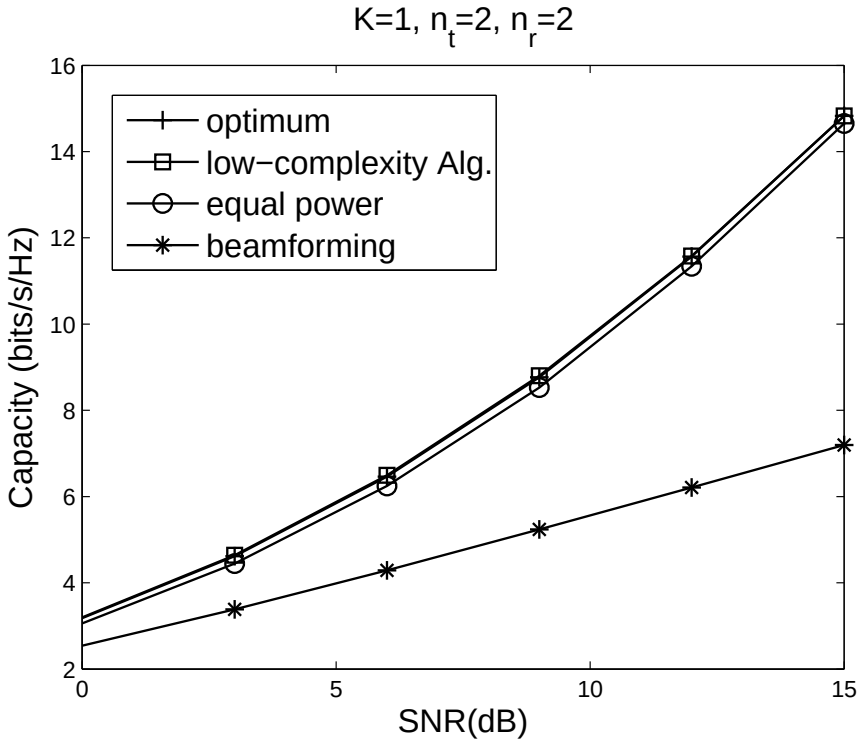


Figure 2.7. Comparisons of capacities achieved by different schemes in the single-user case with $n_t = n_r = 5$. The variance matrix is the transpose of V

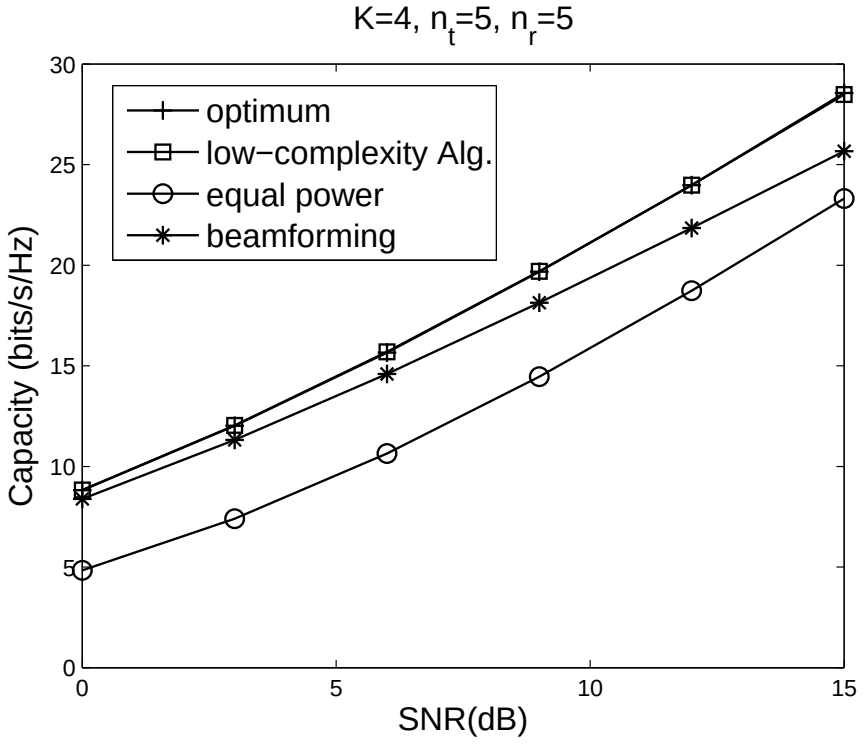


Figure 2.8. Comparisons of capacities achieved by different schemes in the four-user case with $n_t = n_r = 5$

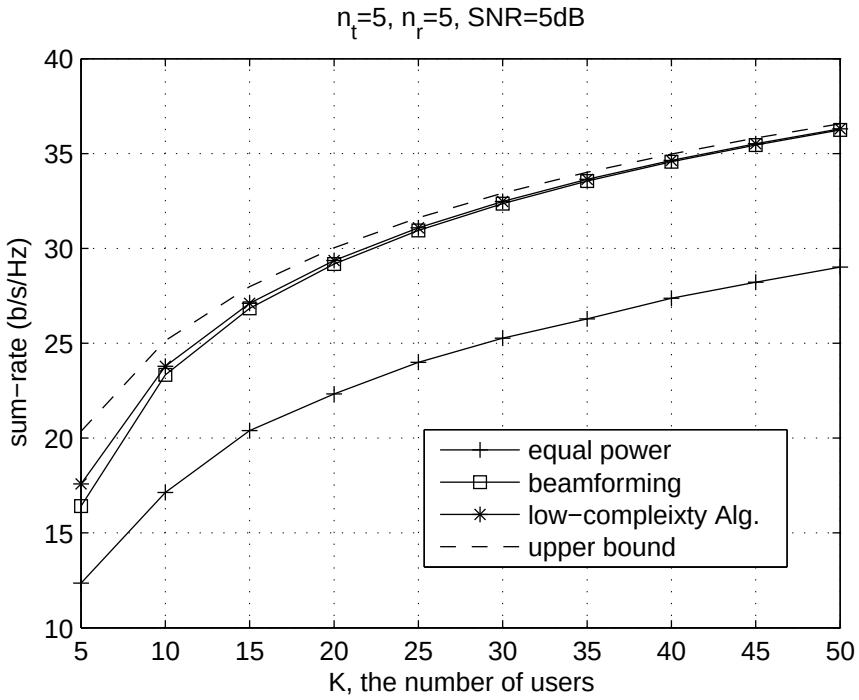


Figure 2.9. Optimality of beamforming for large systems

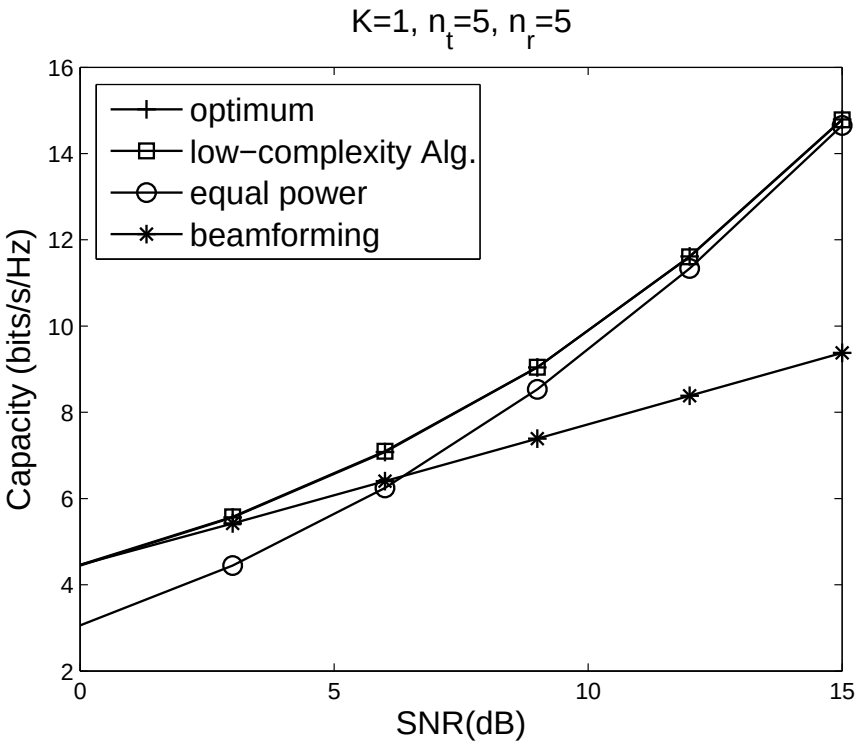


Figure 2.10. Comparisons of capacities achieved by different schemes in the single-user case with $n_t = n_r = 5$, all channel elements follow the Uniform distribution [1]

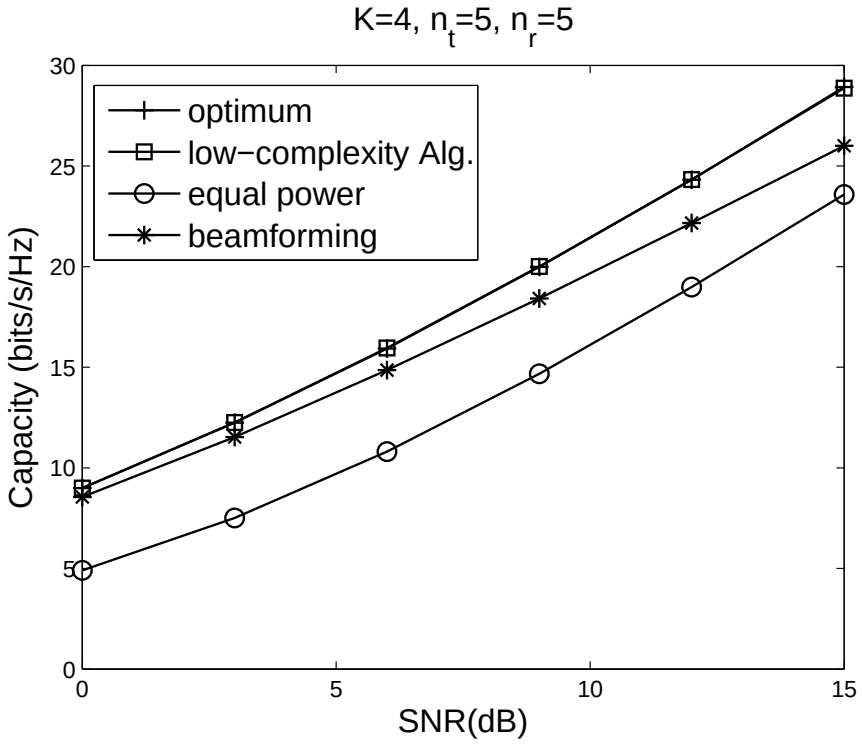


Figure 2.11. Comparisons of capacities achieved by different schemes in the four-user case with $n_t = n_r = 5$, all channel elements follow the Uniform distribution [1]

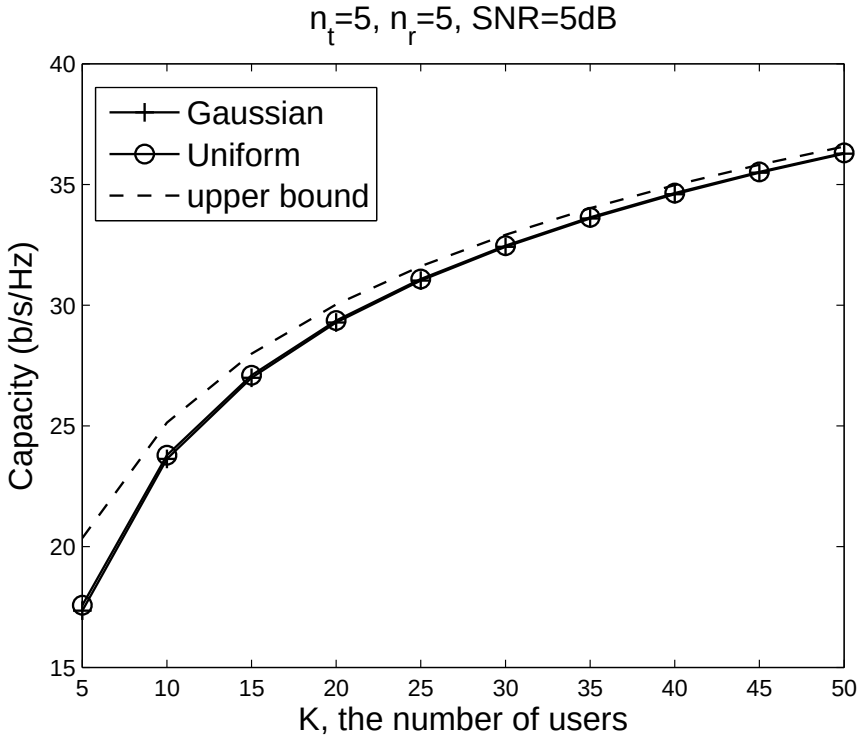


Figure 2.12. Comparison of capacity between the Gaussian distribution and the Uniform distribution for large systems

CHAPTER 3

BEAMFORMING CODEBOOK DESIGN FOR TEMPORALLY CORRELATED FADING CHANNELS

Transmit beamforming provides an efficient mechanism to exploit the diversity gain of the multiple-antenna channel by transmitting signals along the direction of the channel vector. However, this mechanism requires full knowledge of channel state information (CSI) at the transmitter, which is impractical in most wireless systems. Recently, beamforming codebook design based on limited feedback from the receiver has received much interest. The main idea is to construct a predetermined, finite set of beamforming codebook that is known to both the transmitter and the receiver. Then the receiver selects a codeword based on the instantaneous channel vector and feeds back the codeword index to the transmitter for transmission.

Previous work on beamforming codebook designs [17–19] mostly consider identical and independently distributed (i.i.d.) Rayleigh fading channel, or spatially correlated channel. Even though some algorithms [19, 41] are applicable to arbitrary channel distribution, the complexity is high to update the codebook while the channel distribution changes. Meanwhile, temporal correlation channel model is not taken into account in these work. Representative work on developing limited feedback techniques to exploit channel correlation are as follows. In [42] the quantized channel state information (CSI) is fed back to the transmitter for the design of optimal beamforming vector. In [43], a practical feedback scheme based on adaptive delta modulation is employed to track time-varying channels and to maximize capacity. In [44,45], subspace tracking techniques are investigated to exploit temporal or spatial correlation of the channel. A adaptive codebook selection scheme that switches between multiple codebooks to achieve higher performance gain is studied in [46].

Reference [47] proposed a channel adaptive feedback strategy, where the codebook can be derived from one mother codebook by rotation and scaling, to adapt the codebook for arbitrary channel distributions. However, both [46] and [47] require feedback overheads before the transmission of each block, and it costs extra complexity to make selection of codebook from a codeset of large size. A variant rate feedback scheme [48] is proposed to reduce the average feedback rate based on rate distortion theory. All these work above indicate that the limited feedback scheme can benefit from taking temporal correlation into account. This motivates us to develop low-complexity and efficient algorithms for beamforming codebook design under temporally correlated channels based on limited feedback [49].

Our main contributions can be summarized as follows. (1). we propose a low-complexity rotation-based codebook design algorithm exploiting channel temporal correlation without increasing the feedback rate. A root codebook is first designed according to various optimization criteria. Subsequently, the beamforming codebook at each time is obtained by simply rotating the root codebook by an unitary matrix that is determined by the selected codeword from previous time instance. This results in a low-complexity implementation of the adaptive beamforming design. (2). We consider three root codebook designs based on the Lloyd algorithm, aiming to either maximize average received SNR, maximize the capacity, or to minimize BER. While the first two designs follow from similar work in the literature, the BER design is proposed in this paper to optimize the BER performance of a beamforming codebook for a limited feedback system. The three root codebook designs, in conjunction with the proposed rotation-based scheme, are denoted by SNR-R, BER-R, and CAP-R, respectively. (3) We examine both uncoded and channel coded performance of these designs for a first-order autoregressive (AR) model with a low feedback rate of three bits per channel use. For uncoded systems, we show that both SNR-R and CAP-R encounter a "high SNR problem" in that they do not give good BER performance in the high SNR region. The proposed BER-R design is shown to give superior performance in both uncoded and coded systems.

This chapter is organized as follows: In section 3.1 we introduce the channel correlation model and describe the basic system setup for limited feedback. The

proposed rotation-based codebook design is introduced in Section 3.2. In Section 3.3, we consider several root codebook designs based on different optimization criteria. Section 3.4 includes numerical results for both coded and uncoded systems employing various codebooks.

3.1 System Model

We consider a multiple-input single-output (MISO) channel with N_t transmit antenna and a single receive antenna. Assume that the channel is temporally correlated. We model the temporal correlation by a first order autoregressive (AR) process. The channel vector $H^{(n+1)} \in \mathcal{C}^{N_t \times 1}$ at time $n+1$ can be expressed as

$$H^{(n+1)} = aH^{(n)} + gv^{(n)} \tag{3.1}$$

where $v^{(n)}$ is an i.i.d. complex Gaussian vectors that is independent from $H^{(n)}$, and the scalar a denotes the first-order autocorrelation coefficient for each channel gain, i.e.,

$$a = E\big[H^{(n+1)^\dagger} H^{(n)}\big], \tag{3.2}$$

and $(\cdot)^\dagger$ denotes the Hermitian operator. The AR model (3.1) serves as a good approximation of a mobile wireless channel when we let

$$a = J_0(2\pi f_d T), \quad g = \sqrt{1 - |a|^2}, \tag{3.3}$$

where $J_0(\cdot)$ is the zeroth-order Bessel function of the first kind, T is the sampling period, and f_d is the channel maximum doppler frequency.

When transmit beamforming is employed, the transmitted signal at time n can be expressed as

$$X^{(n)} = f^{(n)^\dagger} s^{(n)}, \tag{3.4}$$

where $s^{(n)}$ is a scalar complex symbol for time n that satisfies the average power constraint $E[|s^{(n)}|^2] = \rho$, $f^{(n)} \in \mathcal{C}^{N_t \times 1}$ is the beamforming vector such that $f^{(n)} \in \Omega(N_t, 1)$, where $\Omega(N_t, 1)$ is the set of N_t-dimensional vectors with unit norm. The received signal at time n can be expressed as

$$y^{(n)} = f^{(n)^\dagger} H^{(n)} s^{(n)} + w^{(n)} \tag{3.5}$$

where $w^{(n)}$ is the complex additive white noise with unit variance.

We assume that the receiver has perfect CSI, i.e., it knows the exact $H^{(n)}$, whereas the transmitter only has partial CSI through a low-rate feedback link from the receiver. Here we assume that the feedback link is error-free and delay-free. The problem of beamforming codebook design can be formulated as follows: At time n, assume that a codebook $\mathcal{F}^{(n)} = \{f_1^{(n)}, \ldots, f_{N_b}^{(n)}\}$, where each $f_i^{(n)} \in \Omega(N_t, 1)$, is available at both the transmitter and the receiver. Using the instantaneous channel vector $H^{(n)}$, the receiver chooses the best codeword $\tilde{f}^{(n)} \in \mathcal{F}^{(n)}$ such that

$$\tilde{f}^{(n)} = \arg \max_{f_i^{(n)} \in \mathcal{F}^{(n)}} \left| H^{(n)\dagger} f_i^{(n)} \right|^2. \tag{3.6}$$

The index of the codeword is then fed back to the transmitter, and the transmitter chooses $\tilde{f}^{(n)}$ as the beamforming vector for transmission. We say that a codebook $\mathcal{F}_o^{(n)}$ is optimal if it achieves the maximum average received SNR, i.e.,

$$\mathcal{F}_o^{(n)} = \arg \max_{\mathcal{F}^{(n)}} E\left[\left| H^{(n)\dagger} \tilde{f}^{(n)} \right|^2 \right] \tag{3.7}$$

It is shown in [17] that (3.7) is equivalent to the minimization of average distortion:

$$\begin{aligned} \mathcal{F}_o^{(n)} &= \arg \min_{\mathcal{F}^{(n)}} E\left[1 - \left| \bar{H}^{(n)\dagger} \tilde{f}^{(n)} \right|^2 \right] \\ &= \arg \min_{\mathcal{F}^{(n)}} E\left[d^2(\bar{H}^{(n)\dagger}, \tilde{f}^{(n)}) \right] \end{aligned} \tag{3.8}$$

where $\bar{H}^{(n)} = H^{(n)}/|H^{(n)}|$ denotes the channel direction at time n, $d(w_1, w_2) = \sqrt{1 - |w_1^\dagger w_2|^2}$ denotes the choral distance of two vectors, which equals the sine of the angle between these vectors.

For the memoryless, i.i.d. Rayleigh fading channel, the channel direction is uniformly distributed over the unit sphere $\Omega(N_t, 1)$. Hence, from (3.8), the beamforming codebook design can be viewed a sphere vector quantization (SVQ) problem. Since the channel distribution does not change over time for an i.i.d. channel, a single codebook is always optimal. Therefore, the codebook design in [17] [19] can be performed offline.

3.2 Codebook Design of Temporally Correlated Channel

In this section we describe the proposed beamforming codebook design that explicitly exploits the temporal correlation according to the channel model in (3.1).

For a temporally correlated channel, given $\bar{H}^{(n)}$, $\bar{H}^{(n+1)}$is no longer uniformly distributed over $\Omega(N_t, 1)$, but has a higher probability to be within the neighboring region of $\bar{H}^{(n)}$. Thus, as opposed to the case of i.i.d. channels, the optimal beamforming codebook of a temporally correlated channel is no longer uniformly distributed. Instead, the codewords will be more dense in the regions where the channel directions are more likely to point to. We note that similar observations have been made for spatially correlated channels [50, 51].

In the following we first present a possible approach of codebook design that exploits the temporal correlation according to the channel model in (3.1), then we further propose a equivalent approach with lower complexity.

