國 立 交 通 大 學

電控工程研究所

博 士 論 文

應用於多輸入多輸出系統之最佳收發器設計

Optimal Transceiver Design for MIMO Systems

研 究 生：李建樟

指導教授：林源倍 教授

中 華 民 國 九 十 九 年 七 月

應用於多輸入多輸出系統之最佳收發器設計

Optimal Transceivesr Design for MIMO Systems

研 究 生：李建樟   Student：Chien-Chang Li

指導教授：林源倍   Advisor：Yuan-Pei Lin

國 立 交 通 大 學

電控工程研究所

博 士 論 文

A Dissertation
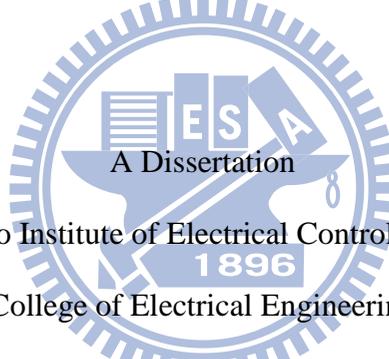
Submitted to Institute of Electrical Control Engineering

College of Electrical Engineering

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Doctor

in

Department of Electrical Engineering

July 2010

Hsinchu, Taiwan, Republic of China

中華民國九十七年七月

應用於多輸入多輸出系統之最佳收發器設計

學生：李建樟　　　　　　　　　　　　指導教授：林源倍

國立交通大學電控工程研究所博士班

摘要

本論文包含兩個部分，在第一個部分我們考慮應用在多輸入多輸出系統的收發器以及位元分配的設計。在早期的文獻當中，收發器的設計都是針對某一個特定的位元配置而設計出來的。在近期的研究中，位元配置也會被考慮在收發器的設計中。在這個部分中，我們考慮的問題是當總功率以及錯誤率有條件限制時，如何最大化傳輸速率。在這問題之下，我們提出了同時設計收發器和位元配置的方法。一般而言，我們推算出來的最佳位元配置都不會是正整數。可是這並不符合實際情況，所以在本篇論文我們也有考慮當位元配置是限制在整數的時候。首先我們先推導出最大化傳輸速率系統和最小化功率系統之間的關係。然後利用這個關係我們可以找到當位元配置被限制在整數的時候，最大化傳輸速率的最佳解。我們也用了很多模擬結果來驗證我們的推導。

在第二個部分，我們考慮的是應用於多載波系統上的傳送與接收窗框之設計。對於多載波系統而言，在設計傳送端和接收端的時候，傳送端濾波器和接收端濾波器的頻譜響應非常重要。對於傳送端而言，如果頻譜響應表現不理想，會導致功率能量的散溢，並且影響到其他的傳輸系統。對於接收端而言，如果頻譜響應表現不理想，受到外界電台的雜訊干擾將會非常嚴重。為了改善頻譜響應，我們會使用所

謂的傳送與接收窗框。在這個部分，我們使用小波理論來設計傳送與接收窗框。我們會引入所謂的子濾波器。我們發現使用小波理論設計的系統在對抗電台雜訊干擾以及抑制功率能量的散溢有較為出色的表現。
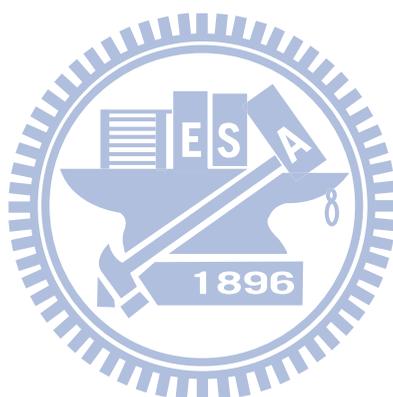
Optimal Transceiver Designs for MIMO Systems

Student: Chien-Chang Li   Advisor: Yuan-Pei Lin

Institute of Electrical Control Engineering
National Chiao Tung University

Abstract

This dissertation consists of two parts. In the first part, we consider the joint design of transceiver and bit allocation for multiple-input multiple-output (MIMO) channels. In the literature, there have been many results on designing transceivers for MIMO channels. In these results, the transceivers are designed for a given bit allocation or designed with real bit allocation. In this thesis, first we jointly optimize the transceiver with real-valued bit allocation for maximizing bit rate over MIMO channels. The optimal transceiver and bit allocation are obtained in a closed form. Second we consider the connection between the power-minimizing and rate-maximizing problems with bit allocation. We will show that if a transceiver is optimal for the power-minimizing problem, it is also optimal for the rate maximizing problem and the converse is true. The result holds whether the bit allocation is integer-constrained or not. Based on the duality, we develop algorithms to find the optimal solution for the power-minimizing problem and rate-maximizing problem with integer bit constraint.

In the second part, we consider the design of transmitting and receiving windows for multicarrier systems. For multicarrier systems, frequency characteristics play an important role in the design of the transmitting and receiving filters. To improve frequency separation, windowing techniques are often used. As these are frequency based characteristics, a filterbank

representation provides a natural and useful way for formulating the problem. In this thesis, we propose a unified filterbank approach to design the transmitting and receiving windows for multicarrier systems. Using the filterbank approach, the frequency separation among the subchannels can be improved.

誌謝

Contents

List of Tables

List of Figures

List of Notations

- Boldfaced lower-case letters represent vectors and boldfaced upper-case letters are reserved for matrices.

- The notation $\mathbf{X}^{\dagger}$ denotes the conjugate transpose of $\mathbf{X}$.

- The notation $\mathbf{X}^{T}$ denotes the transpose of $\mathbf{X}$.

- The notation $\mathbf{I}_M$ is used to represent the $M \times M$ identity matrix.

- The notation $[\mathbf{X}]_{kk}$ denotes the k-th diagonal element of the matrix $\mathbf{X}$. The notation x_k denotes the k-th element of the vector $\mathbf{x}$.

- The notation $\operatorname{diag}(\mathbf{g})$ denotes a diagonal matrix with the elements of $\mathbf{g}$ on its diagonal.

- The notation $\lfloor z \rfloor$ denotes the largest integer that is less than or equal to z.

- The notation $\mathrm{E}[x]$ denotes the expectation of a random variable x.

Chapter 1

Introduction

This thesis consists of two parts. Part I is on the design of bit allocation and transceiver for MIMO channels. Part II is on the design of transmitting and receiving windows for multicarrier systems. The introduction of parts I and II are given in section 1.1 and 1.2 respectively.

1.1 Transceiver designs for MIMO Systems

Multiple-input multiple-output (MIMO) channels arise in applications such as wireless communication systems that use multiple antennas, multicarrier communication systems, and also telephone cables that consist of many twisted pairs. They represent a way to model a wide variety of scenarios. In this part, we focus on the transceiver design with bit allocation for MIMO channels. The design of the MIMO transceivers can be formulated as the optimization problem of an objective function based on the performance of each subchannels.

Transceiver designs for a given bit allocation. For a given bit allocation, many criteria have been considered in the transceiver designs for MIMO channels, e.g., [1]-[17]. Optimal transceivers that maximize the mutual information are proposed in [1]-[5]. Transceiver designs that minimize mean-square error (MMSE) are considered in [6]-[9]. Optimal transceivers that minimize the bit error rate (BER) are derived in [10]-[13]. Optimal transceivers that minimize the transmit

power are proposed in [14][16]. Using the MMSE receiver, unified frameworks for designing MIMO systems with a power constraint are proposed in [17]. A number of useful objective functions can be considered in this framework. For example, the optimal MMSE transceivers that maximize the bit rate and mutual information can be designed using this unified approach.

Transceiver designs with real-valued bit allocation. In [1]-[17], the transceivers are designed for a given bit allocation. Recently, bit allocation is also incorporate in the design of the MIMO system [18]-[27]. Optimal transceivers with bit allocation that minimize the transmit power are proposed in [18][22]. Optimal transceiver with bit allocation designs that use the bit rate maximization criterion are addressed in [23]-[25]. Transceiver designs that consider a number of design criteria are proposed in [26, 27]. For example, power-minimizing transceiver, rate-maximizing transceiver, capacity-maximizing transceiver and BER-minimizing transceiver can be obtained using [26, 27]. For the transceivers designs in [18]-[27], a smaller transmit power or a higher bit rate than the cases without bit allocation can be achieved. Hence bit allocation plays an important role in the power minimization and rate maximization problems. However, the bit allocation obtained in these designs are not integers in general. The MIMO transceiver design for minimizing transmit power or maximizing bit rate with integer bit constraint is not solved and still open.

Integer bit allocation for multicarrier systems. For the multicarrier systems, integer bit allocation has been considered [29]-[36]. The problem of designing integer bit allocation for maximizing bit rate and minimizing transmit power in multicarrier systems is considered in [29]. Algorithms for allocating integer bits to minimize transmit power is proposed in [30]-[31]. In [32], an optimal bit loading algorithm is presented for minimizing BER in multicarrier system. In [33], a bit loading algorithm is proposed to increase the noise margin (additional amount of noise that the system can tolerate). Problems of finding

2

the integer bit allocation for maximizing a concave function is considered in [34], where it is shown that a greedy algorithm can be used to find the optimal solution. In [35], an efficient bit loading algorithm is proposed to minimize an arbitrary convex objective function. The algorithm proposed in [34] and [35] can be used to find the optimal integer bit allocation for both the power-minimizing problem or rate-maximizing problem when a ZF transceiver is given. In [36], an integer bit allocation is proposed to maximize the transmission bit rate in the presence of intercarrier interference. An integer bit allocation for minimizing the quantization error of multiple sources is proposed in [37]. The algorithms in [29]-[37] can be used to find integer bit allocation only when a transceiver is given.

1.2   Multicarrier System

Multicarrier system has attracted considerable attention in recent years as a practical technology for high-speed data transmission over frequency selective channels [45]-[47]. The discrete Fourier transform (DFT) based multicarrier system has been recognized as a very cost-effective realization of multicarrier transceivers. Several important applications of multicarrier system have been found in discrete multitone (DMT) systems such as asymmetric digital subscriber lines (ADSL) [48] and very high speed digital subscriber lines (VDSL) [49][50], and orthogonal frequency division multiplexing (OFDM) systems such as wireless local area network [51] and digital video broadcasting (DVB) [52]. A generic multicarrier system is shown in Fig. 1.1. $F_0(z), F_1(z), \cdots, F_{M-1}(z)$ are the transmitting filters and $H_0(z), H_1(z), \cdots, H_{M-1}(z)$ are the receiving filters. In second part of this thesis, we consider the design of the transmitting and receiving filters of the multicarrier system. In the design of the multicarrier system, the frequency characteristics of the transmitting and receiving filters are important considerations. The stopband attenuation of the transmitting and receiving filters determines how well the subchannels will be separated in the frequency domain. In the con-
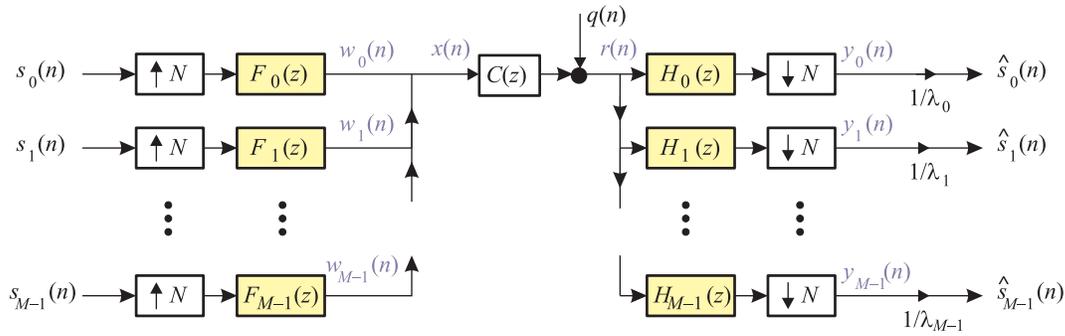
Figure 1.1: Multicarrier system.

ventional multicarrier system, the transmitting and receiving filters come from rectangular windows. Since the sidelobes of the rectangular window are large, the conventional multicarrier system has a poor frequency separation in both the transmitter side and the receiver side.

