Joint Design of Transceivers for Multiple Access Channels using MMSE Decision Feedback Detection

Master of Applied Science (2008)	McMaster University
Electrical and Computer Engineering	Hamilton, Ontario

TITLE:	**Joint Design of Transceivers for Multiple Access Channels using MMSE Decision Feedback Detection**
AUTHOR:	Wenwen Jiang Bachelor of Engineering USTB, China, 2006
SUPERVISOR:	Dr. Kon Max Wong
CO-SUPERVISOR:	Dr. Jian-Kang Zhang
NUMBER OF PAGES:	xiii, 80

Dedications:

To my parents:

Jianlin Jiang and Jingli Chen

Abstract

In this thesis, we consider the joint design of transceivers for a multiple access Multiple Input and Multiple Output (MIMO) system having Inter-Symbol Interference (ISI) channels. The system we consider is equipped with the Minimum Mean Square Error (MMSE) Decision-Feedback (DF) detector. Traditionally, transmitter designs for this system have been based on constraints of either the transmission power or the signal-to-interference-and-noise ratio (SINR) for each user. Here, we explore a novel perspective and examine a transceiver design which is under a fixed sum Gaussian mutual information constraint and minimizes the arithmetic mean square error of the MMSE-decision feedback detection. For this optimization problem, a closed-form solution is obtained. We prove that the optimal solution is achieved if and only if the sum mutual information is uniformly distributed over each individual user per the number of its active subchannels; i.e., user mutual information uniform distribution. Meanwhile, the Gaussian mutual information of the current user under perfect feedback for all the previous users is uniformly distributed over each individual symbol within the block signal of the user; i.e., symbol mutual information uniform distribution. The user mutual information uniform distribution is attained by successively solving a series of inverse (dual) problems of maximizing single user throughput, while the symbol mutual information uniform distribution is maintained by using the equal diagonal QRS decomposition. We also show that such uniform decomposition, in addition to minimizing the arithmetic MSE of MMSE-decision feedback detection, also has another two optimality properties: (a) Both the optimal user-detection order and symbol-detection order are natural orders in terms of signal to interference and noise ratios. (b) The free-distance for the Maximum Likelihood (ML) detector has an asymptotic behavior when the sum Gaussian mutual information tends to large.

Acknowledgements

In the first place the author wishes to express her deep appreciation to her supervisor, Professor Kon Max Wong, for his expert supervision, guidance and encouragement throughout the course of this thesis, which have been inspiring to her growth as a student and have been illuminating the way of her independent research.

The author is very grateful to her co-supervisor Dr. Jian-Kang Zhang for continuously providing helpful suggestions and insightful discussions. She truely appreciates Tingting Liu who studied with her, debated upon questions and research problems, and gave her hand in life. She also express her gratitude to Qiuyuan Huang, Jing Liu, and all her colleagues here in the Signal Processing Group at McMaster University for their friendship and support during occasional stressful times.

She is also deeply indebted to her parents Jianlin Jiang and Jingli Chen, and all other family members for their understanding, love and unceasing faith in her along the way.

Contents

List of Figures

Acronyms

BC	Broadcast Channel
BER	Bit Error Rate
CSI	Channel State Information
DF(E)	Decision Feedback (Equalization)
(I)DFT	(Inverse) Discrete Fourier Transform
DMT	Discrete Multi-Tone
IBI	Inter-Block Interference
ISI	Inter-Symbol Interference
LE	Linear Equalization
MAC	Multi-Access Channel
MIMO	Multi-Input Multi-Output
MISO	Multi-Input Single-Output
ML	Maximum Likelihood
MMSE	Minimum Mean Square Error
MSE	Mean Square Error
OFDM	Orthogonal Frequency Division Multiplexing
QAM	Quadrature Amplitude Modulation
SIMO	Single-Input Multi-Output
SNR	Signal to Noise Ratio
SISO	Single-Input Single-Output
ZF	Zero Forcing

Notations

$E\{\cdot\}$	Expectation
$\mathrm{tr}(\cdot)$	Trace
log	Logarithm function with e base
det	Determinant
$\mathrm{diag}(a_1, a_2, \cdots, a_N)$	A diagonal matrix with diagonal elements $a_1, a_2, \cdots, a_N$
$[\cdot]^T$	Transpose of a matrix or a vector
$[\cdot]^H$	Transpose and complex conjugated of a matrix or a vector (Hermitian)
$[\cdot]^{-1}$	inverse of a matrix
$[\cdot]^\dagger$	Pseudo inverse of a matrix
$\Re\mathbf{A}$	Real part of a matrix
$\Im\mathbf{A}$	Imaginary part of a matrix
$[\mathbf{A}]_k$	The kth diagonal entry of a matrix $\mathbf{A}$
a_{ij}	The ijth element of a matrix $\mathbf{A}$
$\mathbf{A}_{ij}$	The ijth sub-matrix of a block matrix $\mathbf{A}$
$\mathbf{I}_M$	$M \times M$ identity matrix
$\mathbf{E}_{ij}$	The elementary matrix
$\mathbf{A}_{.j}$ or $(\mathbf{A})_{.j}$	The jth column of the matrix $\mathbf{A}$
$\mathbf{A}_{i.}$ or $(\mathbf{A})_{i.}$	The ith row of the matrix $\mathbf{A}$
$\arg\{\cdot\}$	A scaler, vector or matrix which satisfies the conditions enclosed within
∇	Gradient
$\mathcal{Q}[.]$	The quantization process

Text Conventions:

Throughout this thesis, we use the following notation: Matrices and column vectors are denoted by uppercase boldface characters (e.g., $\mathbf{A}$) and lowercase boldface characters (e.g., $\mathbf{b}$), respectively.

Chapter 1

Introduction

During the past years, wireless communication systems have gained significant importance and attention due to the heavy demand of ubiquitous communications in society. While there are only limited resources in a communication system, the increasing amount of information that has to be handled places higher demands on the capacity, reliability and frequency efficiency of the system. Multiple-input multiple-output (MIMO) communication schemes are momentous breakthroughs of communication techniques to meet these recent challenges due to their numerous advantages and potentials including greatly increased channel capacities, as well as diversity and spectral efficiencies.

1.1 What is MIMO

Figure 1.1 illustrates different input-output configurations defining space-time communication systems: 1) Single-input single-output (SISO), 2) Multiple-input single-output (MISO), 3) Single-input multiple-output (SIMO) and 4) Multiple-input multiple-output (MIMO) systems, depending on the number of inputs and outputs at the transmitter and the receiver respectively. MIMO systems, in particular, have attracted much attention in communications, since they offer significant increases in data throughputs and link ranges without additional bandwidths or transmission power. This is achieved by higher spectral efficiency (more bits per second per Hertz of bandwidth) and link reliability or diversity (reduced fading). For these advantages, MIMO is currently a research topic enthusiastically pursued.

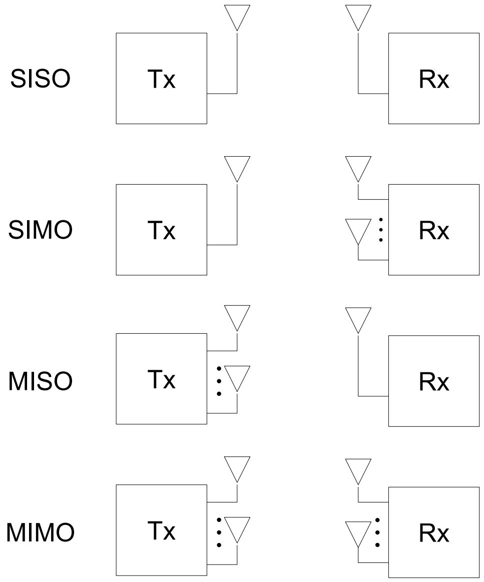

Figure 1.1: Multi-antenna types

To exploit all its advantages, a MIMO system divides its functions into three primary parts:

- *Precoding* – This function of the tranmitter puts appropriate weights on the signals to be transmitted in order to achieve different objectives such as maximizing the link throughtputs (or sum mutual information) at the receiver output, or minimizing the mean square error (MSE) of the detection, or minimizing the bit error rate (BER), etc. Therefore, with the use of precoding, the system performance can be further improved. It should be noted that optimum precoding generally requires full knowledge of the channel state information (CSI) at the transmitter.

- *Spacial multiplexing* – In MIMO systems, this offers a linear increase in the transmission rate (or capacity) for the same bandwidth and with no additional power expenditure [?]. For spatial multiplexing, a high rate signal stream is split into multiple lower rate streams and each of these streams is transmitted from one transmitter antenna in the same frequency channel. If these signals arrive at the receiver with sufficiently different spatial signatures, the receiver can separate these streams, i.e., spacial multiplexing can create parallel channels with knowledge of CSI. The maximum number of spatial streams is limited by the smaller of the numbers of antennas at the transmitter and receiver. In this case, channel knowledge is not required at the transmitter.

- *Diversity* techniques – These are used to increase reliability of transmission especially under fading conditions. There are three main types of diversity: temporal diversity, frequency diversity and spacial diversity, each of which provides replicas of the transmitted signals over time, frequency, and space respectively. As a result, at the receiver end, replicas of the transmitted signals are obtained which can be helpful to recover the original signals. If parts of the signals face a deep fade in the channel and are distorted badly, there are still other replicas of the signals transmitted through different and independent paths or frequencies that can be used for detection. In a MIMO system, due to its multiple antenna facility, spatial diversity through different transmitter and receiver antennas are generally utilized such that the transmitter sends the same signal through multiple paths while the receiver receives multiple replicas of the same transmitted signal. Thus, the higher is the diversity, the better we can combat the fading of a channel. Diversity is characterized by the number of independent fading branches, or paths (routes). These paths are also known as diversity order. Full diversity is achieved when the total degree of freedom (the number of transmitter antennas times the number of receiver antennas) offered in the multi-antenna system is utilized [?].

From the above, it can be seen that the use of multiple dimension at both the transmitter and the receiver brings significant enhancement in spectral efficiency, link reliability as well as a considerable increase of transmission rate.

1.2 Examples of MIMO Communication System

MIMO channels arise in many different scenarios and we will give some typical examples of MIMO applications in this section.

- *A multi-carrier system* in which the available bandwidth is partitioned into L subbands and then each subband is independently used for transmission [?, ?]. Such an approach not only simplifies the communication process but also provides a capacity-achieving structure for a sufficiently large L [?]. If the signals are transmitted using a block transmission together with a cyclic prefix, the corresponding channel model then is represented by a circulant matrix which when combined with an inverse/direct discrete Fourier transform (DFT) at the transmitter/receiver, is transformed into a diagonal matrix with diagonal elements given by DFT coefficients [?].

- *The multi-antenna wireless channel* is a paradigmatic example of a MIMO system (shown in Figure 1.2). This particular system can offer all the main advantages of MIMO systems [?, ?, ?].

- *The wireline digital subscriber line (DSL)* technology has gained popularity as a broadband access technology capable of reliably delivering high data rates over telephone subscriber lines [?]. Modeling a DSL system as a MIMO channel presents many advantages with respect to treating each twisted pair independently [?, ?] which was done three decades ago [?]. The dominant impairment in DSL systems is crosstalk arising from electromagnetic coupling between neighboring twisted-pairs. Near-end crosstalk (NEXT) comprises the signals orig-

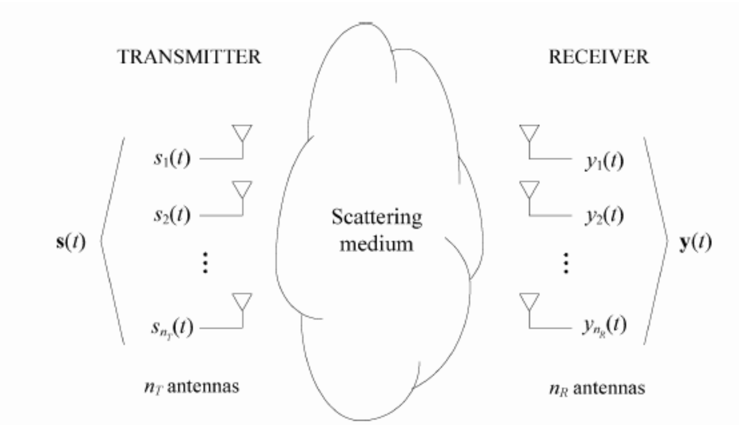

Figure 1.2: Example of a MIMO channel arising in wireless communications

inated in the same side of the received signal and far-end crosstalk (FEXT) includes the signal originated in the opposite side of the received signal. In DSL system, a bundle of twisted pairs is treated as whole. As shown in Figure 1.3, a binder group composed of L users in the same physical location plus some other users that possibly belong to a different service provider and use different types of DSL systems. The MIMO channel represents the communication of the L intended users while the others are treated as interference. DSL chan-

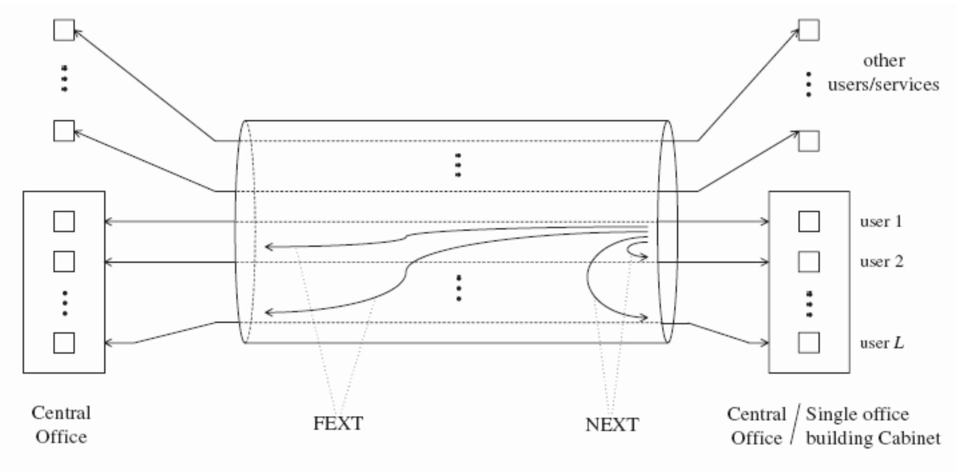

Figure 1.3: Example of a MIMO channel arising in DSL communications

nels are highly frequency-selective, as a consequence, practical communication systems are based on the multicarrier MIMO signal model [?].

1.3 Multi-user MIMO System

Our discussion above focuses on MIMO communication systems with a single user. However, MIMO transmission can also be used by multiuser systems. For a MIMO multi-access channel (MAC), multiple users communicate with one base station. The signals received by the base station is the summation of all the signals from all the users. A similar system can be considered as the broadcast channel (BC), i.e., the downlink channel of the MAC where a common transmitter sends information to distributed receivers. When applying MIMO scheme in a multi-user system, a specific model needs to be established.

A pragmatic approach to deal with multiuser systems consists of employing single-user designs for the users of the network in an iterative manner [?] or iteratively optimizing the receivers and the transmitters [?], but a global optimum may not be reached.

1.4 Motivation and Contribution of the Thesis

Given the increasing importance of the MIMO technology as well as the importance of multi-user communications in practice, a critical issue arises, i.e., the joint design of a transceiver for a multi-access MIMO channel. This is the subject of this thesis. It should be noted that the optimal design solutions for the single user system cannot be directly generalized to a multi-user scenario. The main technical obstacle is, in the case of a multi-user system, the transmitter matrix has a block structure such that each sub-block is constrained in power individually. Thus far, this difficult problem of designing optimal transceiver pairs for a multi-user case has been successfully tackled by minimizing the total MSE in the system employing a linear MMSE receiver [?, ?] and optimal power allocation has been developed for OFDM or DMT system in [?, ?, ?]. Also the capacity in a multi-user system has been maximized in [?, ?] Since minimum mean square error- decision feedback equalization (MMSE-DFE) has superior performance to, as well as other advantages over the linear MMSE equalizer, we will concentrate on the optimum design of a multi-user MIMO transceivers having a MMSE-DFE receiver.

We first considered using individual power constraints in a multiple access communication system with MMSE-DFE when minimizing the arithmetic mean square error (MSE). However this problem faces two main difficulties. The first one is due to the specific block structure for the transmitter matrix. We considered applying the trace-determinant inequality to obtain the lower bound of the MSE. However, the problem of the block structure matrix renders this bound unachievable. The other obstacle is that this optimization problem is not convex and may not be easily solved.

Therefore, we explore a novel perspective of the transceiver pair design for block-by-block intersymbol interference (ISI) multiple access MIMO channels with the MMSE-DF detector such that the arithmetic mean square error for K users is minimized *subject to a fixed sum mutual information constraint.* By using this criterion, we avoid the structural problem of the transmission matrix. Furthermore, by using the dual water-filling solution, the transmission power of each user is simultaneousy minimized and a closed form solution is provided.

To sum up, in this thesis, the optimal design of the transceiver pair for a synchronous multiple access MIMO system in which the K-user data sequences are precoded separately and transmitted block-by-block at full rate over frequency selective ISI channels is taken into account. At the receiver end of this multiple access MIMO system, the MMSE-DF detector is employed to detect the signals. The optimization problem to minimize the arithmetic MSE in a multi-access MIMO communication system subject to fixed sum mutual information is examined, and the closed-form optimal solution is then provided. Simulation results of the performance under this design are presented and compared with that of the multi-access with linear equalization and maximum likelihood detection.

1.5 Organization of the Thesis

The thesis is structured as follows:

- In Chapter 2, the background knowledge of the multi-access MIMO communication system with several detection methods are provided for the reader to better understand this work.
- In Chapter 3, according the system model, we propose a QR interpretation of the decision feedback equalization, which successively cancels the previous detected symbols.
- In Chapter 4, the dual water-filling problem is discussed. In a single user communication system model, an dual water-filling problem is stated and the close-form optimal solution is provided and proved.
- In Chapter 5, the design problem is proposed. By reformulating the objective function and applying the inverse water-filling solution, the final closed form optimal design is obtained. With the aid of QR decomposition, further insight of the optimality is obtained.
- In Chapter 6, we simulate a multi-user MIMO communication system equipped with the optimally designed transceiver. Simulation results are compared to those using a linear receiver proposed in [**?**] and those using maximum likelihood detection.
- In Chapter 7, conclusion on this work and suggestion for future work are presented.

Chapter 2

MIMO System

In this chapter, we present the necessary concepts and theories in adequate depth as a preliminary to this thesis. The mathematical model of the multiple-input multiple-output (MIMO) transmission is first presented. We then discuss the various criteria to measure how the communication system performs. This is followed by the discussion on different detection schemes. Finally, the system model of a multiple access communication system is introduced.

2.1 System Model

A MIMO channel is mathematically represented by a matrix which provides a way to show channels with different natures. The MIMO communication channel with N transmitter and P receiver antennas can be described with the base band model

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \boldsymbol{\xi} \tag{2.1}$$

if the channel is band limited. In this model, $\mathbf{x}$ is the $N \times 1$ signal vector to be transmitted, $\mathbf{H}$ is $P \times N$ channel matrix with the component h_{ij} of the channel matrix is the gain/fading coefficient from the jth transmitter antenna to the ith receiver antenna (as shown in Figure 2.1). The received signals constitute a $P \times 1$

column vector $\mathbf{y}$, where each complex component refers to a receiver antenna. The noise is then denoted by $\boldsymbol{\xi}$.

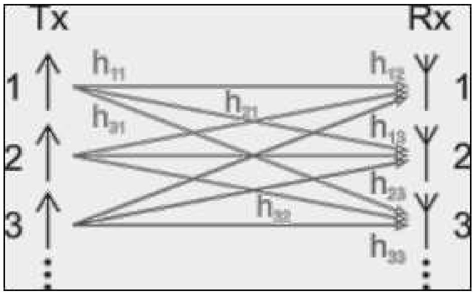

Figure 2.1: MIMO channel model

2.2 Measures of Communication System Performance

There are various ways to measure and compare the performance of different communication systems and transmission schemes. The three main criteria are bit error rate (BER), mean square error (MSE), and channel capacity. Bit error rate is the ratio of the number of incorrectly received bits to the total number of bits sent during a specified time interval. This is a measure of how well the demodulator and encoder perform. More precisely, the average probability of a bit-error at the output of the decoder is a measure of the performance of the system. In general, the probability of error is a function of the code characteristics, the types of waveforms used to transmit the information over the channel, the transmitter power, the characteristics of the channel (i.e., the amount of noise, the nature of the interference, etc.), and the method of demodulation and decoding.