As we know, the beamforming codebook design can be reviewed as a vector quantization problem, where we quantize the channel vector with the best-match codeword within the codebook. Therefore, for any time instance n, we could approximate the channel vector with the current selected codeword, that is, $\bar{H}^{(n)} \approx \tilde{f}^{(n)}$, where $\tilde{f}^{(n)}$ was defined in (3.6). This approximation is asymptotically accurate when N_b goes to infinity. Accordingly, we could design $\mathcal{F}^{(n+1)}$ conditioned upon the knowledge of $\bar{H}^{(n)}$. That is,

$$\mathcal{F}_o^{(n+1)} = \arg\max_{\mathcal{F}^{n+1}} E\Big[\big|H^{(n+1)}\tilde{f}^{(n+1)^\dagger}\big|^2\Big|\bar{H}^{(n)} = \tilde{f}^{(n)}\Big] \tag{3.9}$$

assuming $\bar{H}^{(n)} \approx \tilde{f}^{(n)}$, so that the channel distribution at time $n+1$ is considered to be

$$H^{(n+1)} = \alpha\beta\tilde{f}^{(n)} + gv^{(n)} \tag{3.10}$$

where β is the amplitude of $H^{(n)}$ which is independent of $\bar{H}^{(n)}$. Accordingly, we can rewrite (3.9) into

$$\mathcal{F}_o^{(n+1)} = \arg\max_{\mathcal{F}} E\Big[\big|H\tilde{f}^{\dagger}\big|^2\Big|H = \alpha\beta\tilde{f}^{(n)} + gv\Big] \tag{3.11}$$

Thus, a certain algorithm can be applied for each time instance n to design a good codebook according to (3.11), which, however, costs much high complexity. Meanwhile, it is hard to synchronize the transmitter and the receiver with the same updated codebook at each time instance. Motivated by this, we propose a alternative codebook design which has a much lower complexity.

First of all, using the algorithms proposed previously, we construct a root codebook $\mathcal{F}^r = \arg\max_{\mathcal{F}} E[|H^{(n+1)}\tilde{f}^{\dagger}|^2 \big| \bar{H}^{(n)} = f^r]$ given a randomized $f^r \in \Omega(Nt, 1)$. The key step of our algorithm is to compute $\mathcal{F}^{(n+1)}$ from $\mathcal{F}^r$ through a rotation $U^{(n+1)}$:

$$\mathcal{F}^{(n+1)} = U^{(n+1)}\mathcal{F}^r = \{U^{(n+1)}f_1^r, U^{(n+1)}f_2^r, \cdots, U^{(n+1)}f_{N_b}^r\}, \tag{3.12}$$

where $U^{(n+1)}$ is any unitary matrix such that $\tilde{f}^{(n)} = U^{(n+1)}f^r$. Note that the selected codewords $\tilde{f}^{(n)} \in \mathcal{F}^{(n)}$ is utilized in our design because they are available to the transmitter at time n due to the limited feedback from the receiver. The rotation step (3.12) is based on Theorem 4 below.

Theorem 4 *Under the idealized assumption that the channel direction at time $n-1$: $\bar{H}^{(n-1)} = f^r$ is known to the transmitter at time n. Let $\mathcal{F}_o(f^r) = \{f_1^r, \ldots, f_{N_b}^r\}$ denote an optimal codebook for time n that achieves the maximum average received SNR conditioned upon f^r:*

$$E\Big[\big|H^{(n)}\tilde{f}^{(n)\dagger}\big|^2 \Big| \bar{H}^{(n-1)} = f^r\Big] \tag{3.13}$$

Then if $f' \in \Omega(N_t, 1)$, and $f' = Uf$, where U is a unitary matrix, then we have

$$\mathcal{F}_o(f') = U\mathcal{F}_o(f^r) = \{Uf_1^r, \ldots, Uf_{N_b}^r\} \tag{3.14}$$

must be an optimal codebook conditioned upon $\bar{H}^{(n-1)} = f'$.

Proof.

For any codebook $\mathcal{F}^{(n)}$, the average received SNR conditioned upon $\bar{H}^{(n-1)} = f^r$ equals

$$\begin{aligned} &E\Big[\big|H^{(n)\dagger}\tilde{f}^{(n)}\big|^2 \big| \bar{H}^{(n-1)} = f^r\Big] \\ &= E\Big[\big|\big(UH^{(n)}\big)^{\dagger}\big(U\tilde{f}^{(n)}\big)\big|^2 \big| \bar{H}^{(n-1)} = f^r\Big], \end{aligned} \tag{3.15}$$

where $\tilde{f}^{(n)} = \arg \max_{f_i^{(n)} \in \mathcal{F}^{(n)}} \big|H^{(n)\dagger} f_i^{(n)}\big|^2$.

Conditioned upon $\bar{H}^{(n-1)} = f^r$, we have

$$\begin{aligned} H^{(n)} &= a|H^{(n-1)}|\bar{H}^{(n-1)} + gv^{(n-1)} \\ &= a|H^{(n-1)}|f^r + gv^{(n-1)} \\ &= U^{\dagger}(a|H^{(n-1)}|f' + gz^{(n-1)}), \end{aligned} \tag{3.16}$$

where $z^{(n-1)} = Uv^{(n-1)}$ has the same distribution as $v^{(n-1)}$. It follows that $UH^{(n)} = a|H^{(n-1)}|f' + gz^{(n-1)}$, and thus, the following two conditional distributions are the same:

$$p(UH^{(n)}\big|\bar{H}^{(n-1)} = f^r) \ \sim \ p(H^{(n)}\big|\bar{H}^{(n-1)} = f'). \tag{3.17}$$

Using (3.17), we can rewrite (3.15) as

$$E\Big[|H^{(n)^{\dagger}}\tilde{f}^{(n)}|^2\big|\bar{H}^{(n-1)} = f'\Big], \tag{3.18}$$

where $\tilde{f}^{(n)} = \arg\max_{f_i^n \in U\mathcal{F}^{(n)}} \big|H^{(n)^{\dagger}} f_i^{(n)}\big|^2$. This follows that if $\mathcal{F}_o(f^r)$ is an optimal codebook conditioned upon $\bar{H}^{(n-1)} = f^r$, then the rotated codebook $U\mathcal{F}_o^{(n)}$ must be an optimal codebook conditioned upon $\bar{H}^{(n-1)} = f'$. ■

An immediate consequence of Theorem 4 is as follows: if $\bar{H}^{(n)}$ is available to the transmitter at time $n+1$ and time n, then the optimal codebook $\mathcal{F}_o(\bar{H}^{(n)})$ can be constructed from $\mathcal{F}_o(f^r)$ by a unitary rotation $U^{(n)}$ where $\bar{H}^{(n)} = U^{(n)}f^r$. Now, since only the selected codewords $\tilde{f}^{(n)}$ is available to the transmitter, we approximate the rotation matrix $U^{(n)}$ by $\tilde{f}^{(n)} = U^{(n)}f^r$, assuming that $\tilde{f}^{(n)}$ provide a good approximation of the actual channel directions $\bar{H}^{(n)}$.

In Table 3.1, we summarize the main steps of our codebook design. Our algorithm is initialized by applying GLA or other algorithms to obtain a root codebook $\mathcal{F}_o(f^r)$ from an arbitrary root channel direction f^r. Then we apply proper rotations at each time instance to obtain the appropriate codebook used for the next time slot.

3.3 Root Codebook Design

3.3.1 SNR Based Root Codebook Design

The Lloyd algorithm [19, 52] is a popular numerical method in limited feedback codebook design, it can be applied to arbitrary channel distributions. Here we can use

the generalized Lloyd algorithm (GLA) proposed in [19] to derive our root codebook. The general steps of GLA codebook design are stated as follows:

(1) Nearest neighbor rule: For given codewords, the optimum partition cells satisfy

$$\mathcal{R}_i = \left\{ \left| H^\dagger f_i \right|^2 \geq \left| H^\dagger f_j \right|^2, \forall j \neq i \right\} \tag{3.19}$$

for $i = 1, \cdots, N_b$.

(2) Centroid condition: For a given partition , the optimum codeword vectors satisfy

$$\hat{f}_i = \arg \max_{\substack{f_i \in \mathcal{R}_i \\ \|f_i\|=1}} E\left[\left| H^\dagger f_i \right|^2 | H \in \mathcal{R}_i \right] \tag{3.20}$$

It is easy to find out that $\hat{f}_i$ should be the eigenvector of the dominant eigenvalue of $E\left[HH^\dagger | H \in \mathcal{R}_i\right]$

In practice, a codebook is derived by generating a sufficiently large number of samples (channel realizations) and iterating the above two conditions for sufficient times until it converges. In order to take temporal correlation into account, we implement the GLA as follows. First, we randomly generate a $\bar{H}_r^{(-1)} \in \Omega(N_t, 1)$, which we refer to as the root channel direction. Then we collect L random samples of $\bar{H}_r^{(0)}$ following the AR model $H_r^{(0)} = a|H_r^{(-1)}|\bar{H}_r^{(-1)} + gv^{(-1)}$ and let $\bar{H}_r^{(0)} = H_r^{(0)}/|H_r^{(0)}|$. Using the collection of L samples of $\bar{H}_r^{(0)}$, the optimal beamforming codebook can be computed iteratively following the procedure given in [19].

One straightforward approach is to directly apply this GLA algorithm to design our root codebook, given a randomized f^r. However, we observe a problem of this design. In Figure 3.3, we compare the BER performance of the following codebooks: 'i.i.d. codebook' refers to the GLA designed codebook assuming the channel is i.i.d. distributed; 'SNR-R' codebook denotes the rotated codebook where the root codebook is designed by GLA assuming correlated channel, and the rotation is based on the selected codewords; 'SNR-IR' differs from 'SNR-R' on that the rotation is based on the exact channel direction. We assume $N_t = 4$, $N_b = 8$ and a slow fading channel where $\alpha = 0.9648$ in this figure. As shown in Figure 3.3, SNR-R and SNR-IR performs better than the i.i.d. codebook only at low SNR region. This observation is a little surprising because intuitively the proposed codebook should always has better BER performance since it has a higher average received SNR than the i.i.d. codebook

(Figure 3.2), but actually it is not. A similar problem has been addressed in [53] on the asymptotic performance analysis of the general selection diversity scenarios. Here we will present a particular explanation why this high SNR problem occurs in our scheme.

Denoting the received SNR to be $\gamma = \rho|f^\dagger H|^2 = \rho g$, where $g = |H^\dagger \tilde{f}|^2$ is the channel gain and its distribution is independent of ρ. The bit error rate(assuming BPSK) can be expressed as

$$P_e = \int_0^\infty P_e(\rho, g) f_G(g) dg \tag{3.21}$$

where the conditional error probability $P_e(\rho, g) = Q(\sqrt{2\rho g}) = \int_{\sqrt{2\rho g}}^\infty e^{-\frac{x^2}{2}} dx$. We numerically plot the pdf function $f_{\mathcal{G}}(g)$ of the three codebooks in Figure 3.3.

Proposition 2 *If* $\exists B > 0$, *s.t.* $f_{G,1}(g) \leq f_{G,2}(g)$ *for* $\forall g \leq B$ *and* $\int_0^B f_{G,1}(g)dg < \int_0^B f_{G,2}(g)dg$, *then we have* $\lim_{\rho\to\infty} P_{e,1} < \lim_{\rho\to\infty} P_{e,2}$

proof.

We have

$$\begin{aligned}
\lim_{\rho\to\infty} P_{e,1} &= \lim_{\rho\to\infty} \left(\int_0^B P_e(\rho, g) f_{G,1}(g) dg + \int_B^\infty P_e(\rho, g) f_{G,1}(g) dg\right) \\
&= \lim_{\rho\to\infty} \left(P_e(\rho, t_1 B) f_{G,1}(t_1 B) + \int_B^\infty P_e(\rho, g) f_{G,1}(g) dg\right) \\
&= \lim_{\rho\to\infty} \left(P_e(\rho, t_1 B) f_{G,1}(t_1 B) + O(P_e(\rho, t_1 B))\right)
\end{aligned} \tag{3.22}$$

where we applied the Tylor Theorem to obtain the second line, and $0 < t < 1$. Similarly, we have

$$\lim_{\rho\to\infty} P_{e,2} = \lim_{\rho\to\infty} \left(P_e(\rho, t_2 B) f_{G,2}(t_2 B) + O(P_e(\rho, t_2 B))\right) \tag{3.23}$$

and from the given conditions, we have

$$\begin{aligned}
\int_0^B P_e(\rho, g) f_{G,1}(g) dg &< \int_0^B P_e(\rho, g) f_{G,2}(g) dg \\
P_e(\rho, t_1 B) f_{G,1}(t_1 B) &< P_e(\rho, t_2 B) f_{G,2}(t_2 B)
\end{aligned} \tag{3.24}$$

We complete our proof by taking (3.24) into (3.22) and (3.23). ■

The above proof shows that, P_e is determined by the small value region of g when ρ is large. According to Figure 3.3, $f_{\mathcal{G}}(g)$ of both SNR-R and SNR-IR is greater than that of the i.i.d. codebook below a certain threshold. Therefore, those codebooks would have higher error probability than the i.i.d. codebook when the transmit SNR ρ is high.

From the above analysis, we see that the SNR-R codebook has worse BER performance than the i.i.d. codebook at high SNR region because it has 'a bigger tail' at small value region of g. We propose several approaches to militate this high SNR problem:

(1) Design a codebook with good BER performance. Since the average received SNR is inconsistent with the BER at high SNR region. A straightforward solution would be to directly design a codebook with good BER performance. While it is hard to derive analytical solutions to BER-oriented codebook design, we proposed an algorithm based on optimizing BER in the next subsection, and we will evaluate this new design in our simulation part.

(2) Use modulations with higher constellation size. By increasing the constellation size, we actually draw the transmitted bits to low SNR region. Since with fixed transmit power per symbol, we are increasing the bits within a symbol. Thus, each bit is transmitted at lower power. In this way, we make the high SNR problem occur above a higher threshold, which could be out of the concern in a real communication system.

(3) From Figure 3.3, we see that the inaccurate channel rotation makes the high SNR problem much more severe. We derive the root codebook assuming the channel distribution with given temporal correlation. However, the correlation is actually not that high when we only know the quantized version of channel knowledge. In another word, it would help if we generate the root codebook based on a more accurate channel distribution model. A straightforward method is to employ a smaller correlation coefficient, in order to compensate the quantization error of the channel vector. We tried several coefficients of smaller α and confirmed it really brings improvement. The figures are not posted here for brevity.

(4) Introduce channel coding into the system. As shown in Figure 3.1, $f_{\mathcal{G}}(g)$ of the SNR-R codebook is more flat than that of the i.i.d. codebook, which indicates the proposed codebook has higher temporal diversity. This is the reason why the SNR-R gives worse performance than the i.i.d. codebook for large SNR values. Since at the high SNR region, most errors that occur are due to small values of received SNR (denoted by γ), i.e., when the channel is in outage. As shown in this figure, the probability density function of the received SNR $f(\gamma)$ of SNR-R is greater than that of the i.i.d. codebook when γ is small. This is why even though the SNR-R has an higher average received SNR, it still shows a higher BER (or a higher outage probability) than the i.i.d. codebook. However, by introducing channel coding to exploit the diversity, the system performance would be determined by the mean capacity instead of the outage capacity, in which case SNR-R should outperform the i.i.d. codebook. We will confirm this in our simulation section.

3.3.2 BER Based Codebook Design

As shown above, the criterion of GLA codebook design is to maximize the average received SNR. However, the performance evaluation is usually based on the uncoded bit or symbol error probability for the considered system. According to [41], GLA codebook is suboptimal respect to the BER performance in the uncoded system. Further, the situation is even worse that the GLA codebook shows a high SNR problem under correlated channels, as we will show in our simulation. This promote us to propose new criteria and design codebooks with good BER performance or higher achievable capacity under correlated channels.

First let us reconsider the two steps of the GLA design. With given partitions $R_i, i = 1, \cdots, N_b$ assigned by the nearest neighbor rule, the functionality of the centroid condition is to find the codewords that maximize the average received SNR in each individual partition. That is, for the ith codeword

$$\begin{aligned} \hat{f}_i &= \arg \max_{\substack{f_i \in R_i \\ \|f_i\|=1}} E\left[\left|H^+ f_i\right|^2 \rho | H \in R_i\right] \\ &= \arg \max_{\substack{f_i \in R_i \\ \|f_i\|=1}} E\left[\gamma | H \in R_i\right] \end{aligned} \tag{3.25}$$

where $\gamma = |H^+ f_i|^2 \rho$ denotes the instantaneous received SNR given a specific channel realization $H \in R_i$.

However, our goal is to directly minimize the average BER instead of the average received SNR. Assuming BPSK, we replace the optimization problem in (3.7) by the following criterion

$$\begin{aligned} F_o &= \arg\min_{\mathcal{F}} E\left[P_e(\gamma)\right] \\ &= \arg\min_{\mathcal{F}} E\left[Q(\sqrt{2\gamma})\right] \\ &= \arg\min_{\mathcal{F}} E\left[Q(\sqrt{2\left|H^+\tilde{f}\right|^2\rho})\right] \end{aligned} \tag{3.26}$$

where $P_e(\gamma)$ denotes the conditional BER function given the received SNR γ, and $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$. Accordingly, within the centroid condition, we should find the codeword that minimize the average BER in the given partition. That is,

$$\begin{aligned} \hat{f}_i &= \arg\min_{\substack{f_i \in R_i \\ \|f_i\|=1}} E\left[P_e(\gamma) | H \in R_i\right] \\ &= \arg\min_{\substack{f_i \in R_i \\ \|f_i\|=1}} E\left[Q(\sqrt{2\left|H^+ f_i\right|^2\rho}) | H \in R_i\right] \end{aligned} \tag{3.27}$$

While it is hard to find the optimal solution, we work on some simplifications. Define

$$\begin{aligned} \Delta\gamma &= \gamma^* - \gamma \\ &= \left\|H^\dagger \bar{H}\right\|^2 \rho - \left|H^+ f_i\right|^2 \rho \\ &= \left\|H^\dagger\right\|^2 \rho - \left|H^+ f_i\right|^2 \rho \end{aligned} \tag{3.28}$$

where $\gamma^* = \left\|H^\dagger\right\|^2 \rho$ is the maximal received SNR when we apply the optimal beamforming vector $f = \bar{H}$. Now we expand the Q function to the second order of $\Delta\gamma$

$$\begin{aligned} Q(\sqrt{2\gamma}) &= Q(\sqrt{2(\gamma^* - \Delta\gamma)}) \\ &= Q(\sqrt{2\gamma^*}) + \frac{1}{2\sqrt{\pi\gamma^*}} e^{-\gamma^*} \Delta\gamma + \mathcal{O}(\Delta\gamma) \end{aligned} \tag{3.29}$$

Ignore the high order components and bring it back to (3.27), we have

$$
\begin{aligned}
\hat{f}_i &= \arg \min_{\substack{f_i \in R_i \\ \|f_i\|=1}} E\left[Q(\sqrt{2\left|H^+ f_i\right|^2 \rho})|H \in R_i\right] \\
&\doteq \arg \min_{\substack{f_i \in R_i \\ \|f_i\|=1}} E\left[Q(\sqrt{2\gamma^*}) + \frac{1}{2\sqrt{\pi\gamma^*}} e^{-\gamma^*} \Delta\gamma | H \in R_i\right] \\
&\Leftrightarrow \arg \max_{\substack{f_i \in R_i \\ \|f_i\|=1}} E\left[-\frac{1}{2\sqrt{\pi\left|H\right|^2 \rho}} e^{-|H|^2\rho} \Delta\gamma | H \in R_i\right] \\
&= \arg \max_{\substack{f_i \in R_i \\ \|f_i\|=1}} E[\frac{1}{2\sqrt{\pi\left|H\right|^2 \rho}} e^{-|H|^2\rho} r^* \\
&- \frac{1}{2\sqrt{\pi\left|H\right|^2 \rho}} e^{-|H|^2\rho} r | H \in R_i] \\
&= \arg \max_{\substack{f_i \in R_i \\ \|f_i\|=1}} E\left[\frac{1}{2\sqrt{\pi\left|H\right|^2 \rho}} e^{-|H|^2\rho} \left|H^+ f_i\right|^2 \rho | H \in R_i\right]
\end{aligned}
\tag{3.30}
$$

This is a modified version of the centroid condition. Intuitively it can be viewed as adding a weight factor to each channel realization according to how much this channel realization contributes to BER. Those channel realizations with high error probability would be assigned larger weight factors comparing to those with low error probability, so that $\hat{f}_i$ will end up being closer to them. In this manner, the averaged BER is minimized instead of the average received SNR within the partition.

If we look at the nearest neighbor rule under the BER criterion, it should be

$$
\begin{aligned}
R_i &= \left\{Q(\sqrt{2\left|H^+ f_i\right|^2 \rho}) \leq Q(\sqrt{2\left|H^+ f_i\right|^2 \rho}), \forall j \neq i\right\} \\
\Leftrightarrow R_i &= \left\{\left|H^+ f_i\right| \geq \left|H^+ f_i\right|, \forall j \neq i\right\}
\end{aligned}
\tag{3.31}
$$

which means we can use the exactly same nearest neighbor rule as GLA.

3.3.3 Capacity Based Codebook Design

We could also apply the VQ-based approach [52] which maximizes the average system capacity to design our root codebook. The optimization criterion becomes

$$
\begin{aligned}
F_o &= \arg\max_F E\left[C(\gamma)\right] \\
&= \arg\max_F E\left[\log(1+\gamma)\right] \\
&= \arg\max_F E\left[\log(1+\left|H^+ f_i\right|^2 \rho)\right]
\end{aligned}
\tag{3.32}
$$

The corresponding centroid condition would be

$$\hat{f}_i = \arg \max_{\substack{f_i \in R_i \\ \|f_i\|=1}} E\left[\frac{|H^+ f_i|^2 \rho}{2(1+|H|^2 \rho)} | H \in R_i\right] \tag{3.33}$$

In the simulation part, we will compare the performance of this capacity-based design with other designs that we mentioned before.