Frequency characteristic at the transmitter. At the transmitter side, poor frequency separation leads to significant spectral leakage. In some applications, the PSD (power spectral density) of the transmit signal is required to have a large roll-off in certain frequency bands. Hence poor frequency separation could pose a problem in such applications. For example in some wired transmission application, the PSD of the transmitted signal needs to fall below a threshold in the transmission bands of the opposite direction to avoid interference [48, 49]. The PSD should also be attenuated in amateur radio bands to reduce interference to radio transmission or egress [49].

Frequency characteristic at the receiver. In some applications such as VDSL and ADSL transmission, the multicarrier systems share its spectrum with different types of radio transmission, for example, amplitude-modulation stations and amateur radio [48, 49, 50]. These radio signals can be coupled into telephone wires and interfere with the VDSL signal at the receiving side. This type of noise in a VDSL transmission system is known as radio frequency interference

4

(RFI) ingress [53]. At the receiver side, the spectral roll-off of the receiving filters determine how the tones are affected by RFI interference. A faster roll off means the effect of RFI diminishes faster and fewer neighboring tones are affected. Poor frequency separation results in poor out-of-band rejection. Since the receiving filters of the conventional multicarrier system come from the rectangular window, many neighboring tones can be affected by the RFI ingress.

Improving the Frequency characteristics. In the literature, many methods have been proposed to improve the frequency characteristics of the transmitter. To improve the spectral roll-off of the transmitted signal, a number of continuous-time pulse shaping filters have been proposed, [54]-[59]. Usually continuous-time pulse shapes are designed based on an analog implementation of transmitters and a digital implementation is not admitted [60]. Discrete-time windows have been considered in [61]-[63]. The design of overlapping windows for OFDM with offset QAM (quadrature amplitude modulation) over ISI free channels are studied fully in [62, 63]. More recently, transmitting windows with the cyclic-prefixed property have been proposed in [64, 65] for egress control. Windows that are the inverse of a raised cosine function are optimized in [64], to minimize egress emission. To compensate for the transmitter window, the corresponding receiver requires post-processing equalization [64, 65]. Per-tone windows are proposed in [66] for shaping transmitted spectrum. The shaping of spectrum allows more tones to be used for transmission.

At the receiver side, windowing is also often applied to improve the frequency characteristics. Commonly used windows include Hanning window and Blackman window [75]. In [67], Muschallik use Nyquist windows, which have the property that shifts of the window in the time domain add to a constant, to improve the reception of OFDM systems. Optimal Nyquist windows are considered in [68] to mitigate the effect of additive noise and carrier frequency offsets. To improve RFI suppression, receiver windowing is proposed first in [69] by Spruyt et al. For the

suppression of sidelobes without using extra redundant samples, it is proposed in [70] to use windows that introduced controlled IBI, later removed using decision feedback. To minimize the RFI and channel noise, the receiver windowing is proposed in [71]. The optimal window can be found using the statistics of the received RFI and noise [71]. A combination of raised-cosine window and per tone equalizer are proposed to suppress RFI interference in [72]. However, the channel information is required in these designs. in [73], channel-independent windows are designed by minimizing the sidelobe energy. In this case, ISI (inter symbol interference) is introduced and post processing is required to cancel ISI. Using statistics of channel noise and RFI, a joint design of the TEQ and the receiving window for maximizing bit rates is given in [74].

1.3   Chapter Outline

The designs of transceivers with bit allocation for MIMO channels are discussed in Chapter 2-Chapter 5 and the designs of transmitting and receiving windows for the multicarrier system are discussed in Chapter 6-Chapter 8. Details of the research contributions in each chapter are as follows.

Chapter 2

In this section, we introduce the MIMO systems. We consider both the ZF and MMSE receivers. For the QAM modulation, symbol error rate and bit allocation are also given in this chapter.

Chapter 3

In this chapter, we consider the design of the zero-forcing transceivers for MIMO channels. We jointly optimize the transceiver and bit allocation to maximize the transmission rate for a given target error rate and transmit power. Using the high bit rate assumption, we can simplify the optimization problem and the optimal transceiver can be easily found by the *Hadamard inequality* and the *Poincaré separation theorem*.

Chapter 4

In this chapter, we consider the rate maximizing problem in chapter 4 but the receiver is MMSE. In this design, we do not use the high bit rate assumption. We jointly optimize the transceiver and bit allocation to maximize the bit rate subject to a given error rate and a given transmit power. There are no constraints on the transceiver or the bit allocation. Using the majorization theorem [42], the optimal transceiver and bit allocation can be obtained in a simple close-form and the optimal solution diagonalizes the channel into parallel independent subchannels.

Chapter 5

In this chapter, we study the connections between the power minimization and rate maximization problem. For the problems without integer bit constraint, we will show that these two problems have the same solution. However, the result does not generalize completely to the case with an integer constraint on bit allocation. We show that the power minimization and rate maximization criterion yield the same solution if the statement of problems are modified slightly. Moreover, we also show how to find the optimal solution of the power-minimizing problem and rate-maximizing problem with the integer bit constraint.

Chapter 6

In this chapter, we will give an overview of the multicarrier system. We will derive the filterbank representation of the multicarrier system. We also study the spectral leakage at the transmitter and RFI interference at the receiver.

Chapter 7

In this chapter, we design the receiving windows to improve the frequency separation among the receiving filters. We will consider both the case when the statistics of the interference is available to the receiver (informed receiver) and the case when it is not (uninformed receiver). The frequency responses of the proposed windows achieve a good trade-off in spectral roll-off between high fre-

quency and low frequency than that of rectangular window, Hanning window, Blackman window, Kaiser window and the window design method in [71]. As a result, fewer tones will be dominated by RFI interference. The proposed windows in both cases are channel independent and can be obtained in a closed form solution.

Chapter 8

In this chapter, we propose a unified filterbank framework for the design of windows for multicarrier systems. The approach is more general than the conventional windowing technique. We will use the so-called subfilters to enhance the frequency selectivity of the transmitting and receiving filters while maintaining the orthogonality among the subchannels. For the transmitter side spectral leakage can be reduced and for the receiver side RFI can be further suppressed. When the subfilters form a DFT bank, they can be tied nicely to the conventional windowing such as in [65], [71], and chapter 2. The windows can be optimized through the design of subfilters and frequency separation among the subchannels can be considerably improved.
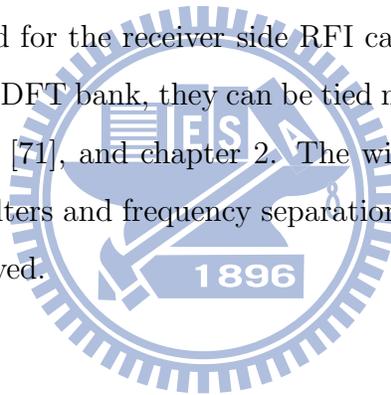
Chapter 2

Overview of MIMO Systems

MIMO systems arise in many different scenarios such as wired-line systems or multi-antenna wireless systems. In this chapter, we will give an overview of the MIMO communication systems.

2.1 Systems Model

A generic MIMO communication system is shown in Fig. 2.1. The MIMO channel is modeled by a $P \times N$ memoryless matrix $\mathbf{H}$. The $P \times 1$ channel noise $\mathbf{q}$ is additive white Gaussian noise with variance N_0. The transmitter matrix $\mathbf{F}$ is of size $N \times M$. The receiver matrix $\mathbf{G}$ is of size $M \times P$. The input of the transmitter is $\mathbf{s}$, an $M \times 1$ vector of modulation symbols.
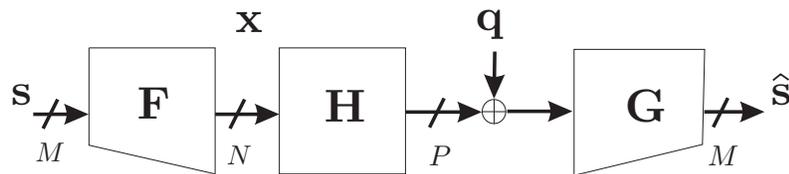
Figure 2.1: MIMO communication system.

The symbols are assumed to be zero mean and uncorrelated; hence the auto-correlation matrix $\mathbf{\Lambda}_s = \mathrm{E}[\mathbf{s}\mathbf{s}^\dagger]$ is a diagonal matrix with $[\mathbf{\Lambda}_s]_{kk} = \sigma_{s_k}^2$, where the notation $\mathbf{X}^\dagger$ denotes the conjugate transpose of $\mathbf{X}$. The total transmit power P

is

$$P = \mathrm{E}\{\mathbf{x}^\dagger \mathbf{x}\} = \mathrm{Tr}(\mathbf{F}\boldsymbol{\Lambda}_s\mathbf{F}^\dagger) = \sum_{k=0}^{M-1}[\mathbf{F}^\dagger\mathbf{F}]_{kk}\sigma_{s_k}^2, \qquad (2.1)$$

where $\mathbf{x}$ is the transmitter output. The output of the receiver is given by

$$\hat{\mathbf{s}} = \mathbf{GHFs} + \mathbf{Gq}, \qquad (2.2)$$

where $\mathbf{e} = \mathbf{Gq}$ is the error vector. Defined the error vector as

$$\mathbf{e} = \mathbf{s} - \hat{\mathbf{s}}. \qquad (2.3)$$

The mean-squared error (MSE) matrix is given by

$$\mathbf{E} = \mathrm{E}[\mathbf{ee}^\dagger], \qquad (2.4)$$

and the error variance of the k-th subchannel is $\sigma_{e_k}^2 = [\mathbf{E}]_{kk}$.

2.2 ZF and MMSE Receivers

In this section, we will introduce the ZF and MMSE receivers for a given transmitter.