MSE is the square of the difference between the detected signals and the transmitted signals. It measures the average square of the distance between the received and the transmitted signal vectors. Therefore, the smaller the MSE is, in general, the less probable it is for the detector to make an error. The mean square error remains a significant parameter for the assessment of the performance of the communication system [?, ?].

Channel capacity is the tightest upper bound on the amount of information that can be reliably transmitted over a communications channel. By the noisy-channel coding theorem [?], the capacity of a given channel is the limiting information rate (in units of information per unit time) that can be achieved with arbitrarily small error probability. The notion of channel capacity defined by Shannon in information theory provides a mathematical quantity by which one can compute it. The key result states that the capacity of the channel is given by the maximum of the mutual information between the input and output of the channel, where the maximization is with respect to the input distribution [?].

When comparison of communication systems are made, other factors such as the transmission power and complexity of implementation should also be taken into consideration.

2.3 Detection Methods

A major problem in data communications arises from the intersymbol interference (ISI) created by a frequency selective channel. ISI is a form of distortion of a signal in which one symbol interferes with subsequent symbols. This is an unwanted phenomenon since the previous symbols have similar effect as noise, thus making the communication less reliable. When the signals are transmitted through a bandlimited channel, wire or wireless, the channel characteristic is usually non-ideal, i.e., the amplitude response is not constant for the pass band and the phase response is not a linear function of frequency. A sequence of the pulses transmitted through the channel will then be distorted and may not be clearly distinguishable at the receiver [?].

This problem, however, can usually be simplified by transmitting the data in a block-based fashion [?, ?]. In particular, effective detection can be performed on a block-by-block basis if the blocks are designed so that there is no inter-block interference (IBI) at the receiver. There are several schemes within this family of block-by-block data communications, the most commonly used being the multi-carrier modulation based Discrete Multi-Tone (DMT) [?, ?] and Orthogonal Frequency Division Multiplexing (OFDM) schemes [?].

To reduce the intersymbol interference(ISI) problem in channels, the signals are often put through an equalizer before making the decision. An equalizer is a device which compensates for the non-ideal frequency response of the channel. In this section, we introduce detection schemes which includes such an equalization process at the receiver.

2.3.1 Maximum Likelihood

From a detection error viewpoint, an optimal transmitter for a single user block-by-block data communication over an ISI channel with Gaussian noise is one that minimizes the detection error probability of the maximum likelihood (ML) detector [?, ?, ?, ?]. If the received signal vector is $\mathbf{y} = \mathbf{H}\mathbf{x} + \boldsymbol{\xi}$ with $\boldsymbol{\xi}$ being the Gaussian white noise, then the joint probability density function (PDF) of the random variable y conditioned on the transmitted sequence $\mathbf{x}$ is

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{(2\pi)^{N/2}\det(\boldsymbol{\Sigma}_{\xi\xi})^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{y}-\mathbf{H}\mathbf{x})^H \boldsymbol{\Sigma}_{\xi\xi}^{-1}(\mathbf{y}-\mathbf{H}\mathbf{x})\right\}$$

for real symbols where $\boldsymbol{\Sigma}_{\xi\xi} = E[\boldsymbol{\xi}\boldsymbol{\xi}^H]$ is the noise covariance matrix. Under the maximum likelihood criterion, we choose $\hat{\mathbf{x}}$ so that it is the value that most likely caused the received value of $\mathbf{y}$ to occur. Thus, the maximum-likelihood estimate, $\hat{\mathbf{x}}_{ML}$, is the one that maximizes this joint probality density, i.e., minimizes the distance between $\mathbf{y}$ and $\mathbf{H}\mathbf{x}$. At high signal-to-noise ratios (SNRs), the average probability of error over all blocks is dominated by the free distance term [?, ?]. (This suggests that maximizing the free distance may be a good transmitter design strategy). In the presence of ISI that spans $(N+1)$ symbols (N interfering components), the ML criterion is equivalent to the problem of estimate a state in the total of M^N states if the information symbol is M-ary. A famous algorithm, the Viterbi algorithm [**?**], is usually used in finding the optimal solution whose basic idea is to search all the possible combination of symbols via the trellis representation of the detection process. This shows that if the symbol length N is large the ML detector is, in general, very complicated to implement (the computational cost grows exponentially with the length of the symbol block) and may not be practical. We will describe two suboptimal channel equalization approaches in the following sections.

2.3.2 Linear Receiver

One suboptimal equalization is linear equalization (LE) which employs a linear transversal filter. The computational complexity (usually it is the complexity of calculating the inverse of the channel matrix) may, in special cases, be as low as a linear function of the channel dispersion length [**?**]. Under such a scheme, decision will be made based on $\hat{\mathbf{x}} = \mathbf{Jy}$, where $\mathbf{J}$ is the equalizer matrix. This equalized signal vector $\hat{\mathbf{x}}$ is then quantized to the nearest symbol to form the estimation such that $\hat{\mathbf{x}}_{LE} = \mathcal{Q}[\hat{\mathbf{x}}]$ where $\mathcal{Q}[.]$ denotes the quantization process. The detection error can then be written as

$$\begin{aligned} \mathbf{e} &= \hat{\mathbf{x}} - \mathbf{x} = \mathbf{Jy} - \mathbf{x} \\ &= \mathbf{J}(\mathbf{Hx} + \boldsymbol{\xi}) - \mathbf{x} = (\mathbf{JH} - \mathbf{I})\mathbf{x} + \mathbf{J}\boldsymbol{\xi} \end{aligned} \tag{2.2}$$

The equalization matrix $\mathbf{J}$ can be designed according to different criteria. Two popular criteria for this purpose are zero-forcing (ZF) criterion and minimum mean square error (MMSE). In zero-forcing, we force the ISI part of the error to be zero, i.e., in Eq. (2.2), we have $\mathbf{JH} - \mathbf{I} = 0$, i.e.,

$$\mathbf{J} = (\mathbf{H})^{\dagger} = (\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H$$

where $(.)^\dagger$ denotes the Moore-Penrose pseudo-inverse of a matrix. The zero-forcing equalizer removes all ISI, and is ideal when the channel is noiseless. However, when the channel is noisy, the zero-forcing (ZF) equalizer may amplify the noise power greatly. A more balanced linear equalizer is the MMSE equalizer which does not aim at eliminating ISI completely, but instead, minimizes the total power of the noise and ISI components at the output. According to Eq. (2.2), the MSE can be obtained as the trace of the error covariance matrix, i.e.,

$$\begin{aligned} \varepsilon &= \mathrm{tr}(E[\mathbf{e}\mathbf{e}^H]) \\ &= \mathrm{tr}\left[(\mathbf{JH}-\mathbf{I})(\mathbf{JH}-\mathbf{I})^H + \mathbf{J}\mathbf{\Sigma}_{\xi\xi}\mathbf{J}^H\right] \end{aligned}$$

and the MMSE linear equalizer matrix $\mathbf{J}$ is the one that minimizes the above MSE, and this minimization can be realized if $E[\mathbf{ey}^H = 0]$. So the equalizer matrix can be expressed as

$$\mathbf{J} = (\mathbf{HH}^H + \mathbf{\Sigma}_{\xi\xi})^{-1}\mathbf{H}^H$$

The advantage of linear equalization is in the simplicity of implementation. However, this is achieved at the expense of loss in accuracy in the sense that the performance of the linear equalizer is worse than ML detection, and under severe ISI, it may not yield acceptable results.

2.3.3 Decision Feedback Equalization

Another effective alternative is to employ decision feedback equalization (DFE) at the receiver which is a good compromise between implementation complexity and overall performance. The DFE is widely used to combats intersymbol interference (ISI) in linear dispersion channels. To disentangle the intersymbol interference, each input symbol based on the entire received sequence is first decoded, and its effect on the remainder of the sequence then subtracted before the decoding of the next symbol begins. The canceling of the interference by the DFE can be effected using either the criterion of ZF or that of the MMSE, and are designated ZF-DFE and MMSE-DFE respectively.

The decision feedback equalizer (DFE) consists of two filters, a feedforward filter **F** and a feedback filter **B**.

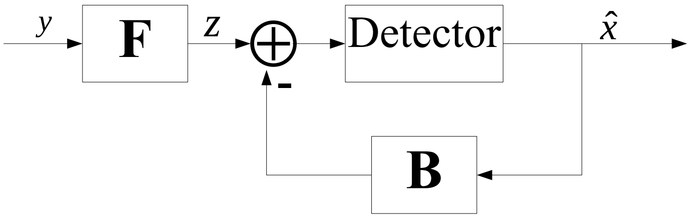

Figure 2.2: A conceptual model for decision feedback equalization

Figure 2.2 shows a conceptual model of the structure of DFE. The input of feedforward section is the received signals $\mathbf{y}$ from which we obtain the output $\mathbf{z} = \mathbf{F}\mathbf{y}$. Therefore in this respect, **F** plays an identical role as the linear equalizer **J** in linear equalization. The functional form of **F** depends on if ZF-DFE or MMSE-DFE is used. The feedback filter has an input which is the sequence of previously detected symbols. These are used to remove the intersymbol interference from the present symbol estimate.

Given a block of N transmitted symbols, the detection proceeds sequentially starting from the Nth symbol by making the decision on $\hat{x}_N = z_N$, and then $\hat{x}_n = z_n - \sum_{i=n+1}^{N} b_{ni}\hat{x}_i$, where b_{ni} is the coefficients in the feedback matrix **B**. Once this block has been estimated, the states of the feedback filter are reset to be zero. Thus we have the structure of **B** as

$$\mathbf{B} = \begin{bmatrix} 0 & b_{12} & b_{13} & \cdots & b_{1N} \\ 0 & 0 & b_{23} & \cdots & b_{2N} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & b_{(N-1)N} \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

which is an upper triangular matrix with diagonal elements being zero. Under the assumption that the previous decisions are all correct, the estimated signal vector can be written as

$$\hat{\mathbf{x}}_{DFE} = \mathbf{Fy} - \mathbf{Bx} = (\mathbf{FH} - \mathbf{B})\mathbf{x} + \mathbf{F}\boldsymbol{\xi} \tag{2.3}$$

Then we can further have the error vector as

$$\mathbf{e} = \hat{\mathbf{x}}_{DFE} - \mathbf{x} = (\mathbf{FH} - \mathbf{B} - \mathbf{I})\mathbf{x} + \mathbf{F}\boldsymbol{\xi} \tag{2.4}$$

For the ZF-DFE, we use the ZF criterion such that $\mathbf{F}_{ZF}\mathbf{H} = \mathbf{B} + \mathbf{I}$, and then the optimal feedforward filter is given by

$$\mathbf{F}_{ZF} = (\mathbf{B} + \mathbf{I})(\mathbf{H})^{\dagger}\boldsymbol{\Sigma}_{\xi\xi}^{-1/2} \tag{2.5}$$

For the MMSE-DFE, we apply the MMSE criterion by exploiting the orthogonality principle $E[\mathbf{e}\mathbf{y}^H] = 0$ [**?**, **?**] so that

$$\mathbf{F}_{MMSE} = (\mathbf{B} + \mathbf{I})\mathbf{R}_{xy}\mathbf{R}_{yy}^{-1} = (\mathbf{B} + \mathbf{I})\mathbf{H}^H(\mathbf{H}\mathbf{H}^H + \boldsymbol{\Sigma}_{\xi\xi})^{-1} \tag{2.6}$$

where

$$\begin{aligned}\mathbf{R}_{yy} &= E[\mathbf{y}\mathbf{y}^H] = E[(\mathbf{Hx} + \boldsymbol{\xi})(\mathbf{Hx} + \boldsymbol{\xi})^H] \\ &= \mathbf{H}\mathbf{H}^H + \boldsymbol{\Sigma}_{\xi\xi}\end{aligned}$$

is the covariance matrix of $\mathbf{y}$, and

$$\begin{aligned}\mathbf{R}_{xy} &= E[\mathbf{x}\mathbf{y}^H] = E[\mathbf{x}(\mathbf{Hx} + \boldsymbol{\xi})^H] \\ &= \mathbf{H}^H = \mathbf{R}_{yx}^H\end{aligned}$$

is the cross-correlation matrix of $\mathbf{x}$ and $\mathbf{y}$.

Since the error of the detection is defined in Eq. (2.4) the error covariance matrix of DFE can be then written as

$$\begin{aligned}\boldsymbol{\Sigma}_{ee} &= E[\mathbf{e}\mathbf{e}^H] = E\left\{[(\mathbf{FH} - \mathbf{B} - \mathbf{I})\mathbf{x} + \mathbf{F}\boldsymbol{\xi}][(\mathbf{FH} - \mathbf{B} - \mathbf{I})\mathbf{x} + \mathbf{F}\boldsymbol{\xi}]^H\right\} \\ &= (\mathbf{FH} - \mathbf{B} - \mathbf{I})(\mathbf{FH} - \mathbf{B} - \mathbf{I})^H + \mathbf{F}\boldsymbol{\Sigma}_{\xi\xi}\mathbf{F}^H \\ &= \mathbf{F}\mathbf{H}\mathbf{H}^H\mathbf{F}^H - (\mathbf{B} + \mathbf{I})\mathbf{H}^H\mathbf{F}^H - \mathbf{FH}(\mathbf{B} + \mathbf{I})^H \\ &\quad + (\mathbf{B} + \mathbf{I})(\mathbf{B} + \mathbf{I})^H + \mathbf{F}\boldsymbol{\Sigma}_{\xi\xi}\mathbf{F}^H\end{aligned} \tag{2.7}$$

Substituting Eq. (2.6), we can obtain

$$\begin{aligned} \mathbf{F}\mathbf{H}\mathbf{H}^H\mathbf{F}^H + \mathbf{F}\boldsymbol{\Sigma}_{\xi\xi}\mathbf{F}^H &= \mathbf{F}(\mathbf{H}\mathbf{H}^H + \boldsymbol{\Sigma}_{\xi\xi})\mathbf{F}^H \\ &= (\mathbf{B}+\mathbf{I})\mathbf{H}^H(\mathbf{H}\mathbf{H}^H + \boldsymbol{\Sigma}_{\xi\xi})^{-1}\mathbf{H}(\mathbf{B}+\mathbf{I})^H \qquad (2.8) \end{aligned}$$

and

$$\begin{aligned} (\mathbf{B}+\mathbf{I})\mathbf{H}^H\mathbf{F}^H &= (\mathbf{B}+\mathbf{I})\mathbf{H}^H(\mathbf{H}\mathbf{H}^H + \boldsymbol{\Sigma}_{\xi\xi})^{-1}\mathbf{H}(\mathbf{B}+\mathbf{I})^H \qquad (2.9) \\ &= \mathbf{F}\mathbf{H}(\mathbf{B}+\mathbf{I})^H \end{aligned}$$

Then substituting Eq. (2.8) and Eq. (2.9) back into Eq. (2.7), the error covariance matrix is further written as

$$\begin{aligned} \boldsymbol{\Sigma}_{ee} &= (\mathbf{B}+\mathbf{I})\left[\mathbf{I} - \mathbf{H}^H(\mathbf{H}\mathbf{H}^H + \boldsymbol{\Sigma}_{\xi\xi})^{-1}\mathbf{H}\right](\mathbf{B}+\mathbf{I})^H \\ &= (\mathbf{B}+\mathbf{I})(\mathbf{I} + \mathbf{H}^H\boldsymbol{\Sigma}_{\xi\xi}^{-1}\mathbf{H})^{-1}(\mathbf{B}+\mathbf{I})^H \qquad (2.10) \end{aligned}$$

Decision feedback equalization offers improved performance over the linear approach while maintaining reasonable complexity. Under the assumption of no error propagation, the MMSE-DFE can achieve the capacity of a Gaussian linear dispersion channel [?]. Even for binary input signals, the capacity achieved by the MMSE decision feedback detector is very close to that achieved by the optimal ML detector at moderate signal to noise ratio region [?]. MMSE equalization has the additional advantage that it combines well with lattice-type codes to achieve the capacity of additive white Gaussian noise channels [?, ?, ?]. Mathematically, the derivative of the mutual information with respect to the signal-to-noise-ratio for an MMSE-DFE is equal to half of the MMSE, regardless of the input statistics [?]. In this thesis, our receiver focuses on the use of MMSE-DFE.

2.4 Transceiver Designs for Single-user MIMO Channels

The term transceiver here corresponds to the combination of the procoding at the transmitter that we discussed before and the equalization parameters at the receiver end. The concept of MIMO transceiver is a transformation applied on the transmitted and received signals to improve the communication performance. Using the channel model in the previous section, the precoded channel model can be described as

$$\mathbf{y} = \mathbf{H}\mathbf{T}\mathbf{x} + \boldsymbol{\xi}$$

where $\mathbf{H}$ is the $P \times M$ channel matrix and $\mathbf{T}$ is an $M \times N$ precoder matrix. The goal to design this transmitter (or precoder) matrix $\mathbf{T}$ is to enhance the communication system in various aspects such as maximizing the channel capacity, minimizing the probability of error, or minimizing the mean square error, etc. At the receiver, the optimal equalization matrices, $\mathbf{J}$ in linear equalization or $\mathbf{F}$ and $\mathbf{B}$ in DFE, can be designed as a function of $\mathbf{T}$. This is the reason why the joint design of transceivers is usually considered for improving the overall performance of the communication system.

Research on transceiver designs for a single user system have been successfully carried out in the past years. For linear receivers the corresponding optimum transmitters designed under different criteria [?, ?, ?, ?, ?] ranging from maximization of mutual information and maximization of SNR to minimization of mean-square error (MSE) and minimization of receiver bit-error-rate (BER). However, compared to use of a ML detector at the receiver, all these show substantial loss in performance.

For DFE receivers, the joint design of the transmitter and MMSE-DFE receiver using a geometric MSE criterion has been obtained in [?]. However, the resulting optimal transmitter does not guarantee to simultaneously minimize the minimized MSE. More recently, closed-form optimal transceivers with ZF-DF [?, ?] and MMSE-DF [?, ?, ?] detectors have been obtained using a newly developed equal diagonal

QRS decomposition of a matrix [?]. It has been shown [?] that with the use of the respective optimum transmitters, the system employing MMSE-DFE is superior in performance to that with the ZF-DFE. Indeed, the MMSE-DFE receiver has been analysed [?, ?, ?, ?] and is referred to as a canonical receiver suggesting that by using the properly designed codes and under the assumption of having no error propagation, reliable communication at rates approaching the capacity of the block transmission system can be achieved by using independent instances of the same (Gaussian) code in each element of block. Thus, for no loss in information and for having a low complexity compared to the ML detector, the MMSE-DFE is a desirable receiver. For this reason, our focus in this thesis is on MIMO systems employing the MMSE-DFE in the receiver.

2.5 Multi-user System Model

Before we address the problem of optimum transceiver design for a multi-user MIMO system, we should first establish a model. The multi-access communication system model considered in this thesis is shown as the following Figure 2.3.

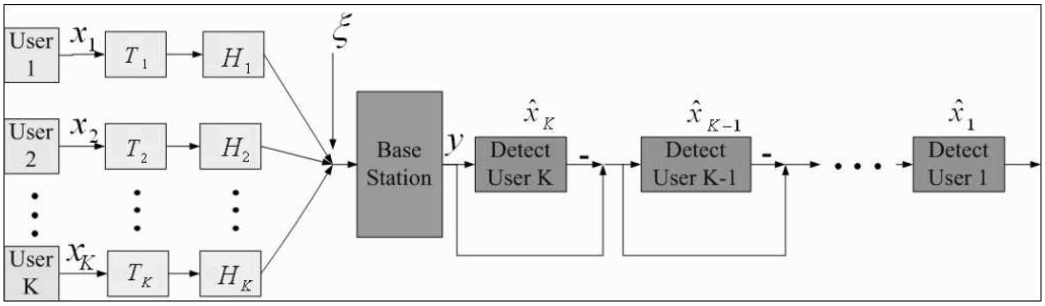

Figure 2.3: Multi-access system model

We consider a block-based synchronous multiple access frequency selective MIMO channel in which the K users' data sequences are precoded separately and are transmitted over distinct ISI channels. Denoting the signal vector for the kth user as $\mathbf{x}_k$,

$k = 1, \cdots, K$, the received signal is given by

$$\mathbf{y} = \sum_{k=1}^{K} \mathbf{H}_k \mathbf{T}_k \mathbf{x}_k + \boldsymbol{\xi} \tag{2.11}$$

where $\mathbf{H}_k$ is a $P \times M$ block Toeplitz tall channel matrix corresponding to zero-padded modulation [?] or an $M \times M$ square block diagonal channel matrix corresponding to DMT modulation [?, ?, ?] for the kth user, and $\mathbf{T}_k$ is an $M \times N_k$ precoder matrix for the kth user. $\mathbf{x}_k$ is the block of N_k transmitted symbols for User k, which is assumed to be zero-mean, white and of identity covariance matrix, and $\boldsymbol{\xi}$ is an $P \times 1$ white Gaussian noise vector with identity covariance matrix and independent of the input signal vector $\mathbf{x}_k$. Assume the channel state information (CSI) is perfectly known for both transmitter and receiver.