3.4 Performance Analysis

In this section, we examine performance of our proposed rotation-based beamforming codebook design. Two fading channels are considered: a slow fading channel with $f_d T = 0.06$ (corresponds to $\alpha = 0.9648$) and a fast fading channel with $f_d T = 0.1$ (corresponds to $\alpha = 0.9037$). We assume that the number of transmit antennas $N_t = 4$ and a single received antenna is used. We choose a 3-bit feedback rate, and all the codebooks we design have a size of $N_b = 8$. We generate a long sequence of correlated channel realizations $\{H^{(0)}, H^{(1)}, \cdots, H^{(n)}, \cdots, \}$, apply various codebooks (the proposed codebook in Table 1, the i.i.d. codebook in, etc.) from which the beamforming vector is selected, and observe the average received SNR, system capacity and bit error rate under both coded and uncoded systems.

3.4.1 Average Received SNR

In Figure 3.2, we measure the average received SNR of SNR-R, CAP-R, BER-R and the previously proposed codebooks in the slow fading channel ($\alpha = 0.9648$). The 'i.i.d. codebook' refers to the GLA codebook [19] assuming the channel is i.i.d. distributed. The 'ideal' curve assumes that the transmitter knows perfect H so that it uses the optimal beamforming vector $f = \bar{H}^{\dagger}$. Figure 3.2 shows that the proposed SNR-R, CAP-R and BER-R codebooks achieves almost the same average received SNR, and outperform the i.i.d. codebook with a roughly $2dB$ gain. We perform the same simulation for the fast fading channel ($\alpha = 0.9037$) in Figure 3.4, and similar observations are obtained, except that the gap between the i.i.d. codebook and the proposed codebooks is reduced to around $1dB$. We also calculate the capacity as another performance measurement. The curves are consistent with the corresponding

received SNR curves in Figure 3.2 and 3.4, so we will not post the figures here for brevity.

3.4.2 BER in the Uncoded System

In this part, we would compare the uncoded BER performance of SNR-R, BER-R and CAP-R codebooks. We assume BPSK modulation in order to be consistent with our derivations in the BER root codebook design section. With the same system configuration $N_t = 4$ and $N_b = 8$ as above, we plot the BER performance in Figure 3.5 for the slow fading channel where $f_dT = 0.06, \alpha = 0.9648.$ and Figure 3.6 for the fast fading channel where $f_dT = 0.1, \alpha = 0.9037$. As shown in Figure 3.5, the SNR-R and CAP-R shows the high SNR problem in the region $\rho \geq 4.5dB$, whereas BER-R codebook always performs better than the i.i.d. codebook. In the low (transmit) SNR region, BER-R achieves approximately $1.5dB$ gain over the i.i.d. codebook, yet gap reduces as ρ increases. It is more obvious in Figure 3.6 that BER-R still exhibits the high SNR problem somehow. There is a crossing that occurs at around $\rho = 10.5dB$. The reason, as we can think of, might be that our proposed BER root codebook design is based on the approximation of the Q function. Meanwhile, the Lloyd algorithm itself is suboptimal. Hence we can not guarantee that our BER root codebook optimize the average BER. However, this design still shows obvious improvement, comparing to SNR-R and CAP-R codebooks, and it shows better BER performance than the i.i.d. codebook at low SNR region.

3.4.3 BER in the Coded Systems

Many real communication systems utilize channel coding to achieve near-capacity performance, so it is important for us to verify the performance of our proposed codebooks in a coded system. During our analysis to the high SNR problem, we expected that the high SNR problem would not exist in a coded system. We confirm this by the following simulations. Assuming $N_t = 4$, $N_b = 8$ and QPSK modulation, we consider a Turbo coded system with $\frac{1}{2}$ coding rate and the coding length of 8192. The generating functions for the recursive encoders are $g_1(D) = 1 + D + D^2$ and $g_2(D) = 1 + D^2$ respectively. Since the performance of the i.i.d. codebook does

not depend on the temporal correlation, we plot all the curves in one figure. The 'SNR-R $f_dT = 0.06$' curve denotes the BER performance of the SNR-R codebook in the slow fading channel, and similarly the 'GLA $f_dT = 0.1$' curve is for the fast fading fading channel. As shown in Figure 3.7 (for the slow fading channel) and Figure 3.8 (for the fast fading channel), SNR-R codebook exhibits a $1.5dB$ gain in the slow fading channel comparing to the i.i.d. codebook, while BER-R and CAP-R codebooks obtain an extra $0.5dB$ gain, in the slow fading channel. In the fast fading channel, the performance of BER-R and CAP-R only degrades slightly while SNR-R codebook has about $1dB$ degradation, comparing to the curves in the slow fading channel.

3.5 Conclusion

In this work, we show that by exploiting channel temporal and spatial correlation, the gain of transmit beamforming can be effectively utilized based on limited feedback. In both slow fading and fast fading channel, the proposed beamforming design demonstrates superior performance over existing designs that do not consider channel correlation. Performance analysis of our design and its extension to more general correlated fading channels including frequency selective channels are directions for future research.

Table 3.1. Steps of proposed beamforming codebook design for a temporally correlated channel

Offline:	
	1.Randomly generate a $N_t \times 1$ root channel direction $f^r \in \Omega(N_t, 1)$; 2. Apply the Generalized Lloyd Algorithm to design a codebook for time 0: $\mathcal{F}_o^{f^r}$, based on channel model (3.1)
Inline:	
Time 0:	
	The receiver selects the best codeword $\tilde{f}^{(0)}$ and feeds back the codeword index to the transmitter. Find an unitary matrix U^0 such that $\tilde{f}^{(0)} = U^{(0)} f^r$. Let $\mathcal{F}^{(1)} = U^{(0)} \mathcal{F}_o f^r$.
Time n-1	...
Time n	The receiver selects the best codeword $\tilde{f}^{(n)}$ for time n and feeds back the codeword index to the transmitter. Let $\mathcal{F}^{(n+1)} = U^{(n+1)} \mathcal{F}_o f^r$ where $\tilde{f}^{(n)} = U^{(n)} f^r$
Time n+1	...

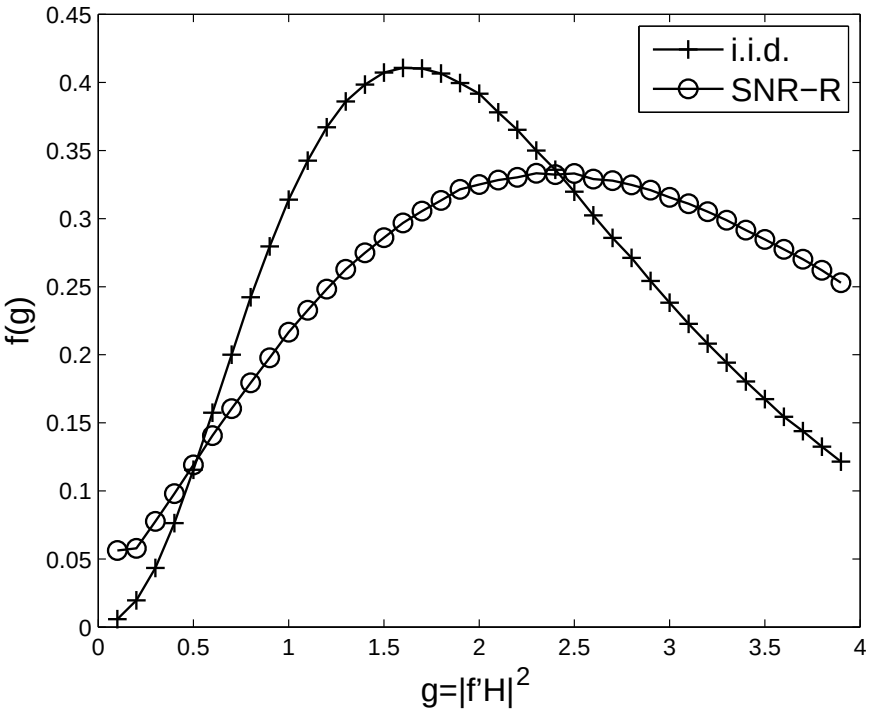


Figure 3.1. the probability density function of the received SNR for the proposed codebook and the i.i.d. codebook

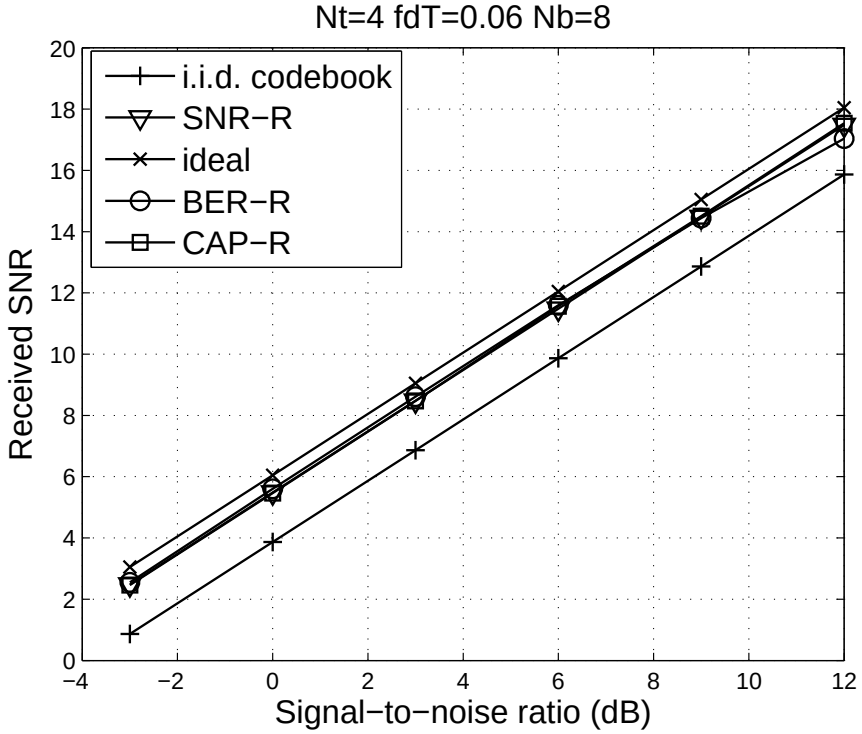


Figure 3.2. Performance comparison of proposed codebook with the i.i.d. codebooks that do not consider temporal correlation in the slow fading channels. $N_t = 4$, $a = 0.9648$.

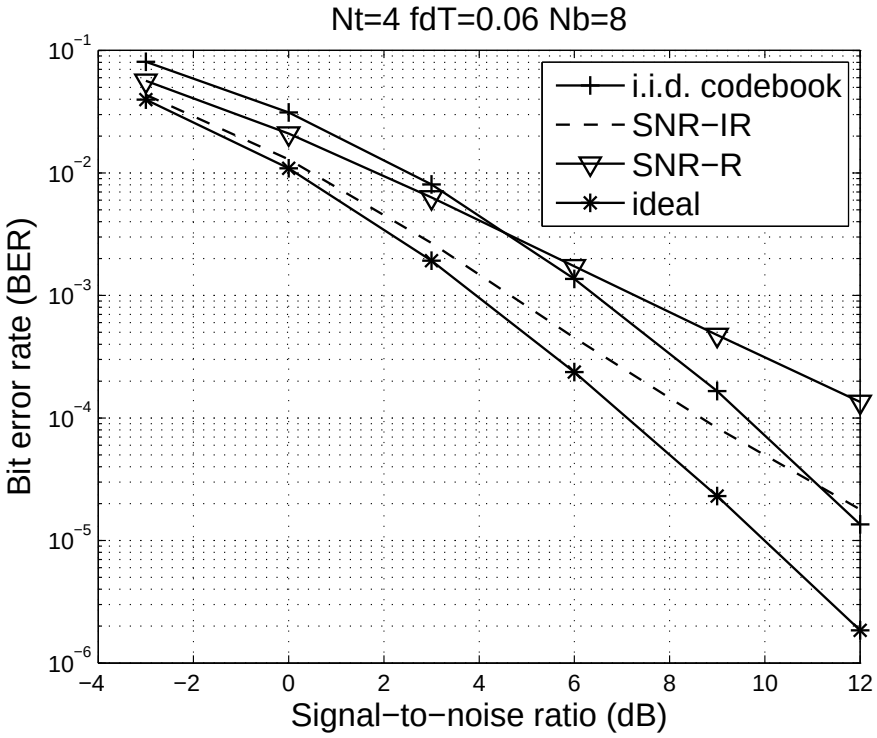


Figure 3.3. BER performance of various codebooks under the slow fading channel

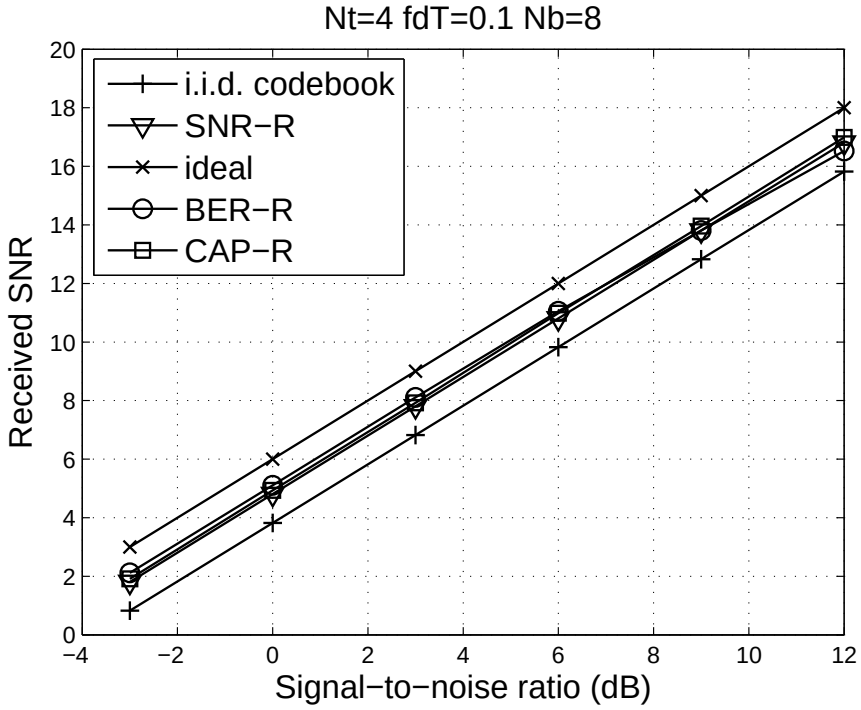


Figure 3.4. Performance comparison of proposed codebook with the i.i.d. codebooks that do not consider temporal correlation in the fast fading channels. $N_t = 4$, $a = 0.9037$.

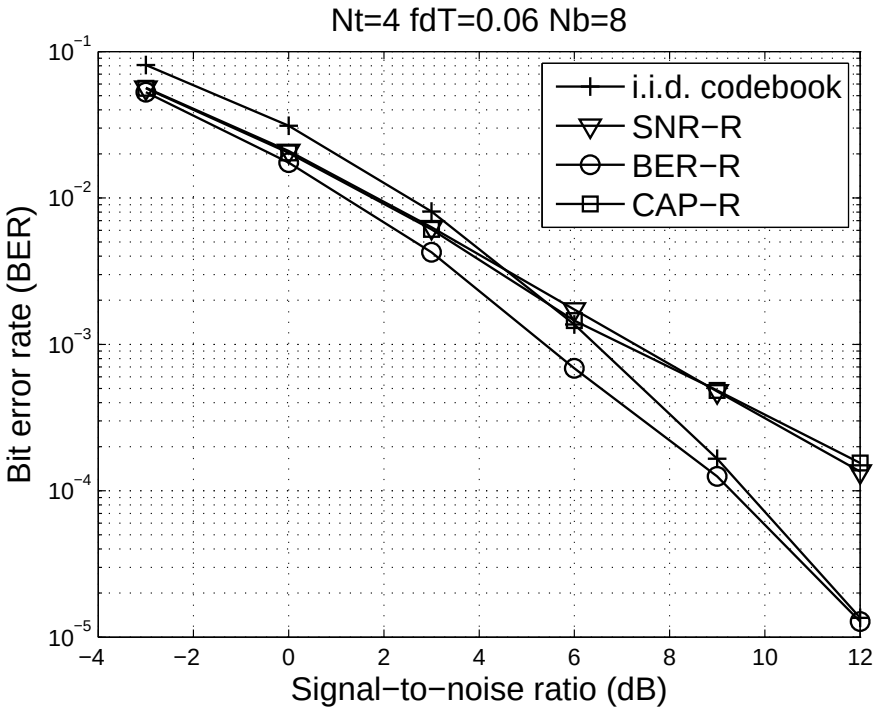


Figure 3.5. BER performance of various codebooks under the slow fading channel

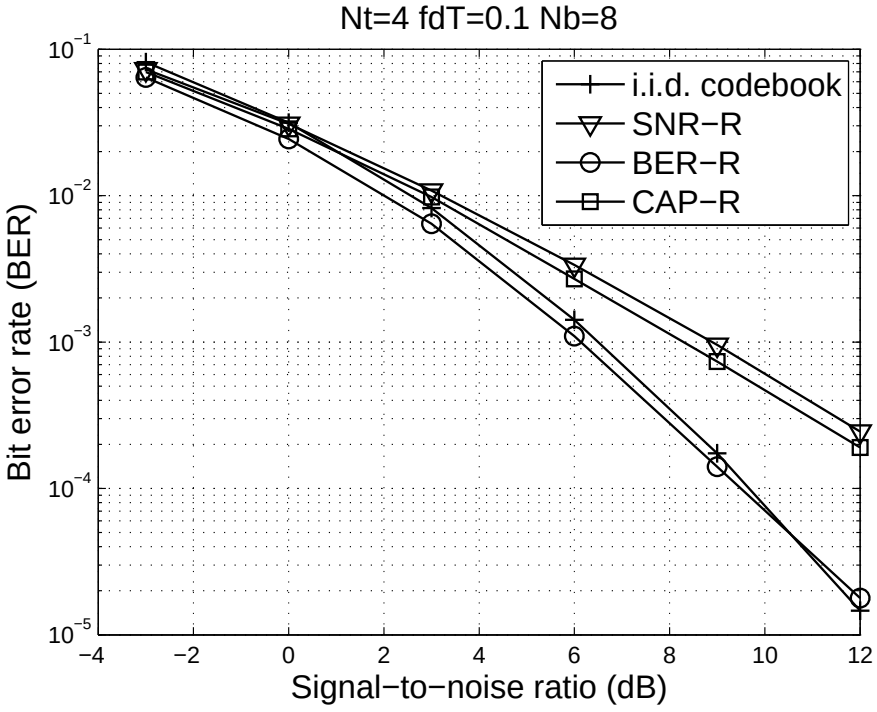


Figure 3.6. BER performance of various codebooks under the fast fading channel

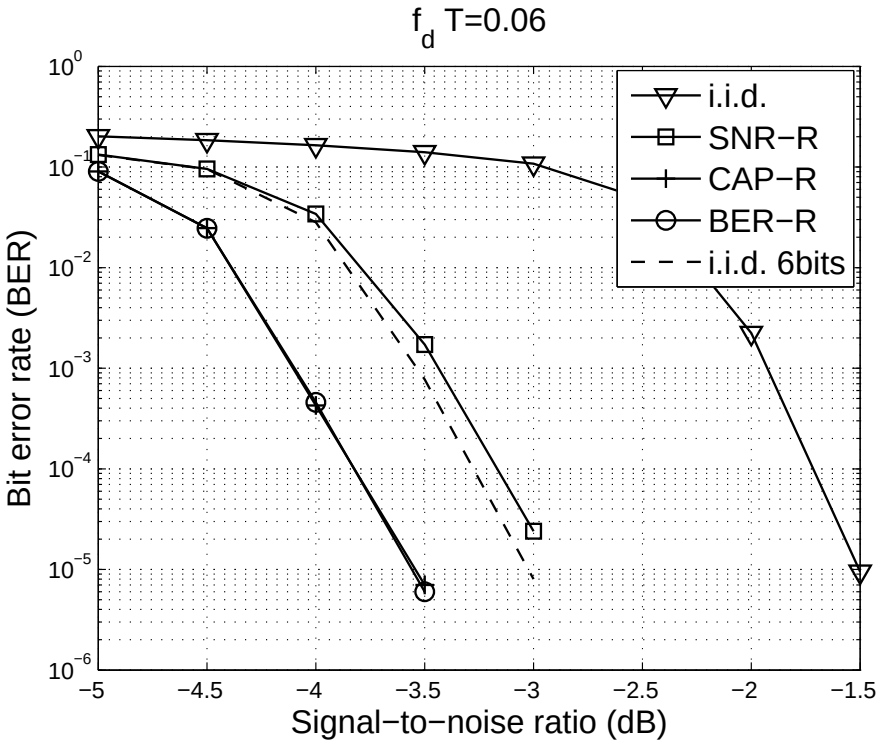


Figure 3.7. BER curves for SNR-R codebook and the i.i.d. codebook in the coded system assuming $f_d T = 0.06$

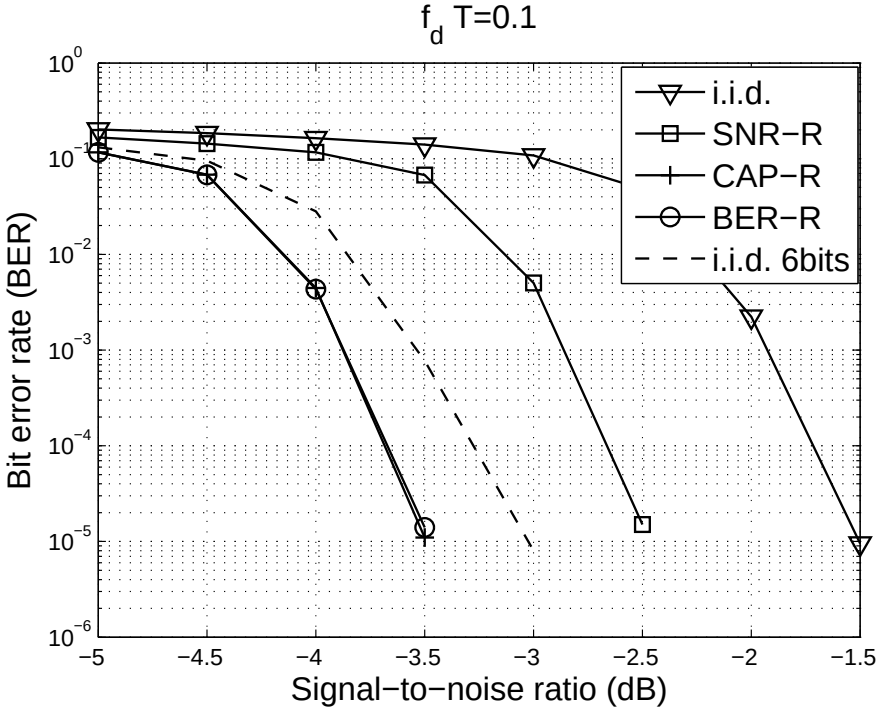


Figure 3.8. BER curves for SNR-R codebook and the i.i.d. codebook in the coded system

CHAPTER 4

MARKOV CHAIN MONTE CARLO DETECTION FOR FREQUENCY SELECTIVE CHANNELS USING LIST CHANNEL ESTIMATES

Turbo equalization (TEQ) is a powerful technique to combat intersymbol interference (ISI) resulting from multipath transmission. In a system that employs turbo equalization, soft information about the transmitted bits is exchanged iteratively between a soft-input/soft-output (SISO) equalizer and a SISO channel decoder such that improved channel equalization can be achieved with the aid of soft information from the channel decoder, and improved decoding thereby also results. The overall performance approaches that of optimal maximum *a posteriori* (MAP) joint equalization and decoding.