ZF Receiver. The zero-forcing condition is given by

$$\mathbf{GHF} = \mathbf{I}_M. \qquad (2.5)$$

To achieve zero-forcing, the rank of $\mathbf{F}$, $\mathbf{H}$, $\mathbf{G}$ must be larger than or equal to M. In Lemma 2.1, we will show that without loss of generality we can choose $\mathbf{G}$ as the pseudo inverse of $\mathbf{HF}$.

Lemma 2.1 *It is no loss of generality to choose $\mathbf{G}$ as the pseudo inverse of $\mathbf{HF}$. That is,*

$$\mathbf{G} = (\mathbf{F}^\dagger\mathbf{H}^\dagger\mathbf{HF})^{-1}\mathbf{F}^\dagger\mathbf{H}^\dagger. \qquad (2.6)$$

In this case, the MSE matrix becomes

$$\mathbf{E} = N_0\mathbf{G}\mathbf{G}^\dagger = N_0(\mathbf{F}^\dagger\mathbf{H}^\dagger\mathbf{H}\mathbf{F})^{-1}. \tag{2.7}$$

Proof: Suppose $(\mathbf{G}, \mathbf{F})$ is a transceiver pair that satisfies the zero-forcing condition in (2.5). Let $\mathbf{G}'$ be the pseudo inverse of $\mathbf{H}\mathbf{F}$, i.e.,

$$\mathbf{G}' = (\mathbf{F}^\dagger\mathbf{H}^\dagger\mathbf{H}\mathbf{F})^{-1}\mathbf{F}^\dagger\mathbf{H}^\dagger. \tag{2.8}$$

Then we have $\mathbf{G}'\mathbf{H}\mathbf{F} = \mathbf{I}_M$. Define $\boldsymbol{\Delta} = \mathbf{G} - \mathbf{G}'$. Since $(\mathbf{G}, \mathbf{F})$ and $(\mathbf{G}', \mathbf{F})$ are both zero-forcing, we have $\boldsymbol{\Delta}\mathbf{H}\mathbf{F} = \mathbf{0}$. It follows that $\boldsymbol{\Delta}\mathbf{G}'^\dagger = \mathbf{0}$. When we use $\mathbf{G}$, the noise variance at k-th subchannel is given by

$$N_0[\mathbf{G}\mathbf{G}^\dagger]_{kk} = N_0[(\mathbf{G}' + \boldsymbol{\Delta})(\mathbf{G}' + \boldsymbol{\Delta})^\dagger]_{kk} \geq N_0[\mathbf{G}'\mathbf{G}'^\dagger]_{kk}, \tag{2.9}$$

where we have used $\boldsymbol{\Delta}\mathbf{G}'^\dagger = \mathbf{0}$ and $[\boldsymbol{\Delta}\boldsymbol{\Delta}^\dagger]_{kk} > 0$. Therefore, we will have smaller subchannel noise variances when we replace $\mathbf{G}$ with $\mathbf{G}'$. Using (2.6), we have the MSE matrix as in (2.7). $\triangle\triangle\triangle$

For the ZF case, the receiver in (2.6) and the MSE matrix in (2.7) depend on the channel matrix $\mathbf{H}$, the transmitter $\mathbf{F}$, and noise variance N_0.

MMSE Receiver. For the MMSE case, the receiver is obtained by minimizing the mean square error [27], i.e.,

$$\mathbf{G} = \arg\min_{\mathbf{G}} \mathrm{E}[\mathbf{e}^\dagger\mathbf{e}]. \tag{2.10}$$

Let $\mathbf{y}$ be the signal received at the receiver, i.e., $\mathbf{y} = \mathbf{H}\mathbf{F}\mathbf{s} + \mathbf{q}$. Using the orthogonality principle [44], we can find $\mathbf{G}$ by solving

$$\mathrm{E}[\mathbf{e}\mathbf{y}^\dagger] = \mathbf{0}. \tag{2.11}$$

Then the MMSE receiver is given by

$$\mathbf{G} = \mathrm{E}[\mathbf{s}\mathbf{y}^\dagger]\mathrm{E}[\mathbf{y}\mathbf{y}^\dagger] \tag{2.12}$$

$$= \boldsymbol{\Lambda}_s\mathbf{F}^\dagger\mathbf{H}^\dagger[\mathbf{H}\mathbf{F}\boldsymbol{\Lambda}_s\mathbf{F}^\dagger\mathbf{H}^\dagger + N_0\mathbf{I}_P]^{-1}. \tag{2.13}$$

Substituting (2.13) into (2.4), the MSE matrix becomes

$$\mathbf{E} = \mathbf{\Lambda}_s - \mathbf{\Lambda}_s \mathbf{F}^\dagger \mathbf{H}^\dagger [\mathbf{HF}\mathbf{\Lambda}_s \mathbf{F}^\dagger \mathbf{H}^\dagger + N_0 \mathbf{I}_P]^{-1} \mathbf{HF}\mathbf{\Lambda}_s. \tag{2.14}$$

For the MMSE case, the receiver in (2.13) and the MSE matrix in (2.14) depend on the channel matrix $\mathbf{H}$, the transmitter $\mathbf{F}$, noise variance N_0, and the signal autocorrelation matrix $\mathbf{\Lambda}_s$. If $\sigma_{s_k}^2 = 0$ for some k, using (2.14) we have $\sigma_{e_k}^2 = 0$.

Reduced system of the MMSE and ZF transceiver. For both the ZF and MMSE transceivers, the signal power $\sigma_{s_i}^2$ assigned to the i-th subchannel may be equal to zero and thus $s_i = 0$. In this case the autocorrelation matrix $\mathbf{\Lambda}_s$ is not invertible. Suppose M_r subchannels are assigned nonzero power. Let $\mathbf{s}_r$ be the $M_r \times 1$ vector obtained by deleting the entries of $\mathbf{s}$ that are assigned with zero power. Let $\mathbf{F}_r$ be the $N \times M_r$ matrix obtained by deleting the columns of $\mathbf{F}$ that correspond to the subchannels assigned with zero power. Then the transmitter output $\mathbf{x}$ is

$$\mathbf{x} = \mathbf{Fs} = \mathbf{F}_r \mathbf{s}_r. \tag{2.15}$$

When we consider the transmitter $\mathbf{F}_r$ with input $\mathbf{s}_r$, the transmitter output is the same as the original system. Hence we can consider only the subchannels assigned with nonzero power. Let $\mathbf{\Lambda}_r$ be the $M_r \times M_r$ diagonal matrix obtained by deleting the columns and rows of $\mathbf{\Lambda}_s$ with zero power. Since $\mathbf{\Lambda}_r$ is invertible, the reduced $M_r \times M_r$ MSE matrix becomes

$$\mathbf{E}_r = \begin{cases} [N_0^{-1} \mathbf{F}_r^\dagger \mathbf{H}^\dagger \mathbf{HF}_r + \mathbf{\Lambda}_r^{-1}]^{-1}, & \text{for the MMSE receiver;} \\ N_0 (\mathbf{F}_r^\dagger \mathbf{H}^\dagger \mathbf{HF}_r)^{-1}, & \text{for the ZF receiver.} \end{cases} \tag{2.16}$$

2.3 Symbol Error Rate

For the QAM modulation, suppose the power allocation $\mathbf{\Lambda}_s$, transmitter $\mathbf{F}$, and the number of bits loaded on k-th subchannel b_k are given. Then the symbol

error rate ϵ_k of the k-th subchannel can be approximated by [45]

$$\epsilon_k \approx 4\left(1 - \frac{1}{2^{b_k/2}}\right)Q\left(\sqrt{\frac{3\beta_k}{(2^{b_k} - 1)}}\right), \tag{2.17}$$

where

$$\beta_k = \begin{cases} \sigma^2_{s_k}/\sigma^2_{e_k}, & \text{for the ZF receive;} \\ \sigma^2_{s_k}/\sigma^2_{e_k} - 1, & \text{for the MMSE receiver.} \end{cases} \tag{2.18}$$

The function $Q(x)$ is the area under a Gaussian tail, i.e., $Q(x) = (1/\sqrt{2\pi}) \int_x^\infty e^{-u^2/2} du$.

2.4 Bit Allocation

Suppose the power allocation $\mathbf{\Lambda}_s$ and the transmitter $\mathbf{F}$ are given. Then β_k in (2.18) can be determined. For the QAM modulation, equation (2.17) relates the error rate to β_k. It can be used to obtain the number of bits that can be loaded on the k-th subchannel for a given β_k and target symbol error rate ϵ_k [38]. By rearranging the terms in (2.17), we get

$$b_k = \log_2\left(1 + \frac{\beta_k}{\Gamma_k}\right), \tag{2.19}$$

where $\Gamma_k = \frac{1}{3}[Q^{-1}(\epsilon_k/4)]^2$. The total number of bits that can be transmitted in one block is

$$B = \sum_{k=0}^{M-1} b_k = \sum_{k=0}^{M-1} \log_2\left(1 + \frac{\beta_k}{\Gamma_k}\right). \tag{2.20}$$

2.5 Summary

In this section, we gave an overview of a generic MIMO communication system. We have introduced the ZF and MMSE receivers when the transmitter and power allocation, and channel are given. We also introduced the symbol error rate and bit allocation when the QAM modulation is used.

Chapter 3

Rate-Maximizing Zero-Forcing Transceivers with Bit Allocation

In this chapter, we will *jointly* design the transceiver and bit allocation for maximizing bit rate for the ZF transceiver. Using the high bit rate assumption, we can simplify the optimization problem. The solutions are obtained in two steps. Firstly, we design the optimal bit and power allocation for a given transceiver and a given power constraint. Secondly, we design the optimal transceiver that maximizes the bit rate based on the optimal bit and power allocation. In the second step, the optimal transceiver can be easily found by the Hadamard inequality and the Poincaré separation theorem. The optimal transceiver and bit allocation can be obtained in a closed form.

3.1 Problem Formulation

Suppose the target error rate of all the subchannels are equal to ϵ. Using the high bit rate assumption, i.e., $2^{b_k} \gg 1$, the bit allocation b_k in (2.19) is approximated by

$$b_k = \log_2 \left(\frac{\sigma_{s_k}^2}{\sigma_{e_k}^2 \Gamma} \right). \tag{3.1}$$

We will see in section 3.2 that such an assumption facilitates the derivation of the optimal transceiver. Using the high bit rate assumption, the problem of

maximizing bit rate subject to a zero-forcing constraint and a total transmit power constraint P_0 can be formulated as

$$
\begin{aligned}
&\underset{\mathbf{G},\,\mathbf{F},\,\{\sigma^2_{s_k}\}}{\text{maximize}} \quad B = \sum_{k=0}^{M-1} \log_2\left(\frac{\sigma^2_{s_k}}{\sigma^2_{e_k}\Gamma}\right),\ \sigma^2_{e_k} = N_0[\mathbf{G}\mathbf{G}^\dagger]_{kk} \\
&\text{subject to} \quad
\begin{cases}
\text{Tr}(\mathbf{F}\boldsymbol{\Lambda}_{\mathrm{s}}\mathbf{F}^\dagger) \le \mathrm{P}_0 \\
\mathbf{G}\mathbf{H}\mathbf{F} = \mathbf{I}_M \\
\sigma^2_{s_k} \ge 0,\, k = 0, 1, \cdots, M-1.
\end{cases}
\end{aligned}
\tag{3.2}
$$

In section 3.2, we derive the optimal bit allocation and transmitter for the rate maximization problem.