Transceiver designs for multi-user system is not an easy extension of those for single users and it is a difficult problem which may not even have a closed form solution. In [?],transceiver optimization for multi-user system using linear equalization is discussed. The objective there is to minimize the total MSE under the individual power constraint. [?] formulates the original problem and transforms it into a convex optimization problem which is then solved by using the numerical method. In [?], the transmitter of a multi-user system is optimized by considering the minimization of the transmission power under rate region constraint. The inverse problem is also posed and solved. When the rate or power is constrained in certain regions, the optimization problem can be formulated into a convex problem. Only numerical algorithm and solution are provided in the paper. In this thesis, we consider a multi-access system and exploit the MMSE-DFE to detect the received signals. Our task is to obtain an optimum design for all the K transceivers to minimize the arithmetic MSE under a novel constraint of fixed sum Gaussian mutual information.

Chapter 3

QR Interpretation of MMSE Decision Feedback Equalization

As mentioned in Chapter 2, the MMSE-DFE is a desirable receiver reasonable in implementation complexity and excellent in performance. Here in this section, we provide a new interpretation of MMSE-DF detection from the viewpoint of QR decomposition in linear algebra. For simplicity in illustration, we examine its operation in the case of a general single user.

Let $\mathbf{x} = [x_1 \ x_2 \ \cdots \ x_N]^T$ be an $N \times 1$ vector of symbols to be transmitted over a noisy channel. Each symbol x_i is chosen from a finite-size alphabet $\mathcal{X}$. Consider a general matrix channel

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \boldsymbol{\xi} \tag{3.1}$$

where $\mathbf{H}$ is a $P \times N$ channel matrix (known to the receiver), $\boldsymbol{\xi} = [\xi_1 \ \xi_2 \ \cdots \ \xi_P]^T$ is a noise vector with a covariance matrix $E(\boldsymbol{\xi}\boldsymbol{\xi}^H) = \boldsymbol{\Sigma}_{\xi\xi}$, and $\mathbf{y} = [y_1 \ y_2 \ \cdots \ y_P]^T$ is the observed received vector. Our task is to detect (estimate) the vector $\mathbf{x} \in \mathcal{X}^N$ given the noisy observation $\mathbf{y} = [y_1 \ y_2 \ \cdots \ y_P]^T$. We denote the estimate of $\mathbf{x}$ by $\hat{\mathbf{x}} = [\hat{x}_1 \ \hat{x}_2 \ \cdots \ \hat{x}_N]^T$. The matrix $\mathbf{H}$ here can be of a general format. for example, if we let $\mathbf{H} = [\mathbf{H}_1\mathbf{T}_1 \ \mathbf{H}_2\mathbf{T}_2 \ \cdots \ \mathbf{H}_K\mathbf{T}_K]$ and $\mathbf{x} = [\mathbf{x}_1^T \ \mathbf{x}_2^T \ \cdots \ \mathbf{x}_K^T]^T$, the model in (3.1) becomes that of the multi-user system in Eq. (2.11).

3.1 The Feedforward and Feedback Filter Matrix in MMSE-DFE

As we discussed in Section 2.3.3, the DF receiver, assuming perfect feedback, makes successive decision on the vector $\mathbf{z} = \mathbf{Fy} - \mathbf{Bx}$ [?, ?, ?, ?, ?], where $\mathbf{F}$ and $\mathbf{B}$ are the feedforward and feedback matrices, respectively. For the MMSE-DF receiver, the feedforward matrix in Eq. (2.6) can be written as

$$\mathbf{F}_{MMSE} = (\mathbf{B}+\mathbf{I})\mathbf{R}_{xy}\mathbf{R}_{yy}^{-1} = (\mathbf{B}+\mathbf{I})\mathbf{G}^{-1}\mathbf{H}^{H}\mathbf{\Sigma}_{\xi\xi}^{-1} \tag{3.2}$$

For notational convenience, we also denote

$$\mathbf{G} = \mathbf{I} + \mathbf{H}^{H}\mathbf{\Sigma}_{\xi\xi}^{-1}\mathbf{H} \tag{3.3}$$

We will call $\mathbf{G}^{1/2}$ the mutual information matrix. Applying the QR decomposition to $\mathbf{G}^{1/2}$ such that $\mathbf{G}^{1/2} = \mathbf{QR}$, then we have

$$\mathbf{G} = (\mathbf{QR})^{H}(\mathbf{QR}) = \mathbf{R}^{H}\mathbf{R} = (\mathbf{D}^{-1/2}\mathbf{R})^{H}\mathbf{D}(\mathbf{D}^{-1/2}\mathbf{R}) \tag{3.4}$$

where $\mathbf{D}$ is a positive diagonal matrix $\mathbf{D} = \mathrm{diag}(d_1, d_2, \cdots, d_N)$ whose ith diagonal entry is equal to the square of the corresponding diagonal element of the R-factor $\mathbf{R}$, i.e.,

$$d_i = [\mathbf{R}]_i^2 \tag{3.5}$$

Therefore the matrix $\mathbf{D}^{-1/2}\mathbf{R}$ is an upper triangular matrix with unit diagonal entries. From Eq. (2.10) and Eq. (3.4), the error covariance matrix of MMSE-DFE can be written as

$$\begin{aligned}
\mathbf{\Sigma}_{ee} &= (\mathbf{B}+\mathbf{I})(\mathbf{I}+\mathbf{H}^{H}\mathbf{\Sigma}_{\xi\xi}^{-1}\mathbf{H})^{-1}(\mathbf{B}+\mathbf{I})^{H} \\
&= (\mathbf{B}+\mathbf{I})\mathbf{G}^{-1}(\mathbf{B}+\mathbf{I})^{H} \\
&= (\mathbf{B}+\mathbf{I})\left[(\mathbf{D}^{-1/2}\mathbf{R})^{H}\mathbf{D}(\mathbf{D}^{-1/2}\mathbf{R})\right]^{-1}(\mathbf{B}+\mathbf{I})^{H}
\end{aligned} \tag{3.6}$$

Since $(\mathbf{D}^{-1/2}\mathbf{R})^{H}$ is then a lower triangular matrix, $(\mathbf{D}^{-1/2}\mathbf{R})^{H}\mathbf{D}(\mathbf{D}^{-1/2}\mathbf{R})$ is the Cholesky decomposition of $\mathbf{G}$. Denote $\mathbf{L} = (\mathbf{D}^{-1/2}\mathbf{R})^{H}$, then we can have Eq. (3.6) as

$$\begin{aligned} \mathbf{\Sigma}_{ee} &= [(\mathbf{B}+\mathbf{I})\mathbf{L}^{-H}]\mathbf{D}^{-1}[\mathbf{L}^{-1}(\mathbf{B}+\mathbf{I})^{H}] \\ &= \mathbf{Z}\mathbf{D}^{-1}\mathbf{Z}^{H} \end{aligned} \tag{3.7}$$

where $\mathbf{Z} = (\mathbf{B}+\mathbf{I})\mathbf{L}^{-H}$. Since both $\mathbf{B}+\mathbf{I}$ and $\mathbf{L}^{-H} = \mathbf{R}^{-1}\mathbf{D}^{1/2}$ are upper triangular matrices with unit diagonal entries, $\mathbf{Z}$ is an upper triangular matrix with unit diagonal entries. The mean square error is defined as

$$\mathrm{tr}(\mathbf{\Sigma}_{ee}) = \mathrm{tr}(\mathbf{Z}\mathbf{D}^{-1}\mathbf{Z}^{H}) = \mathrm{tr}(\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^{H}) \tag{3.8}$$

where $\tilde{\mathbf{Z}} = \mathbf{Z}\mathbf{D}^{-1/2}$. From the definition of trace, Eq. (3.8) can be written as

$$\begin{aligned} \mathrm{tr}(\mathbf{\Sigma}_{ee}) &= \mathrm{tr}(\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^{H}) \\ &= \sum_{n=1}^{N} \tilde{\mathbf{Z}}_{n\cdot}[\tilde{\mathbf{Z}}^{H}]_{\cdot n} \\ &= \sum_{n=1}^{N} \tilde{\mathbf{Z}}_{n\cdot}[\tilde{\mathbf{Z}}^{*}_{n\cdot}]^{T} \end{aligned} \tag{3.9}$$

Therefore, the MSE is the summation of the inner product of the columns of $\tilde{\mathbf{Z}}$ and their conjugate. Since $\mathbf{Z}$ has unit diagonal entries, we can rewrite Eq. (3.9) as

$$\begin{aligned} \mathrm{tr}(\mathbf{\Sigma}_{ee}) &= \sum_{n=1}^{N} \tilde{\mathbf{Z}}_{n\cdot}[\tilde{\mathbf{Z}}^{*}_{n\cdot}]^{T} \\ &= \sum_{n=1}^{N} d_n^{-1} + \delta \end{aligned}$$

where δ is a non-negative number. Then we can have

$$\mathrm{tr}(\mathbf{\Sigma}_{ee}) \geq \sum_{n=1}^{N} d_n^{-1}$$

where equality holds if and only if $\delta = 0$, i.e, $\mathbf{Z}$ is an identity matrix. Therefore, to minimize mean square error of DFE, we obtain that the feedback matrix must satisfy

$$(\mathbf{B}+\mathbf{I})\mathbf{L}^{-H} = (\mathbf{B}+\mathbf{I})(\mathbf{D}^{-1/2}\mathbf{R})^{-1} = \mathbf{I} \tag{3.10}$$

i.e,

$$\mathbf{B}+\mathbf{I}=\mathbf{D}^{-1/2}\mathbf{R} \tag{3.11}$$

Therefore the feedforward matrix can be re-written as

$$\begin{aligned}\mathbf{F}_{MMSE} &= \mathbf{D}^{-1/2}\mathbf{R}\mathbf{G}^{-1}\mathbf{H}^{H}\mathbf{\Sigma}_{\xi\xi}^{-1}\\ &= \mathbf{D}^{-1/2}\mathbf{R}^{-H}\mathbf{H}^{H}\mathbf{\Sigma}_{\xi\xi}^{-1}\\ &= \mathbf{D}^{-1}(\mathbf{B}+\mathbf{I})^{-H}(\mathbf{\Sigma}_{\xi\xi}^{-1/2}\mathbf{H})^{H}\mathbf{\Sigma}_{\xi\xi}^{-1/2} \qquad (3.12)\end{aligned}$$

Examining the feedforward filter matrix in Eq. (3.12), we see that the architecture of the feed-forward matrix $\mathbf{F}$ consists of four parts:

1. Whitening filter $\mathbf{\Sigma}_{\xi\xi}^{-1/2}$ – This process matches the noise covariance and whitens the spatially correlated noise.

2. Matched filter $(\mathbf{\Sigma}_{\xi\xi}^{-1/2}\mathbf{H})^{H}$ – This process matches the channel and noise and thus functions as a matched filter.

3. De-correlating filter $(\mathbf{B}+\mathbf{I})^{-H}$ – The purpose is to de-correlate the detection error so that the resulting error covariance matrix is diagonal.

4. Scaling process $\mathbf{D}^{-1}$ – Here, different subchannels are scaled by different coeeficients.

3.2 The Error Covariance Matrix in MMSE-DFE

Using Eqs. (3.11) and (3.7), the error covariance matrix of MMSE-DFE in Eq. (2.10),can be written as

$$\begin{aligned}\mathbf{\Sigma}_{ee} &= \mathbf{D}^{-1/2}\mathbf{R}\mathbf{G}^{-1}\mathbf{R}^{H}\mathbf{D}^{-1/2}\\ &= \mathbf{D}^{-1/2}\mathbf{R}\mathbf{R}^{-1}\mathbf{R}^{-H}\mathbf{R}^{H}\mathbf{D}^{-1/2}\\ &= \mathbf{D}^{-1} \qquad (3.13)\end{aligned}$$

where we can see that after MMSE-DF equalization, the error vector is uncorrelated but not white. From Eq. (3.5), the error covariance matrix in Eq. (3.13) can be rewritten in terms of the diagonal entries in the R-factor such that

$$\mathrm{E}[\mathbf{e}\mathbf{e}^H] = \mathrm{diag}\left([\mathbf{R}]_1^{-2}, [\mathbf{R}]_2^{-2}, \cdots, [\mathbf{R}]_N^{-2}\right) \tag{3.14}$$

and thus, the arithmetic MSE of MMSE-DF detection can be expressed in terms of the diagonal entries of the R-factor of the mutual information matrix $\mathbf{G}^{1/2}$ as

$$\mathcal{E} = \frac{1}{N}\sum_{n=1}^{N}[\mathbf{R}]_n^{-2} \tag{3.15}$$

3.3 QR Decomposition and MMSE-DFE

Returning to the QR decomposition of the mutual information matrix $\mathbf{G}^{1/2}$ such that $\mathbf{G}^{1/2} = \mathbf{QR}$, where $\mathbf{Q}$ denotes a $N \times N$ orthonormal matrix and $\mathbf{R}$ denotes an $N \times N$ upper triangular matrix, we have

$$\mathbf{R} = \begin{pmatrix} r_{11} & r_{12} & \dots & r_{1N} \\ 0 & r_{22} & \dots & r_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & r_{NN} \end{pmatrix}, r_{kk} > 0 \qquad \text{for } k = 1, 2, \cdots, N$$

In addition, we notice that

$$\mathbf{H}^H\mathbf{\Sigma}_{\xi\xi}^{-1}\mathbf{H} = \mathbf{G} - \mathbf{I} = \mathbf{R}^H\mathbf{R} - \mathbf{I} \tag{3.16}$$

Applying the feedforward filter matrix $\mathbf{F}_{MMSE}$ in Eq. (3.12) to the received signal vector $\mathbf{y}$ and using Eqs. (3.11) and (3.16) yields

$$\mathbf{D}^{-1/2}\mathbf{R}^{-H}\mathbf{H}^H\mathbf{\Sigma}_{\xi\xi}^{-1}\mathbf{y} = \mathbf{D}^{-1/2}(\mathbf{R} - \mathbf{R}^{-H})\mathbf{x} + \mathbf{D}^{-1/2}\mathbf{R}^{-H}\mathbf{H}^H\mathbf{\Sigma}_{\xi\xi}^{-1}\boldsymbol{\xi} \tag{3.17}$$

Therefore, after having been processed by $\mathbf{F}_{MMSE}$, the original channel in Eq. (3.1) is transformed into the following channel model,

$$\widetilde{\mathbf{y}} = \widetilde{\mathbf{R}}\mathbf{x} + \widetilde{\boldsymbol{\xi}} \tag{3.18a}$$

where

$$\widetilde{\mathbf{y}} = \mathbf{D}^{-1/2}\mathbf{R}^{-H}\mathbf{H}^{H}\boldsymbol{\Sigma}_{\xi\xi}^{-1}\mathbf{y} \tag{3.18b}$$

$$\widetilde{\boldsymbol{\xi}} = \mathbf{D}^{-1/2}\underbrace{\left(\mathrm{diag}([\mathbf{R}]_1^{-1},\cdots,[\mathbf{R}]_N^{-1}) - \mathbf{R}^{-H}\right)\mathbf{x}}_{\text{Interference}} + \mathbf{D}^{-1/2}\mathbf{R}^{-H}\mathbf{H}^{H}\boldsymbol{\Sigma}_{\xi\xi}^{-1}\boldsymbol{\xi} \tag{3.18c}$$

$$\widetilde{\mathbf{R}} = \mathbf{D}^{-1/2}(\mathbf{R} - \mathrm{diag}([\mathbf{R}]_1^{-1},\cdots,[\mathbf{R}]_N^{-1})) \tag{3.18d}$$

Notice that the reason that we subtract the diagonal elements of $\mathbf{R}^{-H}$ from $\mathbf{R}$ in Eq. (3.18d) is that we want to obtain an unbiased estimation of $\mathbf{x}$ (see more details in [**?**]). So the covariance matrix of $\widetilde{\boldsymbol{\xi}}$ is determined by

$$\boldsymbol{\Sigma}_{\widetilde{\xi}\widetilde{\xi}} = \mathbf{D}^{-1}\big[\mathrm{diag}([\mathbf{R}]_1^{-2},\cdots,[\mathbf{R}]_N^{-2}) - \mathrm{diag}([\mathbf{R}]_1^{-1},\cdots,[\mathbf{R}]_N^{-1})\big(\mathbf{R}^{-1}+\mathbf{R}^{-H}\big)+\mathbf{I}\big] \tag{3.19}$$

Thus, $[\boldsymbol{\Sigma}_{\widetilde{\xi}\widetilde{\xi}}]_k = [\mathbf{R}]_k^{-2}(1-[\mathbf{R}]_k^{-2})$. This shows that the signal to interference and noise ratio for the k-th symbol x_k is

$$\mathrm{SINR}_k = \frac{[\mathbf{R}]_k^{-2}\left([\mathbf{R}]_k - [\mathbf{R}^{-1}]_k\right)^2}{[\boldsymbol{\Sigma}_{\widetilde{\xi}\widetilde{\xi}}]_k} = [\mathbf{R}]_k^{2} - 1 \tag{3.20}$$

which is consistent with the result given in [**?**, **?**].

The above discussion establishes the equivalence between the MMSE-DFE detection and the QR decomposition. Thus, the following Algorithm 1 provides an interpretation of MMSE-DF detection for the equivalent channel model in Eq. (3.18).

Algorithm 1 (QR interpretation of MMSE-DFE):

1. *QR-decomposition.* Perform the QR-decomposition of the mutual information matrix, $\mathbf{G}^{1/2} = \mathbf{QR}$ to form the upper triangular matrix $\widetilde{\mathbf{R}}$ defined by Eq. (3.18d). We have

$$\begin{pmatrix} \widetilde{y}_1 \\ \widetilde{y}_2 \\ \vdots \\ \widetilde{y}_N \end{pmatrix} = \begin{pmatrix} \widetilde{r}_{11} & \widetilde{r}_{12} & \dots & \widetilde{r}_{1N} \\ 0 & \widetilde{r}_{22} & \dots & \widetilde{r}_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \widetilde{r}_{NN} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_L \end{pmatrix} + \begin{pmatrix} \tilde{\xi}_1 \\ \tilde{\xi}_2 \\ \vdots \\ \tilde{\xi}_N \end{pmatrix} \tag{3.21}$$

Equation (3.21) is equivalently written as

$$\tilde{y}_k = [\widetilde{\mathbf{R}}]_k x_k + \sum_{m=k+1}^{N} \widetilde{r}_{km} x_m + \tilde{\xi}_k$$

2. *Hard decision.* From the last row in Eq. (3.21) we first estimate the symbol x_N by making the minimum-error-probability hard decision $\hat{x}_N = \mathcal{Q}\left[\tilde{y}_N/[\widetilde{\mathbf{R}}]_N\right]$. The $\mathcal{Q}[x]$ operation here is defined as choosing the closest symbol to x to be the estimation $\hat{x}$.

3. *Cancelation.* Substitute the estimated symbol $\hat{x}_N$ back into the $(N-1)$-th row in Eq. (3.21) so as to remove the interference term in $\tilde{y}_{N-1}$ and then estimate x_{N-1}. Continue this procedure until we obtain the estimate of the first symbol x_1. The above procedure is described by the following recursive algorithm,

$$\begin{aligned} \hat{x}_N &= \mathcal{Q}\left[\frac{\tilde{y}_N}{[\widetilde{\mathbf{R}}]_N}\right] \\ \hat{x}_k &= \mathcal{Q}\left[\frac{\tilde{y}_k - \sum_{m=k+1}^{N} \widetilde{r}_{k,m}\hat{x}_m}{[\widetilde{\mathbf{R}}]_k}\right] \qquad \text{for } k = N-1, N-2, \cdots, 1 \end{aligned}$$

Chapter 4

The Dual water-filling Solution

The theme of this thesis is on multi-access MIMO communication systems and the goal is to obtain an optimum transceiver design for the system. The water-filling solution has been derived as the optimum solution for such designs for single-user cases, especially when the channel capacity is to be maximized subject to a power constraint. In this chapter, we examine the water-filling problem from a different perspective by considering its inverse problem. In other words, we seek a transmitter design that minimizes the total transmission power of the input signal subject to a fixed Gaussian mutual information constraint. A closed-form optimal solution is obtained by allotting the total information to each eigen-subchannel according to water-filling. This information loading scheme also provides a novel interpretation to the water-filling solution of the original problem of maximizing the Gaussian mutual information.