Most of the SISO equalizers that have been proposed in the literature are either trellis-based [54–56], which aim to approximate the optimal MAP detection, or are based on soft ISI cancellation [57] and linear filtering [58, 59]. For channels with long memory or when the size of signal constellation is large, the complexity of the trellis-based approach becomes prohibitive due to the exponential number of states in the trellis. While it is possible to reduce the complexity of such algorithms using state-reduction techniques [54, 60–64], much of the study is still limited to channels with a moderate number of states in the trellis representation. The linear filtering based SISO equalizers, such as the minimum mean-square-error (MMSE) turbo equalizer [58, 59], have been widely studied in the literature due to their excellent performance and lower complexity compared to that of the trellis-based approaches. Such equalizers have been successfully applied to channels with long delay spread, such as the underwater acoustic (UWA) channels where the ISI can

span tens or even hundreds of symbol periods. The MMSE turbo equalizer, however, and any suboptimal SISO equalizer, can incur performance loss compared to that using the optimal MAP equalizer [58, 59]. Recent work that applies TEQ techniques to joint channel estimation and data detection include [65–68] for terrestrial channels, and [69–71] for UWA channels. These approaches largely adopt the linear filtering approach to combat ISI for channels with long memory.

In this work, we investigate a different class of SISO equalizers based on Markov Chain Monte Carlo (MCMC) techniques [23, 26]. The MCMC detector adopts a statistical method, called Gibbs sampling, to search for a small (to keep the complexity low) but important (to achieve good performance) sample set containing the most likely transmitted symbol vectors. The key point that makes MCMC attractive for SISO detection over ISI channels is that, unlike the optimal MAP detector, its complexity does not grow exponentially with the channel memory, yet its performance remains close to that of the optimum MAP detector. Original versions of SISO MCMC detectors have been developed in [27] for stationary frequency-selective channels, assuming perfect knowledge of the channel impulse response (CIR) at the receiver. It is shown in [27] that the MCMC detector can better approximate the optimal MAP detector than the MMSE turbo equalizer [58,59] and achieve substantial performance gains, which amounted to around 2 dB for the various ISI channels and conditions considered.

In this work, we extend the work of [27] to develop new MCMC detectors for the more challenging time-varying frequency-selective channels. Such channels require a different design of MCMC detectors beyond those of [27] to facilitate joint channel tracking and data detection. A main contribution of this paper is the development of a new MCMC detector, termed MCMC with list channel estimates (MCMC-LCE), for time-varying frequency-selective channels. A salient feature of the MCMC-LCE detector is that it computes the log-likelihood ratio (LLR) of the coded bits based on a list of estimates of the CIR. This allows for additional degrees of freedom in the estimation of the CIR, and hence, can significantly improve the robustness of the detector to channel uncertainty. Compared to existing detectors that operate based on a single CIR estimate (which can be inaccurate), the MCMC-LCE detector

demonstrates superior channel tracking capability and data detection performance. The MCMC-LCE detector is driven by an adaptive variable step-size least mean square (VSLMS) algorithm for channel estimation. The VSLMS algorithm does not require prior information of the probability distribution of the CIR, and hence, is particularly attractive for situations where a good statistical channel model is not available. This is the case for UWA channels, whose study has been one of the main motivations for the developments in this paper.

To demonstrate the effectiveness of the proposed MCMC-LCE detector, we provide performance comparisons with MMSE detectors over a set of synthetic ISI channels and real UWA channel measurements from at-sea experiments. The UWA channels feature large delay spread, frequency-dependent broadband Doppler, and high time variability, and are considered to be one of the most challenging communication channels in use today [72]. Our results confirm that, for both synthetic channels and UWA channels, the MCMC-LCE can offer significant performance improvements over the MMSE-based detector. Various versions of MMSE-based detectors exist in the literature, even though they all follow similar principles of TEQ. The differences among them are mostly due to their specific implementation of the channel estimator, e.g., the MMSE approaches of [66, 70, 71, 73] adopt symbol-wise channel refinement, or employ a direct adaptive equalizer structure in which the channel is not explicitly estimated but rather the coefficients of an equalizer are adaptively estimated, as in an adaptive linear or decision feedback equalizer. The adaptive channel estimation algorithms used may also vary, e.g., [66] uses a recursive least squares (RLS) algorithm, while [70, 71] adopt the LMS algorithm. In [74, 75], the channel estimate is assumed to be constant over a block of data. Since detailed comparisons of these MMSE equalizers are outside the scope of this paper, we simply present performance comparisons with the MMSE detectors similar to those in [66, 71, 73], with minor modifications to optimize their performance for the channels considered in this work.

Applications of blind MCMC detection to stationary frequency-selective channels have been considered previously in [76] for a single-carrier system, and later extended to multicarrier systems [77, 78]. In [76], the authors adopted a Gibbs sampler to generate a sample sequence of the transmitted data and estimated CIR in an alternating

fashion. A probability distribution of the CIR is assumed to be known *a priori*, based on which samples of the CIR are drawn. For a stationary four-tap ISI channel with binary phase shift keying, [76] showed that good performance can be achieved with a small number of samples. Moreover, the *a posterior probabilities* (APP) of the bits are computed following a bit-counting approach. The MCMC-LCE detector considered in this work is different from that considered in [76]. First, we are interested in a time-varying channel with a potentially long CIR and in which *a priori* statistics might not be available. This makes it difficult to directly apply the Bayesian framework developed in [76]. Second, in order to reduce the number of samples required for satisfactory detection performance under long CIR, we compute the LLR values of the bits based on importance sampling, rather than bit-counting. It has been shown in prior work [23, 26] that the former approach improves the reliability of the LLR values with a greatly reduced number of samples. Third, due to the time-variability of the channel, our design is adaptive in nature to enable reliable channel tracking.

The remainder of this chapter is organized as follows. In Section 4.1, we present the system setup used herein. Section 4.2 presents a review of the VSLMS algorithm and the rationale for adopting it to the specific application of interest in this paper. The main contribution of the paper, namely, the proposed MCMC-LCE detector for time-varying channels, is presented in Section 4.3. Numerical results and performance analysis for the synthetic channel and the UWA channels are presented in Section 4.4.

Throughout this chapter, we use lower case bold letters to represent vectors, and capital bold letters to represent matrices. The notation $(\cdot)^T$ denotes the transpose operator, and $(\cdot)^*$ denotes conjugate transpose. We also let $\mathbf{x}_{i:j} = (x_i, x_{i+1}, \cdots, x_j)^T$ represent a partial sequence of vector $\mathbf{x}$.

4.1 System Setup

4.1.1 Transmitter Structure

Figure 4.1(a) presents a block diagram of the transmitter of the communication system considered in this paper. The vector $\mathbf{b}$ contains a sequence of (uncoded) information bits. This is passed to a channel encoder that adds redundant bits to $\mathbf{b}$. The coded bits are passed through an interleaver whose output is the interleaved coded

bit vector $\mathbf{c}$. We consider a 2^Q-ary symbol alphabet $\mathcal{A} = \{v_1, v_2, \cdots, v_{2^Q}\}$. Each symbol v_k consists of bits $\{v_{k,1}, v_{k,2}, \cdots, v_{k,Q}\}$, where $v_{k,j} \in \{0, 1\}$. The modulation block converts $\mathbf{c}$ to a vector of complex-valued symbols from the symbol alphabet $\mathcal{A}$. Periodically inserted pilot symbols are added and the result that we call $\mathbf{x}$ is passed to the channel for transmission. We refer to $\mathbf{x}$ as a data packet. As presented in Figure 4.1(b), each packet consists of T frames, and each frame consists of a segment of pilot symbols, followed by several segments of data symbols. We assume that the channel variation over time is slow enough so that it can be approximated by a constant matrix over the duration of each segment.

4.1.2 Channel Model

The communication link is a single-input multiple-output (SIMO) multi-path channel with a single transmit element and K receive elements. The received signal at the k-th receive element at time n, denoted by $y_n^{(k)}$, is given by

$$y_n^{(k)} = \sum_{p=0}^{L} h_{n,p}^{(k)} x_{n-p} + w_n^{(k)}, \quad k = 1, \cdots, K, \tag{4.1}$$

where n and p are time and path indices, respectively, $h_{n,p}^{(k)}$ is the channel gain of the p-th path between the transmitter and the k-th receive element, x_n is the transmitted signal, $w_n^{(k)}$ is the channel noise.

For brevity, we let $\mathbf{x}$ denote the transmitted symbol vector, and denote the received signal from the K receive elements as $\mathbf{Y} = (\mathbf{y}^{(1)}, \cdots, \mathbf{y}^{(K)})$, where N is the number of transmitted symbols and $\mathbf{y}^{(k)} = (y_0^{(k)}, y_1^{(k)}, \cdots, y_{N-1}^{(k)})^T$.

4.1.3 Receiver Architecture

A block diagram of the receiver architecture is shown in Figure 4.2. Based on the packet format shown in Figure 4.1, we divide $\mathbf{Y}$ into a number of segments, say, $\mathbf{Y}_0$, $\mathbf{Y}_1$, $\mathbf{Y}_2$, $\cdots$, with $\mathbf{Y}_i$ being the received signal of the i-th segment. The MCMC-LCE is applied to each segment for joint data detection and channel estimation. The MCMC-LCE is a soft-in soft-out detector. For segment l, its input is $\boldsymbol{\gamma}$, the LLR values of coded bits, provided by the channel decoder, and the initial channel estimate $\hat{\mathbf{H}}_l$. We assume that the channel variation over time is slow enough so that it can be

approximated by a constant matrix over segment l. Thus $\hat{\mathbf{H}}_l$ is a constant channel gain matrix given by $\hat{\mathbf{H}}_l = (\mathbf{h}_l^{(1)}, \cdots, \mathbf{h}_l^{(K)})$, where $\mathbf{h}^{(k)} = (h_0^{(k)}, h_1^{(k)}, \cdots, h_L^{(k)})^T$ is a column vector representing the CIR vector between the transmit element and the k-th receive element. The derivation of $\hat{\mathbf{H}}_l$ and the details of MCMC-LCE are presented in Section 4.3. We apply MCMC-LCE over each segment l sequentially to generate updated extrinsic LLR values of all coded bits, denoted by $\boldsymbol{\alpha}$, and these are passed, after de-interleaving, to the channel decoder. Subsequently, after channel decoding, the updated $\boldsymbol{\gamma}$ values are fed back to the MCMC-LCE to facilitate the next iteration of joint data detection, channel estimation, and decoding.

4.2 Channel Estimation

Successful implementation of a TEQ to a great extent depends on the quality of the channel estimator. Ideally, one wishes to use a perfect estimate of the channel. However, this cannot be the case in practice, particularly, when the channel is time-varying and noisy. An adaptive algorithm need to collect sufficient samples of the underlying signals to find the channel that fit best to the signals statistics. This introduces a lag in adaptation process, since the channel will be changed by the time that the signals samples/statistics are collected, hence, the channel estimator will suffer from an adaptation lag. The larger the adaptation lag, the worse will be the channel estimate. To reduce the adaptation lag, one may reduce the number of samples based on which the channel estimate is calculated. But, using a smaller number of signal samples lead to a less accurate estimate of the channel. Hence, one should choose the adaptation parameters to strike a balance between the adaptation lag and the misadjustment arising from the lack of sufficient statistics. This concept is well understood and been widely addressed in the literature within the general frame of adaptive filters theory, e.g., see [79–81].

To elaborate the above points further, with direct reference to the common adaptive filtering algorithms, recall the least-mean square (LMS) algorithm update equation

$$\mathbf{w}_{n+1} = \mathbf{w}_n + 2\mu e_n \mathbf{x}_n^* \tag{4.2}$$

where $\mathbf{w}_n$ and $\mathbf{x}_n$ are, respectively, the tap-weight and tap-input vectors, and μ is the step-size parameter. A smaller value of μ results in a longer memory in considering the underlying signals statistics, hence, a larger adaptation lag. Also, assuming that the signals statistics (accordingly, the channel impulse response, in the case of interest to this paper) are time-invariant, a smaller μ results in a channel estimate with a lower misadjustment. Misadjustment arises because of the use of noisy samples of the gradient in the LMS update equation. However, in the case of time-varying channels, μ has to be chosen to strike a balance between the adaptation lag (i.e., the tracking capability of the algorithm) and the misadjustment of the channel estimate.

For a given statistics of the gradient vector $2e_n\mathbf{x}_n^*$ in (4.2), a number of studies have been performed to find the optimum choice of the step-size parameter μ that optimizes the balance between the adaptation lag and misadjustment; e.g., [79, 82]. Moreover, a variable step-size LMS (VSLMS) algorithm that adaptively finds this optimum choice of μ has been proposed, [82]. In [83], it has been noted that in some cases, such as multipath channels, the statistics of the elements of the gradient vector $2e_n\mathbf{x}_n^*$ may vary significantly and thus a VSLMS algorithm with different step-size parameters for various filter taps was proposed. Further development of this algorithm was later reported in [84]. In our study of UWA channels, we have found that the VSLMS algorithms of [83, 84] the best match. The impulse response of a UWA channel typically consists of a few sparse multipaths (or cluster of multipths) each with a different fading rate. Many taps in the impulse response may be zero, and thus that are associated with the line-of-sight path or arise from reflections from the sea-bed may vary slowly. Only the taps that are associated with the reflection from the sea surface will be fast-fading. The use of VSLMS algorithms of [83, 84] will lead to a self-optimizing channel estimator that without any prior knowledge of the channel adapts to the various segments of the channel impulse response. In other words, the VSLMS algorithm, in a very effective way, learns about the sparsity of the channel as well as the fading rate of the various taps in the channel and adapts its step-size parameters for a near optimum tracking of the channel impulse response.

4.2.1 VSLMS Channel Estimation

A number of choices of the VSLMS algorithms have been suggested in [84]. Following the discussion and suggestions made in [84], we adopt the following in this paper. We update the channel estimates as follows

$$\hat{h}_{n+1,p}^{(k)} = \hat{h}_{n,p}^{(k)} + \mu_{n,p}^{(k)} e_n^{(k)} x_{n-p}^*. \tag{4.3}$$

where $\mu_{n,p}^{(k)}$ is the step-size parameter at time n for the pth tap of the kth channel in the UWA SIMO channel. The step-size parameter $\mu_{n,p}^{(k)}$ is updated as

$$\mu_{n,p}^{(k)} = (1 + \rho \, \mathrm{Re}\{e_n^{(k)} x_{n-p}^* \phi_{n,p}^{(k)*}\}) \mu_{n-1,p}^{(k)}, \tag{4.4}$$

where

$$\phi_{n,p}^{(k)} = \alpha \phi_{n-1,p}^{(k)} + e_{n-p} x_{n-p-1}^*, \tag{4.5}$$

and ρ and α are the algorithm parameters that should be optimized empirically.

To give some insights to the mechanism of the VSLMS algorithm, we note that $\phi_{n,p}^{(k)}$ indicates the average direction of the stochastic gradient $e_n x_{n-p}^*$ over the past iterations. The update equation (4.4), effectively, compares the signs of the real and imaginary parts of $e_n x_{n-p}^*$ and $\phi_{n,p}^{(k)}$, and increases the step-size parameter if on average these terms have similar signs, and decreases the step-size parameter, otherwise; see [83] for my insight. When the VSLMS algorithm has converged to a point close to the respective optimum tap weight, the signs of the real and imaginary parts of $e_n x_{n-p}^*$ change more frequently and thus the step-size parameter is decreased.

4.2.2 VSLMS Under Decision-Directed Mode

For the data segment, the VSLMS algorithm operates based on soft estimates of the transmitted symbols. We should replace x_i in (4.4) by $\bar{x}_i$, which is defined as

$$\bar{x}_i = E(x_i) = \sum_{v_k \in \mathcal{A}} v_k P(x_i = v_k). \tag{4.6}$$

The computation of the symbol a posteriori probabilities $P(x_i = v_k)$ depends on the prior LLR for the bits $\{x_{i,j}, j = 1, \cdots, Q\}$ that constitute x_i. Let

$$\lambda_{i,j} = \ln \frac{P(x_{i,j} = 0)}{P(x_{i,j} = 1)}, \tag{4.7}$$

where $\lambda_{i,j}$ is provided by the channel decoder, or by the MCMC-LCE. We obtain [59]

$$P(x_i = v_k) = \prod_{j=1}^{Q} P(x_{i,j} = v_{k,j}) = \prod_{j=1}^{Q} \frac{1}{2}(1 + \tilde{v}_{k,j} \tanh(\lambda_{i,j})), \tag{4.8}$$

where

$$\tilde{v}_{k,j} = \begin{cases} +1, & v_{k,j} = 0 \\ -1, & v_{k,j} = 1 \end{cases} \tag{4.9}$$

4.3 MCMC Detector Based on List Channel Estimate

In this section, we present the proposed MCMC-LCE detector. The MCMC-LCE detector generalizes and improves upon earlier versions of the MCMC detectors presented in prior work [2, 85, 86]. Our study of the synthetic channels and UWA channels shows that the MCMC-LCE detector developed here yields superior and more robust performance as compared to these earlier versions.

4.3.1 MCMC-LCE as a Low-Complexity Approximation to MAP Detection

First, we briefly review the optimal MAP detection and explain why MCMC-LCE provides a low-complexity approximation to MAP detection. Let us consider MAP detection over an arbitrary data segment l. For ease of presentation, we represent x_l, which is the transmitted symbol vector for segment l, by its corresponding bit vector c_l. For brevity, we ignore the subscript l in c_l, and simply denote it by c. Assume that the length of the bit vector c is $B = QS$, where S is the number of coded symbols in the i-th segment, and Q is the number of bits in each constellation point. The prior LLR for each bit in $\mathbf{c}$, provided by the channel decoder, is given by $\gamma = (\gamma_0, \cdots, \gamma_{B-1})$, where $r_m = \ln \frac{P(c_m=0)}{P(c_m=1)}$, for every $0 \leq m \leq (B-1)$. Assume that $\mathbf{H_l}$, the channel CIR for segment l, is perfectly known. The MAP detector computes the extrinsic LLR of each bit c_m as follows.

$$\begin{aligned}\hat{\alpha}_m &= \ln \frac{P(c_m = 0|\mathbf{Y_l}, \gamma, \mathbf{H_l})}{P(c_m = 1|\mathbf{Y_l}, \gamma, \mathbf{H_l})} - \gamma_m \\ &= \ln \frac{\sum\limits_{\mathbf{c:c_m=0}} P(\mathbf{c}|\mathbf{Y_l}, \gamma, \mathbf{H_l})}{\sum\limits_{\mathbf{c:c_m=1}} P(\mathbf{c}|\mathbf{Y_l}, \gamma, \mathbf{H_l})} - \gamma_m \\ &= \ln \frac{\sum\limits_{\mathbf{c:c_m=0}} p(\mathbf{Y_l}|\mathbf{c}, \mathbf{H_l})\mathbf{P}(\mathbf{c}|\gamma)}{\sum\limits_{\mathbf{c:c_m=1}} p(\mathbf{Y_l}|\mathbf{c}, \mathbf{H_l})\mathbf{P}(\mathbf{c}|\gamma)} \end{aligned} \tag{4.10}$$

Each summation in (4.10) is over a total of 2^{B-1} combinations of $\mathbf{c}$. For typical values of B which are in the order of at least few hundreds, this clearly is a prohibitive complexity. The main idea of the MCMC-LCE is to use Gibbs samplers to find a list of most likely pairs of transmitted sequences and CIR, and then approximate the numerator and the denominator in (4.10) by summing over a much smaller number of samples belonging to the list. The complexity of the MCMC-LCE is controlled by the list size. In this work we will show that a very small list size is sufficient for the MCMC-LCE to achieve good performance.