3.2   Optimal Zero-Forcing Transceiver

First, we will find the power allocation that maximizes the bit rate for a given zero-forcing transceiver. To this end, we use the Karush-Kuhn-Tucker (KKT) condition [77]. Let $\sigma^{2*}_{s_k}$ be a local maximum for the optimization problem in (4.6). Then there exists constants α and μ_k, for $k = 0, 1, \cdots, M-1$ such that:

1. $\alpha \le 0$.

2. $\mu_k \le 0$, for $k = 0, 1, \cdots, M-1$.

3. $\left.\dfrac{\partial}{\partial \sigma^2_{s_k}}\left(\sum_{k=0}^{M-1} \log_2\left(\dfrac{\sigma^2_{s_k}}{\sigma^2_{e_k}\Gamma}\right) + \alpha(\text{Tr}(\mathbf{F}\boldsymbol{\Lambda}_{\mathrm{s}}\mathbf{F}^\dagger) - \mathrm{P}_0) + \sum_{\mathrm{k}=0}^{\mathrm{M}-1} \mu_{\mathrm{k}}(-\sigma^2_{\mathrm{s_k}}) \right)\right|_{\sigma^2_{\mathrm{s_k}}=\sigma^{2*}_{\mathrm{s_k}}} = 0.$

4. $\left.\alpha(\text{Tr}(\mathbf{F}\boldsymbol{\Lambda}_{\mathrm{s}}\mathbf{F}^\dagger) - \mathrm{P}_0)\right|_{\sigma^2_{\mathrm{s_k}}=\sigma^{2*}_{\mathrm{s_k}}} = 0.$

5. $\left.\mu_k(-\sigma^2_{s_k}) = 0\right|_{\sigma^2_{\mathrm{s_k}}=\sigma^{2*}_{\mathrm{s_k}}} = 0$, for $k = 0, 1, \cdots, M-1$.

By solving the above conditions, the optimal power allocation is given by

$$
\sigma^2_{s_k} = \frac{P_0}{M[\mathbf{F}^\dagger\mathbf{F}]_{kk}}.
\tag{3.3}
$$

From (3.3), we can see that the power allocation depends only on the transmitter for the given P_0 and M. Using (3.3), the bit rate is given by

$$B = \sum_{k=0}^{M-1} \log_2 \left(\frac{P_0}{M\Gamma[\mathbf{F}^\dagger\mathbf{F}]_{kk}\sigma_{e_k}^2} \right) \tag{3.4}$$

$$= \log_2 \left(\prod_{k=0}^{M-1} \frac{P_0}{M\Gamma[\mathbf{F}^\dagger\mathbf{F}]_{kk}\sigma_{e_k}^2} \right). \tag{3.5}$$

Next, we will design the optimal zero-forcing transceiver that maximizes the bit rate in (3.5). Suppose the $P \times N$ channel matrix $\mathbf{H}$ has rank K. Let the singular value decomposition of $\mathbf{H}$ be

$$\mathbf{H} = \mathbf{U} \begin{bmatrix} \mathbf{\Lambda} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{V}^\dagger, \tag{3.6}$$

where the $K \times K$ diagonal matrix $\mathbf{\Lambda}$ contains the nonzero singular values of $\mathbf{H}$. The $P \times P$ matrix $\mathbf{U}$ and the $N \times N$ matrix $\mathbf{V}$ are unitary. We assume that the elements of $\mathbf{\Lambda}$ are in nonincreasing order and $K \geq M$ so that solutions of zero-forcing transceivers exist.

Lemma 3.1 *Without loss of generality, we can express* $\mathbf{F}$ *to be of the following form:*

$$\mathbf{F} = \mathbf{V} \begin{bmatrix} \mathbf{A} \\ \mathbf{0} \end{bmatrix}, \tag{3.7}$$

for appropriate $K \times M$ *matrix* $\mathbf{A}$ *of rank* M.

Proof: Suppose $(\mathbf{G}, \mathbf{F})$ is a transceiver pair that satisfies the zero-forcing condition. As $\mathbf{V}$ is unitary, $\mathbf{F}$ can always be represented as

$$\mathbf{F} = \mathbf{V} \begin{bmatrix} \mathbf{A} \\ \mathbf{A}_1 \end{bmatrix}, \tag{3.8}$$

where $\mathbf{A}$ is a $K \times M$ matrix, $\mathbf{A}_1$ is an $(N-K) \times M$ matrix, and the notation T denotes the transpose. Define a new transceiver $\mathbf{F}'$ as

$$\mathbf{F}' = \mathbf{V} \begin{bmatrix} \mathbf{A} \\ \mathbf{0} \end{bmatrix}. \tag{3.9}$$

Then we have

$$\mathbf{GHF}^{'} = \mathbf{GHF}. \tag{3.10}$$

Therefore, when we replace the transmitter by $\mathbf{F}^{'}$, the new system still satisfies the zero-forcing condition $\mathbf{GHF} = \mathbf{I}_M$. As the receiver is not changed, the new system has the same subchannel noise variances and hence the same bit rate performance. Now, let us compare the transmit power of $\mathbf{F}$ and $\mathbf{F}^{'}$ for the same $\mathbf{\Lambda}_s$. The transmit power when we use $\mathbf{F}$ is

$$\text{Tr}(\mathbf{F}\mathbf{\Lambda}_{\rm s}\mathbf{F}^{\dagger}) = \text{Tr}(\mathbf{A}\mathbf{\Lambda}_{\rm s}\mathbf{A}^{\dagger}) + \text{Tr}(\mathbf{A}_1\mathbf{\Lambda}_{\rm s}\mathbf{A}_1^{\dagger}). \tag{3.11}$$

Note that the transmit power with $\mathbf{F}^{'}$ is $\text{Tr}(\mathbf{F}^{'}\mathbf{\Lambda}_{\rm s}\mathbf{F}^{'\dagger}) = \text{Tr}(\mathbf{A}\mathbf{\Lambda}_{\rm s}\mathbf{A}^{\dagger}) \leq \text{Tr}(\mathbf{F}\mathbf{\Lambda}_{\rm s}\mathbf{F}^{\dagger})$. This means a transmitter of the form in (3.7) is no loss of generality. $\triangle\triangle\triangle$

Using Lemma 3.1 and Lemma 2.1, the receiver $\mathbf{G}$ is given by

$$\mathbf{G} = (\mathbf{A}^{\dagger}\mathbf{\Lambda}^2\mathbf{A})^{-1}[\; \mathbf{A}^{\dagger}\mathbf{\Lambda} \quad \mathbf{0} \;]\mathbf{U}^{\dagger}, \tag{3.12}$$

where $\mathbf{A}$ is the matrix given in (3.7). In this case, the noise variance at the k-th subchannel becomes

$$\sigma_{e_k}^2 = N_0[\mathbf{GG}^{\dagger}]_{kk} = N_0[(\mathbf{A}^{\dagger}\mathbf{\Lambda}^2\mathbf{A})^{-1}]_{kk}. \tag{3.13}$$

Note that the transmitter and receiver in (3.7) and (3.12) have the same form as those in [28] and [14]. The transceivers in [28] and [14] are designed for minimizing the transmit power for a given bit allocation, while we jointly design the optimal transceiver and bit allocation for maximizing the transmission rate. Lemma 3.1 lead us to conclude that the matrix $\mathbf{A}$ in (3.7) is the only part of the transceiver left to be designed. Using the expression of $\mathbf{F}$ in Lemma 3.1 and the expression of $\sigma_{e_k}^2$ in (3.13), the bit rate in (3.5) becomes

$$B = \log_2\left[(\frac{P_0}{MN_0\Gamma})^M \frac{1}{\Phi}\right], \tag{3.14}$$

where $\Phi = \prod_{k=0}^{M-1}[\mathbf{A}^{\dagger}\mathbf{A}]_{kk}[(\mathbf{A}^{\dagger}\mathbf{\Lambda}^2\mathbf{A})^{-1}]_{kk}$. To maximize b, we need to find $\mathbf{A}$ that minimizes Φ.

Optimal structure of A. Applying the Hadamard inequality [41], we have

$$\Phi = \prod_{k=0}^{M-1} [\mathbf{A}^\dagger \mathbf{A}]_{kk} [(\mathbf{A}^\dagger \mathbf{\Lambda}^2 \mathbf{A})^{-1}]_{kk} \tag{3.15}$$

$$\geq \det[\mathbf{A}^\dagger \mathbf{A}] \det[(\mathbf{A}^\dagger \mathbf{\Lambda}^2 \mathbf{A})^{-1}]. \tag{3.16}$$

The equality holds if and only if the matrix $\mathbf{A}$ satisfies the following two conditions: 1) $\mathbf{A}^\dagger \mathbf{A}$ is diagonal, and 2) $\mathbf{A}^\dagger \mathbf{\Lambda}^2 \mathbf{A}$ is diagonal. The first condition means that the columns of $\mathbf{A}$ are orthogonal, while the second means that the columns of $\mathbf{\Lambda} \mathbf{A}$ are orthogonal. As $\mathbf{\Lambda} \mathbf{A}$ is orthogonal, we can express it as $\mathbf{\Lambda} \mathbf{A} = \mathbf{Q} \mathbf{D}$, for some $K \times M$ unitary matrix $\mathbf{Q}$ such that $\mathbf{Q}^\dagger \mathbf{Q} = \mathbf{I}_M$, and some $M \times M$ nonsingular diagonal matrix $\mathbf{D}$. As $\mathbf{\Lambda}$ is nonsingular, we can write $\mathbf{A} = \mathbf{\Lambda}^{-1} \mathbf{Q} \mathbf{D}$. Then the product of the two determinant quantities in (3.16) becomes $\det[\mathbf{A}^\dagger \mathbf{A}] \det[(\mathbf{A}^\dagger \mathbf{\Lambda}^2 \mathbf{A})^{-1}] = \det(\mathbf{Q}^\dagger \mathbf{\Lambda}^{-2} \mathbf{Q})$. Hence the bit rate in (3.14) is simplified to

$$B = \log_2 \left[\left(\frac{P_0}{M N_0 \Gamma} \right)^M \frac{1}{\det(\mathbf{Q}^\dagger \mathbf{\Lambda}^{-2} \mathbf{Q})} \right]. \tag{3.17}$$

Note that the bit rate in (3.17) is independent of $\mathbf{D}$. Without loss of generality, we can choose $\mathbf{D}$ to be any $M \times M$ nonsingular diagonal matrix. For example, we can choose $\mathbf{D} = \mathbf{I_M}$. To achieve the maximal bit rate, we need to find $\mathbf{Q}$ that minimizes $\det(\mathbf{Q}^\dagger \mathbf{\Lambda}^{-2} \mathbf{Q})$.

Optimal Q: We can find $\mathbf{Q}$ with the help of the Poincaré separation theorem [41], which is presented below for convenience.