4.1 The Water-filling Solution

The basic model for a MIMO system is given by Eq. (2.1). We now employ a precoder $\mathbf{T}$ at the transmission end to process the data before sending out. Thus, the discrete-time baseband model for the received signal is given by

$$\mathbf{y} = \mathbf{H}\mathbf{T}\mathbf{x} + \boldsymbol{\xi} \tag{4.1}$$

where $\mathbf{H}$ is an $P \times M$ complex matrix that models the channel, $\mathbf{T}$ is an $M \times N$ linear precoding matrix ($M \geq N$), $\mathbf{x}$ is the block of N transmitted symbols, which is assumed to be zero-mean, white and of identity covariance matrix, and $\boldsymbol{\xi}$ is an $P \times 1$ white Gaussian noise vector with identity covariance matrix and independent of the input signal vector $\mathbf{x}$. It is well known that if the channel matrix $\mathbf{H}$ is known at both the transmitter and receiver, the Gaussian mutual information of model (4.1) is given by [**?**]

$$\mathcal{I}_{\mathbf{H}} = \log\det\ (\mathbf{I} + \mathbf{T}^H\mathbf{H}^H\mathbf{H}\mathbf{T}) \tag{4.2}$$

Hence, for a given transmission power constraint $\mathrm{tr}(\mathbf{T}^H\mathbf{T}) \leq p$, the Gaussian mutual information $\mathcal{I}_{\mathbf{H}}$ is maximized when the transmitter $\mathbf{T}$ is the water-filling solution [**?**, **?**]. Therefore, the capacity-achieving input for the channel model (4.1) is obtained by solving the following optimization problem:

Problem 4.1. *(***Water-filling problem***) Find a transmitter* $\mathbf{T}$ *such that the Gaussian mutual information* $\mathcal{I}_{\mathbf{H}}$ *is maximized, i.e.,*

$$\max_{\mathbf{T}} \log\det\ (\mathbf{I} + \mathbf{T}^H\mathbf{H}^H\mathbf{H}\mathbf{T})$$

subject to a total transmission power constraint,

$$\mathrm{tr}(\mathbf{T}^H\mathbf{T}) \leq p$$

If the eigenvalue decomposition of $\mathbf{H}^H\mathbf{H}$ is $\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^H$ with eigenvalues λ_m for $m = 1, \cdots, M$ arranged in a non-increasing order, then, the optimal transmitter is obtained according to the water-filling principal over the eigenvalues. More specifically, the solution selects the largest integer k not exceeding M such that

$$\frac{1}{\lambda_k} < \frac{1}{k}\left(p + \sum_{j=1}^{k} \lambda_j^{-1}\right)$$

Therefore, the optimal $\mathbf{T}$ is given by $\mathbf{T} = \mathbf{U}_k \mathbf{\Phi} \mathbf{S}$, where $\mathbf{S}$ is an arbitrary unitary matrix, $\mathbf{U}_k$ consists of the first k columns of the unitary matrix $\mathbf{U}$, and $\mathbf{\Phi}$ is a diagonal matrix with diagonal entries being

$$|\phi_{ii}|^2 = \frac{1}{k}\left(p + \sum_{j=1}^{k} \lambda_j^{-1}\right) - \lambda_i^{-1}$$

Thus, the maximum Gaussian mutual information; i.e., channel capacity, is given by

$$C = \log\left[k^{-k}\left(p + \sum_{j=1}^{k} \lambda_j^{-1}\right)^{k} \prod_{i=1}^{k} \lambda_i^{-1}\right]$$

Note that the channel capacity is not affected by the unitary matrix $\mathbf{S}$. Therefore, we have an extra degree of freedom provided by the unitary matrix $\mathbf{S}$ within the water-filling solution family which can be designed so as to improve other aspects of system performance [**?**, **?**, **?**].

4.2 The Dual water-filling Problem and its Solution

After reviewing the classic water-filling principle, we now consider the dual of the problem of maximizing the throughput.

4.2.1 Problem Statement and Necessary Lemmas

Our inverse problem of the capacity-achieving input can now be formally stated as

Problem 4.2. *(***Dual water-filling problem***) Find a transmitter* $\mathbf{T}$ *such that the total transmitted power is minimized subject to a fixed Gaussian mutual information,* *i.e.,*

$$\min_{\mathbf{T}} tr(\mathbf{T}^H\mathbf{T})$$

subject to the following Gaussian mutual information constraint:

$$\log\det\ (\mathbf{I} + \mathbf{T}^H\mathbf{H}^H\mathbf{H}\mathbf{T}) = \mathcal{I}_{\mathbf{H}}$$

It is clear that if $\mathcal{I}_{\mathbf{H}} = 0$, then, $\mathbf{T} = \mathbf{0}$ is one of the optimal solution to Problem 4.2. Therefore, in the following we only need to consider the case where $\mathcal{I}_{\mathbf{H}} \neq 0$.

In order to solve this optimization problem, we first establish the following three lemmas, the proofs of which are given in Appendices A.

Lemma 4.1. *For $\mathcal{I}_{\mathbf{H}} > 0$, let $\{a_k\}_{k=1}^n$ be a positive decreasing sequence, i.e., $a_1 \geq a_2 \geq \cdots \geq a_n$, and r_a be the largest positive integer not exceeding n such that*

$$a_k > \left(\frac{\prod_{i=1}^{r_a} a_i}{2^{\mathcal{I}_{\mathbf{H}}}}\right)^{1/r_a} \qquad \text{for } k = 1, 2, \cdots, r_a \tag{4.3}$$

Let the positive sequence $\{b_k\}_{k=1}^n$ satisfy the following two conditions,

1. *$a_1 \geq b_1 \geq a_2 \geq b_2 \geq \cdots \geq a_n \geq b_n$.*

2. *Let r_b be the maximal positive integer such that*

$$b_k > \left(\frac{\prod_{i=1}^{r_b} b_i}{2^{\mathcal{I}_{\mathbf{H}}}}\right)^{1/r_b} \qquad \text{for } k = 1, 2, \cdots, r_b. \tag{4.4}$$

Then, we have

$$\frac{1}{r}\sum_{i=1}^{r}\left(b_i^{-1} - a_i^{-1}\right) \leq \left(\frac{2^{\mathcal{I}_{\mathbf{H}}}}{\prod_{i=1}^{r} b_i}\right)^{1/r} - \left(\frac{2^{\mathcal{I}_{\mathbf{H}}}}{\prod_{i=1}^{r} a_i}\right)^{1/r}$$

where $r = \min\{r_a,\ r_b\}$, and the equality holds when $b_i = a_i$ for $i = 1, 2, \cdots, r$.

Lemma 4.2. *Let $c_1 \geq c_2 \geq \cdots \geq c_q > 0$ and r_c be the greatest integer not exceeding q such that*

$$c_k > \left(\frac{\prod_{n=1}^{r_c} c_n}{2^{\mathcal{I}_{\mathbf{H}}}} \right)^{\frac{1}{r_c}} \qquad \text{for } k = 1, 2, \cdots, r_c.$$

Then, the sequence

$$P_k = k \left(\frac{2^{\mathcal{I}_{\mathbf{H}}}}{\prod_{n=1}^{k} c_n} \right)^{\frac{1}{k}} - \sum_{n=1}^{k} c_n^{-1}$$

for $k = 1, 2, \cdots, r_c$ is strictly decreasing.

Lemma 4.3. *For any complex matrix $\mathbf{T}$ and Hermitian matrices $\mathbf{A}$ and $\mathbf{B}$, we can obtain the following derivative*

$$\frac{\partial \log \det(\mathbf{A} + \mathbf{T}^H \mathbf{B} \mathbf{T})}{\partial \mathbf{T}} = \left[\mathbf{B}\mathbf{T}(\mathbf{A} + \mathbf{T}^H \mathbf{B} \mathbf{T})^{-1} \right]^* \tag{4.5}$$

We also need the following lemma, the proof of which is given in [?]

Lemma 4.4. *Let $\mathbf{M}$ be an $M \times N$ matrix and $\bar{\mathbf{M}}_n$ be the remaining matrix by deleting the nth column from $\mathbf{M}$. If we let $\{\sigma_i\}$ and $\{\bar{\sigma}_i\}$ denote the singular value sequences of $\mathbf{M}$ and $\bar{\mathbf{M}}_n$, respectively, both arranged in nonincreasing order, then, we have the following two statements:*

1. *If $M \geq N$, then,*

$$\sigma_1 \geq \bar{\sigma}_1 \geq \sigma_2 \geq \bar{\sigma}_2 \geq \cdots \geq \bar{\sigma}_{N-1} \geq \sigma_N \geq 0$$

2. *If $M < N$, then,*

$$\sigma_1 \geq \bar{\sigma}_1 \geq \sigma_2 \geq \bar{\sigma}_2 \geq \cdots \geq \sigma_M \geq \bar{\sigma}_M \geq 0$$

4.2.2 The Optimal Solution

Now, we are in a position to formally state our main result.

Theorem 4.1. *Let the Gaussian mutual information $\mathcal{I}_{\mathbf{H}}$ for the channel model (4.1) be given by Eq. (4.2) and the eigen-value decomposition of $\mathbf{H}^H\mathbf{H}$ be*

$$\mathbf{H}^H\mathbf{H} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$$

where $\mathbf{U}$ is an $M \times M$ $(M \geq 1)$ unitary matrix and $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_M)$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_M > 0$. If we let r be the maximal positive integer not exceeding M such that

$$\lambda_m > \left(\frac{\prod_{i=1}^{r}\lambda_i}{2^{\mathcal{I}_{\mathbf{H}}}}\right)^{1/r} \qquad \text{for } m = 1, 2, \cdots, r \tag{4.6}$$

then, the optimal solution, $\mathbf{T}_{\mathrm{opt}}$, of Problem 4.2 is an $M \times r$ tall matrix, given by

$$\mathbf{T}_{\mathrm{opt}} = \mathbf{U}_r\mathbf{\Gamma}\mathbf{V}^H \tag{4.7}$$

where $\mathbf{U}_r$ is an $M \times r$ matrix consisting of the first r columns of the unitary matrix $\mathbf{U}$, $\mathbf{V}$ is an arbitrarily $r \times r$ unitary matrix and $\mathbf{\Gamma} = \mathrm{diag}(\gamma_1, \gamma_2, \cdots, \gamma_r)$ with each γ_m determined by

$$\gamma_m = \sqrt{\left(\frac{2^{\mathcal{I}_{\mathbf{H}}}}{\prod_{i=1}^{r}\lambda_i}\right)^{1/r} - \lambda_m^{-1}} \tag{4.8}$$

The minimum power is determined by

$$P = r\left(\frac{2^{\mathcal{I}_{\mathbf{H}}}}{\prod_{i=1}^{r}\lambda_i}\right)^{1/r} - \sum_{m=1}^{r}\lambda_m^{-1} \tag{4.9}$$

Proof: Introducing the Lagrangian function, $\mathcal{L}(\mathbf{T})$, of the original Problem 4.2

$$\mathcal{L}(\mathbf{T}) = \mathrm{tr}(\mathbf{T}^H\mathbf{T}) - \rho \log\det(\mathbf{I} + \mathbf{T}^H\mathbf{H}^H\mathbf{H}\mathbf{T})$$

where ρ is the Lagrange multiplier, and requiring that the gradient of $\mathcal{L}(\mathbf{T})$ vanishes, using Lemma 4.3 we have

$$\mathbf{T}^* - \rho\left[\mathbf{H}^H\mathbf{H}\mathbf{T}(\mathbf{I} + \mathbf{T}^H\mathbf{H}^H\mathbf{H}\mathbf{T})^{-1}\right]^* = \mathbf{0}$$

i.e.,

$$\mathbf{T} - \rho\,\mathbf{H}^H\mathbf{H}\mathbf{T}(\mathbf{I} + \mathbf{T}^H\mathbf{H}^H\mathbf{H}\mathbf{T})^{-1} = \mathbf{0} \tag{4.10}$$

Left-multiplying both sides of Eq. (4.10) by $\mathbf{T}^H$ yields

$$\mathbf{T}^H\mathbf{T} + \rho(\mathbf{I} + \mathbf{T}^H\mathbf{H}^H\mathbf{H}\mathbf{T})^{-1} = \rho\mathbf{I} \tag{4.11}$$

Let the singular value decomposition of $\mathbf{T}$ be

$$\mathbf{T} = \mathbf{W}\mathbf{\Gamma}\mathbf{V}^H$$

where $\mathbf{W}$ is an $M \times N$ column-wise orthonormal matrix, $\mathbf{V}$ is an $N \times N$ matrix and $\mathbf{\Gamma}$ is a diagonal matrix $\mathbf{\Gamma} = \text{diag}(\gamma_1, \gamma_2, \cdots, \gamma_N)$ with $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_N > 0$. Substituting this decomposition into Eq.(4.11) results in

$$\mathbf{\Gamma}^H\mathbf{\Gamma} + \rho(\mathbf{I} + \mathbf{\Gamma}^H\mathbf{W}^H\mathbf{H}^H\mathbf{H}\mathbf{W}\mathbf{\Gamma})^{-1} = \rho\mathbf{I} \tag{4.12}$$

Since $\mathbf{T}$ is required to be of full column rank, $\mathbf{\Gamma}$ must be invertible. From Eq. (4.12), since all the other matrices are diagonal $\mathbf{W}^H\mathbf{H}^H\mathbf{H}\mathbf{W}$ must be diagonal. Let $\mathbf{W}^H\mathbf{H}^H\mathbf{H}\mathbf{W} = \mathbf{\Theta} = \text{diag}(\theta_1, \theta_2, \cdots, \theta_N)$ with each θ_k being non-negative. Then, Eq. (4.12) can be rewritten as

$$\mathbf{\Gamma}^H\mathbf{\Gamma} + \rho(\mathbf{I} + \mathbf{\Gamma}^H\mathbf{\Theta}\mathbf{\Gamma})^{-1} = \rho\mathbf{I}$$

Since both $\mathbf{\Gamma}$ and $\mathbf{\Theta}$ are diagonal matrices, we can easily equate the diagonal elements resulting in

$$\gamma_n = \sqrt{\rho - \theta_n^{-1}} \tag{4.13}$$

In this case, the Gaussian mutual information constraint can be expressed in terms of γ_n and θ_n as

$$\det(\mathbf{I} + \mathbf{\Gamma}^H\mathbf{\Theta}\mathbf{\Gamma}) = \prod_{n=1}^{N}(1 + \gamma_n^2\theta_n) = 2^{\mathcal{I}_{\mathbf{H}}}$$

Combining this with Eq. (4.13) yields

$$\rho = \left(\frac{2^{\mathcal{I}_{\mathbf{H}}}}{\prod_{n=1}^{N} \theta_n} \right)^{1/N}$$

and as a result, the power in the mth subchannel is given by

$$\gamma_m^2 = \left(\frac{2^{\mathcal{I}_{\mathbf{H}}}}{\prod_{n=1}^{N} \theta_n} \right)^{1/N} - \theta_m^{-1} \tag{4.14}$$

Therefore, the resulting total power, $\mathcal{F}(\theta_1, \theta_2, \cdots, \theta_N)$, is given by

$$\mathcal{F}(\theta_1, \theta_2, \cdots, \theta_N) = N \left(\frac{2^{\mathcal{I}_{\mathbf{H}}}}{\prod_{n=1}^{N} \theta_n} \right)^{1/N} - \sum_{m=1}^{N} \theta_m^{-1}$$

where $\theta_1, \theta_2, \cdots, \theta_N$ must satisfy the following two constraints:

1. Positivity of the power of each subchannel:

$$\theta_n > \left(\frac{\prod_{i=1}^{N} \theta_i}{2^{\mathcal{I}_{\mathbf{H}}}} \right)^{1/N} \qquad \text{for} \;\; n = 1, 2, \cdots, N$$

2. $\lambda_1 \geq \theta_1 \geq \lambda_2 \geq \theta_2 \geq \cdots \geq \lambda_{N-1} \geq \theta_{N-1} \geq \lambda_N \geq \theta_N$, as a result of Lemma 4.4.

Thus, the proof of Theorem 4.1 is reduced to finding an optimal $\boldsymbol{\theta}$ that minimizes $\mathcal{F}(\theta_1, \theta_2, \cdots, \theta_N)$ subject to the above two constraints. Now, from Lemma 4.1, for r_λ to be the largest integer that satisfies

$$\lambda_n > \left(\frac{\prod_{i=1}^{r_\lambda} \lambda_i}{2^{\mathcal{I}_{\mathbf{H}}}} \right)^{1/r_\lambda}, \qquad \text{for } n = 1, 2, \cdots, r_\lambda$$

we can obtain $\mathcal{F}(\theta_1, \theta_2, \cdots, \theta_r) \geq \mathcal{F}(\lambda_1, \lambda_2, \cdots, \lambda_r)$, where $r = \min\{N,\ r_\lambda\}$ and where equality holds when $\lambda_m = \theta_m$, for $i = 1,\ 2, \cdots, r$. Then applying Lemma 4.2, we can obtain the last integer variable N by observing that the total power P is a decreasing function with respect to r, P is minimized when $N = r_\lambda$. This completes the proof of Theorem 4.1.

4.2.3 Further Discussion

While providing a closed form solution of the optimum transmitter and the minimum transmission power for the dual water-filling Problem, Theorem 4.1 also provides us with a novel outlook of the water-filling solution. Suppose the power is constrained on the transmitter matrix $\mathbf{T}$, i.e., $\mathrm{tr}(\mathbf{T}^H\mathbf{T}) \le p$. We attempt to maximize the Gaussian mutual information $\mathcal{I}_{\mathbf{H}}$. From Eq. (4.9), the power constraint on the transmitter can be changed to

$$P = r\left(\frac{2^{\mathcal{I}_{\mathbf{H}}}}{\prod_{n=1}^{r}\lambda_n}\right)^{1/r} - \sum_{n=1}^{r}\lambda_n^{-1} \le p$$

which leads to

$$2^{\mathcal{I}_{\mathbf{H}}} \le \left(\frac{p+\sum_{n=1}^{r}\lambda_n^{-1}}{r}\right)^{r}\prod_{n=1}^{r}\lambda_n \tag{4.15}$$

Therefore, the Gaussian mutual information $\mathcal{I}_{\mathbf{H}}$ is maximized when equality holds in Eq. (4.15). In other words, the channel capacity is achieved and equal to

$$C = r\log\left(\frac{p+\sum_{n=1}^{r}\lambda_n^{-1}}{r}\right) + \sum_{n=1}^{r}\log\lambda_n \tag{4.16}$$

Substituting this maximum information into the power loading solution in Eq. (4.8) yields

$$\gamma_n^2 = \frac{p+\sum_{k=1}^{r}\lambda_k^{-1}}{r} - \lambda_n^{-1}$$

which is exactly the water-filling solution.

4.3 Why Use the Inverse Problem?

The maximum Gaussian mutual information, i.e., channel capacity is the fundamental limit for reliable data communications [?, ?, ?]. From an information theoretic viewpoint, maximizing throughput is a major concern and is thus an important design criterion for the transmitter. Therefore, this criterion has been used by many researchers to design capacity-achieving transmitters not only for a variety of single user and deterministic channel models [?, ?, ?], but for different kinds of multiple users [?, ?], multiple input and multiple output random channel models as well [?, ?, ?, ?, ?].

From the inverse perspective, if we fix the mutual information of the communication system, we are proposing a requirement of the channel capacity. Since the objective is for the total transmission power, this design solves the problem of: "what is the minimum power needed to guarantee such amount of information transmitted reliably through the channel". From the results, we find that, the optimal solution of the dual water-filling solution performs in a similar way to the famous water-filling solution. However, in the inverse problem, it is the mutual information that is allocated to the different subchannels.