4.3.2 General Description of MCMC-LCE

A block diagram of the MCMC-LCE detector is presented in Figure 4.3. Here, we consider the operations of MCMC-LCE over an arbitrary data segment l. The initial channel estimate for the l-th segment is given by $\hat{\mathbf{H}}_l$. Details for computing $\hat{\mathbf{H}}_l$ is given in Section 4.3.4. Based on the received signal $\mathbf{Y}_l$, the initially estimated CIR $\hat{\mathbf{H}}_l$, and the soft feedback γ from the channel decoder, the MCMC-LCE employs G parallel Gibbs samplers (GS) to collect a list containing pairs of most likely transmitted vectors and matching CIRs $\{\mathbf{x_l^{(i)}}, \hat{\mathbf{H}}_\mathbf{l}^{(\mathbf{i})}\}$. Assume that we run a total of I iterations within each GS, then a maximum of $G \cdot I$ sample pairs are collected. This list is used to compute the LLRs of the transmitted bits, based on which we obtain a soft symbol estimate $\bar{\mathbf{x}}_l$. The extrinsic LLRs, α, will be passed to the channel decoder. Based on $\bar{\mathbf{x}}_l$, the VSLMS algorithm is applied to obtain an updated channel estimate $\hat{\mathbf{H}}_l^f$. This channel estimate is then passed as an input to the MCMC-LCE detector for segment $l + 1$. Next, we explain various blocks of the MCMC-LCE detector shown in Figure 4.3 in detail.

4.3.3 Generation of a List Using Gibbs Sampler

We first explain how the Gibbs sampler works. At the beginning, the Gibbs sampler generates an initial vector $\mathbf{c}^{(0)} = (c_0^{(0)}, c_1^{(0)}, \cdots, c_{B-1}^{(0)})$ randomly. Then, during the i-th iteration, the Gibbs sampler starts from the sample vector $\mathbf{c}^{(i-1)}$, the estimated CIR $\tilde{\mathbf{H}}_l^{(i-1)}$, found during the $(i-1)$-th iteration, and updates each of the B bits sequentially to obtain a new vector $\mathbf{c}^{(i)}$. Specifically, the m-th bit, $c_m^{(i)}$, is generated according to its *a posteriori* probability (APP) distribution conditioned upon the newly updated bits $(c_0^{(i)}, \cdots c_{m-1}^{(i)})$, that have already been generated during the same iteration, and also bits $(c_{m+1}^{(i-1)}, \cdots c_{B-1}^{(i-1)})$ obtained from the $(i-1)$-th iteration. Let

$$\bar{\mathbf{c}}_m = (c_0^{(i)}, \cdots, c_{m-1}^{(i)}, c_{m+1}^{(i-1)}, \cdots, c_{B-1}^{(i-1)}).$$

We draw a sample $c_m^{(i)}$ based on the LLR provided by the Gibbs sampler, denoted by $\beta_m^{(i)}$, defined as

$$\beta_m^{(i)} = \ln \frac{P(c_m = 0 | \bar{\mathbf{c}}_m, \mathbf{Y}_l, \gamma_m, \tilde{\mathbf{H}}_l^{(i-1)})}{P(c_m = 1 | \bar{\mathbf{c}}_m, \mathbf{Y}_l, \gamma_m, \tilde{\mathbf{H}}_l^{(i-1)})}. \tag{4.11}$$

Next, we describe how to compute $\beta_m^{(i)}$. For each $a = 0, 1$, we define

$$\mathbf{c}^{[a]} = \{c_0^{(i)}, \cdots, c_{m-1}^{(i)}, a, c_{m+1}^{(i-1)}, \cdots, c_{B-1}^{(i-1)}\},$$

and let $\mathbf{x}^{[a]}$ denote the symbol vector corresponding to $\mathbf{c}^{[a]}$. Assume that bit c_m is mapped to symbol x_j. Also, let $\mathbf{x}_{n-L:n} = (x_{n-L}, x_{n-L+1}, \cdots, x_n)$. Let $\tilde{\mathbf{h}}^{(i,k)}$ denote the k-th column of $\tilde{\mathbf{H}}_l^{(i)}$. For each $a = 0, 1$, we obtain

$$\beta_m^{(i)} = \ln \frac{\prod_{k=1}^{K} p(\mathbf{y}^{(k)} | \mathbf{x}^{[0]}, \tilde{\mathbf{h}}^{(i,k)}) P(\mathbf{x}^{[0]})}{\prod_{k=1}^{K} p(\mathbf{y}^{(k)} | \mathbf{x}^{[1]}, \tilde{\mathbf{h}}^{(i,k)}) P(\mathbf{x}^{[1]})} = \ln \frac{\prod_{k=1}^{K} \prod_{n=j}^{j+L} p(y_n^{(k)} | \mathbf{x}_{n-L:n}^{[0]}, \tilde{\mathbf{h}}^{(i,k)})}{\prod_{k=1}^{K} \prod_{n=j}^{j+L} p(y_n^{(k)} | \mathbf{x}_{n-L:n}^{[1]}, \tilde{\mathbf{h}}^{(i,k)})} + \ln \frac{P(c_m = 0)}{P(c_m = 1)}. \tag{4.12}$$

In (4.12), the subscripts l in $\mathbf{y}^{(k)}$ and $\tilde{\mathbf{h}}^{(i,k)}$ are suppressed for brevity.

In the following, we explain how to evaluate the probability density function in (4.12). The key point is to compute an effective noise variance $\tilde{\sigma}_k^2$ which takes into account both channel noise and channel estimation error. Detailed derivations are given as follows. We assume that the channel is approximately constant within segment l. Thus, even the actual channel is time-varying, the p-th tap of the estimated

CIR at iteration i for channel k, denoted by $\tilde{h}_p^{(i,k)}$, does not depend on the time index n. We then have

$$h_{n,p}^{(k)} = \tilde{h}_p^{(i,k)} + e_{n,p}^{(i,k)}, \tag{4.13}$$

where $e_{n,p}^{(i,k)}$ is the channel estimation error. Accordingly, the received signal can be written as

$$y_n^{(k)} = \sum_{p=0}^{L} h_{n,p}^{(k)} x_{n-p} + w_n^{(k)} = \sum_{p=0}^{L} \tilde{h}_p^{(i,k)} x_{n-p} + w_n^{(k)} + \sum_{p=0}^{L} e_{n,p}^{(i,k)} x_{n-p}. \tag{4.14}$$

The effective noise becomes $\tilde{w}_n^{(i,k)} = w_n^{(k)} + \sum_{p=0}^{L} e_{n,p}^{(i,k)} x_{n-p}$. We then estimate the variance of the effective noise $\tilde{\sigma}_k^2 = \mathrm{Var}(\tilde{w}_n^{(i,k)})$ from the pilot segment as

$$\tilde{\sigma}_k^2 \approx \frac{1}{N_p} \sum_{n=1}^{N_p} \left| y_n^{(k)} - \sum_{l=0}^{L} \tilde{h}_p^{(i,k)} x_{n-p} \right|^2, \tag{4.15}$$

where $x_n, n = 1, \cdots, N_p$ are pilot symbols. Assume that $\tilde{w}_n^{(i,k)}$ is a complex Gaussian random variable with zero mean and variance $\tilde{\sigma}_k^2$, we can simplify (4.12) as

$$\beta_m^{(i)} \approx \sum_{k=1}^{K} \sum_{n=j}^{j+L} \Big(-\frac{1}{\tilde{\sigma}_k^2} \Big| y_n^{(k)} - \sum_{p=0}^{L} \tilde{h}_p^{(i,k)} x_{n-p}^{[0]} \Big|^2 + \frac{1}{\tilde{\sigma}_k^2} \Big| y_n^{(k)} - \sum_{p=0}^{L} \tilde{h}_p^{(i,k)} x_{n-p}^{[1]} \Big|^2 \Big) + \gamma_m. \tag{4.16}$$

By the end of the i-th iteration of the Gibbs sampler, we obtained the LLR values for all the bits in segment l. We then run the VSLMS under the decision-directed mode, replacing the LLR values in (4.7) by $\beta_1^{(i)}, \cdots, \beta_B^{(i)}$ to obtain a sequence of soft symbol estimates $\bar{x}^{(i)}$. The VSLMS algorithm runs over $\bar{x}^{(i)}$, starting from the initial channel estimate $\hat{\mathbf{H}}_l$, to generate $\tilde{\mathbf{H}}_l^{(i)}$. This forms a pair of samples $(\bar{x}^{(i)}, \tilde{\mathbf{H}}_l^{(i)})$. Subsequently, $\tilde{\mathbf{H}}_l^{(i)}$ will be used to generate new samples in the $(i+1)$-th iteration. The operation of the Gibbs sampler are summarized in Algorithm 3.

4.3.4 Computation of LLR Based on the List

Assume that bit c_m is mapped to symbol x_j, the received signals at the k-th receive element that are affected by c_m are $\mathbf{y}_{j:j+L}^{(k)}$. Since $\mathbf{y}_{j:j+L}^{(k)}$ depends only on bits $\{c_n,\ j_1 = Q(j-L) \le n \le Q(j+L+1) - 1 = j_2\}$, we find that when computing the output LLR for c_m, it is sufficient to truncate each sequence in $\mathcal{I}$ to take into account

Algorithm 3: Gibbs sampler combined with VSLMS channel estimation

generate an initial $\mathbf{c^{(0)}}$
for $i = 1$ to I
generate $c_0^{(i)}$ from distribution
$P(c_0^{(i)} = a|c_1^{(i-1)}, c_2^{(i-1)}, \cdots, c_{B-1}^{(i-1)}, \mathbf{Y}_l, \boldsymbol{\gamma}, \tilde{\mathbf{H}}_l^{(i-1)}) \qquad a = 0, 1$
(This distribution is determined by $\beta_0^{(i)}$ using (4.12))
generate $c_1^{(i)}$ from distribution
$P(c_1^{(i)} = a|c_0^{(i)}, c_2^{(i-1)}, \cdots, c_{B-1}^{(i-1)}, \mathbf{Y}_l, \boldsymbol{\gamma}, \tilde{\mathbf{H}}_l^{(i-1)}))$
(This distribution is determined by $\beta_1^{(i)}$ using (4.12))
$\vdots$
generate $c_{B-1}^{(i)}$ from distribution
$P(c_{B-1}^{(i)} = a|c_0^{(i)}, c_1^{(i)}, \cdots, c_{B-2}^{(i)}, \mathbf{Y}_l, \boldsymbol{\gamma}, \tilde{\mathbf{H}}_l^{(i-1)}))$
(This distribution is determined by $\beta_{B-1}^{(i)}$ using (4.12))
Use LLR sequence $\{\beta_1^{(i)}, \cdots, \beta_B^{(i)}\}$ to generate soft symbol sequence $\bar{\mathbf{x}}^{(i)}$
$\tilde{\mathbf{H}}_l^{(i)} = \text{VSLMS}\{\hat{\mathbf{H}}_l, \mathbf{Y}_l, \bar{\mathbf{x}}_l^{(i)}\}$ (Run the VSLMS algorithm under
decision-directed mode using $\mathbf{Y}_l, \bar{\mathbf{x}}_l^{(i)}$ and initial channel estimation $\hat{\mathbf{H}}_l$)
Put $(\mathbf{c_l}, \hat{\mathbf{H}}_\mathbf{l}^{(\mathbf{i})})$ into the list.
end for

only bits $\{c_n,\ j_1 \le n \le j_2\}$. We denote the set that contains the truncated sequences by $\mathcal{I}_{j_1:j_2}$. For each $0 \le m \le B-1$, we construct a larger set $\mathcal{I}_{j_1:j_2}^m$ which includes all sequences in $\mathcal{I}_{j_1:j_2}$, together with new sequences that are obtained by flipping the m-th bit of each sequence in $\mathcal{I}_{j_1:j_2}$. Duplicate sequences are removed from $\mathcal{I}_{j_1:j_2}^m$. Furthermore, we let $\mathcal{I}_{j_1:j_2}^{m,0}$ and $\mathcal{I}_{j_1:j_2}^{m,1}$ denote sequences in $\mathcal{I}_{j_1:j_2}^m$ whose m-th bit equals 0 and 1, respectively. The LLR value for bit c_m is then computed as

$$\alpha_m = \ln \frac{\sum\limits_{\mathbf{c}_{j_1:j_2}^{(i)} \in \mathcal{I}_{j_1:j_2}^{m,0}} \prod\limits_{k=1}^{K} p\big(\mathbf{y}_{j:j+L}^{(k)} | \mathbf{c}_{j_1:j_2}^{(i)}, \tilde{\mathbf{H}}_l^{(i)}\big) \prod\limits_{n=j_1}^{j_2} P\big(c_n|\gamma_n\big)}{\sum\limits_{\mathbf{c}_{j_1:j_2}^{(i)} \in \mathcal{I}_{j_1:j_2}^{m,1}} \prod\limits_{k=1}^{K} p(\mathbf{y}_{j:j+L}^{(k)} | \mathbf{c}_{j_1:j_2}^{(i)}, \tilde{\mathbf{H}}_l^{(i)}) \prod\limits_{n=j_1}^{j_2} P(c_n|\gamma_n)} - \gamma_m. \tag{4.17}$$

4.3.5 Bidirectional Channel Initialization

When processing the l-th segment, the initial channel estimate $\hat{\mathbf{H}}_l$ is set to be $\hat{\mathbf{H}}_l = \frac{\hat{\mathbf{H}}_{l-1}^f + \hat{\mathbf{H}}_{l+1}^b}{2}$, where $\hat{\mathbf{H}}_{l-1}^f$ is the updated channel estimate at the output of the MCMC-LCE for the $(l-1)$-th segment. Detailed derivations $\hat{\mathbf{H}}_{l-1}^f$ of can be found in Section 4.3. The backward channel estimate $\hat{\mathbf{H}}_{l+1}^b$ is obtained by running the VSLMS

algorithm [79] from the neighboring pilot segment to the immediate right of segment l all the way to the left towards the l-th segment. For the data segment, the VSLMS operates under decision-directed mode following techniques described in 4.2.2, where the prior LLR in (4.7) is replaced by $\boldsymbol{\alpha}$, the feedback from channel decoder. Note that for the first iteration of joint detection and decoding, since $\boldsymbol{\alpha}$ is not available, we simply set $\hat{\mathbf{H}}_l = \hat{\mathbf{H}}_{l-1}^f$. Due to the time variation of the channel, we find that the bidirectional channel initialization improves performance compared to the one-sided channel initialization using only $\hat{\mathbf{H}}_{l-1}^f$.

4.3.6 Multiple Runs of MCMC-LCE for the Same Segment

After obtaining $\hat{\mathbf{H}}_l^f$, before moving on to segment $l+1$ directly, we have empirically found that it is often helpful to re-run the MCMC-LCE for the same segment l with $\hat{\mathbf{H}}_l^f$ replacing the initial channel estimate $\hat{\mathbf{H}}_l$. Intuitively, this improves detection performance because in general $\hat{\mathbf{H}}_l^f$ should be more accurate than $\hat{\mathbf{H}}_l$. We note that the advantage of having multiple runs of MCMC-LCE over the same segment is more pronounced as the Doppler rate of the channel increases, because in such scenarios the initial channel estimate $\hat{\mathbf{H}}_l$ is often poor. However, we observe that one should not re-run the MCMC-LCE for too many iterations over the same segment, due to the correlation between the estimated CIR and the data sequences.

4.4 Numerical Results

In this section, we provide performance comparisons of the MCMC-LCE detector with the MMSE detector. For all channels considered, we let $G = I = 10$ for MCMC-LCE. Larger values of G or I do not yield noticeable performance improvement. For MMSE, we test both block-wise channel update and symbol-wise update and present results for the superior version. For channel estimation, a good choice of the step-size parameter μ of LMS algorithm in (4.2) was empirically found to be 0.025 and this value was used for all the results presented here. The parameters of VSLMS are chosen to optimize its performance in this scenario.

4.4.1 Synthetic Channels

First, we consider a synthetic, sparse time-varying ISI channel. The synthetic channel is generated as follows. We assume that there is a total of 12 channel taps ($L = 11$), among which the 1st, 5-th, and 10-th taps are nonzero. Each nonzero tap has an average power of 1/3. The 1st and 5-th taps are constant over each codeword with a fixed magnitude of $\sqrt{1/3}$ and a uniformly generated phase. The 10-th tap is time-varying. It follows a Rayleigh distribution and is generated using the Jakes' model with normalized Doppler rates of $f_dT = 0.004$ or $f_dT = 0.008$.

Here, we use QPSK modulation. A rate 1/2 recursive convolutional code with generator polynomials of $(131, 171)_8$ is used. Each packet contains a total of 14400 coded bits, which spans over 48 frames. The segment size is 50 symbols. Each frame consists of one pilot segment followed by three data segments, i.e., 50 pilot symbols followed by 150 data symbols. We simulate multiple data packets, and the channel is generated independently from packet to packet. Figure 4.4 presents the average bit-error-rate (BER) of MCMC-LCE with VSLMS, MCMC-LCE with LMS, and MMSE with LMS as a function of signal-to-noise ratio (SNR). For $f_dT = 0.004$, we rerun each MCMC-LCE detector 2 times for each iteration of joint detection and decoding. For $f_dT = 0.008$, we increase the number of reruns to 4 to optimize performance. We observe that MCMC-LCE with VSLMS achieves the best performance for both Doppler rates. For $f_dT = 0.004$, the MCMC-LCE with VSLMS is about 2 dB better than MCMC-LCE with LMS, and is about 3 dB better than MMSE with LMS. The performance gap increases significantly for $f_dT = 0.008$. It is shown that the MCMC-LCE with VSLMS is about 3 dB better than MCMC-LCE with LMS, and is more than 6 dB better than MMSE with LMS. This suggests that the VSLMS is superior to LMS in tracking the channel for the Doppler rates considered. Furthermore, the advantage of the list-based joint data detection and channel estimation of the MCMC detector becomes more pronounced with increasing Doppler rate, as shown in the larger gap of the MCMC over MMSE for $f_dT = 0.008$.

In Figure 4.5, we plot the bit error rate over turbo iterations for MCMC-LCE with the normalized doppler to be $f_dT = 0.004$. The figure shows that the curves mostly turn flat after four or five iterations. However, for channels with high doppler,

more number of iterations are required for the BER performance to converge. In Figure 4.6 we examine the convergence behavior of MCMC-LCE at a doppler of $f_dT = 0.008$. In this case, if the convergence can be reached, for instance at SNR = 4dB, at least six iterations are needed to obtain the best performance. We also did simulations for other doppler frequencies and confirmed that transmission can hardly succeed with higher doppler rates. Therefore, in the following simulations, we apply seven turbo iterations which should be sufficient in order to achieve convergence. The figures also indicates that part of the performance gain comes from the first iteration, when MCMC-LCE is capable of making better decisions in the first iteration, which subsequently resulting in better performance over more iterations.

4.4.2 Experimental Data from Underwater Acoustic Channel

Next, we examine the performance of various detectors using data collected from an at-sea underwater experiment, conducted off the coast of Martha's Vineyard, MA during Oct. 14th - Nov. 2nd, 2008. We processed the received data from both 60 and 200 meters of transmission. The 16QAM modulation is used. For 60-meter transmission, one acoustic transducer and $K = 4$ receive hydrophones are used; for 200-meter transmission, one acoustic transducer and $K = 2$ receive hydrophones are used. The data was modulated using a carrier frequency of 13 kHz, and a symbol rate 9.77k sym/sec. We compare the performance of different detectors over multiple data files, which are transmitted 2 hours apart. Due to the change in weather conditions, wind speed, wave height and other environmental factors during the course of the data transmission, the channel conditions corresponding to each file can vary significantly, resulting in variable BER performance for different files. We examine a total of 23 data files, with index from 1 to 23, corresponding to epoch 8 to 30 out of 149 total epoches in the experiment [87]. The environmental data are exhibited in Table 4.1. The table indicates that, data file 11-13 endure tougher weather than other files. As we will confirm in the following simulations, weather conditions have significant impact on the UWA transmission quality.

Some transmission parameters are listed in Table 4.2. The transmission format is as follows. Each file contains seven independently coded packets. Each packet

contains a total of 28800 coded bits, which spans six frames. The segment size is 400 symbols. Each frame consists of one pilot segment followed by three data segments, i.e., 400 pilot symbols followed by 1200 data symbols.

In Figure 4.7, we present the BER of MCMC-LCE with VSLMS, and MMSE with LMS after seven iterations of data detection and decoding. We rerun the MCMC-LCE detector 2 times within each iteration. The x-axis is the index of the data file, and the y-axis is the BER of each data file. As shown in Figure 4.7 (a), the MCMC-LCE detector dramatically outperforms the MMSE detector for files 11-13, for which the MMSE has high BER, and the MCMC-LCE has low BERs of below 5×10^{-4}. Furthermore, in order to show the performance difference of the two detectors for other data files with low BER, a zoom-in figure is provided in Figure 4.7 (b) where in the y-axis the largest value of the BER is set to be 2×10^{-3}. As shown in Figure 4.7 (b), for those data files that both detectors perform well (with a BER below 2×10^{-3}), MCMC-LCE performs better than the MMSE detector for files 1 2, 9, 15, and is only slightly worse than the MMSE for files 7,10. The BER values are listed in Table 4.3.

To provide some insight into such performance difference, the estimated CIR of data file 12, where the MCMC-LCE significantly outperforms MMSE, and the estimated CIR of data file 1, where both MCMC-LCE and CR-MMSE perform equally well, are presented in Figure 4.5. It can be seen that the CIR of data file 12 varies faster than that of data file 1, and the energy of the CIR is also more spread out in time. The advantage of the MCMC-LCE becomes more pronounced for such a channel because the list channel estimation provides superior channel tracking capability compared to the MCMC, in which channel estimation is performed based on a single estimated data sequence. In Figure 4.9, we plot the average mean square error (MSE) of the estimated CIR with the genie-aided CIR as a function of each iteration of joint detection and channel decoding. Here , the y-axis is the channel estimate misalignment $E(|\hat{h}_l^k - \bar{h}_l^k|^2)/E(|\bar{h}_l^k|^2)$, where $\bar{h}_l^k$ is the genie-aided channel estimation for the l-th segment and the k-th channel, obtained by running the VSLMS over the perfectly known data sequence. The expectation is over all data segments and all channels for data file 12. Note that within each each iteration, we run the MCMC-LCE twice to improve performance. It is shown that, for data set 12, within a

few turbo iterations, the MCMC-LCE with VSLMS with achieves a MSE of less than 7%. In comparison, for the MMSE detector, the MSE of the estimated CIR remains as high as 38% after several iterations. The MCMC-LCE with LMS achieves a MSE of about 15%, which is slightly worse than that of MCMC-LCE with VSLMS, but is much better than the MMSE with LMS. This is in agreement with results shown in Figure 4.7 that the BER of MMSE is very high (about 23%), while the MCMC-LCE with VSLMS achieves almost zero error. The performance of MCMC-LCE with LMS is only slightly worse than that of MCMC-LCE with VSLMS (not shown in the Figure), but still much better than the MMSE.