Poincaré separation theorem [41]: *Let $\mathbf{B}$ be an $n \times n$ Hermitian matrix and $\mathbf{C}$ be an $n \times r$ unitary matrix with $\mathbf{C}^\dagger \mathbf{C} = \mathbf{I}_r$. Then we have $\rho_i(\mathbf{B}) \leq \rho_i(\mathbf{C}^\dagger \mathbf{B} \mathbf{C}) \leq \rho_{n-r+i}(\mathbf{B})$, $i = 0, 1, \cdots, r-1$, where the notation $\rho_i(\mathbf{X})$ denotes the i-th smallest eigenvalue of $\mathbf{X}$.*

By the Poincaré separation theorem, we have $[\mathbf{\Lambda}^{-2}]_{ii} \leq \rho_i(\mathbf{Q}^\dagger \mathbf{\Lambda}^{-2} \mathbf{Q})$, $i = 0, 1, \cdots, M-1$, where we have used the property that the diagonal elements of $\mathbf{\Lambda}$ are in nonincreasing order. Since the diagonal matrix $\mathbf{\Lambda}$ is nonsingular,

we know that $\mathbf{\Lambda}^{-2}$ is positive definite and $[\mathbf{\Lambda}^{-2}]_{ii} > 0$ for $i = 0, 1, \cdots, K - 1$. Therefore, we have

$$\det(\mathbf{Q}^\dagger \mathbf{\Lambda}^{-2} \mathbf{Q}) = \prod_{i=0}^{M-1} \rho_i(\mathbf{Q}^\dagger \mathbf{\Lambda}^{-2} \mathbf{Q}) \tag{3.18}$$

$$\geq \prod_{i=0}^{M-1} [\mathbf{\Lambda}^{-2}]_{ii} = \det(\mathbf{\Lambda}_{\mathrm{M}}^{-2}), \tag{3.19}$$

where $\mathbf{\Lambda}_M$ is an $M \times M$ diagonal matrix whose diagonal elements consist of the M largest singular values of $\mathbf{H}$. The inequality in (3.19) becomes an equality when we choose

$$\mathbf{Q} = \left[\begin{array}{c} \mathbf{I}_M \\ \mathbf{0} \end{array} \right]. \tag{3.20}$$

With this choice of $\mathbf{Q}$ and $\mathbf{D} = \mathbf{I}_M$, we have

$$\mathbf{A} = \mathbf{\Lambda}^{-1} \mathbf{Q} \mathbf{D} = \left[\begin{array}{c} \mathbf{\Lambda}_M^{-1} \\ \mathbf{0} \end{array} \right]. \tag{3.21}$$

Using (3.7) and (3.12), the optimal transceiver is given by

$$\mathbf{F} = \mathbf{V} \left[\begin{array}{c} \mathbf{\Lambda}_M^{-1} \\ \mathbf{0} \end{array} \right], \quad \mathbf{G} = \left[\begin{array}{cc} \mathbf{I}_M & \mathbf{0} \end{array} \right] \mathbf{U}^\dagger . \tag{3.22}$$

The maximal bit rate in (3.17) is given by $b = \log_2[(\frac{P_0}{MN_0\Gamma})^M \det(\mathbf{\Lambda}_{\mathrm{M}}^2)]$. Substituting (3.22) into (3.3) and (3.13), we have

$$\sigma_{s_k}^2 = \frac{P_0}{M}[\mathbf{\Lambda}_M^2]_{kk} \text{ and } \sigma_{e_k}^2 = N_0. \tag{3.23}$$

Using (2.19), the number of bits allocated to the k-th subchannel becomes

$$b_k = \log_2 \left[1 + (\frac{P_0}{MN_0\Gamma})[\mathbf{\Lambda}_M^2]_{kk} \right]. \tag{3.24}$$

We can see that more bits are assigned to subchannels that correspond to larger singular values of the channel.

Remarks:

1. Note that if we choose $\mathbf{D} = \boldsymbol{\Lambda}_M$, we have

$$\mathbf{F} = \mathbf{V} \begin{bmatrix} \mathbf{I}_M \\ \mathbf{0} \end{bmatrix}, \quad \mathbf{G} = [\, \boldsymbol{\Lambda}_M^{-1} \quad \mathbf{0}\,] \mathbf{U}^\dagger. \tag{3.25}$$

   In this case, $\sigma_{s_k}^2 = P_0/M$ and all subchannels are assigned the same power. The bit allocation and bit rate are the same as the case when we choose $\mathbf{D} = \mathbf{I}_M$. This is because the signal to noise ratio $\sigma_{s_k}^2 / \sigma_{e_k}^2$ is not affected by $\mathbf{D}$. Therefore, bit assignment and hence bit rate performance will be the same.

2. In (4.50), the bits are not integers in general. We can use truncation, i.e., $\tilde{b}_k = \lfloor b_k \rfloor$, where the notation $\lfloor z \rfloor$ denotes the largest integer that is less than or equal to z. Zero bits may be assigned to some subchannels ($\tilde{b}_k = 0$ if $P_0[\boldsymbol{\Lambda}_M^2]_{kk} < MN_0\Gamma$) and the power allocated to these subchannels is wasted. In this case, we will remove the worst subchannel and compute bit and power allocation in the remaining subchannel. We continue like this until all the power is used by subchannels with nonzero bits. The iterative bit allocation algorithm is given below.

   Integer bit allocation algorithm:

   Let M_0 be the number of subchannels that will be assigned nonzero bits. Initially, set $M_0 = M$.

   (a) Compute $\xi_k = \frac{P_0[\boldsymbol{\Lambda}_M^2]_{kk}}{M_0 N_0 \Gamma}$ for $k = 0, 1, \cdots, M_0 - 1$.

   (b) If $\xi_k \geq 1$ for $k = 0, 1, \cdots, M_0 - 1$, then go to step (c). Else, if $\xi_k < 1$ for some subchannels, set $M_0 = M_0 - 1$ and go to step (a).

   (c) Compute the bit allocation by $b_k = \lfloor \log_2(1 + \xi_k) \rfloor$ for $0 \leq k < M_0$. For $M_0 \leq k < M$, we set $b_k = 0$.

3.3 Simulations

In the simulation, we evaluate the performance of the proposed method. The number of subchannels M is 4. The channel used is a 4×4 MIMO channel ($P = N = 4$). The elements of $\mathbf{H}$ are complex Gaussian random variables whose real and imaginary parts are independent with zero mean and variance $1/2$. The noise vector $\mathbf{q}$ is assumed to be complex Gaussian with $\mathrm{E}[\mathbf{q}\mathbf{q}^\dagger] = \mathbf{I}_4$. QAM modulation is used for the input symbols. Optimal zero-forcing transceiver in (3.22) is used for the proposed method. Although the high bit rate assumption ($b_k \gg 1$) is used in the derivation of the optimal transceivers, the assumption is not used in the computation of transmission bit rate in the simulations. We will use the integer bit allocation in remark 2 instead.

Fig. 3.1 shows the transmission rates for different transmit power to noise ratio (P_0/N_0). The symbol error rates are 10^{-2} for all the subchannels. The transmission rates are averaged over 10^6 random channel realizations. For comparison, we have also shown the results of three zero-forcing systems: the maximum signal to noise ratio (MSNR) transceiver in [8], the unit noise variance (UNV) transceiver in [14], and the SVD-waterfilling solution in [3], and also the results of two optimal transceivers in [13] that are designed using a minimum mean square error (MMSE) criterion and a maximum mutual information (MMI) criterion. Both of the MMSE [13] and MMI [13] systems use MMSE reception. In the UNV [14] and MSNR [8] systems, as all the subchannels have the same signal to noise ratios, the same bits are assigned for all subchannels. For the MMSE and MMI systems, we use the bit loading method mentioned in equation (46) of [13]. Fig. 3.1 shows that the proposed method can achieve a bit rate considerably higher than MMSE [13], UNV [14], and MSNR [8], and slightly better than MMI [13] and SVD-waterfilling [3]. We should note that, although the proposed system is zero-forcing, it is still better than the two MMSE systems in [13], in which the noise statistics is also

taken into consideration. In Fig. 3.2, we plot the bit error rates averaged over 10^6 random channel realizations when the total number of bits per block is fixed to eight for the same six systems. For the proposed method, we compute the bit allocation that is obtained when $P_0/N_0 = 12$ dB (the corresponding bits per block is eight in Fig. 3.1) and the same bit allocation is used in generating the plot in Fig. 3.2. Similarly, we allocate bits for the other five system such that the number of total bits is eight. In Fig. 3.2, we can see that the proposed method has the smallest bit error rate. The bit error rate of the proposed zero-forcing system is even smaller than the MMI [13], which use a MMSE receiver.
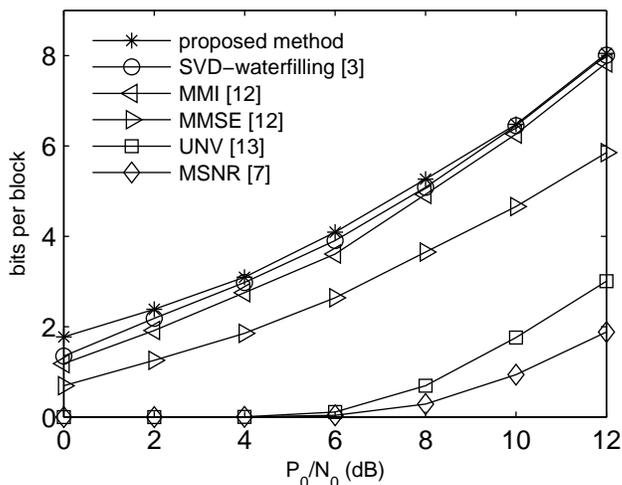
Figure 3.1: Transmission bit rates for a fixed error rate.