It should be noted that similar research work to Problem 4.2 can be found in [?] where the delay-limited capacity (DLC) for the general class of fading channel in MISO, SIMO and MIMO was derived, and the impact of the mean component and spatial correlation on the bounds of DLC was characterized. Also, in [?], the power region and capacity region are characterized under rate and power constraints for the fading multi-access channels and fading broadcast channels with multiple transmitter and receiver antennas. In general, there is no closed-form analytic solution for these optimal power and rate allocation problems. Therefore, efficient numerical methods have been developed and provided in [?].

Chapter 5

Joint Design of Transceivers for Multiple Access Channel

With the dual water-filling solution in mind, we now consider the joint design of the transceivers for a multiple access ISI MIMO system equipped with the MMSE-DFE. The goal of our design is to minimize the arithmetic MSE for the K users subject to a fixed sum Gaussian mutual information constraint.

5.1 Problem Statement

Since the design objective is to minimize the arithmetic MSE of the K users, let us first form our design problem into a certain optimization problem.

5.1.1 Arithmetic MSE and the Original Problem

The multi-user MIMO system model is shown in Eq.(2.11) of Section 2.5 such that

$$\mathbf{y} = \sum_{k=1}^{K} \mathbf{H}_k \mathbf{T}_k \mathbf{x}_k + \boldsymbol{\xi} \tag{5.1}$$

In general, for a DF receiver, signals are detected in the reverse order of the symbol index. Here, we follow this reverse order for user detection, i.e., we first detect the signal from User K, then User $K-1$, and so on under the assumption the base station has the signals of all the K users at all time. Based on this detection order, we thus re-write the received signal in Eq. (5.1) as

$$\begin{aligned}\mathbf{y} - \sum_{i=k+1}^{K} \mathbf{H}_i \mathbf{T}_i \mathbf{x}_i &= \mathbf{H}_k \mathbf{T}_k \mathbf{x}_k + \underbrace{\sum_{\ell=1}^{k-1} \mathbf{H}_\ell \mathbf{T}_\ell \mathbf{x}_\ell + \boldsymbol{\xi}}_{\boldsymbol{\zeta}_k} \\ &= \mathbf{H}_k \mathbf{T}_k \mathbf{x}_k + \boldsymbol{\zeta}_k \qquad \text{for } k = K, 2, \cdots, 1 \qquad (5.2)\end{aligned}$$

where $\boldsymbol{\zeta}_k$ is the kth interference-plus-noise vector. In Eq. (5.2), the MMSE-DFE is used to detect $\mathbf{x}_k$ from the received signal $\mathbf{y}$ by successively canceling the previously detected user signals. Therefore, the resulting error vector for the kth user of the MMSE successive cancelation detection is defined as $\mathbf{e}_k = \hat{\mathbf{x}}_k - \mathbf{x}_k$. Then, $\mathbf{e}_k = \hat{\mathbf{x}}_k - \mathbf{x}_k = (\mathbf{F}_k \mathbf{H}_k - \mathbf{I} - \mathbf{B}_k)\mathbf{x}_k + \mathbf{F}_k \boldsymbol{\xi}_k$, where $\mathbf{F}_k$ and $\mathbf{B}_k$ are the feedforward and feedback matrices for the kth user respectively. With the results in Chapter 3, using the matrix inversion lemma [?] leads to the following error covariance matrix for User k [?, ?, ?, ?]:

$$\begin{aligned}\boldsymbol{\Sigma}_{ee_k} &= E[\mathbf{e}_k \mathbf{e}_k^H] \\ &= (\mathbf{B}_k + \mathbf{I})(\mathbf{G}_k)^{-1}(\mathbf{B}_k + \mathbf{I})^H \\ &= \text{diag}([\mathbf{R}_k]_1^{-2}, [\mathbf{R}_k]_2^{-2}, \cdots, [\mathbf{R}_k]_{N_k}^{-2}) \qquad (5.3)\end{aligned}$$

where

$$\mathbf{G}_k = \mathbf{I} + (\mathbf{H}_k \mathbf{T}_k)^H (\boldsymbol{\Sigma}_k)^{-1} \mathbf{H}_k \mathbf{T}_k$$

and $\boldsymbol{\Sigma}_k$ is the covariance of the interference and noise. Since the independence of signals from different users and noise, and the assumption about the noise, i.e. $E[\mathbf{x}_i \mathbf{x}_k] = 0$, $E[\mathbf{x}_k \boldsymbol{\xi}_k] = 0$ and $E[\boldsymbol{\xi}\boldsymbol{\xi}^H] = \mathbf{I}$, we can obtain that

$$\begin{aligned}\boldsymbol{\Sigma}_k &= E[\boldsymbol{\zeta}_k \boldsymbol{\zeta}_k^H] = \mathbf{I} + \sum_{\ell=1}^{k-1} \mathbf{H}_\ell \mathbf{T}_\ell (\mathbf{H}_\ell \mathbf{T}_\ell)^H \\ \boldsymbol{\Sigma}_1 &= \mathbf{I}\end{aligned}$$

for $k = 1, 2, \cdots, K$. $\mathbf{R}_k$ is the upper triangular matrix in the QR-decompostion of $\mathbf{G}_k^{1/2}$. If we define the average MSE of the K users of the successive cancellation detector as

$$\mathcal{E} \triangleq \frac{1}{N}\sum_{k=1}^{K} \operatorname{tr}\left(\mathrm{E}[\mathbf{e}_k\mathbf{e}_k^H]\right) = \frac{1}{N}\sum_{k=1}^{K} \operatorname{tr}\left(\mathbf{\Sigma}_{ee_k}\right) \tag{5.4}$$

where $N = \sum_{k=1}^{K} N_k$, then, our optimization problem can be formally stated as follows:

Problem 5.1. *Let* $\operatorname{rank}(\mathbf{H}_k) = L_k$, $k = 1, 2, \cdots, K$. *Then, given K non-negative integers $N_1, N_2, \cdots, N_K$ with $N_k \leq L_k$, where N_k is the length of the transmitted signal vector x_k, find the matrix sequence $\{\mathbf{T}_k\}_{k=1}^K$ such that*

1. *the MMSE for the K users of the MMSE-DF detection is first minimized, subject to a fixed sum mutual information constraint, i.e.,*

$$\{\widetilde{\mathbf{T}}_k\}_{k=1}^K = \arg \min_{\{\mathbf{T}_k\}_{k=1}^K} \mathcal{E} \tag{5.5}$$

s.t.

$$\mathcal{I} = \log\det\left(\mathbf{I} + \sum_{k=1}^{K} \mathbf{H}_k\mathbf{T}_k\mathbf{T}_k^H\mathbf{H}_k^H\right) \tag{5.6}$$

2. *then, with respect to all the remaining free parameters, the transmission power for each user is minimized sucessively.*

5.1.2 Problem Reformulation

In order to solve Problem 5.1, we employ the inequality relationship between the trace and determinant of a square matrix: for any positive semi-definite matrix $\mathbf{M}$, we have the relationship that $\operatorname{tr}(\mathbf{M}) \geq \det(\mathbf{M})$, so that the total system error of the MMSE-DFE in Eq. (5.4) is lower-bounded by

$$\mathcal{E} \geq \frac{1}{N}\sum_{k=1}^{K} N_k \det\left(\mathbf{G}_k^{-1/N_k}\right) \tag{5.7a}$$

$$= \frac{1}{N}\sum_{k=1}^{K} N_k \det\left(\mathbf{I} + \mathbf{T}_k^H\mathbf{H}_k^H(\mathbf{\Sigma}_k)^{-1}\mathbf{H}_k\mathbf{T}_k\right)^{-1/N_k} \tag{5.7b}$$

For matrices $\mathbf{A}$ and $\mathbf{B}$ of compatible dimensions, we have the Sylvester's determinant theorem $\det(\mathbf{I}+\mathbf{AB}) = \det(\mathbf{I}+\mathbf{BA})$ and property such that $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$ [**?**], for each $\det(\mathbf{G}_k)$ in Eq. (5.7a), we have

$$\begin{aligned} & \det\left(\mathbf{I}+\mathbf{T}_k^H\mathbf{H}_k^H(\mathbf{\Sigma}_k)^{-1}\mathbf{H}_k\mathbf{T}_k\right) \\ = & \det\left(\mathbf{I}+\mathbf{H}_k\mathbf{T}_k\mathbf{T}_k^H\mathbf{H}_k^H(\mathbf{\Sigma}_k)^{-1}\right) \\ = & \det\left(\mathbf{\Sigma}_k+\mathbf{H}_k\mathbf{T}_k\mathbf{T}_k^H\mathbf{H}_k^H\right)\det(\mathbf{\Sigma}_k)^{-1} \\ = & \frac{\det(\mathbf{\Sigma}_{k+1})}{\det(\mathbf{\Sigma}_k)} \end{aligned} \tag{5.8}$$

Also, we apply the inequality relationship between the geometric and arithmetic mean such that $\frac{1}{N}\sum_{n=1}^{N} x_n \geq \left(\prod_{n=1}^{N} x_N\right)^{1/N}$ where the equality holds if and only if $x_1 = x_2 = \cdots = x_N$. Thus, the following inequality applies to the right-hand side of Eq. (5.7b)

$$\frac{1}{N}\sum_{k=1}^{K} N_k \frac{\det(\mathbf{\Sigma}_k)^{1/N_k}}{\det(\mathbf{\Sigma}_{k+1})^{1/N_k}} \geq \det(\mathbf{\Sigma}_{K+1})^{-1/N} = 2^{-\frac{\mathcal{I}}{N}} \tag{5.9}$$

Equality in Eq. (5.7a) holds if and only if matrices $\mathbf{G}_k^{1/2}$ have equal diagonal R-factors, i.e.,

$$[\mathbf{R}_k]_1 = [\mathbf{R}_k]_2 = \cdots = [\mathbf{R}_k]_{N_k} \tag{5.10}$$

Hence $\mathcal{E}$ reaches its minimum value when the condition in Eq. (5.10) holds. These equal diagonal entries, in the DF receiver, mean that the mutual information of the currently detected user is uniformly distributed over each individual symbol within the block signal of the user when all the previous user signals have been perfectly detected. Equality in Eq. (5.9) holds if and only if $\det(\mathbf{\Sigma}_k)$ constitutes a geometrical sequence, i.e.,

$$\left(\frac{\det(\mathbf{\Sigma}_1)}{\det(\mathbf{\Sigma}_2)}\right)^{1/N_1} = \cdots = \left(\frac{\det(\mathbf{\Sigma}_K)}{\det(\mathbf{\Sigma}_{K+1})}\right)^{1/N_K} \tag{5.11}$$

which means the averaged sum mutual information is uniformly distributed over each individual user if the mutual information of each user is defined as $\frac{1}{N_k}\log\det(\mathbf{G}_k)$ and is equivalent to

$$\det\left(\mathbf{G}_k\right) = 2^{\frac{N_k}{N}\mathcal{I}} \tag{5.12}$$

Therefore, solving Problem 5.1 is finally reduced to solving the following optimization problem:

Problem 5.2. *For any given K non-negative integers $N_1, N_2, \cdots, N_K$ with $N_k \leq L_k$, find a sequence of matrices $\{\mathbf{T}_k\}_{k=1}^K$ such that*

1. *the total power for the kth user is minimized subject to the constraints that the mutual information for User k is $\mathcal{I}_k = \log\det(\mathbf{G}_k) = \frac{N_k}{N}\mathcal{I}$.*
2. *within the space of the remaining parameters, Condition in Eq. (5.10) is satisfied.*

5.2 The Optimal Solution

Examining the reformulated problem stated in the foregoing section, the first requirement in Problem 5.2 can be satisfied by using the result in the dual water-filling for the single-user system in Chapter 4. After this, the lower bound of the average MSE is fixed, i.e., the inequality in Eq. (5.9) holds with equality. To meet the second requirement, we need to exploit a property of the optimal solution. If we modify the matrix $\mathbf{G}_k$ in Eq. (5.7b) by attaching a unitary matrix $\mathbf{S}_k$ to $\mathbf{T}_k$ such that $\tilde{\mathbf{T}}_k = \mathbf{T}_k\mathbf{S}_k$, then, we have

$$\begin{aligned}
\det(\tilde{\mathbf{G}}_k)^{-1} &= \det(\mathbf{I} + (\mathbf{T}_k\mathbf{S}_k)^H\mathbf{H}_k^H\mathbf{\Sigma}_k^{-1}\mathbf{H}_k\mathbf{T}_k\mathbf{S}_k)^{-1} \\
&= \det\left(\mathbf{S}_k^H(\mathbf{I} + \mathbf{T}_k^H\mathbf{H}_k^H\mathbf{\Sigma}_k^{-1}\mathbf{H}_k\mathbf{T}_k)\mathbf{S}_k\right)^{-1} \\
&= \det(\mathbf{I} + \mathbf{T}_k^H\mathbf{H}_k^H\mathbf{\Sigma}_k^{-1}\mathbf{H}_k\mathbf{T}_k)^{-1} = \det(\mathbf{G}_k)^{-1}
\end{aligned}$$

Therefore, attaching a unitary matrix to $\mathbf{T}_k$ does not affect the value of the lower bound of the average MSE, but the trace of the error covariance matrix would change with different choices of the unitary matrix. Applying the QRS decomposition [**?**] to the mutual information matrix $\mathbf{G}_k$ yields $\mathbf{G}_k^{1/2}\mathbf{S}_k = \mathbf{Q}_k\mathbf{R}_k$ with $\mathbf{R}_k$ having equal diagonal elements. Thus, the condition of Eq. (5.10) is met.

Therefore, the above development for finding the optimal solution of Problem 5.1 can be summarized to form the following theorem:

Theorem 5.1. *Given any K non-negative integers $N_1, N_2, \cdots, N_K$ with $N_k \leq L_k$, let*

$$\mathcal{A}_k = \mathbf{H}_k^H(\mathbf{\Sigma}_k)^{-1}\mathbf{H}_k \qquad \text{for } k = 1, 2, \cdots, K$$

and let the eigen-decomposition of $\mathcal{A}_k$ be $\mathcal{A}_k = \mathbf{U}_k\mathbf{\Lambda}_k(\mathbf{U}_k)^H$ with the diagonal elements in $\mathbf{\Lambda}_k$ arrange in non-increasing order. Then, the optimal solution to Problem 5.1 is given by

$$\widetilde{\mathbf{T}}_k \ = \ \mathbf{U}_{N_k,k}(\mathbf{\Gamma}_k)^{1/2}\mathbf{S}_k, \qquad k = 1, 2, \cdots, K \tag{5.13}$$

where $\mathbf{U}_{N_k,k}$ is the first N_k columns of $\mathbf{U}_k$, $\mathbf{S}_k$ is an $N_k \times N_k$ unitary matrix denoting the S-factors of the QRS decomposition of $\mathbf{G}_k^{1/2}$, and N_k is a pre-assigned subchannel number for the kth user. For the k-th user, let r_k be the maximal positive integers such that

$$\lambda_{n,k} > \left(\frac{\prod_{i=1}^{r_k}\lambda_{i,k}}{2^{\mathcal{I}_k}}\right)^{1/r_k} \qquad \textit{for } n = 1, 2, \cdots, r_k \tag{5.14}$$

If $N_k \leq r_k$, the diagonal entries of $\mathbf{\Gamma}_k$ are determined by

$$\gamma_{n,k} = \left(\frac{2^{\mathcal{I}_k}}{\prod_{i=1}^{N_k}\lambda_{i,k}}\right)^{1/N_k} - (\lambda_{n,k})^{-1} \tag{5.15}$$

for $n = 1, 2, \cdots, N_k$. If $N_k > r_k$, the diagonal entries of Γ_k are assigned by

$$\gamma_{n,k} = \begin{cases} \left(\frac{2^{\mathcal{I}_k}}{\prod_{i=1}^{r_k}\lambda_{i,k}}\right)^{1/r_k} - (\lambda_{n,k})^{-1} & n = 1, \cdots, r_k \\ 0 & n = r_k + 1, \cdots, N_k \end{cases}$$

Theorem 5.1 tells us that the optimal solution of Problem 5.1 is achieved if and only if

1. the mutual information of each user per active subchannel is uniformly distributed among all users, i.e., user mutual information uniform distribution

2. the mutual information of each user under perfect feedback is uniformly distributed among individual symbols within the signal block of the user transmitted over the active subchannels; i.e., symbol mutual information uniform distribution.

A more detail explanation on this is given in ensuing section.

5.3 Optimality Discussion

In this section, we will further explain the two optimality conditions stated above. Then, we will show that such uniform distribution of the sum mutual information has two optimality properties.

5.3.1 Decomposition of Sum Gaussian Mutual Information

To decompose the sum Gaussian mutual information, we need to first establish the following lemma.

Lemma 5.1. *Let* $\mathbf{H} = [\mathbf{H}_1 \ \mathbf{H}_2 \ \cdots \ \mathbf{H}_K]$*. Then, the sum mutual information matrix* $\mathbf{G}^{1/2} = (\mathbf{I} + \mathbf{H}^H\mathbf{H})^{1/2}$ *of* $\mathbf{H}$ *can be decomposed as*

$$\mathbf{G}^{1/2} = \mathbf{QR} \tag{5.16}$$

where $\mathbf{R}$ *is an upper triangular matrix with the* (i, j)*th block matrix being*

$$\mathbf{R}_{ij} = \begin{cases} \mathbf{G}_i^{1/2} & \textit{if } i = j \\ \mathbf{G}_i^{-1/2}\mathbf{H}_i^H\mathbf{\Sigma}_i^{-1}\mathbf{H}_j & \textit{if } i < j \end{cases} \tag{5.17}$$

with $\mathbf{\Sigma}_i = \mathbf{I} + \sum_{k=0}^{i-1}\mathbf{H}_k\mathbf{H}_k^H$ $(\mathbf{\Sigma}_1 = \mathbf{I})$ *and* $\mathbf{G}_i = \mathbf{I} + \mathbf{H}_i^H\mathbf{\Sigma}_i^{-1}\mathbf{H}_i$*.*

The proof of Lemma 5.1 is in Appendix B.1. Although it is a specific application of a block QR or the block Cholesky decomposition, Lemma 5.1 gives us a closed-form block R-factor for the sum mutual information matrix and provides a simple and clear relationship between the sum mutual information and the mutual information of each individual user with the MMSE-DF detector. This will help us easily understand the optimal solution of Problem 5.1 given in Theorem 5.1 from the viewpoint of information theory.

Under the assumption that the channel matrix $\mathbf{H}$ is known to both the receiver and the transmitter, the Gaussian sum mutual information for the precoded channel model in Eq. (2.11) is given by [**?**],

$$I_G(\mathbf{x};\mathbf{y}) = \log\det\big(\mathbf{I} + \sum_{k=1}^{K} \mathbf{H}_k\mathbf{T}_k(\mathbf{H}_k\mathbf{T}_k)^H\big) \tag{5.18}$$

In order to give interpretation of the optimal transmitter pairs derived in Section 5.2 from an information theoretic viewpoint, we rewrite channel model in Eq. (2.11) as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \boldsymbol{\xi} \tag{5.19}$$

where $\mathbf{H} = [\mathbf{H}_1\mathbf{T}_1 \ \mathbf{H}_2\mathbf{T}_2 \ \cdots \ \mathbf{H}_K\mathbf{T}_K]$ represents the precoded channel. Therefore the original channel model, Eq. (2.11), can be mathematically treated as the virtual MIMO channel model Eq. (5.19). Correspondingly, the Gaussian sum mutual information expressed in Eq. (5.18) can be regarded as the Gaussian mutual information of Eq. (5.19) with a white Gaussian input signal vector $\mathbf{x}$. Therefore, we can employ the results in [**?**] as the following lemma.

Lemma 5.2. *Let $\mathbf{R}$ denote the R-factor of $\mathbf{H}$. Then, under an assumption of error-free feedback, the mutual information between the $(N-i)$th symbol (or user) x_{N-i} and $\mathbf{y}$ conditional on $\mathbf{x}_N^{N-i+1} = [x_N, x_{N-1}, \cdots, x_{N-i+1}]$ for the model in Eq. (2.11) can be expressed as [**?**]*

$$I(x_{N-i};\mathbf{y}|\mathbf{x}_N^{N-i+1}) = \log([\mathbf{R}^2]_{N-i}) \tag{5.20}$$

for $i = 0, 1, \cdots, N-1$.