In Figure 4.10, for data set 12, we plot the magnitude of the average step size of the VSLMS for each channel tap. We observe that the step sizes are larger around tap 5, tap 45 and tap 60. These correspond to taps that have larger time variations (larger Doppler), as shown in Figure 4.5. For the taps with slow/no variation, the step-size has approached a minimum value, set in the implementation of the VSLMS algorithm.

In Figure 4.11, the bit error rate performance is plotted with the increase of turbo iterations for the 60 meter transmission. Comparing to Figure 4.5 and Figure 4.6 for synthetic channels, an interesting observation is that the two MCMC detectors shows similar performance at the first iteration, and the gap enlarges over multiple turbo iterations. This indicates that MCMC-LCE has better error correction capability over iterations than SEM-MCMC when it starts out with similar erroneous detection results.

In Figure 4.12, we plot the BER of the MCMC-LCE and MMSE detectors after seven iterations of data detection and decoding for a 200 meter transmission. Similar to that of the 60 meter transmission, MCMC-LCE performs significantly better than the MMSE detector. As shown in Figure 4.12 (a), MCMC-LCE reduces the BER substantially for data files 7, 11,12, for which the MMSE has high BER. It is also observed from Figure 4.12 (b) that the MCMC-LCE performs better than the MMSE for most of the other files where the BERs of both detectors fall below 2×10^{-3}. The BER values are listed in Table 4.4.

In Figure 4.13, the bit error rate performance is plotted with the increase of turbo iterations for the 200 meter transmission. The setting is the same as in Figure 4.12. The figure clearly shows that the two MCMC detectors have the same larger diversity gain than the MMSE detector. After seven iterations, MCMC-LCE gives an order of magnitude gain over the MMSE detector.

4.5 Conclusion

In this work, we developed a new MCMC-LCE detector for iterative channel estimation and data detection over time-varying ISI channels. The approach taken in this paper operates based on a list of data and channel estimates, and is driven by a VSLMS algorithm for channel estimation. These lead to improved channel tracking and data detection performance over existing approaches. The MCMC-LCE yields superior performance to state-of-art turbo MMSE equalizers for both synthetic channels and UWA channels. The proposed approach differs from existing works in the literature that are largely trellis based, or are based on linear filtering. Hence, this represents a new venue for turbo equalization through statistical data detection and channel estimation. In particular, the approach of data detection based on a list of channel estimates enables reliable channel tracking for fast time-varying channels.

Table 4.1. Experiment environments and weather conditions

File index	Wave height (m)	Wave period (sec)	Wind speed (m/s)
1	0.63	6.97	1.10
2	0.70	7.00	0.87
3	0.57	6.80	1.40
4	0.57	7.20	1.20
5	0.60	7.53	1.30
6	0.67	7.93	4.10
7	0.63	6.37	5.70
8	0.70	5.00	6.37
9	0.77	4.37	7.47
10	0.93	4.27	8.37
11	**1.17**	**4.17**	**9.07**
12	**1.10**	4.30	5.77
13	**1.00**	4.37	3.50
14	0.87	4.90	6.23
15	0.63	5.87	5.63
16	0.60	6.13	6.00
17	0.50	6.60	4.10
18	0.47	8.00	3.90
19	0.43	8.13	6.33
20	0.40	8.57	6.50
21	0.47	6.27	7.40
22	0.43	5.47	6.90
23	0.40	7.37	4.70

Table 4.2. Parameter setting for the SPACE08 experiment

Carrier frequency	13 KHz
Sampling rate	39 KHz
Symbol rate	9.77 KHz
Symbol interval	0.1 ms
Frame duration	0.16 s
Packet duration	0.98 s
No. pilots/frame	400
No. data/frame	1200
No. frame/codeword	6
Codeword length	28800

Table 4.3. BER of 60 meter transmission over UWA channels, 16QAM modulation, one acoustic transmit transducer and four receive hydrophones. seven turbo iterations.

File index	MMSE	SEM-MCMC	MCMC-LCE
1	1.7×10^{-4}	0	0
2	7×10^{-4}	2.9×10^{-4}	2.8×10^{-4}
$3 \sim 6$	0	0	0
7	4.9×10^{-5}	1.1×10^{-4}	1.3×10^{-4}
8	0	0	0
9	7.9×10^{-5}	0	0
10	2.9×10^{-5}	2.7×10^{-4}	2.2×10^{-4}
11	**0.12**	7×10^{-4}	4.8×10^{-4}
12	**0.23**	1.8×10^{-2}	2.9×10^{-4}
13	4.9×10^{-2}	0	0
14	0	0	0
15	1.3×10^{-3}	3.0×10^{-5}	0
$16 \sim 23$	0	0	0

Table 4.4. BER of 200 meter transmission over UWA channels, 16QAM modulation, one acoustic transmit transducer and two receive hydrophones. seven turbo iterations.

File index	**MMSE**	**SEM-MCMC**	**MCMC-LCE**
1	3.0×10^{-5}	3.0×10^{-5}	0
$2 \sim 3$	0	0	0
4	3.0×10^{-4}	3.5×10^{-4}	1.8×10^{-4}
5	5.0×10^{-5}	2.0×10^{-4}	4.0×10^{-5}
6	2.8×10^{-4}	6.0×10^{-4}	2.2×10^{-4}
7	1.4×10^{-2}	2.3×10^{-3}	5.4×10^{-3}
8	3.8×10^{-3}	5.7×10^{-4}	7.0×10^{-4}
9	1.3×10^{-4}	0	3.1×10^{-4}
10	6.0×10^{-4}	2.3×10^{-4}	4.7×10^{-4}
11	4.7×10^{-2}	2.1×10^{-4}	1.3×10^{-3}
12	4.7×10^{-2}	7.9×10^{-5}	6.0×10^{-4}
13	9.6×10^{-4}	3.0×10^{-4}	4.2×10^{-4}
14	8.9×10^{-5}	6.0×10^{-5}	1.1×10^{-4}
15	9.9×10^{-5}	3.6×10^{-4}	1.3×10^{-4}
16	1.7×10^{-4}	2.0×10^{-4}	1.1×10^{-4}
17	0	0	0
18	1.4×10^{-4}	2.4×10^{-4}	2.5×10^{-4}
19	2.2×10^{-3}	2.2×10^{-3}	3.1×10^{-3}
20	4.8×10^{-4}	4.3×10^{-4}	3.1×10^{-4}
21	6.9×10^{-5}	0	3.0×10^{-5}
22	1.3×10^{-4}	1.1×10^{-4}	1.3×10^{-4}
23	0	0	0

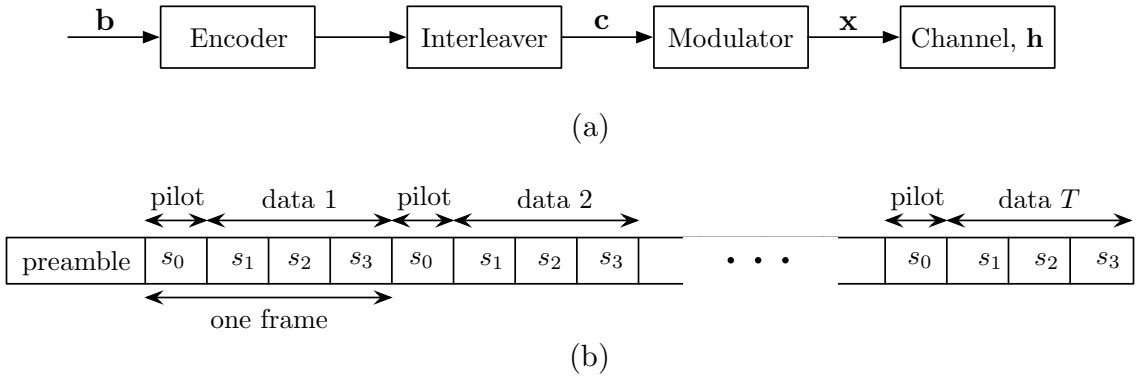


Figure 4.1. Illustration of the packet transmission structure. (a) Transmitter block diagram. (b) Packet format. A packet consists of T frames. Each frame is divided into one pilot segment and several data segments. In this example, each frame contains a pilot segment s_0 and three data segments s_1, s_2, s_3.

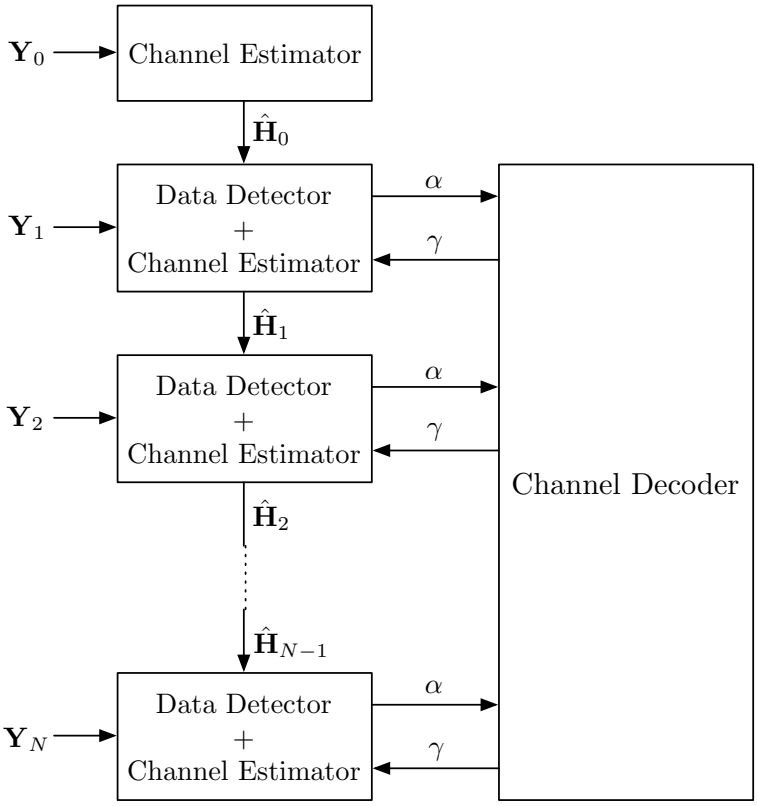


Figure 4.2. Receiver flow diagram for MCMC-LCE.

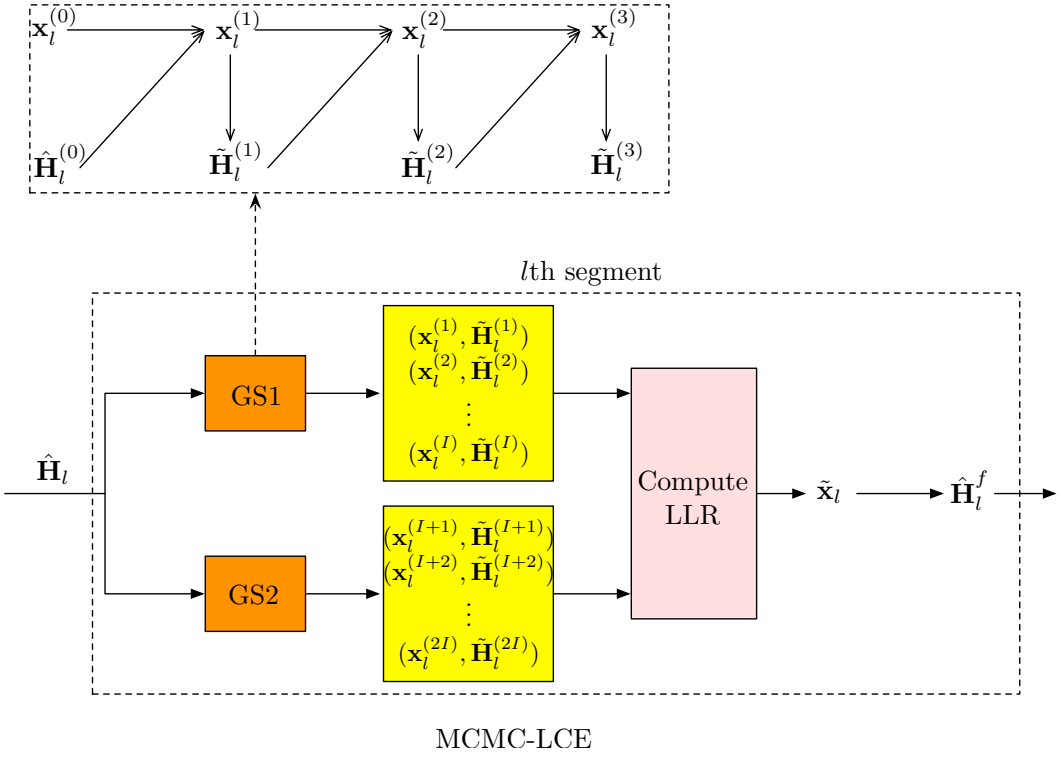


Figure 4.3. Illustration of the MCMC-LCE detector

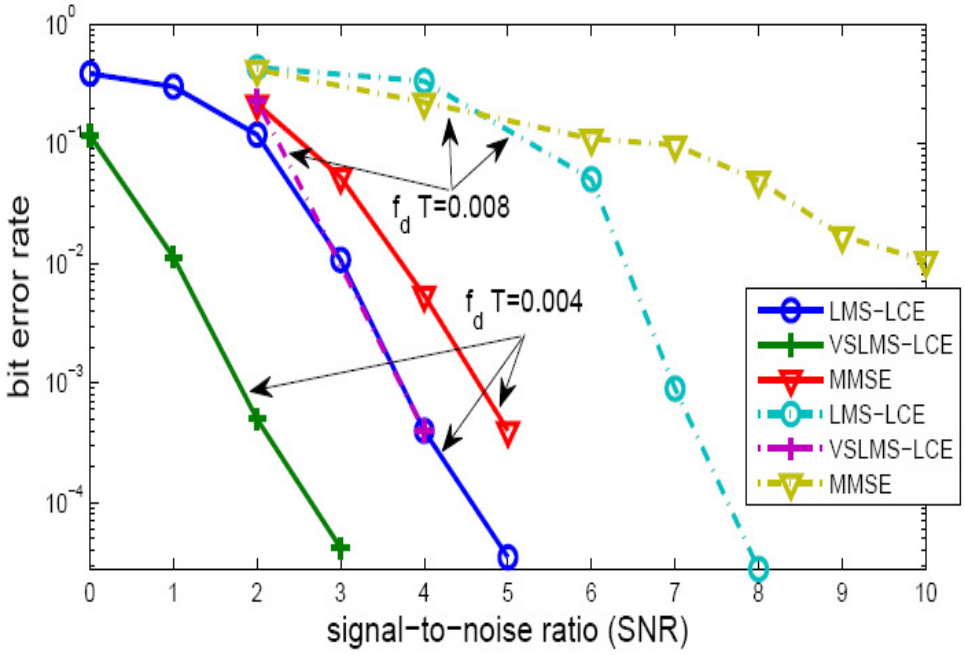


Figure 4.4. Comparisons of MCMC-LCE and MMSE detectors over a synthetic Rayleigh ISI channel with $L = 11$.

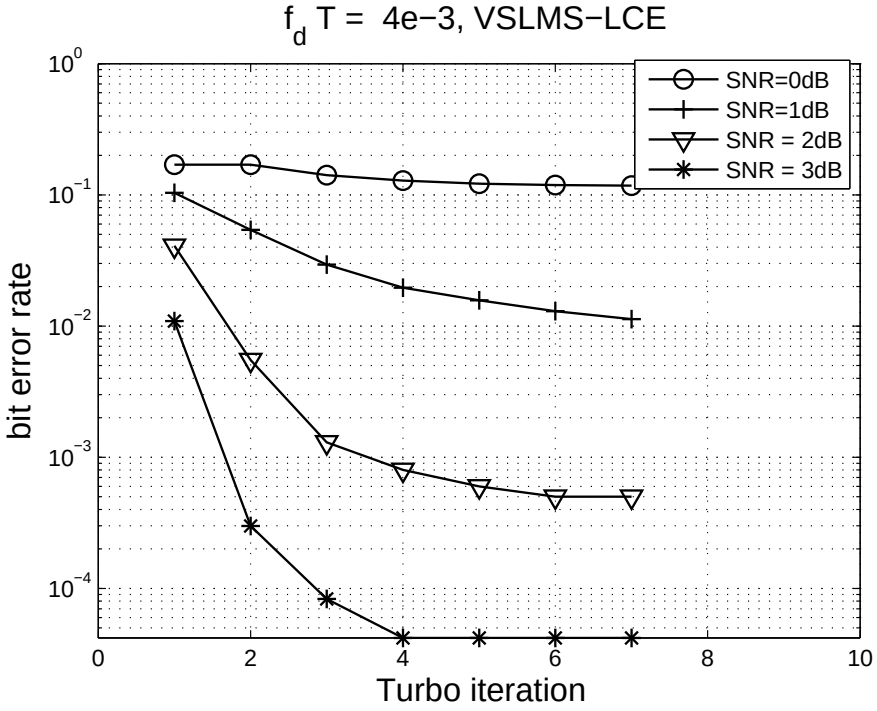


Figure 4.5. Convergence analysis on VSLMS-LCE at different SNR points, with $f_d T = 4 \times 10^{-3}$.

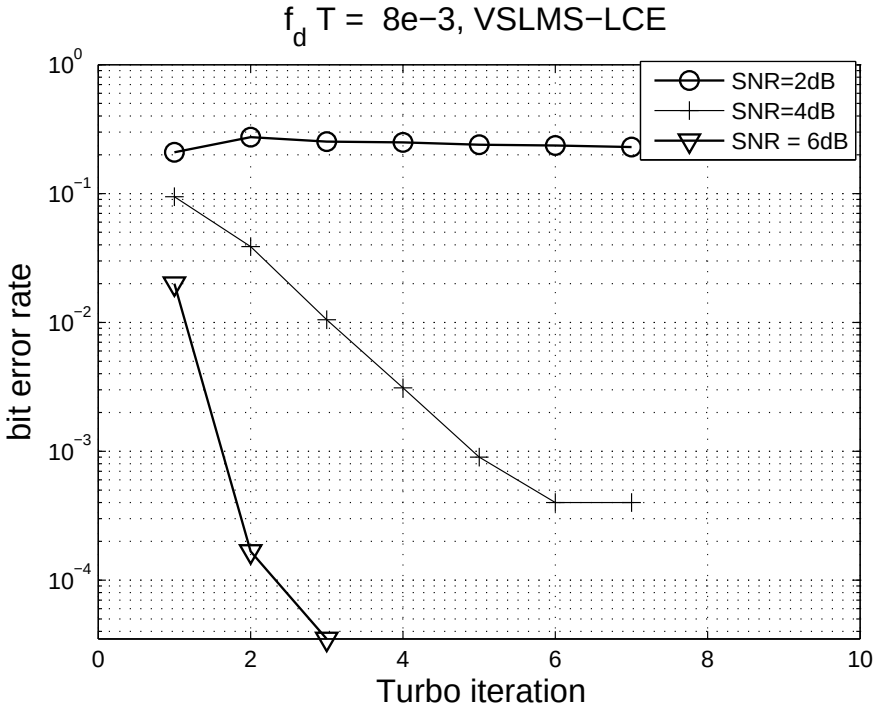
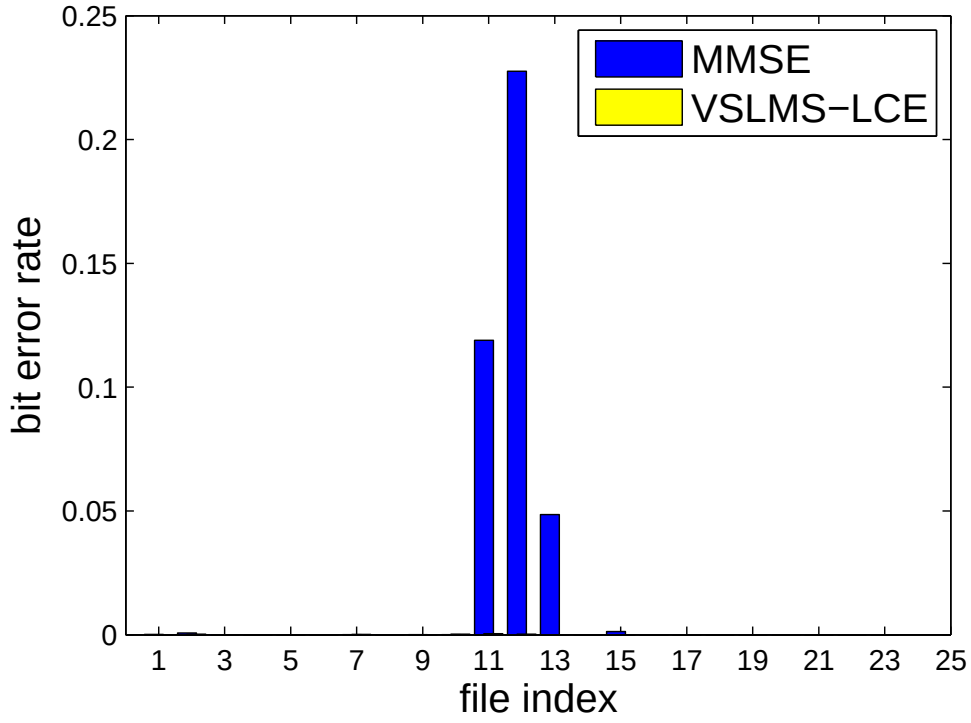
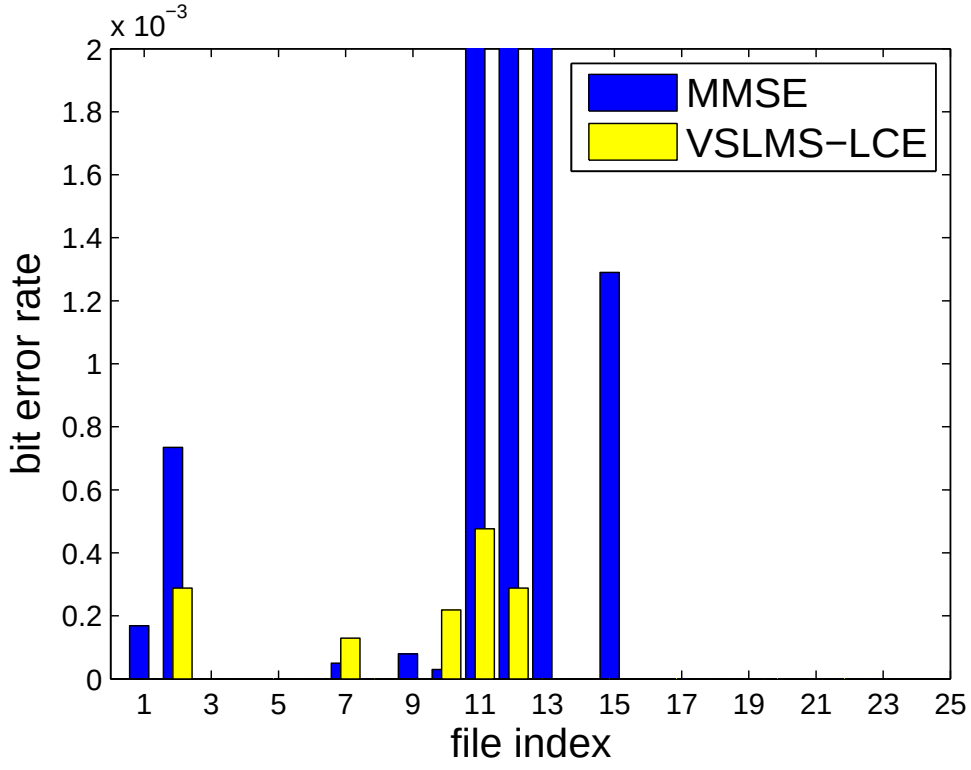


Figure 4.6. Convergence analysis on VSLMS-LCE at different SNR points, with $f_d T = 8 \times 10^{-3}$.