3.4 Summary

In this chapter, we have designed the transceiver over an MIMO channel for maximizing transmission rate. The bit allocation and transceiver were jointly optimized subject to a total power constraint for a fixed error rate. Using a high bit rate assumption, we showed that we can simultaneously obtain the optimal bit allocation and transceiver easily. We have demonstrated through simulations

22
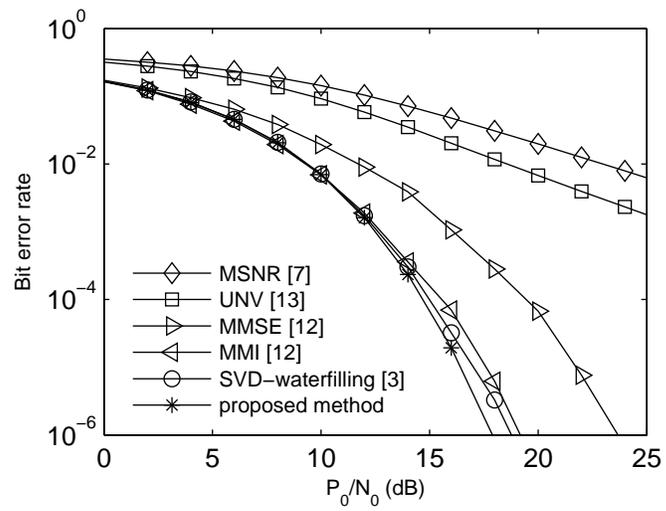
Figure 3.2: Bit error rate performance.

that the proposed method can indeed achieve a higher transmission rate although the high bit rate assumption is not used in the computation of bit allocation.
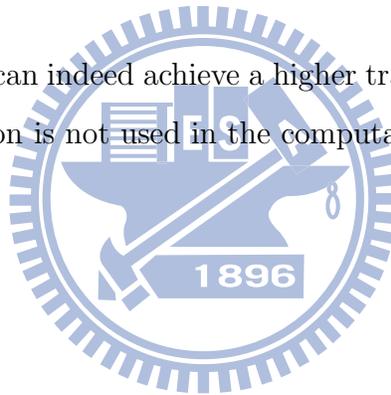
Chapter 4

Optimal MMSE Transceivers with Bit Allocation to Maximize Bit Rate

In chapter 3, we have designed the ZF transceiver for maximizing bit rate using the high bit rate assumption. In this chapter, we will design the rate-maximizing transceiver for the MMSE receiver. In this design, we do not use the high bit rate assumption as in chapter 3. We will find the optimal solution using the majorization theory. We will show the optimal MMSE receiver is in fact zero-forcing. Based on the optimal solution, we can also develop an algorithm to find the optimal integer bit allocation.

4.1 Preliminaries

In this chapter, we will use the majorization theorem to solve the optimization problem. Some related notation and results from [42] are given in this section.

Definition 4.1 [42] *Let* $\mathbf{x}$, $\mathbf{y}$ *be* $n \times 1$ *vectors, and the elements of* $\mathbf{x}$ *and* $\mathbf{y}$ *be in nonincreasing order. We say* $\mathbf{x}$ *is majorized by* $\mathbf{y}$ *(or* $\mathbf{y}$ *majorizes* $\mathbf{x}$*) if*

$$
\begin{array}{rcl}
\sum_{l=0}^{k} x_l & \leq & \sum_{l=0}^{k} y_l, \quad 0 \leq k < n-1 \\
\sum_{l=0}^{n-1} x_l & = & \sum_{l=0}^{n-1} y_l,
\end{array}
\tag{4.1}
$$

and it is denoted by $\mathbf{x} \prec \mathbf{y}$.

Definition 4.2 [42] *A real valued function ϕ defined on an n-dimensional space Ω is said to be Schur-convex on Ω if*

$$\phi(\mathbf{x}) \leq \phi(\mathbf{y}), \qquad (4.2)$$

whenever $\mathbf{x} \prec \mathbf{y}$, for all $\mathbf{x}, \mathbf{y} \in \Omega$.

Proposition 4.1 [42] *Let $\mathbf{X}$ be an $N \times N$ Hermitian matrix with diagonal elements denoted by the vector $\mathbf{d}$ and eigenvalues denoted by the vector $\boldsymbol{\lambda}$. Then we have*

$$\mathbf{d} \prec \boldsymbol{\lambda}. \qquad (4.3)$$

Proposition 4.2 [42] *Schur-convex linear combination. Let*

$$f(x_0, x_1, \cdots, x_{P-1}) = \sum_{l=0}^{P-1} a_l g(x_l), \qquad (4.4)$$

where $x_0 \leq x_1 \leq \cdots \leq x_{P-1}$ and assume the following conditions:

1. *$a_{P-1} \geq a_{P-2} \geq \cdots \geq a_1 \geq a_0 > 0$.*

2. *$\frac{dg(x)}{dx} \leq 0$ ($g(x)$ monotone decreasing).*

3. *$\frac{d^2 g(x)}{dx^2} \geq 0$ ($g(x)$ convex).*

Then $f(x_0, x_1, \cdots, x_{P-1})$ is Schur-convex on $\{x_0, x_1, \cdots, x_{P-1}\}$.

Proposition 4.3 [42] *Let a_i, b_i, $i = 1, \cdots, n$, be two sets of numbers. Let the nonincreasing arrangement of a_i and b_i be $\hat{a}_i$ and $\hat{b}_i$ respectively, i.e., $\hat{a}_1 \geq \hat{a}_2 \geq \cdots \geq \hat{a}_n$ and $\hat{b}_1 \geq \hat{b}_2 \geq \cdots \geq \hat{b}_n$. Then we have*

$$\sum_{i=1}^{n} a_i b_i \geq \sum_{i=1}^{n} \hat{a}_i \hat{b}_{n-i+1}. \qquad (4.5)$$

4.2 Problem Formulation

In this section, we will formulate the problem of designing the optimal transceiver for maximizing bit rate. Assume the symbol error rates (SER) are the same for all the subchannels. The problem of maximizing bit rate subject to a total transmit power constraint P_0 can be formulated as

$$\underset{\mathbf{G},\mathbf{\Lambda}_s}{\text{maximize}} \quad b = \sum_{k=0}^{M-1} \log_2 \left(1 + \left(\frac{\sigma_{s_k}^2}{\sigma_{e_k}^2} - 1 \right) / \Gamma \right) \tag{4.6}$$
$$\text{subject to} \quad \text{Tr}(\mathbf{F}\mathbf{\Lambda}_s\mathbf{F}^\dagger) \leq P_0.$$

The following Lemma shows that without loss of generality we can assume the diagonal elements of $\mathbf{\Lambda}_s$ is either 0 or 1.

Lemma 4.1 *For the bit rate maximization problem in* (4.6)*, there is no loss of generality to assume that $\sigma_{s_k}^2 \in \{0, 1\}$.*

Proof: Suppose the system $(\mathbf{F}, \mathbf{\Lambda}_s)$ is optimal for the bit rate maximization problem in (4.6). In general, $\mathbf{\Lambda}_s$ is not invertible. If $\sigma_{s_k}^2 = 0$ for some k, using (2.14) we have $\sigma_{e_k}^2 = 0$. Let M_r be the number of nonzero elements in $\mathbf{\Lambda}_s$. From section 2.2, we can consider only the subchannels assigned with nonzero power. Let the reduced transmitter be $\mathbf{F}_r$ and the reduced power allocation be $\mathbf{\Lambda}_r$, where $\mathbf{\Lambda}_r$ is invertible. Now, consider a new MMSE system $(\tilde{\mathbf{F}}, \tilde{\mathbf{\Lambda}}_s)$ which is given by $\tilde{\mathbf{\Lambda}}_s = \mathbf{I}_{M_r}$, and

$$\tilde{\mathbf{F}} = \mathbf{F}_r\mathbf{\Lambda}_r^{1/2}. \tag{4.7}$$

Then the transmit power when we use $\tilde{\mathbf{F}}$ and $\tilde{\mathbf{\Lambda}}_s$ is given by

$$\text{Tr}(\tilde{\mathbf{F}}\tilde{\mathbf{\Lambda}}_s\tilde{\mathbf{F}}^\dagger) = \text{Tr}(\mathbf{F}_r\mathbf{\Lambda}_r\mathbf{F}_r^\dagger). \tag{4.8}$$

Clearly, the transmit power of the new system is the same as the original system. Now let's compare the bit rate of the original system with the new system. For the new system, the MSE matrix is $\tilde{\mathbf{E}} = \mathbf{\Lambda}_r^{-1/2}\mathbf{E}_r\mathbf{\Lambda}_r^{-1/2}$, and hence the k-th subchannel signal to noise ratio is

$$\frac{\tilde{\sigma}_{s_k}^2}{\tilde{\sigma}_{e_k}^2} = \frac{1}{\sigma_{e_k}^2 \sigma_{s_k}^{-2}} = \frac{\sigma_{s_k}^2}{\sigma_{e_k}^2}, \tag{4.9}$$

26

which is the same as the original system. As the bit loading formulation in (2.20) depends only the subchannel signal to noise ratios, we can conclude that the bit rate of the new system is the same as the original system. Therefore, without loss of generality we can assume $\sigma_{s_k}^2 \in \{0, 1\}$. $\triangle\triangle\triangle$

In the next section, we will assume $\sigma_{s_k}^2 \in \{0, 1\}$ and find the optimal MMSE transceiver that maximize the transmission rate in (4.6).

4.3   Optimal Transceiver Design

Suppose the $P \times N$ channel matrix $\mathbf{H}$ has rank K. Let the singular value decomposition of $\mathbf{H}$ be

$$\mathbf{H} = \mathbf{U} \begin{bmatrix} \mathbf{\Lambda} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{V}^\dagger, \qquad (4.10)$$

where the $K \times K$ diagonal matrix $\mathbf{\Lambda}$ contains the nonzero singular values of $\mathbf{H}$ in nonincreasing order, i.e., $\lambda_0 \geq \lambda_1 \geq \cdots \geq \lambda_{K-1}$. The $P \times P$ matrix $\mathbf{U}$ and the $N \times N$ matrix $\mathbf{V}$ are unitary.

Lemma 4.2 *Without loss of generality, the transmitter can be expressed as*

$$\mathbf{F} = \mathbf{V} \begin{bmatrix} \mathbf{A} \\ \mathbf{0} \end{bmatrix}, \qquad (4.11)$$

for appropriate $K \times M$ matrix $\mathbf{A}$. For the choice of $\mathbf{F}$ in (4.11), the reduced MSE matrix $\mathbf{E}_r$ in (2.16) is given by

$$\mathbf{E}_r = [N_0^{-1}\mathbf{A}_r{}^\dagger\mathbf{\Lambda}^2\mathbf{A}_r + \mathbf{I}_{M_r}]^{-1}, \qquad (4.12)$$

where $\mathbf{A}_r$ is obtained by removing the columns of $\mathbf{A}$ that correspond to the subchannels assigned with zero power.

Proof: As $\mathbf{V}$ is unitary, $\mathbf{F}$ can always be represented as

$$\mathbf{F} = \mathbf{V} \begin{bmatrix} \mathbf{A} \\ \mathbf{A}_1 \end{bmatrix}, \qquad (4.13)$$

where $\mathbf{A}$ is a $K \times M$ matrix and $\mathbf{A}_1$ is an $(N - K) \times M$ matrix. Define a new transmitter $\mathbf{F}'$ as

$$\mathbf{F}' = \mathbf{V} \begin{bmatrix} \mathbf{A} \\ \mathbf{0} \end{bmatrix}. \tag{4.14}$$

Then we have $\mathbf{HF} = \mathbf{HF}'$ Using (2.14), we can see that the new MSE matrix $\mathbf{E}'$ is equal to $\mathbf{E}$, i.e., $\mathbf{E}'=\mathbf{E}$. Therefore, the new system has the same subchannel error variances and hence the same bit rate performance. Now, let us compare the transmit power. The transmit power when we use $\mathbf{F}$ is

$$\mathrm{Tr}(\mathbf{F}\mathbf{\Lambda}_s\mathbf{F}^\dagger) = \mathrm{Tr}(\mathbf{A}\mathbf{\Lambda}_s\mathbf{A}^\dagger) + \mathrm{Tr}(\mathbf{A}_1\mathbf{\Lambda}_s\mathbf{A}_1^\dagger). \tag{4.15}$$

Note that the transmit power with $\mathbf{F}'$ is

$$\mathrm{Tr}(\mathbf{F}'\mathbf{\Lambda}_s\mathbf{F}'^\dagger) = \mathrm{Tr}(\mathbf{A}\mathbf{\Lambda}_s\mathbf{A}^\dagger) \leq \mathrm{Tr}(\mathbf{F}\mathbf{\Lambda}_s\mathbf{F}^\dagger). \tag{4.16}$$

This means a transmitter of the form in (4.11) is no loss of generality. We can verify that when the transmitter is as in (4.11), the reduced MSE matrix is as given in (4.2).

$$\triangle\triangle\triangle$$

Note that the transmitter in (4.11) has the same form as in (3.7) for the ZF case. Using Lemma 4.2, the problem is reduced to the design of the matrix $\mathbf{A}$ only. The following Lemma shows that the optimization of $\mathbf{A}$ can be further simplified to that of a unitary matrix and a diagonal matrix.