Therefore, by Lemma 5.1 we have

$$\begin{aligned} I_G(\mathbf{x};\mathbf{y}) &= \sum_{i=0}^{N-1} I(x_{N-i};\mathbf{y}|\mathbf{x}_N^{N-i+1}) \\ &= \sum_{k=1}^{K} \sum_{i=\widetilde{N}_{k-1}}^{\widetilde{N}_k-1} I(x_{N-i};\mathbf{y}|\mathbf{x}_N^{N-i+1}) \qquad (5.21) \\ &= \sum_{k=1}^{K} \sum_{i=\widetilde{N}_{k-1}}^{\widetilde{N}_k-1} \log([\mathbf{R}^2]_{N-i}) \end{aligned}$$

where $\widetilde{N}_k = \sum_{\ell=K}^{K-k+1} N_\ell$ and $\widetilde{N}_0 = 0$. Since $\mathbf{R}$ is a block upper triangular matrix with the diagonal matrix being the R-factor of the QR decomposition of $\mathbf{G}_k$, we can obtain that

$$\sum_{i=\widetilde{N}_{k-1}}^{\widetilde{N}_k-1} \log([\mathbf{R}^2]_{N-i}) = \log\det(\mathbf{G}_k)$$

which indicates each user's mutual information can be decomposed into the summation of each subchannel's mutual information without any loss, and further

$$I_G(\mathbf{x};\mathbf{y}) = \sum_{k=1}^{K} \log\det(\mathbf{G}_k) = \log\det\left(\mathbf{I} + \sum_{k=1}^{K} \mathbf{H}_k\mathbf{T}_k(\mathbf{H}_k\mathbf{T}_k)^H\right)$$

This shows that the sum mutual information is decomposed into the summation of each user's mutual information. For a given matrix $\mathbf{H}$, its singular values are fixed under any unitary transformation and hence, its eigen-subchannel mutual information does not change. However, the R-factor diagonal values of the mutual information matrix change with the unitary transformation. As a result, the capacity of each R-factor-value subchannel in Eq. (3.18) for the MMSE-DF detector will change too. In other words, different unitary transmitters lead to different R-factors and hence, different R-factor value subchannel capacities and different detection error performances for the MMSE-DF detector.

5.3.2 Even Distribution of Mutual Information

From the above discussion, a natural question arises: What is the optimal (in terms of minimizing the mean square error) distribution of mutual information among the R-factor value subchannels in Eq. (3.18) for the MMSE-DF detector? The answer is: from both the information theoretic viewpoint and the signal detection error viewpoint, the condition of uniformly distributed mutual information is optimal. This uniformity of distribution is effected by applying the S-factor of the QRS decomposition to the mutual information matrix. Therefore, we have the following statement.

Property 5.1. *(Uniform decomposition of mutual information for the MMSE-DF detector) Under the assumption of error-free feedback, the sum Gaussian mutual information for a block-by-block precoded multiple access MIMO channel in Eq. (2.11) can be uniformly decomposed into the sum of each R-factor value subchannel in Eq. (3.18) with $\mathbf{H} = \widetilde{\mathbf{H}}$ for the MMSE-DF detector by rotating the input signal vector with the S-factor of each user's mutual information matrix $\widetilde{\mathbf{G}}_k^{1/2}$.*

Uniform decomposition of the sum Gaussian mutual information, in addition to minimizing the MSE of MMSE-decision feedback detection described by Theorem 5.1, also has the following two optimality properties. Suppose we wish to use the VBLAST detector [**?**] based on the MMSE-DF detector for the optimal system designed in the previous section. A natural question is: What is the optimal detection order?

Property 5.2. *If the mutual information matrix of a channel matrix has an equal-diagonal R-factor, the optimal detection order (that ensures that the high SINR components are detected first) is the natural order, i.e., $x_N \rightarrow x_{N-1} \rightarrow \cdots \rightarrow x_1$, in other words, the i-th symbol to be detected is the symbol x_{N+1-i}.*

Proof: Let the QR decomposition of $\mathbf{G}^{1/2}$ be $\mathbf{G}^{1/2} = \mathbf{QR}$. Then, according to the QR interpretation of MMSE-DF detection given in Chapter 3, we know that the SINR_k of k-th symbol is $\text{SINR}_k = [\mathbf{R}]_k^2 - 1$. In addition, the QR interpretation of the optimally ordered successive cancelation detection in Section 3.3 tells us that to prove Property 1, we only need to prove that if we permute the last column with any other column of $\mathbf{G}^{1/2}$, the corresponding diagonal entries of the resulting R-factors do not increase. The detail of this proof can be found in [**?**].

Definition 5.1. *Define the minimum distance of a finite constellation $\mathcal{X}$ as*

$$d_{\min}(\mathcal{X}) = \min_{x \neq x', x, x' \in \mathcal{X}} |x - x'| = \sqrt{\min_{\mathbf{x}, \mathbf{x}' \in \mathcal{X}^N, \mathbf{x} \neq \mathbf{x}'} ||\mathbf{x} - \mathbf{x}'||^2} \tag{5.22}$$

Definition 5.2. *Define the free distance of an $M \times N$ channel matrix $\mathbf{H}$ as*

$$d_{\text{free}}(\mathbf{H}) = \sqrt{\min_{\mathbf{x}, \mathbf{x}' \in \mathcal{X}^N, \mathbf{x} \neq \mathbf{x}'} (\mathbf{x} - \mathbf{x}')^H \mathbf{H}^H \mathbf{H} (\mathbf{x} - \mathbf{x}')} \tag{5.23}$$

The following property, whose proof is given in Appendix B.2 shows the asymptotic behavior of the free distance for a channel with an equal-diagonal R-factor mutual information matrix.

Property 5.3. *If the mutual information matrix $\mathbf{G}^{1/2}$ of $\mathbf{H}$ have an equal-diagonal R-factor, then,*

$$\lim_{I \to \infty} \frac{d_{\text{free}}(\mathbf{H})}{2^I - 1} = d_{\min}(\mathcal{X}) \tag{5.24}$$

Chapter 6

Simulation Results

In this chapter, we verify the performance of our optimal transceiver design using computer simulations. Here, we present four examples in which each element of the transmitted signal vectors is independently and equally likely selected from the 4-quadrature amplitude modulation constellation.

6.1 Example 1: A Two-user Scenario

In this example, we consider the scenario of a two-user system. Two users communicate with a base station independently, and each user employs a DMT modulation having 32 available subcarriers. The number of subchannels N_k allocated to each of the users is predetermined. The channel is modeled as an FIR filter with 10 taps and the tap coefficients are generated independently from a zero-mean circular complex Gaussian distribution. The signals are selected with equal probabilities from a 4-QAM constellation. All the three cases use the designed decision feedback equalization. If a subchannel is used by both of the users, then it is called a shared subchannel. Let N_k indicates the number of subchannels User k will uses. Since there are, in total, 32 subchannels in this system, if $N_1 + N_2 \leq 32$, then each of the two users can use separate subchannels without sharing. However, if some subchannels are of bad condition, they may not be used by any of the users, in which case, the better subchannels may have to be shared. If $N_1 + N_2 > 32$, then there must be some subchannels which have to be shared by the two users. Figure 6.1 shows the BER against the sum Gaussian mutual information averaged over 1000 channel realizations. For each realization, the additive Gaussian noise is also generated independently from a zero-mean circular complex Gaussian distribution, and is normalized to unit energy. Three cases are simulated:

- The number of subcarriers assigned to User 1 and User 2 is 16 each ($N_1 = 16$ and $N_2 = 16$), if all the subchannels are good, then there is no shared subcarriers between the two users;
- $N_1 = 16$ and $N_2 = 17$, i.e., there is at least one shared subchannel;
- $N_1 = 17$ and $N_2 = 17$, i.e., there are at least two subchannels shared by these two users.

From Figure 6.1, it is observed that the BER decreases with the amount of sum Gaussian mutual information. In general, we find that when the number of shared subchannels grows, for the same mutual information, the average bit error rate increases, and this phenomenon is more obvious in the high sum Gaussian mutual information part. On the other hand, for the same BER, the amount of sum Gaussian mutual information increases with the number of subchannels shared.

6.2 Example 2: A Three-user Scenario

A three-user scenario is modeled and simulated. Again, each user employs a DMT modulation having 32 subcarriers. Here, the environment and the system are the same as those in Example 1. Figure 6.2 shows the BER against the sum mutual information averaged over 1000 channel realizations. Three cases are studied:

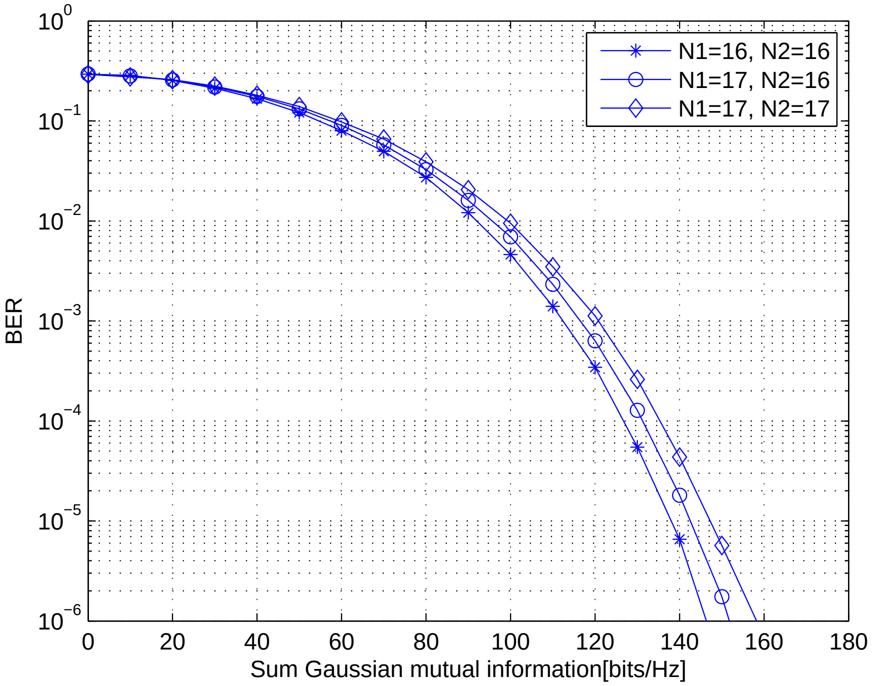

Figure 6.1: BER vs the sum Gaussian information in two-user scenario

- $N_1 = 11$, $N_2 = 11$, and $N_3 = 10$, again, if all the channels are good, there will be no shared subcarriers among the three users;

- $N_1 = 11$, $N_2 = 11$, and $N_3 = 11$, i.e., at least one subchannel is shared;

- $N_1 = 12$, $N_2 = 11$, and $N_3 = 11$, i.e., at least two subchannels are shared.

In Figure 6.2, we obtain similar results as those shown in Figure 6.1. This confirms the expectation that when the signals from more users are transmitted through the same subchannel, the more errors will occur.

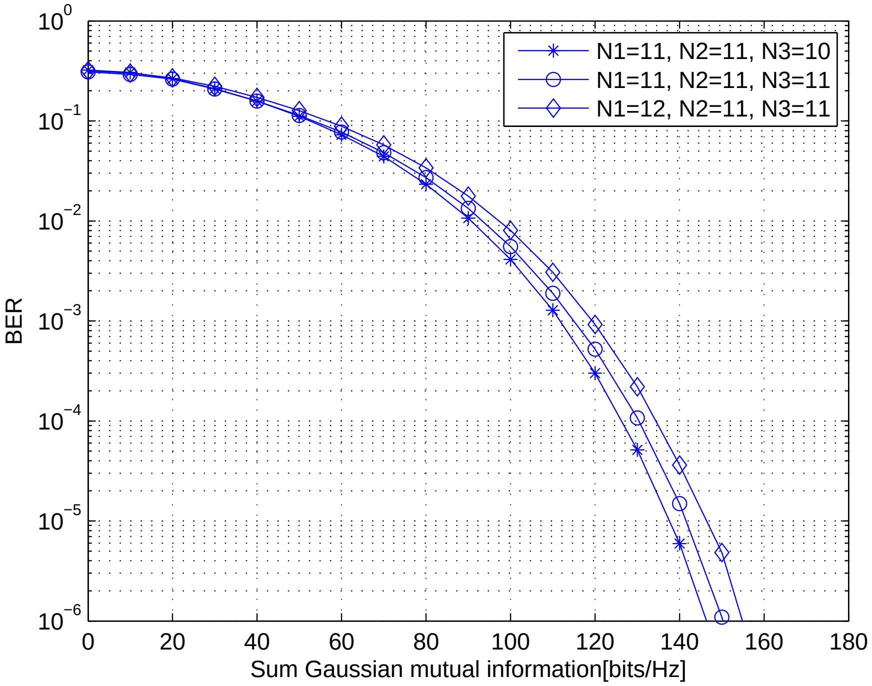

Figure 6.2: BER vs the sum Gaussian information in three-user scenario

6.3 Example 3: Comparison with Linear Equalization

In this example, we compare the performance of the transceiver design proposed in this paper with the linear transceiver design proposed in [?]. The simulation environment is the same as in Example 1. To ensure a fair comparison, the sum Gaussian mutual information $\mathcal{I}$ and numbers of subcarriers N_1 and N_2 assigned to each user in our design are calculated from the algorithm in [?] with a fixed power constraint. Then with these sum Gaussian mutual information and block length for each user, our proposed solution in Chapter 5 is run to design the transceiver in DFE system, after which the transmission power is calculated. 200 channel realizations are simulated and taken average over the sum Gaussian mutual information. Figure 6.3 shows the average bit error rate against the averaged sum Gaussian mutual information, Figure 6.4 shows the average bit error rate against the averaged signal to noise ratio, and Figure 6.5 shows SNR vs the amount of sum Gaussian mutual information. In Figure 6.5, the vertical axis label SNR means the transmitted signal power to noise power ratio.

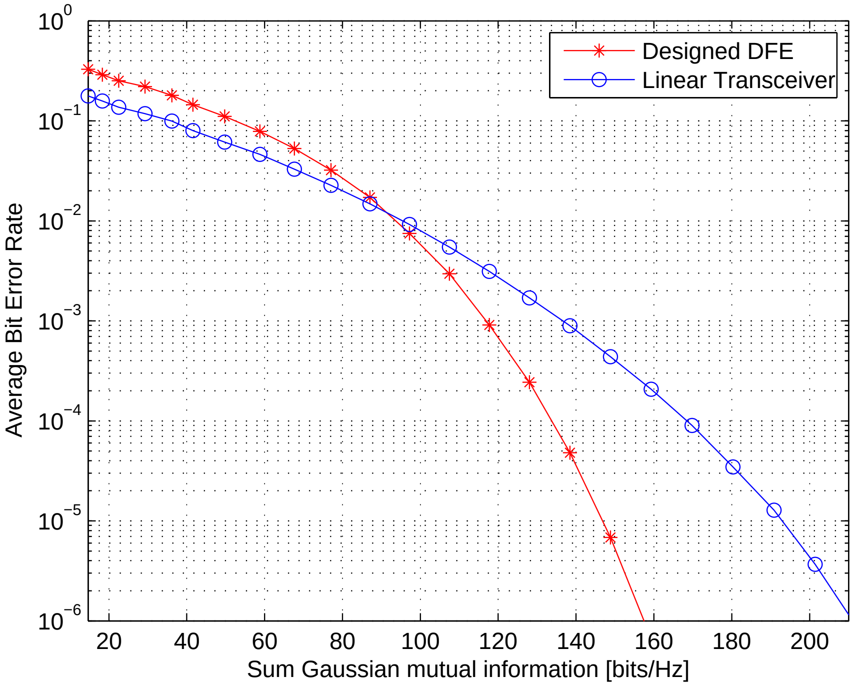

Figure 6.3: BER vs the sum Gaussian information: compared with linear MMSE detection

It can be observed from Figure 6.3 and Figure 6.4 that a significant gain over the linear receiver is obtained when the sum Gaussian mutual information is greater than 90 bits per Hz. However, it is also observed that the performance of our DFE is worse than that of the linear receiver when the sum Gaussian mutual information is less than 90 bits per Hz. A reasonable cause is that propagation of errors occurs in the successive cancelation detection. It can also be observed that in the Figure 6.5

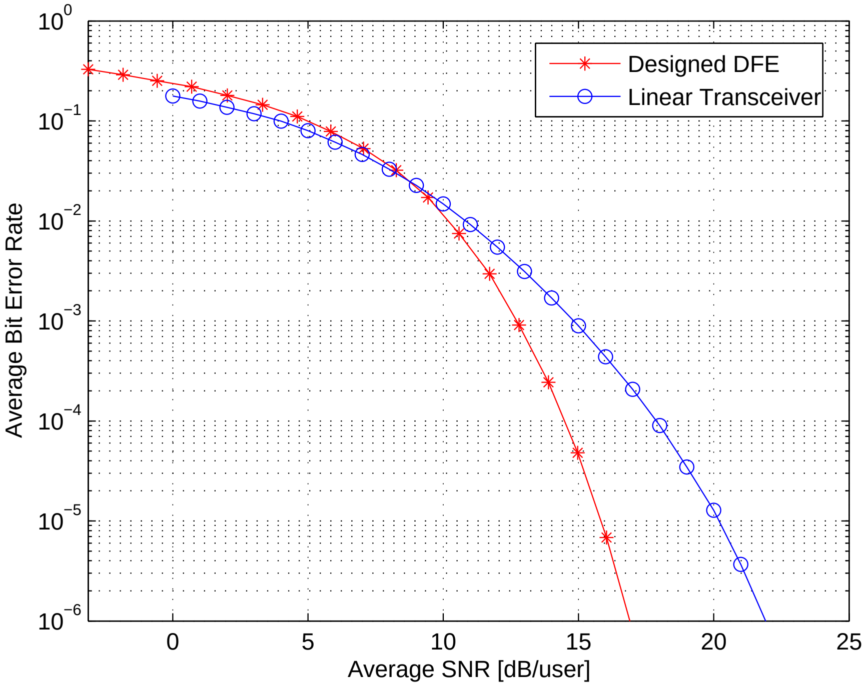

Figure 6.4: BER vs average SNR: compared with linear MMSE detection

when the sum Gaussian mutual information is low, the SNR (at a fixed noise power) for MMSE-DFE is about $2dB$ lower for each user than that of linear transceiver. This shows that our systems requires lower power to achieve the same amount of sum Gaussian mutual information.

6.4 Example 4: Comparison with ML Detection

In this example, we compare our designed MMSE-DFE transceiver with both the maximum likelihood detector (MLD) and MMSE linear equalization in [**?**]. Again, we consider a two user case. However, different from the previous example, a DMT modulation having only 4 available subcarriers is employed.