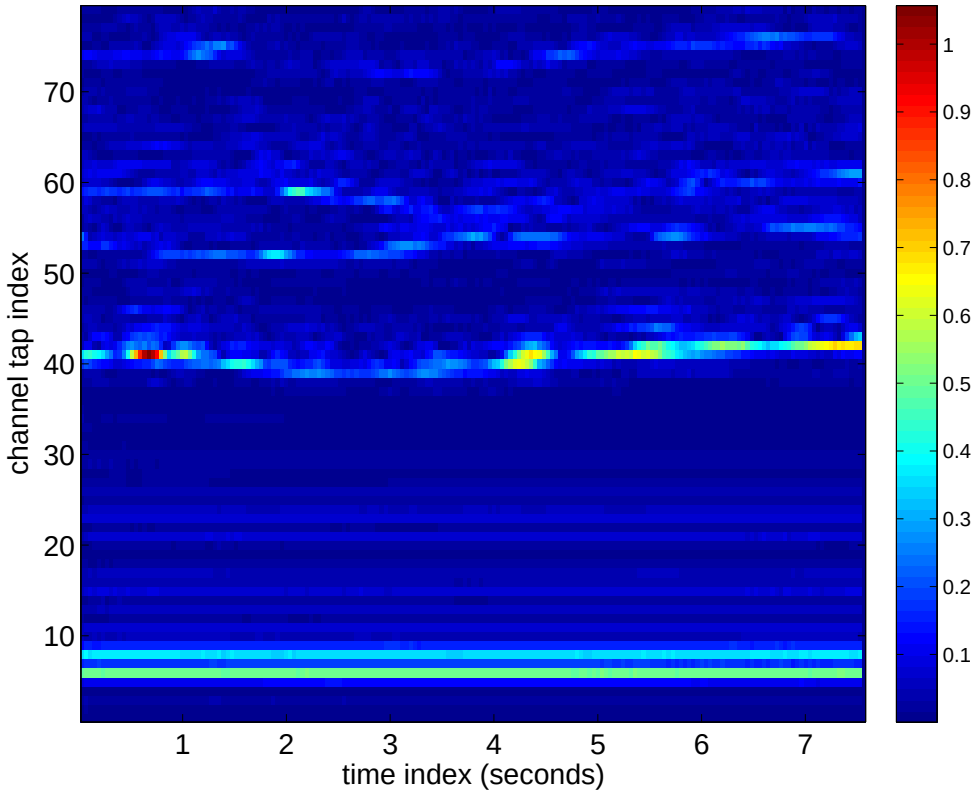


(a) BER of all data files

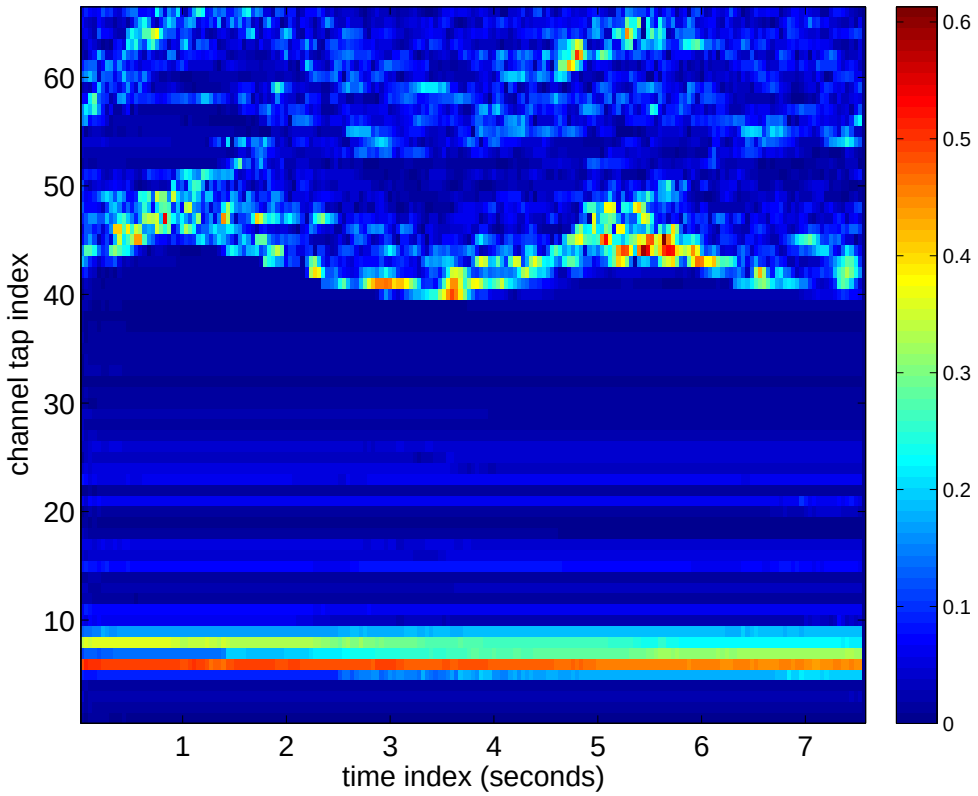


(b) Zoom-in figure for the lower part of Figure 4.7 (a) for BER below 2×10^{-3}

Figure 4.7. BER of 60 meter transmission over a UWA channel, 16QAM modulation, one acoustic transmit transducer and four receive hydrophones. Each data file contains 7 packets.



(a) Data set 1 under mild channel condition



(b) Data set 12 under more difficult channel condition

Figure 4.8. Estimated channel impulse response for the 60 meter distance UWA channel.

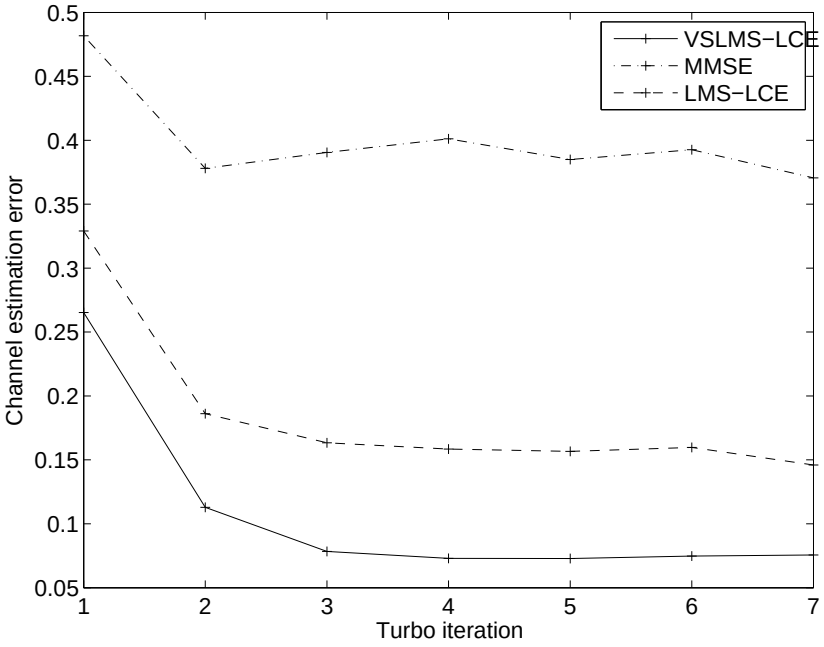


Figure 4.9. Quality of channel estimation for data set 12

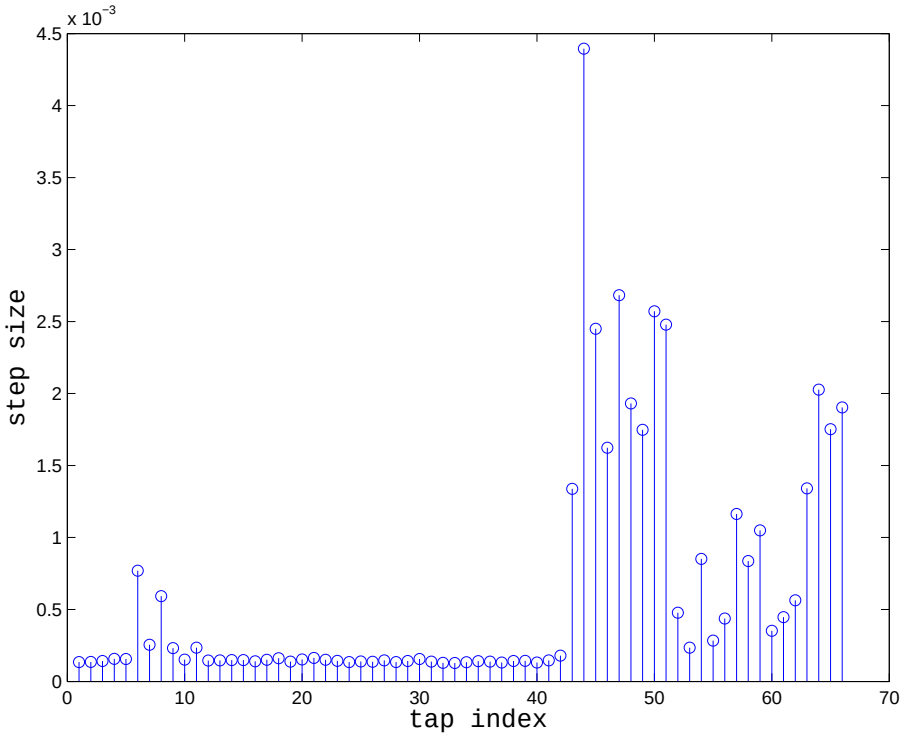


Figure 4.10. Step sizes for data set 12.

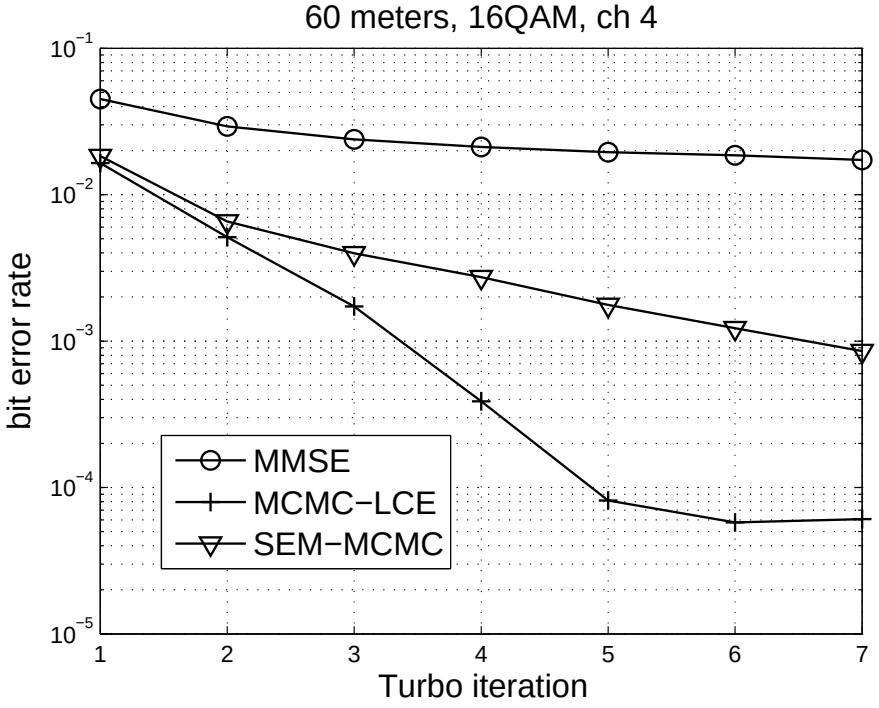
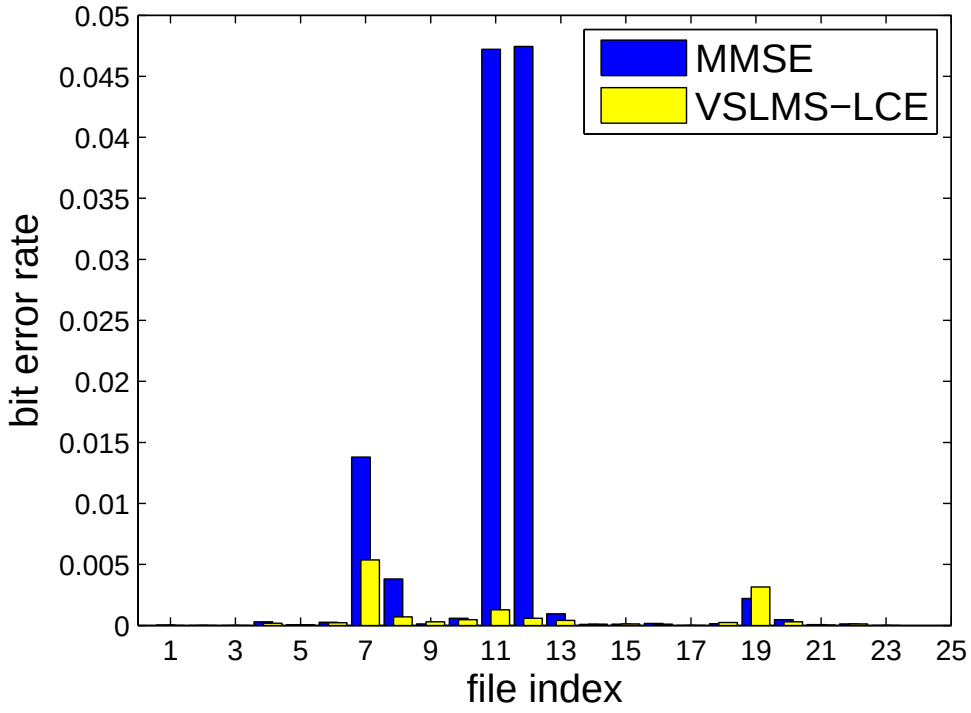
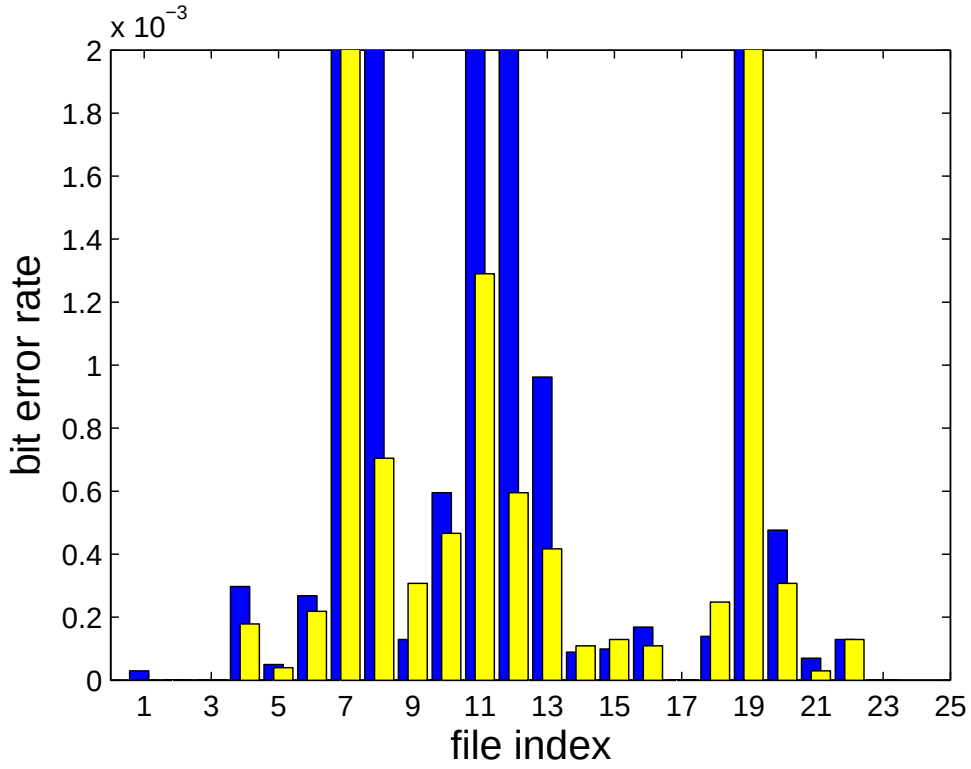


Figure 4.11. Bit error rate versus the number of turbo iterations. 60 meter transmission over a UWA channel, 16QAM modulation, one transducer and four receive hydrophones, seven iterations of joint data detection and channel decoding.



(a) BER of all data files



(b) Zoom-in figure for the lower part of Figure 4.12 (a) for BER below 2×10^{-3}

Figure 4.12. BER of 200 meter transmission over a UWA channel, 16QAM modulation, one acoustic transmit transducer and two receive hydrophones. Each data file contains 7 packets.

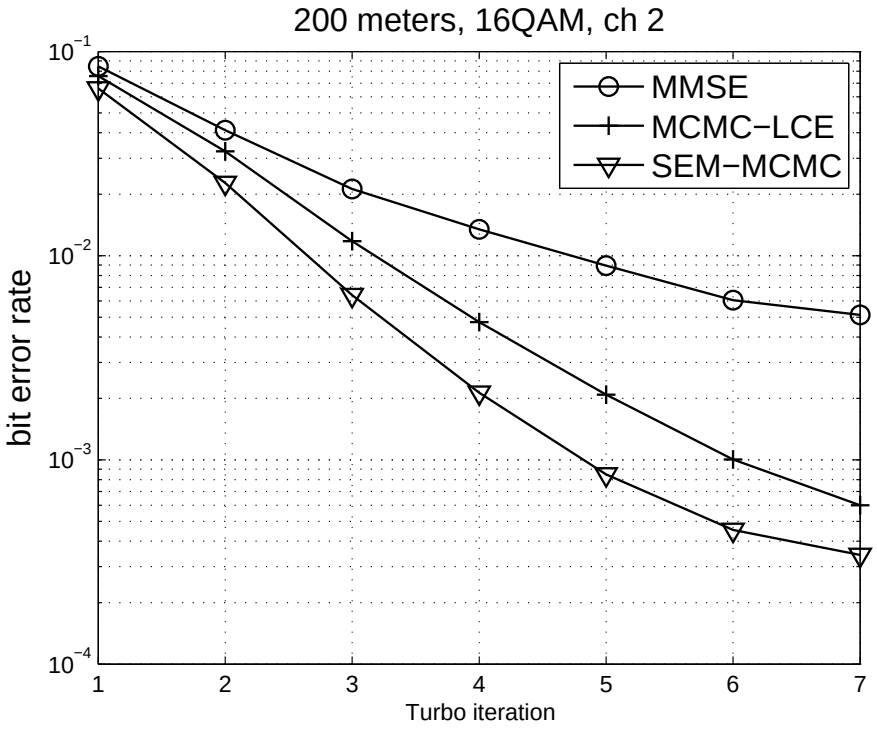


Figure 4.13. Bit error rate versus the number of turbo iterations. 200 meter transmission over a UWA channel, 16QAM modulation, one transducer and four receive hydrophones, seven iterations of joint data detection and channel decoding.

CHAPTER 5

CONCLUSIONS AND FUTURE WORK

In this thesis, we have conducted information-theoretical analysis of the capacity and signaling design for the MIMO MAC channels under one of the most general and practical channel models – the virtual representation model. The analysis developed in this work is without the common simplifying assumption of Gaussian statistics and Kronecker correlation for the channel matrix elements. We have provided practical guidelines for signaling design based on beamforming, which requires only knowledge of channel statistics at the transmitter. This effectively reduces the amount of channel information that needs to be fed back to the transmitter for optimal signaling and suggests that beamforming would be more robust to channel estimation errors at the receiver. It is of interest to further study this robustness and compare optimal signaling and beamforming in the presence of estimation errors. We also note that the virtual representation model studied in the thesis is restricted to uniform linear arrays of antennas at the transmitter and receiver. Extensions of the capacity analysis and signaling design to arbitrary array geometries is of great interests.