Lemma 4.3 *When the the transmission rate in* (4.6) *is maximized, the MSE matrix* $\mathbf{E}$ *is diagonal. Then the matrix* $\mathbf{A}$ *in lemma* 4.2 *is of the form*

$$\mathbf{A} = \mathbf{\Lambda}^{-1}\mathbf{Q}\mathbf{D}, \tag{4.17}$$

for some $K \times M$ *unitary matrix* $\mathbf{Q}$ *such that* $\mathbf{Q}^\dagger\mathbf{Q} = \mathbf{I}_M$, *and some* $M \times M$ *diagonal matrix* $\mathbf{D}$.

Proof: Let $g(x) = \log_2(1 + (x^{-1} - 1)/\Gamma)$ and $x > 0$. Then we have $\frac{\partial g}{\partial x} \leq 0$ and

$\frac{\partial^2 g}{\partial x^2} \geq 0$ for $x > 0$. Suppose $(\mathbf{\Lambda}_s, \mathbf{F})$ is optimal for (4.6). Let M_r denote the number of nonzero elements in the optimal power allocation. Let $(\mathbf{F}_r, \mathbf{\Lambda}_r)$ be the reduced system obtained for the optimal solution. Without loss of generality, we can assume $\{\sigma^2_{e_{r,k}}\}$ is in nondecreasing order[1]. Using Proposition 2, the transmission rate

$$b(\{\sigma^2_{e_{r,k}}\}) = \sum_{k=0}^{M-1} \log_2 \left(1 + \left(\frac{1}{\sigma^2_{e_{r,k}}} - 1 \right)/\Gamma \right) = \sum_{k=0}^{M-1} g(\sigma^2_{e_{r,k}}) \tag{4.18}$$

is a schur-convex function on $\{\sigma^2_{e_{r,k}}\}$. Suppose $\mathbf{E}_r$ is not diagonal. Let the eigenvalue decomposition of $\mathbf{E}_r$ be $\mathbf{E}_r = \mathbf{T}\mathbf{\Lambda}_e\mathbf{T}^\dagger$, where $\mathbf{T}$ is unitary and the diagonal elements of $\mathbf{\Lambda}_e$, denoted by $\lambda_{e,k}$, are in nondecreasing order. Now consider a new transmitter $\tilde{\mathbf{F}} = \mathbf{F}_r\mathbf{T}$. The new transmission power $\mathrm{Tr}(\tilde{\mathbf{F}}\tilde{\mathbf{F}}^\dagger)$ is the same as the case when we use $\mathbf{F}_r$. Using (2.16), the MSE matrix of the new system is given by

$$\tilde{\mathbf{E}} = [N_0^{-1}\tilde{\mathbf{F}}^\dagger\mathbf{H}^\dagger\mathbf{H}\tilde{\mathbf{F}} + \mathbf{I}]^{-1} \tag{4.19}$$

$$= [N_0^{-1}\mathbf{T}^\dagger\mathbf{F}_r^\dagger\mathbf{H}^\dagger\mathbf{H}\mathbf{F}_r\mathbf{T} + \mathbf{I}]^{-1} \tag{4.20}$$

$$= \mathbf{T}^\dagger\mathbf{E}_r\mathbf{T} \tag{4.21}$$

$$= \mathbf{\Lambda}_e \tag{4.22}$$

The new subchannel noise will be decorrelated when we use $\tilde{\mathbf{F}}$ and the subchannel noise variances are $\lambda_{e,0}, \cdots, \lambda_{e,M-1}$. By Proposition 1, we have that $\{\lambda_{e,k}\} \succ \{\sigma^2_{e_k}\}$. Then by Definition 2, we have

$$b(\{\lambda_{e,k}\}) \geq b(\{\sigma^2_{e_k}\}). \tag{4.23}$$

That is, a higher bit rate can be achieved when the subchannel noise are decorrelated. This is a contradiction, so $\mathbf{E}_r$ must be diagonal, which implies $\mathbf{E}$ is

---

[1]Assume $\mathbf{F}_r$ is optimal for the problem in (4.6) and $\sigma^2_{e_{r,k}} = [\mathbf{E}_r]_{kk}$ is not in nondecreasing order. Let the new transmitter be $\tilde{\mathbf{F}} = \mathbf{F}_r\mathbf{P}$, where $\mathbf{P}$ is a permutation matrix. Then the new MSE matrix $\tilde{\mathbf{E}}$ is $\tilde{\mathbf{E}} = \mathbf{P}^T\mathbf{E}_r\mathbf{P}$. Let $\mathbf{P}$ be chosen such that $\tilde{\sigma}^2_{e_k} = [\tilde{\mathbf{E}}]_{kk}$ is in nondecreasing order. We can verify that the transmit power and bit rate of new system are the same as the case when we use $\mathbf{F}_r$. Therefore, it is no loss of generality to assume $\sigma^2_{e_{r,k}}$ in nondecreasing order.

diagonal. Using the expression of $\mathbf{E}_r$ in (4.12), we know that $\mathbf{A}_r^\dagger \mathbf{\Lambda}^2 \mathbf{A}_r$ is diagonal. Since $\mathbf{A}_r^\dagger \mathbf{\Lambda}^2 \mathbf{A}_r$ is diagonal, the columns of $\mathbf{\Lambda}\mathbf{A}_r$ are orthogonal. Let M_r be the number of nonzero elements in $\mathbf{\Lambda}_s$. We can express $\mathbf{\Lambda}\mathbf{A}_r$ as

$$\mathbf{\Lambda}\mathbf{A}_r = \mathbf{Q}_0 \mathbf{D}_0, \tag{4.24}$$

for some $K \times M_r$ unitary matrix $\mathbf{Q}_0$ such that $\mathbf{Q}_0^\dagger \mathbf{Q}_0 = \mathbf{I}_{M_r}$, and some nonsingular $M_r \times M_r$ diagonal matrix $\mathbf{D}_0$. Rearranging (4.24), we can write $\mathbf{A}_r$ as

$$\mathbf{A}_r = \mathbf{\Lambda}^{-1} \mathbf{Q}_0 \mathbf{D}_0. \tag{4.25}$$

Note that $\mathbf{A}_r$ is obtained by removing some columns of $\mathbf{A}$. Since the columns removed from $\mathbf{A}$ do not affect the transmit power and bit rate, without loss of generality we can assume these columns are zero vectors. Hence $\mathbf{A}$ can be expressed as

$$\mathbf{A} = \mathbf{\Lambda}^{-1} \mathbf{Q} \mathbf{D}, \tag{4.26}$$

where $\mathbf{Q}$ is a $K \times M$ unitary matrix such that $\mathbf{Q}_0$ can be obtained by removing the columns of $\mathbf{Q}$ and $\mathbf{D}$ is an $M \times M$ diagonal matrix whose diagonal elements consists of the diagonal elements of $\mathbf{D}_0$ and zero. $\triangle\triangle\triangle$

Using the expression of the matrix $\mathbf{A}$ in (4.17), the transmit power can be written as

$$\begin{aligned}
\mathrm{Tr}(\mathbf{F}\mathbf{\Lambda_s}\mathbf{F}^\dagger) &= \mathrm{Tr}(\mathbf{A}\mathbf{\Lambda_s}\mathbf{A}^\dagger) \\
&= \mathrm{Tr}(\mathbf{D}^\dagger \mathbf{Q}^\dagger \mathbf{\Lambda^{-2}} \mathbf{Q} \mathbf{D} \mathbf{\Lambda_s}) \\
&= \sum_{k=0}^{M-1} \sigma_{s_k}^2 |d_k|^2 [\mathbf{Q}^\dagger \mathbf{\Lambda^{-2}} \mathbf{Q}]_{kk},
\end{aligned} \tag{4.27}$$

where d_k denotes the k-th diagonal element of $\mathbf{D}$. In this case, the k-th subchannel error variance is

$$\sigma_{e_k}^2 = \begin{cases} \dfrac{1}{N_0^{-1}|d_k|^2 + 1}, & \text{if } \sigma_{s_k}^2 = 1; \\ 0, & \text{if } \sigma_{s_k}^2 = 0, \end{cases} \tag{4.28}$$

which depends on d_k only but not $\mathbf{Q}$. The bit allocation becomes

$$b_k = \log_2 \left(\frac{|d_k|^2}{N_0 \Gamma} + 1 \right) \tag{4.29}$$

Hence the problem in (4.6) becomes

$$\begin{array}{ll} \underset{\mathbf{Q},\, |d_k|^2}{\text{maximize}} & b = \sum_{k=0}^{M-1} \log_2 \left(\frac{|d_k|^2}{N_0 \Gamma} + 1 \right) \\ \text{subject to} & \left\{ \begin{array}{l} \sum_{k=0}^{M-1} |d_k|^2 [\mathbf{Q}^\dagger \mathbf{\Lambda}^{-2} \mathbf{Q}]_{kk} \leq P_0, \\ |d_k|^2 \geq 0, \end{array} \right. \quad \text{for } k = 0, 1, \cdots, M-1. \end{array}$$

$$(4.30)$$

In (4.30), d_k and the unitary matrix $\mathbf{Q}$ are the only free parameters left to be determined. As the subchannel noise variances do not depend on the unitary matrix $\mathbf{Q}$, changing $\mathbf{Q}$ affects only the transmission power but not the bit rate. The following lemma shows us how to find the optimal $\mathbf{Q}$.

Lemma 4.4 *One optimal choice of* $\mathbf{Q}$ *for the problem in* (4.30) *is*

$$\mathbf{Q} = \begin{bmatrix} \mathbf{I}_M \\ \mathbf{0} \end{bmatrix}. \qquad (4.31)$$

In this case, the transmit power can be written as

$$\sum_{k=0}^{M-1} |d_k|^2 [\mathbf{Q}^\dagger \mathbf{\Lambda}^{-2} \mathbf{Q}]_{kk} = \sum_{k=0}^{M-1} |d_k|^2 [\mathbf{\Lambda}_M^{-2}]_{kk}, \qquad (4.32)$$

where $\mathbf{\Lambda}_M$ *is an* $M \times M$ *diagonal matrix whose diagonal elements consists of the* M *largest singular value of* $\mathbf{H}$.