Figure 6.6 and Figure 6.7 show the average bit error rates at different sum Gaussian mutual information and average SNRs respectively. The sum Gaussian mutual

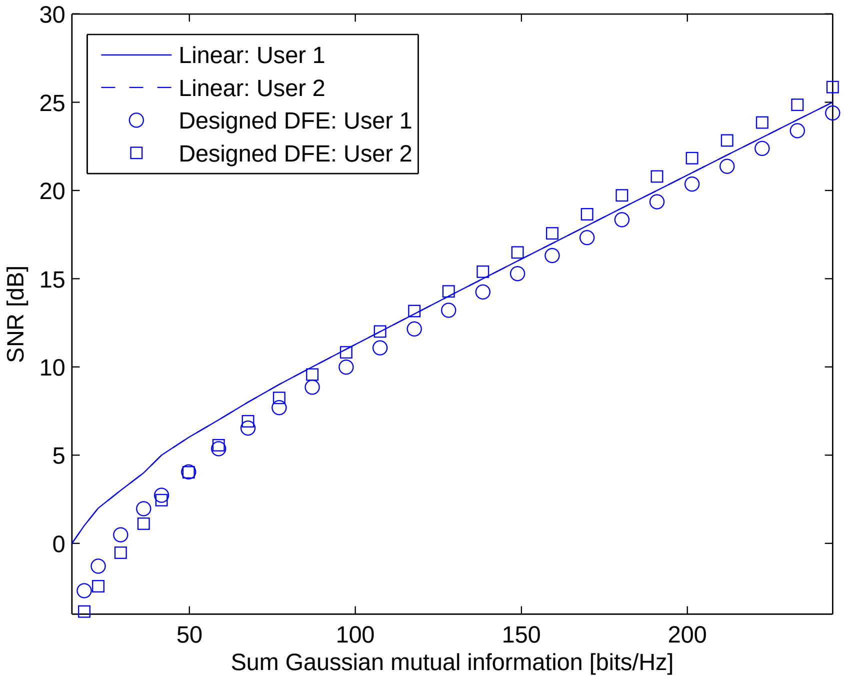

Figure 6.5: SNR vs the sum Gaussian information: compared with linear MMSE detection

information in the case of the linear receiver is calculated at each SNR from 0 to $20dB$. The number of subchannels assigned to each user, N_k is also calculated. For the MMSE-DFE system, at each value of sum Gaussian mutual information and noise power, we obtain our optimal transceiver design for a particular channel realization. The SNR for each channel realization is then calculated. The mean SNR is then obtained by averaging the SNR over 100 channel realizations. For the case of MLD, two scenarios are examined: the first one applies the optimum transmitter from our design and uses ML for detection; the other does not use any precoder at the transmitter but only ML detection at the receiver. From Figures 6.6 and 6.7, we can see that the performance of our optimum transceiver approaches that of the precoder-MLD combination. It can be observed that there is only a small gap between these two

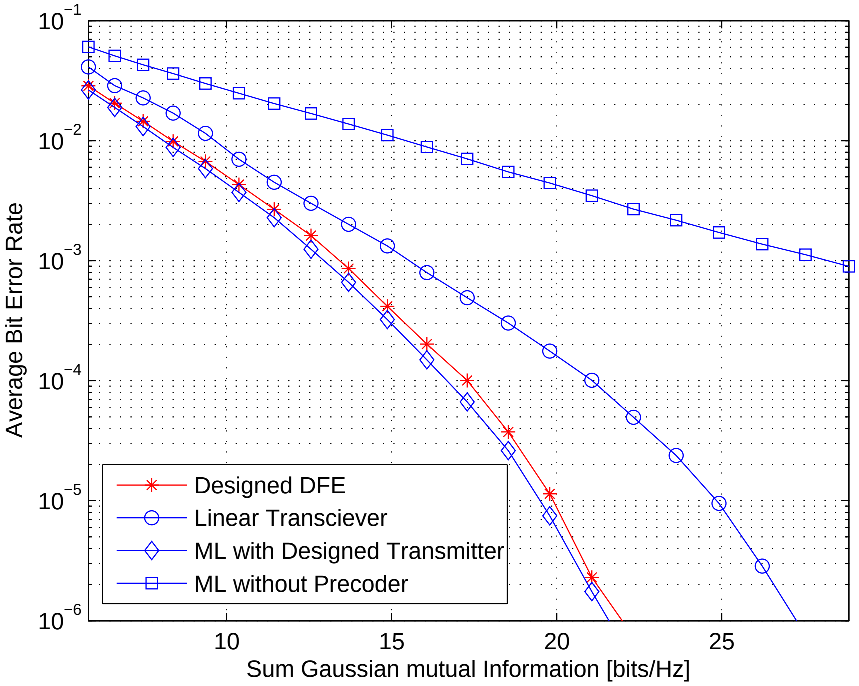

Figure 6.6: BER vs the sum Gaussian mutual information: compared with linear MMSE and MLD

performance curves. However, the MLD without precoder performs very poorly. The reason is in DMT modulation the channel matrix is diagonal, so there is less diversity in transmission if no precoding is used. For a specific channel realization, if the channel coefficients are small compared with noise coefficients, the channel is dominated by the noise. ML detection can be impaired badly due to the diagonal structure of the channel. As a result, the overall average of the bit error rate of the ML detection without precoder is dominated by these several bad cases. Also we find that the cross-over point among the BER curves moves to a much lower SNR compared to that in the previous example, which probably is due to the shorter transmission block and less error propagation. This makes the interpretation of error propagation more reasonable.

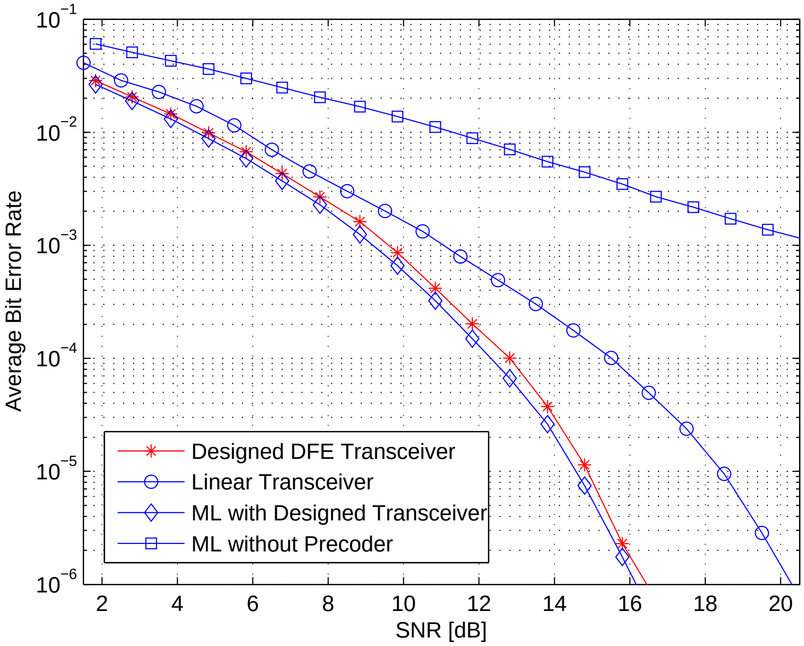

Figure 6.7: BER vs average SNR: compared with linear MMSE and MLD

For a further comparison, we examine the SNR of the fours cases. To ensure a fair comparison between MLD and our DFE, transmission power of the case of MLD with no precoder is made the same with that of DFE system and the number of transmitted symbols within a block is the same in these four schemes. The noise power is also the same in DFE and MLD. Therefore the transmitted signal power to noise power ratio of the ML detection is the same as that of the designed DFE transmitter. Figure 6.8 indicates the SNR in linear, DFE and ML systems. The vertical axis label SNR means the transmitted signal power to noise power ratio. It can be noticed that the SNR is almost the same in the communication system with designed MMSE-DFE transceiver as that in the system equipped with linear transceiver.

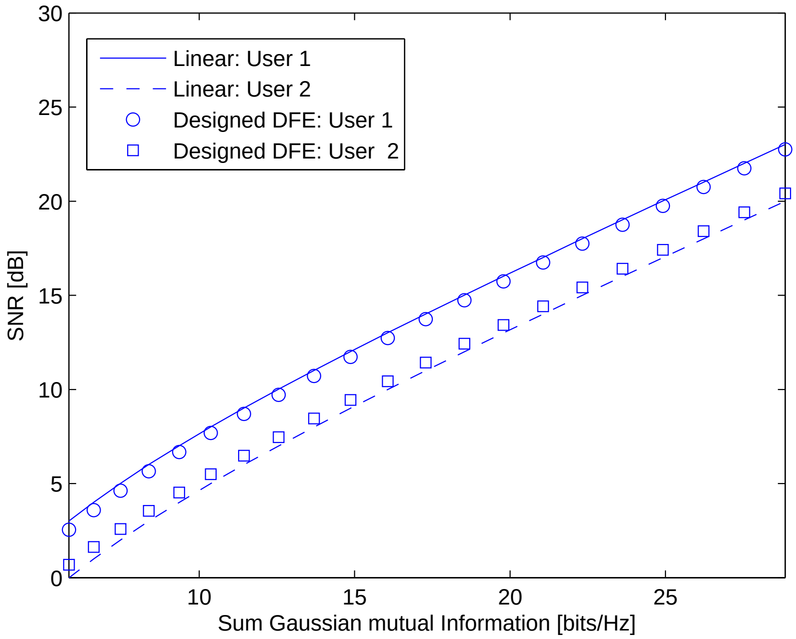

Figure 6.8: SNR vs the sum Gaussian mutual information: compared with linear MMSE and ML detection

Chapter 7

Conclusion and Future Work

7.1 Conclusion

In this thesis, we have jointly designed the precoder, the feedforward and the feedback matrix of a block-by-block transmission scheme for an ISI multiple-access MIMO communication system equipped with the MMSE-DF receiver. The design minimizes the average MSE under a fixed sum Gaussian mutual information. Through the development, we assumed that channel state information is perfectly known at both the transmitters and receivers. The optimal closed-form solution is obtained by the following two steps:

1. Find an optimal transmitter that minimizes the total power for a single user case subject to a fixed Gaussian mutual information, i.e, solve a dual problem of maximizing single user throughput. Therefore, by successively solving these dual problems user after user, the total Gaussian mutual information can be uniformly distributed over each user with the MMSE-DF detector.

2. Properly choose unitary matrices within the dual water-filling solution family using the equal diagonal QRS decomposition, so that the Gaussian mutual information of each user can be uniformly distributed into each active subchannels

of the user.

In addition to minimizing the arithmetic MSE of MMSE-decision feedback detection, the optimal systems are further revealed to possess the following two optimality properties in the precoded MAC MIMO detection theory:

1. Both the optimal user-detection order and symbol-detection order are natural orders in terms of signal to interference and noise ratios.

2. The free-distance for the ML detector has an asymptotic behavior when the sum Gaussian mutual information tends to large.

On the other hand, despite the fact that our attention here was restricted on a specific design of minimizing the arithmetic MSE of MMSE-decision feedback detection for an MAC, the methodology developed in this paper can be extended into the following fairly general optimization problem. Given are a matrix $\mathbf{H} = [\mathbf{H}_1 \ \mathbf{H}_2 \ \cdots \ \mathbf{H}_K]$, a non-negative constant $\mathcal{I}$, and K non-negative integers $N_1, N_2, \cdots, N_K$. Subject a constraint $\log\det\big(\mathbf{I} + \sum_{k=1}^{K} \mathbf{H}_k \mathbf{T}_k^H \mathbf{T}_k \mathbf{H}_k\big) = \mathcal{I}$, we need to find each matrix $\mathbf{T}_k$ that achieves the minimum

$$\min_{\mathbf{T}_k} \sum_{k=1}^{K} \sum_{n=1}^{N_k} \mathcal{F}\big([\mathbf{R}_k]_n\big)$$

with $[\mathbf{R}_k]_n$ being the nth diagonal entry of the R-factor of QR decomposition of the matrix $\mathbf{G}_k^{1/2}$, where $\mathbf{G}_k = \mathbf{I} + (\mathbf{H}_k\mathbf{T}_k)^H (\mathbf{\Sigma}_k)^{-1} \mathbf{H}_k \mathbf{T}_k$ and $\mathbf{\Sigma}_k = \mathbf{I} + \sum_{\ell=1}^{k-1} \mathbf{H}_\ell \mathbf{T}_\ell (\mathbf{H}_\ell \mathbf{T}_\ell)^H$. In addition, function $\mathcal{F}(2^t)$ with respect to t is assumed to be convex. This class of optimization problems has a closed-from solution, which can be attained from our presented technique in this paper. Thus, the solution strategy depends only on the features of the MMSE-DF receiver, but does not depend on the specific structure of the objective function $\mathcal{F}(\cdot)$. As a result, a single user case where an asymptotic bit error rate of the MMSE-DF detector was minimized [**?**] can be generalized in straightforward way to the multiple users scenario.

However, we must point out here that although the power of each user can be explicitly obtained by separately and efficiently solving the individual dual water-filling problem, the resulting optimal solution is not optimal in the sense of total power minimization of all users. That is to say, we still have freedom of optimally allocating the power of each user such that the total power of all users is minimized while maintaining the optimum value of the original objective.

7.2 Future Work

Following this thesis, we can extend the work along several directions. We give a few examples for potential future research as follows:

- As we said in the above conclusion, in this work, the transmission power is minimized individually for each user, which does not indicate the total transmission power is minimized. Therefore, minimizing the total transmission power can be considered in the future development.

- People may question what the practical meaning for using the sum Gaussian mutual information as a constrain is. In this thesis, especially in the simulation part in Chapter 6, we fix the signal constellation, which means the transmission rate is fixed. Therefore the sum mutual information can reflect the total SNR to some extent. Then an interesting extension of this work is to minimize the arithmetic MSE subject to individual user power constraint.

- Even though the mean square error is a good criterion to design the transceivers, the bit error rate is more accurately to reflect the communication system's performance. Unfortunately, for block based MSE-DFE there is not a closed-form solution for the probability of error, we can not find an exact expression for the BER. However, an approximation of the BER for DFE detection has been given in [**?**, **?**, **?**, **?**, **?**]. We can further use Jensen's inequality to obtain some lower bound of the approximation and find ways to minimize it. Similarly the optimization constraints can be the sum Gaussian mutual information, single user's mutual information, individual user's transmission power or the total transmission power.

Appendix A

Proof of Lemmas in Chapter 4

A.1 Proof of Lemma 4.1

We consider the following two cases.

Case 1: $r = 1$. In this case, we only need to prove

$$b_1^{-1} - a_1^{-1} \leq \frac{2^{\mathcal{I}_{\mathbf{H}}}}{b_1} - \frac{2^{\mathcal{I}_{\mathbf{H}}}}{a_1}$$

Since $\mathcal{I}_{\mathbf{H}} \geq 0$, the above inequality is always true, and the equality holds when $a_1 = b_1$.

Case 2: $r > 1$. For the given $c > 0$, let $f_c(t) = crt - t^r$. Since the first derivative of $f_c(t)$ with respect to t is given by $f_c'(t) = r(c - t^{r-1})$, $f_c'(t) > 0$ when $0 \leq t < c^{1/(r-1)}$ and hence, $f_c(t)$ is increasing. On the other hand, notice that condition (4.4) is equivalent to

$$b_m^{1-\frac{1}{m}} > \left(\frac{\prod_{i=1}^{m-1} b_i}{2^{\mathcal{I}_{\mathbf{H}}}} \right)^{1/m} \qquad \text{for } m = 1, 2, \cdots, r_b$$

which, in turn, is equivalent to

$$b_m > \left(\frac{\prod_{i=1}^{m-1} b_i}{2^{\mathcal{I}_{\mathbf{H}}}} \right)^{1/(m-1)} \qquad \text{for } m = 1, 2, \cdots, r_b \tag{A.1}$$

Let

$$\begin{aligned}\mathcal{F}(b_1, b_2, \cdots, b_r) &= r\left(\frac{2^{\mathcal{I}_{\mathbf{H}}}}{\prod_{i=1}^{r} b_i}\right)^{1/r} - \sum_{i=1}^{r} b_i^{-1} \\ &= f_c(t_r) - \sum_{i=1}^{r-1} b_i^{-1}\end{aligned}$$

where $c = \left(\frac{2^{\mathcal{I}_{\mathbf{H}}}}{\prod_{i=1}^{r-1} b_i}\right)^{1/r}$, $r = \min\{r_a,\ r_b\}$ and $t_r = b_r^{-1/r}$. From condition (A.1), if $m = r$, we have $b_r > \left(\frac{\prod_{i=1}^{r-1} b_i}{2^{\mathcal{I}_{\mathbf{H}}}}\right)^{1/(r-1)}$. This results in $b_r^{-1/r} < \left(\frac{2^{\mathcal{I}_{\mathbf{H}}}}{\prod_{i=1}^{r-1} b_i}\right)^{1/r(r-1)}$, and thus, $0 \leq t < c^{1/(r-1)}$. Since $a_r \geq b_r$, we can obtain

$$a_r^{-1/r} < b_r^{-1/r} < \left(\frac{2^{I_{\mathbf{H}}}}{\prod_{i=1}^{r-1} b_i}\right)^{1/r(r-1)}$$

Using the monotonicity of $f_c(t)$, we have

$$\mathcal{F}(b_1, b_2, \cdots, b_r) \geq \mathcal{F}(b_1, b_2, \cdots, b_{r-1}, a_r)$$

Continuing this process until we obtain

$$\mathcal{F}(b_1, b_2, \cdots, b_r) \geq \mathcal{F}(a_1, a_2, \cdots, a_r) \tag{A.2}$$

with the equality holding when $b_i = a_i$ for $i = 1, 2, \cdots, r$. This completes the proof of Lemma 4.1.

A.2 Proof of Lemma 4.2

For any positive integer k with $k + 1 \leq r_c$, we have

$$\begin{aligned}P_{k+1} - P_k &= (k+1)\left(\frac{2^{\mathcal{I}_{\mathbf{H}}}}{\prod_{n=1}^{k+1} c_n}\right)^{\frac{1}{k+1}} - \sum_{n=1}^{k+1} c_n^{-1} - k\left(\frac{2^{\mathcal{I}_{\mathbf{H}}}}{\prod_{n=1}^{k} c_n}\right)^{\frac{1}{k}} + \sum_{n=1}^{k} c_n^{-1} \\ &= k\left(\left(\frac{2^{\mathcal{I}_{\mathbf{H}}}}{\prod_{n=1}^{k+1} c_n}\right)^{\frac{1}{k+1}} - \left(\frac{2^{\mathcal{I}_{\mathbf{H}}}}{\prod_{n=1}^{k} c_n}\right)^{\frac{1}{k}}\right) \\ &\quad + \left(\left(\frac{2^{\mathcal{I}_{\mathbf{H}}}}{\prod_{n=1}^{k+1} c_n}\right)^{\frac{1}{k+1}} - c_{k+1}^{-1}\right)\end{aligned} \tag{A.3}$$

Since the first term can be rewritten as

$$\begin{aligned} & k\left(\left(\frac{2^{\mathcal{I}_{\mathbf{H}}}}{\prod_{n=1}^{k+1} c_n}\right)^{\frac{1}{k+1}} - \left(\frac{2^{\mathcal{I}_{\mathbf{H}}}}{\prod_{n=1}^{k} c_n}\right)^{\frac{1}{k}}\right) \\ = \; & k\left(\frac{2^{\mathcal{I}_{\mathbf{H}}}}{\prod_{n=1}^{k} c_n}\right)^{\frac{1}{k+1}}\left(c_{k+1}^{-\frac{1}{k+1}} - \left(\frac{2^{\mathcal{I}_{\mathbf{H}}}}{\prod_{n=1}^{k} c_n}\right)^{\frac{1}{k(k+1)}}\right) \end{aligned} \tag{A.4}$$

and the second term can be represented by

$$\left(\frac{2^{\mathcal{I}_{\mathbf{H}}}}{\prod_{n=1}^{k+1} c_n}\right)^{\frac{1}{k+1}} - c_{k+1}^{-1} = c_{k+1}^{-\frac{1}{k+1}}\left(\left(\frac{2^{\mathcal{I}_{\mathbf{H}}}}{\prod_{n=1}^{k} c_n}\right)^{\frac{1}{k+1}} - c_{k+1}^{-\frac{k}{k+1}}\right) \tag{A.5}$$

For simplicity, let

$$a = \left(\frac{2^{\mathcal{I}_{\mathbf{H}}}}{\prod_{n=1}^{k} c_n}\right)^{\frac{1}{k(k+1)}}, \qquad b = c_{k+1}^{-\frac{1}{k+1}}$$

Then, from Eqs.(A.4) and (A.5) we have

$$P_{k+1} - P_k \;=\; ka^k(b-a) + b(a^k - b^k) \tag{A.6}$$

Since c_{k+1} satisfies the following inequality, $c_{k+1} > \left(\frac{\prod_{n=1}^{k+1} c_n}{2^{\mathcal{I}_{\mathbf{H}}}}\right)^{\frac{1}{k+1}}$. Hence, $c_{k+1}^{-1} < \left(\frac{2^{\mathcal{I}_{\mathbf{H}}}}{\prod_{n=1}^{k+1} c_n}\right)^{\frac{1}{k+1}}$. This is equivalent to

$$c_{k+1}^{-1} - \left(\frac{2^{\mathcal{I}_{\mathbf{H}}}}{\prod_{n=1}^{k+1} c_n}\right)^{\frac{1}{k+1}} \;<\; 0$$

As a consequence, we obtain $a^k > b^k$ and hence, $a > b$. Combing this with Eq.(A.6) yields

$$\begin{aligned} & P_{k+1} - P_k \\ = \; & (a-b)(a^{k-1}b + a^{k-2}b^2 + \cdots + ab^{k-1} + b^k - ka^k) \\ < \; & (a-b)(a^k + \cdots + a^k - ka^k) = 0 \end{aligned}$$

This completes the proof of Lemma 4.2.