A main contribution of this thesis is the development and validation of low-complexity statistical detection methods for frequency-selective channels with inter-symbol interference to approach the performance of the optimal maximum a posteriori probability detection. The proposed MCMC detectors are high performance, low-complexity detectors that can significantly outperform state-of-the-art methods, particularly, under the most challenging operating environments of UWA channels. Our work demonstrates that MCMC detectors have the potential to revolutionize receiver design in current UWA modems and achieve near-optimal detection that would otherwise be infeasible using traditional, deterministic detection algorithms. Directions for future work include further exploration of MCMC detectors for highly

dynamic environments, in which joint MCMC detection, doppler compensation, and adaptive channel tracking algorithms need to be developed jointly to further improve receiver performance. Our preliminary results using high mobility data collected from at-sea experiments have shown great promise along this line of research. While in this thesis we have restricted our attention to single-carrier systems, in future work we plan to extend our design to multi-carrier transmissions. It is expected that MCMC detection will play an important role in improving the transmission range, rate, and quality of underwater communications.

REFERENCES

[1] A. M.Tulino, A. Lozano, and S. Verdu, "Impact of antenna correlation on the capacity of multiantenna channels," *IEEE Trans. on Information Theory*, vol. 51, pp. 2491–2509, July 2005.

[2] H. Wan, R.-R. Chen, J. W. Choi, A. Singer, J. Preisig, and B. Farhang-Boroujeny, "Stochastic expectation maximization algorithm for long-memory fast-fading channels," in *Proc. IEEE Global Communication Conference (Globecom), 2010*, 2010.

[3] E. Telatar, "Capacity of multi-antenna gaussian channels," *European Trans. Telecomm.*, vol. 10, no. 6, pp. 585–595, Nov.-Dec. 1999.

[4] S. Jafar and A. Goldsmith, "Transmitter optimization and optimality of beamforming for multiple antenna systems," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 1165–1175, July 2004.

[5] Z. Wang and G. B. Giannakis, "Outage mutual information rate of space-time MIMO channels," *IEEE Trans. on Information Theory*, vol. 50, no. 4, pp. 657–662, Sep. 2001.

[6] V. V. Veeravalli, Y. Liang, and A. M. Sayeed, "Correlated mimo wireless channels: Capacity, optimal signaling, and asymptotics," *IEEE Trans. on Information Theory*, vol. 51, pp. 2058–2072, 2005.

[7] A. M. Tulino, A. Lozano, and S. Verdu, "Impact of antenna correlation on the capacity of multiantenna channels," *IEEE Trans. Information Theory*, vol. 51, pp. 2491–2509, July 2005.

[8] E. Jorswieck and H. Boche, "Channel capacity and capacity range of beamforming in mimo wireless systems under correlated fading with covariance feedback," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 1543–1553, Sep. 2004.

[9] W. Rhee, W. Yu, and J. Cioffi, "The optimality of beam-forming in uplink multiuser wireless systems," *IEEE Trans. Wireless Comm.*, vol. 3, no. 1, pp. 86–96, Jan. 2004.

[10] S. Jafar and A. Goldsmith, "Transmitter optimization and optimality of beamforming for multiple antenna systems," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 1165–1175, July 2004.

[11] H. Shin and J. H. Lee, "Capacity of multiple-antenna fading channels: Spatial fading correlation, double scattering, and keyhole," *IEEE Trans. on Info. Theory*, vol. 49, pp. 2636–2647, Oct. 2003.

[12] A. Soysal and S. Ulukus, "Optimality of beamforming in fading multiple access channels," *IEEE Trans. on Communications*, pp. 1171–1183, April, 2009.

[13] ——, "Optimum power allocation for single-user mimo and multi-user mimo-mac with partial csi," *IEEE Trans. on Selected Areas. in Communications*, no. 7, pp. 1402–1412, Sep. 2007.

[14] V. V. Veeravalli, Y. Liang, and A. M. Sayeed, "Correlated mimo wireless channels: Capacity, optimal signaling, and asymptotics," *IEEE Transactions on information theory*, vol. 51, pp. 2058–2072, 2005.

[15] A. M. Sayeed, "Deconstructing multi-antenna fading channels," *IEEE Trans. Signal Processing*, pp. 2563–2579, Oct. 2002.

[16] H. Wan, R.-R. Chan, and Y. Liang, "Optimality of beamforming for mimo multiple access channels via virtual representation," *IEEE Trans. Signal Processing*, vol. 58, no. 10, pp. 5458–5462, Oct. 2010.

[17] D. J. Love, R. W. Heath, and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Trans. on Information Theory*, vol. 49, pp. 2735–2747, Oct. 2003.

[18] D. J. Love, R. W. H. Jr., and T. Strohmer, "Limited feedback precoding for orthogonal space-time block codes," *IEEE Global Telecom. Conf.*, Dec.2004.

[19] P. Xia and G. B. Giannakis, "Design and analysis of transmit-beamforming based on limited-rate feedback," *IEEE Trans. on Signal Proc.*, vol. 54, pp. 1853–1863, May 2006.

[20] R. Chen, J. Liu, and X. Wang, "Convergence analyses and comparisons of markov chain monte carlo algorithms in digital communications," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 255–270, Feb. 2002.

[21] A. Doucet and X. Wang, "Monte carlo methods for signal processing," *IEEE Signal Processing Magazine*, vol. 22, no. 6, pp. 152–170, Nov. 2005.

[22] R. Peng, R.-R. Chen, and B. Farhang-Beroujeny, "Low complexity markov chain monte carlo detector for channels with intersymbol interference," *IEEE Trans. on Signal Processing*, vol. 58, no. 4, pp. 2206–2217, April 2010.

[23] R.-R. Chen, R. Peng, A. Ashikhmin, and B. Farhang-Beroujeny, "Approaching mimo capacity using bitwise markov chain monte carlo detection," *To appear IEEE Trans. Commun.*, Feb. 2010.

[24] R.-R. Chen, R. Peng, and B. Farhang-Beroujeny, "Markov chain monte carlo: Applications to mimo detecton and channel equalization," *Information Theory and Applications Workshop*, pp. 44–49, Feb. 2009.

[25] Z. Yang, B. Lu, and X. Wang, "Bayesian monte carlo multiuser receiver for space-time coded multicarrier cdma systems," *IEEE VTS-Fall VTC 2000*, vol. 1, pp. 429–433, Sep. 2000.

[26] B. Farhang-Boroujeny, H. Zhu, and Z. Shi, "Markov chain monte carlo algorithms for cdma and mimo communication systems," *IEEE Trans. Signal. Process.*, vol. 54, no. 5, pp. 1896–1909, May 2006.

[27] R. Peng, R.-R. Chen, and B. Farhang-Beroujeny, "Low complexity markov chain monte carlo detector for channels with intersymbol interference," *IEEE Trans. on Signal Processing*, vol. 58, no. 4, pp. 2206–2217, April 2010.

[28] I. E. Telatar, "Capacity of multi-antenna gaussian channels," *Eur. Trans. Telecom*, vol. 10, pp. 585–595, Nov. 1999.

[29] E. Jorswieck and H. Boche, "Channel capacity and capacity range of beamforming in mimo wireless systems under correlated fading with covariance feedback," *IEEE Trans. on Wireless Commu.*, vol. 3, pp. 1543–1553, Sep. 2004.

[30] A. Soysal and S. Ulukus, "Optimum power allocation for single-user mimo and multi-user mimo-mac with partial csi," *IEEE Trans. on Selected Areas in Communication*, no. 7, pp. 1402–1412, Sep. 2007.

[31] A. M. Sayeed, "Deconstructing multi-antenna fading channels," *IEEE Trans. Signal Processing*, pp. 2563–2579, Oct. 2002.

[32] H. Wan, R.-R. Chen, and Y. Liang, "Optimality of beamforming in MIMO multi-access channels vis virtual representation," *in Proc. IEEE International Symposium on Information Theory (ISIT'08)*, pp. 2573–2577, July 2008.

[33] X. Li, S. Jin, X. Gao, M. Mckay, and K. Wong, "Transmitter optimization and beamforming optimality conditions for double-scattering MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 7, no. 9, pp. 3647–3654, Sep. 2008.

[34] D. Gesbert, H. Bolcskei, D. A. Gore, and A. J. Paulraj, "Outdoor MIMO wireless channels: Models and performance prediction," *IEEE Trans. Commun.*, vol. 50, no. 12, pp. 1886–1888, Dec. 2002.

[35] W. Dai, B. Rider, and Y. Liu, "Joint beamforming for multiaccess MIMO systems with finite rate feedback," *IEEE Transactions on Wireless Communications*, vol. 8, no. 5, pp. 2618–2628, May. 2009.

[36] A. M.Tulino, A. Lozano, and S. Verdu, "Capacity-achieving input covariance for correlated multi-antenna channels," *Proc. 2003 Allerton Conf. Computing, Control and Communications*, pp. 242–251, Oct. 2003.

[37] V. L. Girko, "A refinement of the central limit theorem for random determinants," *Theory Probab. Its Applic.*, vol. 42, no. 1, pp. 121–129, 1998.

[38] D. Chizhik, G. J. Foschini, M. J. Gans, and R. A. Valenzuela, "Keyholes, correlations, and capacities of multielement transmit and receive antennas," *IEEE Trans. Wireless Commun.*, vol. 1, pp. 361–368, Apr. 2002.

[39] W. Rhee, W. Yu, and J. M. Cioffi, "The optimality of beamforming in uplink multiuser wireless systems," *IEEE Trans. on Wireless Commu.*, vol. 3, no. 1, pp. 86–96, Jan. 2004.

[40] S. Lasaulce, A. Suarez, M. Debbah, and L. Cottatellucci, "Power allocation game for fading MIMO multiple access channels with antenna correlation," *Conference on GameComm*, Oct. 2007.

[41] B. L. Shengli Zhou, "BER criterion and codebook construction for finite-rate precoded spatial multiplexing with linear receivers," *IEEE Trans. on Sig. Proc.*, vol. 54, pp. 1653–1665, May 2006.

[42] A. Narula, M. J. Lopez, M. D. Trott, and G. W. Wornell, "Efficient use of side information in multiple-antenna data transmission over fading channels," *IEEE Journal on Sel. Areas Commu.*, vol. 16, pp. 1423–1436, Oct. 1998.

[43] J. C. Roh and B. D. Rao, "An efficient feedback method for mimo systems with slowly time-varying channels," *WCNC*, pp. 760–764, 2004.

[44] B. C. Banister and J. R. Zeidler, "Feedback assisted transmission subspace tracking for mimo systems," *IEEE Journal on Sel. Areas Commu.*, vol. 21, pp. 452–463, Apr. 2003.

[45] J. Yang and D. B. Williams, "Transmission subspace tracking for mimo systems with low-rate feedback," *IEEE Trans. on Commu.*, vol. 55, pp. 1629–1639, Aug. 2007.

[46] B. Mondal and R. W. H. Jr., "Adaptive feedback for mimo beamforming systems," *in Proc. of IEEE 5th Workshop on Sig. Proc. Adv. Wireless Commu.*, pp. 213–217, July 2004.

[47] J. Roopsha Samanta, Robert W. Heath, "Codebook adaption for quantized mimo beamforming systems," *IEEE Conf. on Signals, Systems and Computers*, pp. 376–380, Oct. 2005.

[48] S. H. Ting, K. Sakaguchi, and K. Araki, "Differential quantization of eigenmodes for mimo eigenmode transmission systems," *IEEE Commu. Letters*, vol. 9, pp. 697–699, Aug. 2005.

[49] H. Wan and R.-R. Chen, "Beamforming codebook design for temporally correlated fading channels," *Proc. Allerton Conference on Communication, Control, and Computing*, pp. 1347–1353, Sept. 2008.

[50] D. J. Love and R. W. Heath, "Limited feedback diversity techniques for correlated channels," *IEEE Trans. on Veh. Tech.*, vol. 55, pp. 718–722, Mar. 2006.

[51] V. Raghavan, A. M. Sayeed, and V. V. Veeravalli, "Limited feedback precoder design for spatially correlated mimo channels," *IEEE*, 2007.

[52] B. D. June Chul Roh, "Transmit beamforming in multiple-antenna systems with finite rate feedback: A vq-based approach," *IEEE Trans. on Inf. Theory*, vol. 52, pp. 1101–1112, March 2006.

[53] Z. Wan and G. B. Giannakis, "A simple and general parameterization quantifying performance in fading channels," *IEEE Trans. on Commu.*, vol. 51, pp. 1389–1398, Aug.2003.

[54] C. Douillard, M. Jezequel, C. Berrou, A. Picart, P. Didier, and A. Glavieux, "Iterative correction of intersymbol interference: Turbo equalization," *Eur. Trans. Telecommun.*, vol. 6, pp. 507–511, Sept.-Oct. 1995.

[55] L. M. Davis, I. B. Collings, and P. Hoeher, "Joint MAP equalization and channel estimation for frequency-selective and frequency-flat fastfading channels," *IEEE Trans. on Commun.*, vol. 49, p. 2106?114, Dec. 2001.

[56] J. Garcia-Frias and J. D. Villasenor, "Combined turbo detection and decoding for unknown ISI channels," *IEEE Trans. on Commun.*, vol. 51, p. 79?5, Jan. 2003.

[57] A. Glavieux, C. Laot, and J. Labat, "Turbo equalization over a frequency selective channel," *in Proc. Int. Symp. Turbo Codes, Related Topics, Brest, France*, pp. 96–102, Sept. 1997.

[58] M. Tuchler, R. Koetter, and A. Singer, "Turbo equalization: principles and new results," *IEEE Trans. Commun.*, vol. 50, no. 5, pp. 754–767, May 2002.

[59] M. Tuchler, A. C. Singer, and R. Koetter, "Minimum mean squared error equalization using a priori information," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 673–683, March 2002.

[60] T. Hashimoto, "A list-type reduced-constraint generalization of the Viterbi algorithm," *IEEE Trans. on Inform. Theory*, vol. 33, no. 6, pp. 866–876, Nov. 1987.

[61] R. E. Kamel and Y. Bar-Ness, "Reduced-complexity sequence estimation using state partitioning," *IEEE Trans. on Commun.*, vol. 44, no. 9, pp. 1057–1063, Sep. 1996.

[62] R. Raheli, A. Polydoros, and C.-K. Tzou, "Per-survivor processing: A general approach to MLSE in uncertain enviroments," *IEEE Trans. on Commun.*, vol. 43, no. 2/3/4, pp. 354–364, Feb./Mar./April 1995.

[63] J. Hagenauer and C. Kuhn, "The list-sequential (liss) algorithm and its application," *IEEE Trans. on commun.*, vol. 55, no. 5, pp. 918–928, May 2007.

[64] V. Franz and J. B. Anderson, "Concatenated decoding with a reduced-search bcjr algorithm," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 2, pp. 186–195, Feb. 1998.

[65] A. Anastasopoulos and K. M. Chugg, "Adaptive soft-input soft-output algorithms for iterative detection with parametric uncertainty," *IEEE Trans. Commun.*, vol. 48, pp. 1638–1649, Oct. 2000.

[66] R. Otnes and M. Tuchler, "Iterative channel estimation for turbo equalization of time-varying frequency-selective channels," *IEEE Trans. on Wireless Commun.*, vol. 3, no. 6, pp. 1918–1923, Nov. 2004.

[67] M. Nissila and S. Pasupathy, "Soft-input soft-output equalizers for turbo receivers: a statistical physics perspective," *IEEE Trans. Commun.*, vol. 55, pp. 1300–1307, July 2007.

[68] K. Fang, L. Rugini, and G. Leus, "Low-complexity block turbo equalization for OFDM systems in time-varying channels," *IEEE Trans. Signal Processing*, vol. 56, pp. 5555–5566, Nov. 2008.

[69] O. T. Nilsson, B. Olofsson, N. Nordenvaad, and E. M.L. Sangfelt, "Underwater communication link with iterative equalization," *OCEANS 2006*, pp. 1–6, Sept. 2006.

[70] R. Otnes and T. H. Eggen, "Underwater acoustic communications: long-term test of turbo equalization in shallow water," *IEEE Trans. on Oceanic Engineering*, vol. 33, pp. 321–334, July 2008.

[71] J. W. Choi, R. J.Drost, A. C.Singer, and J. Preisig, "Iterative multi-channel equalization and decoding for high frequency underwater acoustic channels," *Sensor Array and Multichannel Signal Processing Workshop*, July 2008.

[72] A. Singer, J. Nelson, and S. Kozat, "Signal processing for underwater acoustic communications," *IEEE communication magazine*, vol. 47, no. 1, pp. 90–96, Jan. 2009.

[73] J. Choi, T. J. Riedl, K. Kim, A. Singer, and J. Presig, "Practical application of turbo equalization to underwater acoustic communications," *Proc. 7th International Symposimum on Wireless Communication System (ISWCS)*, pp. 601–605, July 2010.

[74] K. D. Kammeyer, V. Khn, and T. Petermann, "Blind and nonblind turbo estimation for fast fading GSM channels," *IEEE J. Select. Areas Commun.*, vol. 19, p. 1718?728, Sept 2001.

[75] B. L. Yeap, C. H. Wong, and L. Hanzo, "Reduced complexity in-phase/quadrature-phase M-QAM turbo equalization using iterative channel estimation," *IEEE Trans. Wireless Commun.*, vol. 2, no. 2, pp. 2–10, Jan 2003.

[76] X. Wang and R. Chen, "Blind turbo equalization in Gaussian and Impulsive Noise," *IEEE Trans. on Vehicular Technology*, vol. 50, no. 4, pp. 1092–1105, July 2001.

[77] B. Lu and X. Wang, "Bayesian Blind Turbo Receiver for Coded OFDM Systems with Frequency Offset and Frequency-Selective Fading," *IEEE Journal on Selected Areas on Communications*, vol. 19, no. 12, pp. 2516–2527, Dec. 2001.

[78] Z. Yang and X. Wang, "A sequential monte carlo blind receiver for ofdm systems in frequency-selective fading channel," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 271–280, Feb. 2002.

[79] B. Farhang-Boroujeny, *Adaptive filters: theory and application.* Chichester, U.K.: John Wiley & Sons, 1998.

[80] S. Haykin, *Adaptive filter theory*. Upper Saddle Reiver, N.J.: Prentice-Hall, 1996.

[81] A. H. Sayed, *Fundamentals of adaptive filtering*. Hoboken, N.J.: Wiley & Sons : IEEE Press, 2003.

[82] V. Mathews and Z. Xie, "Stochastic gradient adaptive filters with gradient adaptive step size," *IEEE Trans. on Signal Processing*, vol. SP-41, no. 6, pp. 2075–2087, Jun. 1993.

[83] B. Farhang-Boroujeny, "Variable step-size lms algorithm, new development and experiments," *IEE Proceedings, Vision, Image and Signal Processing*, pp. 311–317, 1994.

[84] W. Ang and B. Farhang-Boroujeny, "A new class of gradient adaptive step-size lms algorithms," *IEEE Trans. on Signal Processing*, vol. 49, pp. 805–810, April. 2001.

[85] H. Wan, R.-R. Chen, J. W. Choi, A. Singer, J. Preisig, and B. Farhang-Boroujeny, "Markov Chain Monte Carlo detection for underwater acoustic channels," in *Proc. Information theory and applications workshop(ITA)'10*, 2010 (Invited paper).

[86] ——, "Joint channel estimation and markov chain monte carlo detection for frequency-selective channels," in *Proc. the 6th IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM), 2010*, 2010.

[87] J. W. Choi, "Digital communication receiver algorithms and architectures for reduced complexity and high throughput," Ph.D. dissertation, University of Illinois at Urbana-Champaign, 2010.

PUBLICATIONS

Journal papers

- **Hong Wan**, Rong-Rong Chen, Jun Won Choi, Andrew C. Singer, James C. Presig, and Behrouz Farhang-Boroujeny, "Markov chain monte carlo detection for frequency-selective channels using list channel estimation", IEEE Trans. of Selected Topics in Signal Processing, Volume 5, Dec. 2011.

- **Hong Wan**, Rong-Rong Chen, and Yingbin Liang, "Optimality of beamforming for MIMO multiple access channels via virtual representation", IEEE Trans. on Signal Processing, Nov. 2010.

Conference papers

- **Hong Wan**, Rong-Rong Chen, Jun Won Choi, Andrew Signer, and Behrouz Farhang-Boroujeny, "Joint channel estimation and markov chain monte carlo detection for frequency-selective channels", the 6th IEEE sensor array and multichannel signal processing workshop (SAM), 2010. **(Invited paper)**

- **Hong Wan**, Rong-Rong Chen, Jun Won Choi, Andrew Singer, James Preisig, and Behrouz Farhang-Boroujeny, "Stochastic expectation maximization algorithm for long-memory fast-fading channels", IEEE Global Communication Conference (Globecom), 2010.

- **Hong Wan**, Rong-Rong Chen, Jun Won Choi, Andrew Singer, James Preisig, and Behrouz Farhang-Boroujeny, "Markov chain monte carlo detection over time-varying frequency-selective channels", Proc. Allerton Conference on Communication, Control, and Computing, 2010.

- **Hong Wan**, Rong-Rong Chen, Jun Won Choi, Andrew Singer, James Preisig, and Behrouz Farhang-Boroujeny, "Markov Chain Monte Carlo detection for underwater acoustic channels", in Proc. Information theory and applications workshop (ITA), 2010. **(Invited paper)**

· **Hong Wan** and Rong-Rong Chen, "Beamforming codebook design for temporally correlated fading channels", Proc. Allerton Conference on Communication, Control, and Computing, 2008.

· **Hong Wan**, Rong-Rong Chen, and Yingbin Liang, "Optimality of beamforming in MIMO multi-access channels via virtual representation", Proc. IEEE International Symposium on Information Theory (ISIT), 2008.