A.3 Proof of Lemma 4.3

The derivative of a complex matrix $\mathbf{T}$ is defined as

$$\frac{\partial f(\mathbf{T})}{\partial \mathbf{T}} = \frac{1}{2}\left(\frac{\partial f(\mathbf{T})}{\partial \Re \mathbf{T}} - j\frac{\partial f(\mathbf{X})}{\partial \Im \mathbf{T}}\right) \tag{A.7}$$

Here in our problem, $f(\mathbf{T}) = \log\det(\mathbf{A} + \mathbf{T}^H\mathbf{B}\mathbf{T})$ where $\mathbf{T}$ is an $M \times N$ matrix, $\mathbf{B}$ is an $M \times M$ Hermitian matrix, and $\mathbf{A}$ is an $N \times N$ Hermitian matrix. Applying the formula of derivative of scalar functions of a matrix with respect to the matrix defined in [?], we can first obtain

$$\begin{aligned}
&\frac{\partial \log\det(\mathbf{A} + \mathbf{T}^H\mathbf{B}\mathbf{T})}{\partial \Re \mathbf{T}} \\
=\ & \frac{1}{\det(\mathbf{A} + \mathbf{T}^H\mathbf{B}\mathbf{T})}\frac{\partial \det(\mathbf{A} + \mathbf{T}^H\mathbf{B}\mathbf{T})}{\partial \Re \mathbf{T}} \\
=\ & \frac{1}{\det(\mathbf{A} + \mathbf{T}^H\mathbf{B}\mathbf{T})}\sum_{ij}\mathbf{E}_{ij}\frac{\partial \det(\mathbf{A} + \mathbf{T}^H\mathbf{B}\mathbf{T})}{\partial \Re t_{ij}}
\end{aligned} \tag{A.8}$$

where $\mathbf{E}_{ij}$ denotes the $N \times M$ elementary matrix which has a unity in the ijth position and all the other elements are zero. By using the general form of the derivative of a determinant with respect to a scalar that is stated in [?], we can further expand Eq. (A.8) into

$$\begin{aligned}
&\frac{1}{\det(\mathbf{A} + \mathbf{T}^H\mathbf{B}\mathbf{T})}\sum_{ij}\mathbf{E}_{ij}\det(\mathbf{A} + \mathbf{T}^H\mathbf{B}\mathbf{T})\mathrm{tr}\left[(\mathbf{A} + \mathbf{T}^H\mathbf{B}\mathbf{T})^{-1}\frac{\partial(\mathbf{A} + \mathbf{T}^H\mathbf{B}\mathbf{T})}{\partial \Re t_{ij}}\right] \\
=\ & \sum_{ij}\mathbf{E}_{ij}\mathrm{tr}\left[(\mathbf{A} + \mathbf{T}^H\mathbf{B}\mathbf{T})^{-1}\frac{\partial(\mathbf{T}^H\mathbf{B}\mathbf{T})}{\partial \Re t_{ij}}\right]
\end{aligned}$$

Let $\mathbf{Y} = (\mathbf{A} + \mathbf{T}^H\mathbf{B}\mathbf{T})$. Then the above equation can be written as

$$\begin{aligned}
&\sum_{ij}\mathbf{E}_{ij}\mathrm{tr}\left[(\mathbf{A} + \mathbf{T}^H\mathbf{B}\mathbf{T})^{-1}\frac{\partial(\mathbf{T}^H\mathbf{B}\mathbf{T})}{\partial \Re t_{ij}}\right] \\
=\ & \sum_{ij}\mathbf{E}_{ij}\mathrm{tr}\left[\mathbf{Y}^{-1}(\mathbf{E}_{ij}^H\mathbf{B}\mathbf{T} + \mathbf{T}^H\mathbf{B}\mathbf{E}_{ij})\right]
\end{aligned} \tag{A.9}$$

If $\mathbf{A}$ and $\mathbf{B}$ are both Hermitian , then $\mathbf{Y}$ is also a Hermitian matrix, i.e., $\mathbf{Y}^H = \mathbf{Y}$. If we denote $\mathbf{C} = \mathbf{E}_{ij}^H\mathbf{BT}$, then $\mathbf{T}^H\mathbf{BE}_{ij} = \mathbf{C}^H$. Hence Eq. (A.9) can be rewritten as

$$\begin{aligned}
&\sum_{ij}\mathbf{E}_{ij}\mathrm{tr}\left[\mathbf{Y}^{-1}(\mathbf{E}_{ij}^H\mathbf{BT}+\mathbf{T}^H\mathbf{BE}_{ij})\right]\\
=&\sum_{ij}\mathbf{E}_{ij}\mathrm{tr}\left[\mathbf{Y}^{-1}\mathbf{C}+\mathbf{Y}^{-1}\mathbf{C}^H\right]\\
=&\sum_{ij}\mathbf{E}_{ij}\left[\mathrm{tr}(\mathbf{CY}^{-1})+\mathrm{tr}(\mathbf{Y}^{-1}\mathbf{C}^H)\right]
\end{aligned}$$

According to the definition of trace, we have

$$\begin{aligned}
\mathrm{tr}(\mathbf{CY}^{-1}) &= \sum_{k=1}^{N}(\mathbf{C}_{k\cdot}\mathbf{Y}_{\cdot k}^{-1})\\
\mathrm{tr}(\mathbf{Y}^{-1}\mathbf{C}^H) &= \sum_{k=1}^{N}(\mathbf{Y}_{k\cdot}^{-1}(\mathbf{C}_{\cdot k})^H)
\end{aligned}$$

with $\mathbf{C}_{k\cdot}$ and $\mathbf{Y}_{k\cdot}^{-1}$ denotes the kth row of the matrix $\mathbf{C}$ and $\mathbf{Y}^{-1}$, $\mathbf{Y}_{\cdot k}^{-1}$ and $\mathbf{C}_{\cdot k}^H$ denotes the kth column of the matrix $\mathbf{Y}^{-1}$ and $\mathbf{C}^H$, respectively. Since

$$\begin{aligned}
\mathbf{C} &= \mathbf{E}_{ij}^H\mathbf{BT} = \begin{bmatrix}\mathbf{0}\\ \vdots\\ (\mathbf{BT})_{i\cdot}\\ \vdots\\ \mathbf{0}\end{bmatrix}\\
\mathbf{C}^H &= \mathbf{T}^H\mathbf{BE}_{ij} = \begin{bmatrix}\mathbf{0} & \cdots & (\mathbf{T}^H\mathbf{B})_{\cdot i} & \cdots & \mathbf{0}\end{bmatrix} \qquad \text{(A.10)}
\end{aligned}$$

the only non-zero row in matrix $\mathbf{C}$ is the jth row which is the ith row of matrix $\mathbf{BT}$ and the only non-zero column in $\mathbf{C}^H$ is the jth column which is nothing but the ith column in $\mathbf{T}^H\mathbf{B}$. Thus $\mathbf{C}_{k\cdot}\mathbf{Y}_{\cdot k}^{-1}$ and $\mathbf{Y}_{k\cdot}^{-1}\mathbf{C}_{\cdot k}$ are not zero when $k = j$, and

$$\begin{aligned}
&\sum_{ij}\mathbf{E}_{ij}\left[\mathrm{tr}(\mathbf{CY}^{-1})+\mathrm{tr}(\mathbf{Y}^{-1}\mathbf{C}^H)\right]\\
=&\sum_{ij}\mathbf{E}_{ij}(\mathbf{BT})_{i\cdot}\mathbf{Y}_{\cdot j}^{-1}+\sum_{ij}\mathbf{E}_{ij}\mathbf{Y}_{j\cdot}^{-1}(\mathbf{T}^H\mathbf{B})_{\cdot i}\\
=&\ \mathbf{BTY}^{-1}+[\mathbf{Y}^{-1}\mathbf{T}^H\mathbf{B}]^T\\
=&\ \mathbf{BT}(\mathbf{A}+\mathbf{T}^H\mathbf{BT})^{-1}+[(\mathbf{A}+\mathbf{T}^H\mathbf{BT})^{-1}\mathbf{T}^H\mathbf{B}]^T \qquad \text{(A.11)}
\end{aligned}$$

That is to say,

$$\frac{\partial \log\det(\mathbf{A}+\mathbf{T}^H\mathbf{B}\mathbf{T})}{\partial \Re\mathbf{T}} = \mathbf{B}\mathbf{T}(\mathbf{A}+\mathbf{T}^H\mathbf{B}\mathbf{T})^{-1} + [(\mathbf{A}+\mathbf{T}^H\mathbf{B}\mathbf{T})^{-1}\mathbf{T}^H\mathbf{B}]^T \quad \text{(A.12)}$$

Similarly, we can find

$$\frac{\partial \log\det(\mathbf{A}+\mathbf{T}^H\mathbf{B}\mathbf{T})}{\partial \Im\mathbf{T}} = j\mathbf{B}\mathbf{T}(\mathbf{A}+\mathbf{T}^H\mathbf{B}\mathbf{T})^{-1} - j[(\mathbf{A}+\mathbf{T}^H\mathbf{B}\mathbf{T})^{-1}\mathbf{T}^H\mathbf{B}]^T \quad \text{(A.13)}$$

Substituting Eq. (A.12) and Eq. (A.13) into equation (A.7), we can obtain that

$$\begin{aligned} & \frac{\partial \text{log } \det(\mathbf{A}+\mathbf{T}^H\mathbf{B}\mathbf{T})}{\partial \mathbf{T}} \\ = \;& [(\mathbf{A}+\mathbf{T}^H\mathbf{B}\mathbf{T})^{-1}\mathbf{T}^H\mathbf{B}]^T \\ = \;& \left[\mathbf{B}\mathbf{T}(\mathbf{A}+\mathbf{T}^H\mathbf{B}\mathbf{T})^{-1}\right]^* \end{aligned} \quad \text{(A.14)}$$

Appendix B

Proof of Lemmas and Properties in Chapter 5

B.1 Proof of Lemma 5.1

We know from [**?**] that there exists a unitary matrix $\mathbf{Q}$ such that matrix $\mathbf{G}^{1/2}$ can be decomposed into $\mathbf{G}^{1/2} = \mathbf{Q}\mathbf{R}$, where $\mathbf{R}$ is a block triangular matrix; i.e.,

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} & \dots & \mathbf{R}_{1K} \\ 0 & \mathbf{R}_{22} & \dots & \mathbf{R}_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{R}_{KK} \end{pmatrix}.$$

Therefore, the (i, j)th block matrix $(i \leq j)$ of $\mathbf{R}^H\mathbf{R}$ is $\sum_{k=1}^{i} \mathbf{R}_{ki}^H \mathbf{R}_{kj}$. On the other hand, the (i, j)th block matrix $(i \leq j)$ of $\mathbf{G}$ is $\mathbf{H}_i^H\mathbf{H}_j$ for $i < j$ and $\mathbf{I} + \mathbf{H}_i^H\mathbf{H}_i$ for $i = j$. Hence, we have

$$\sum_{k=1}^{i} \mathbf{R}_{ki}^H \mathbf{R}_{kj} = \begin{cases} \mathbf{I} + \mathbf{H}_i^H\mathbf{H}_i & \text{if } i = j \\ \mathbf{H}_i^H\mathbf{H}_j & \text{if } i < j \end{cases} \tag{B.1}$$

Now, we use mathematical induction on the row number of $\mathbf{R}$ to prove Eq. (5.17). Therefore, $\mathbf{R}_{11}^H\mathbf{R}_{11} = \mathbf{I} + \mathbf{H}_1^H\mathbf{H}_1$, from which we get

$$\mathbf{R}_{11} = \left(\mathbf{I} + \mathbf{H}_1^H\mathbf{H}_1\right)^{1/2}. \tag{B.2}$$

As a result, we have

$$\mathbf{R}_{1j} = \left(\mathbf{I} + \mathbf{H}_1^H\mathbf{H}_1\right)^{-1/2}\mathbf{H}_1^H\mathbf{H}_j \qquad \text{for } j = 2, 3, \cdots, N \tag{B.3}$$

Therefore, for the first row, Statement in Eq. (5.17) is true. We now assume that Statement in Eq. (5.17) is true for $i < L$; i.e.,

$$\mathbf{R}_{ij} = \begin{cases} \mathbf{G}_i^{1/2} & \text{if } i = j \\ \mathbf{G}_i^{-1/2}\mathbf{H}_i^H\mathbf{\Sigma}_i^{-1}\mathbf{H}_j & \text{if } i < j \end{cases} \tag{B.4}$$

In the following, we are going to prove that this statement is also true for $i = L$. From Eq. (B.1) with $i = j = L$ and using the induction assumption in Eq. (B.4) we have

$$\begin{aligned} \mathbf{R}_{LL}^H\mathbf{R}_{LL} &= \mathbf{I} + \mathbf{H}_L^H\mathbf{H}_L - \sum_{i=1}^{L-1}\mathbf{R}_{iL}^H\mathbf{R}_{iL} \\ &= \mathbf{I} + \mathbf{H}_L^H\mathbf{H}_L - \mathbf{H}_L^H\left(\sum_{i=1}^{L-1}\mathbf{\Sigma}_i^{-1}\mathbf{H}_i\mathbf{G}_i^{-1}\mathbf{H}_i^H\mathbf{\Sigma}_i^{-1}\right)\mathbf{H}_L \end{aligned} \tag{B.5}$$

Using the Matrix Inverse Lemma $(\mathbf{H}+\mathbf{C}\mathbf{B}^{-1}\mathbf{D})^{-1} = \mathbf{H}^{-1}-\mathbf{H}^{-1}\mathbf{C}(\mathbf{B}+\mathbf{D}\mathbf{H}^{-1}\mathbf{C})^{-1}\mathbf{D}\mathbf{H}^{-1}$ [**?**], we have

$$\begin{aligned} \mathbf{\Sigma}_i^{-1/2}\mathbf{H}_i\mathbf{G}_i^{-1}\mathbf{H}_i^H\mathbf{\Sigma}_i^{-1/2} &= \mathbf{\Sigma}_i^{-1/2}\mathbf{H}_i\left(\mathbf{I} + \mathbf{H}_i^H\mathbf{\Sigma}_i^{-1}\mathbf{H}_i\right)^{-1}\mathbf{H}_i^H\mathbf{\Sigma}_i^{-1/2} \\ &= \mathbf{I} - \left(\mathbf{I} + \mathbf{\Sigma}_i^{-1/2}\mathbf{H}_i\mathbf{H}_i^H\mathbf{\Sigma}_i^{-1/2}\right)^{-1} \end{aligned} \tag{B.6}$$

Substituting Eq. (B.6) into Eq. (B.5) yields

$$\begin{aligned} \mathbf{R}_{LL}^H\mathbf{R}_{LL} &= \mathbf{I} + \mathbf{H}_L^H\mathbf{H}_L - \mathbf{H}_L^H\sum_{i=1}^{L-1}(\mathbf{\Sigma}_i^{-1} - \mathbf{\Sigma}_{i+1}^{-1})\mathbf{H}_L \\ &= \mathbf{I} + \mathbf{H}_L^H\mathbf{\Sigma}_L^{-1}\mathbf{H}_L = \mathbf{G}_L \end{aligned} \tag{B.7}$$

Similarly, from Eq. (B.1) with $i = L < j$ and using the induction assumption in Eq. (B.4) we have

$$\begin{aligned}\mathbf{R}_{LL}^H\mathbf{R}_{Lj} &= \mathbf{H}_L^H\mathbf{H}_j - \sum_{k=1}^{L-1}\mathbf{R}_{kL}^H\mathbf{R}_{kj} \\ &= \mathbf{H}_L^H\mathbf{H}_j - \mathbf{H}_L^H\sum_{k=1}^{L-1}\mathbf{\Sigma}_k^{-1}\mathbf{H}_k\mathbf{G}_k^{-1}\mathbf{H}_k^H\mathbf{\Sigma}_k^{-1}\mathbf{H}_j \\ &= \mathbf{H}_L^H\mathbf{\Sigma}_L^{-1}\mathbf{H}_j \qquad \text{(B.8)}\end{aligned}$$

Combining Eq. (B.8) with Eq. (B.7), we have shown that Statement in Eq. (5.17) is true for $i = L$.

B.2 Proof of Property 5.3

We first note that

$$\mathbf{H}^H\mathbf{H} = (\mathbf{G} - \mathbf{I})$$

Now consider two different signal vectors: $\mathbf{x} = [x_1, x_2, \cdots, x_N]^T$ and $\mathbf{x}' = [x_1', x_2', \cdots, x_N']^T$. If $x_k = x_k'$ for $k = 2, \cdots, N$, but $x_1 \neq x_1'$. Then

$$\begin{aligned}(\mathbf{x}-\mathbf{x}')^H\mathbf{H}^H\mathbf{H}(\mathbf{x}-\mathbf{x}') &= \left((\mathbf{x}-\mathbf{x}')^H\mathbf{G}(\mathbf{x}-\mathbf{x}') - |x_1 - x_1'|^2\right) \\ &= ([\mathbf{R}]_1^2 - 1)|x_1 - x_1'|^2 \qquad \text{(B.9)}\end{aligned}$$

Hence, by taking the minima of both sides of Eq. (B.9), we obtain

$$d_{\text{free}}^2(\mathbf{H}) \leq \min_{x_1, x_1' \in \mathcal{X}, x_1 \neq x_1'} ([\mathbf{R}]_1^2 - 1)\cdot|x_1 - x_1'|^2 = (2^I - 1)\cdot d_{\min}^2(\mathcal{X})$$

which leads to

$$d_{\text{free}}^2(\mathbf{H}) \leq (2^I - 1)\cdot d_{\min}^2(\mathcal{X}) \qquad \text{(B.10)}$$

On the other hand, we note that

$$\begin{aligned}(\mathbf{x}-\mathbf{x}')^H\mathbf{H}^H\mathbf{H}(\mathbf{x}-\mathbf{x}') &= \left((\mathbf{x}-\mathbf{x}')^H\mathbf{G}(\mathbf{x}-\mathbf{x}') - \|\mathbf{x}-\mathbf{x}'\|^2\right) \\ &= \left(\sum_{i=1}^N\left|\sum_{j=i}^N[\mathbf{R}]_{ij}\cdot(x_j - x_j')\right|^2 - \|\mathbf{x}-\mathbf{x}'\|^2\right) \qquad \text{(B.11)}\end{aligned}$$

Assume $\mathbf{x} \neq \mathbf{x}'$. Let k be an integer such that $x_i = x'_i$, for $i > k$, but $x_k \neq x'_k$. Then, from Eq. (B.11), using the upper triangularity of $\mathbf{R}$, we have

$$\begin{aligned}(\mathbf{x}-\mathbf{x}')^H\mathbf{H}^H\mathbf{H}(\mathbf{x}-\mathbf{x}') &= \left(\sum_{i=1}^{k}\left|\sum_{j=i}^{k}[\mathbf{R}]_{ij}(x_j - x'_j)\right|^2 - \|\mathbf{x}-\mathbf{x}'\|^2\right) \\ &\geq \left(([\mathbf{R}]_1^2 - 1)|x_k - x'_k|^2 - \|\mathbf{x}-\mathbf{x}'\|^2\right) \\ &\geq \left((2^I - 1)\cdot d_{\min}^2(\mathcal{X}) - \|\mathbf{x}-\mathbf{x}'\|^2\right) \qquad \text{(B.12)}\end{aligned}$$

Taking the minima of both sides of Eq. (B.12) yields

$$d_{\text{free}}^2(\mathbf{H}) \geq \left((2^I - 1)\cdot d_{\min}(\mathcal{X}) - \|\mathbf{x}-\mathbf{x}'\|_{\max}^2\right) \qquad \text{(B.13)}$$

Since constellation $\mathcal{X}$ is finite, quantity $\|\mathbf{x}\|_{\max}^2$ is bounded and as a result, we can obtain from Eq. (B.13) that

$$\lim_{I\to\infty}\frac{d_{\text{free}}(\mathbf{H})}{2^I - 1} \geq d_{\min}(\mathcal{X}) \qquad \text{(B.14)}$$

Combining (B.10) with (B.14), we complete the proof of (5.24). Moreover, we know from [**?**] that

$$\lim_{\text{snr}\to\infty}\frac{\ln \mathrm{P}_{\text{MLD}}(\text{snr})}{\ln Q\left(\frac{\sqrt{\text{snr}}\, d_{\text{free}}(\mathbf{H})}{2}\right)} = 1 \qquad \text{(B.15)}$$