Proof: We first establish a lower bound on the expression of the transmit power in (4.27) for any given $\mathbf{Q}$ and d_k. That is,

$$\sum_{k=0}^{M-1} |d_k|^2 [\mathbf{Q}^\dagger \mathbf{\Lambda}^{-2} \mathbf{Q}]_{kk} \geq \sum_{k=0}^{M-1} |d_k|^2 [\mathbf{\Lambda}_M^{-2}]_{kk}, \qquad (4.33)$$

where $\mathbf{\Lambda}_M$ is an $M \times M$ diagonal matrix whose diagonal elements consists of the M largest singular value of $\mathbf{H}$. The lower bound can be achieved by choosing $\mathbf{Q}$ as in (4.31). To prove (4.33), for convenience, we extend the M-point sequence d_k to a K-point sequence $\tilde{d}_k$ by zero padding, i.e.,

$$\tilde{d}_k = \left\{ \begin{array}{ll} d_k, & 0 \leq k \leq M-1 \\ 0, & M \leq k \leq K-1. \end{array} \right. \qquad (4.34)$$

31

Without loss of generality, we can assume that $|d_k|$ is in nonincreasing order[2], and thus so is $|\tilde{d}_k|$. Let $\mathbf{Q}_1$ be a $K \times (K - M)$ matrix such that the $K \times K$ matrix $\mathbf{Q}_0 = [\; \mathbf{Q} \quad \mathbf{Q}_1 \;]$ is unitary. Then the transmit power in (4.27) can be rewritten as

$$\sum_{k=0}^{M-1} |d_k|^2 [\mathbf{Q}^\dagger \mathbf{\Lambda}^{-2} \mathbf{Q}]_{kk} = \sum_{k=0}^{K-1} |\tilde{d}_k|^2 [\mathbf{Q}_0^\dagger \mathbf{\Lambda}^{-2} \mathbf{Q}_0]_{kk} = \sum_{k=0}^{K-1} |\tilde{d}_k|^2 \alpha_k, \qquad (4.35)$$

where $\alpha_k = [\mathbf{Q}_0^\dagger \mathbf{\Lambda}^{-2} \mathbf{Q}_0]_{kk}$. Let $\{\tilde{\alpha}_k\}$ be the nondecreasing arrangement of $\{\alpha_k\}$. Then by Proposition 4.3 we have

$$\sum_{k=0}^{K-1} |\tilde{d}_k|^2 \alpha_k \geq \sum_{k=0}^{K-1} |\tilde{d}_k|^2 \tilde{\alpha}_k. \qquad (4.36)$$

Now, let us define a function ϕ as

$$\phi(\{\tilde{\alpha}_k\}) = -\sum_{k=0}^{K-1} |\tilde{d}_k|^2 \tilde{\alpha}_k. \qquad (4.37)$$

Note that the function $\phi(\{\tilde{\alpha}_k\})$ is schur-convex on $\{\tilde{\alpha}_k\}$. To see this, let $g(x) = -x$, $x > 0$. Because $\frac{\partial g}{\partial x} \leq 0$ and $\frac{\partial^2 g}{\partial x^2} \geq 0$ for $x > 0$, by Proposition 4.2 we know that $\phi(\{\tilde{\alpha}_k\}) = \sum_{k=0}^{K-1} |\tilde{d}_k|^2 g(\tilde{\alpha}_k)$ is schur-convex on $\{\tilde{\alpha}_k\}$. Let γ_k be the k-th eigenvalues of $\mathbf{Q}_0^\dagger \mathbf{\Lambda}^{-2} \mathbf{Q}_0$ in nondecreasing order, i.e., $\gamma_k = [\mathbf{\Lambda}^{-2}]_{kk}$. By Proposition 4.1, we know $\{\gamma_k\} \succ \{\tilde{\alpha}_k\}$, which implies

$$\phi(\{\tilde{\alpha}_k\}) \leq \phi(\{\gamma_k\}) \qquad (4.38)$$

as ϕ is schur-convex. This means

$$\sum_{k=0}^{K-1} |\tilde{d}_k|^2 \tilde{\alpha}_k \geq \sum_{k=0}^{K-1} |\tilde{d}_k|^2 \gamma_k. \qquad (4.39)$$

Using (4.36), (4.39), and the facts that the last $K - M$ elements of $\{d_k\}$ are zeros and $\gamma_k = [\mathbf{\Lambda}^{-2}]_{kk}$, we have the inequality in (4.33). Now, we will use (4.33) to

---

[2]For the case that $\{|d_k|\}$ is not in nonincreasing order, let $\mathbf{D}' = \mathbf{P}\mathbf{D}\mathbf{P}^T$ and $\mathbf{Q}' = \mathbf{Q}\mathbf{P}^T$, where $\mathbf{P}$ is the permutation matrix such that $|d_k'|$ is in nonincreasing order. We can verify the new transmission rate and the new transmit power for $\mathbf{D}'$ and $\mathbf{Q}'$ are the same as the case when we use $\mathbf{D}$ and $\mathbf{Q}$. Therefore, it is no loss of generality to assume that $\{|d_k|\}$ is in nonincreasing order

show that one optimal choice of $\mathbf{Q}$ is as given in (4.31). Suppose $\mathbf{Q}^*$ and d_k^* are optimal. Then using (4.33) we have

$$\sum_{k=0}^{M-1} |d_k^*|^2 [\mathbf{Q}^{*\dagger} \boldsymbol{\Lambda}^{-2} \mathbf{Q}^*]_{kk} \geq \sum_{k=0}^{M-1} |d_k^*|^2 [\boldsymbol{\Lambda}_M^{-2}]_{kk}. \tag{4.40}$$

Suppose $\sum_{k=0}^{M-1} |d_k^*|^2 [\mathbf{Q}^{*\dagger} \boldsymbol{\Lambda}^{-2} \mathbf{Q}^*]_{kk} > \sum_{k=0}^{M-1} |d_k^*|^2 [\boldsymbol{\Lambda}_M^{-2}]_{kk}$. Consider the new $\tilde{\mathbf{Q}}$ and $\tilde{d}_k$ given by

$$\tilde{\mathbf{Q}} = \begin{bmatrix} \mathbf{I}_M \\ \mathbf{0} \end{bmatrix}, \text{ and } \tilde{d}_k = \left(\frac{\sum_{k=0}^{M-1} |d_k^*|^2 [\mathbf{Q}^{*\dagger} \boldsymbol{\Lambda}^{-2} \mathbf{Q}^*]_{kk}}{\sum_{k=0}^{M-1} |d_k^*|^2 [\boldsymbol{\Lambda}_M^{-2}]_{kk}} \right)^{1/2} d_k^*. \tag{4.41}$$

The transmit power of the new system is

$$\sum_{k=0}^{M-1} |\tilde{d}_k|^2 [\tilde{\mathbf{Q}}^{\dagger} \boldsymbol{\Lambda}^{-2} \tilde{\mathbf{Q}}]_{kk} = \sum_{k=0}^{M-1} |d_k^*|^2 [\mathbf{Q}^{*\dagger} \boldsymbol{\Lambda}^{-2} \mathbf{Q}^*]_{kk}, \tag{4.42}$$

which is the same as the optimal solution. Since $\tilde{d}_k > d_k^*$, the bit rate of the new system is is larger than that of the optimal system. This contradicts the assumption that $\mathbf{Q}^*$ and d_k^* are optimal for the problem in (4.30). Hence for the optimal solution $\mathbf{Q}^*$ and d_k^*, the equality in (4.40) must hold and one optimal choice of $\mathbf{Q}^*$ is as given in (4.31). $\triangle\triangle\triangle$

Using the expression of transmit power in (4.32), the problem in (4.30) can be simplified as

$$
\begin{array}{ll}
\underset{\{|d_k|^2\}}{\text{maximize}} & \sum_{k=0}^{M-1} \log_2 \left(\frac{|d_k|^2}{N_0 \Gamma} + 1 \right) \\
\text{subject to} & \begin{cases} \sum_{k=0}^{M-1} |d_k|^2 [\boldsymbol{\Lambda}_M^{-2}]_{kk} \leq P_0, \\ |d_k|^2 \geq 0, \end{cases} \quad k = 0, 1, \cdots, M-1.
\end{array} \tag{4.43}
$$

To solve (4.43), only $|d_k|^2$ remain to be designed. We can use the Karush-Kuhn-Tucker (KKT) condition [77]. Let $|d_k^*|^2$ be a local maximum. Then there exists constants α and β_k, for $k = 0, 1, \cdots, M-1$ such that:

1. $\alpha \leq 0$.

2. $\beta_k \leq 0$, for $k = 0, 1, \cdots, M-1$.

3. $\frac{\partial}{\partial |d_k|^2}\left(\sum_{k=0}^{M-1} \log_2 \left(1 + \frac{|d_k|^2}{N_0\Gamma} \right) + \alpha(\sum_{k=0}^{M-1} |d_k|^2[\mathbf{\Lambda}_M^{-2}]_{kk} - P_0) + \sum_{k=0}^{M-1} \beta_k(-|d_k|^2) \right) \Bigg|_{|d_k|^2 = |d_k^*|^2}$
   $= 0.$

4. $\alpha(\sum_{k=0}^{M-1} |d_k|^2[\mathbf{\Lambda}_M^{-2}]_{kk} - P_0) \Bigg|_{|d_k|^2 = |d_k^*|^2} = 0.$

5. $\beta_k(-|d_k|^2) = 0 \Bigg|_{|d_k|^2 = |d_k^*|^2} = 0$, for $k = 0, 1, \cdots, M-1.$

By solving condition 2, we have

$$\frac{1}{(|d_k^*|^2 + N_0\Gamma) \log_e 2} + \alpha[\mathbf{\Lambda}_M^{-2}]_{kk} - \beta_k = 0. \qquad (4.44)$$

Suppose $\alpha = 0$. Since $|d_k^*|^2$, $[\mathbf{\Lambda}_M^{-2}]_{kk}$, and $\log_e 2$ are all positive, using (4.44) we have

$$\beta_k = \frac{1}{(|d_k^*|^2 + N_0\Gamma) \log_e 2} > 0. \qquad (4.45)$$

This contradicts condition 2. Hence we have $\alpha < 0$. As $\alpha \neq 0$, condition 3 is reduced to

$$\sum_{k=0}^{M-1} |d_k^*|^2[\mathbf{\Lambda}_M^{-2}]_{kk} = P_0. \qquad (4.46)$$

If $\beta_k < 0$, then using condition 5 we have $|d_k|^2 = 0$. If $\beta_k = 0$, by (4.44) we have

$$|d_k^*|^2 = \frac{-1}{\alpha \log_e 2[\mathbf{\Lambda}_M^{-2}]_{kk}} - N_0\Gamma, \qquad (4.47)$$

Thus the optimal $|d_k^*|^2$ is given by

$$|d_k^*|^2 = \left(\frac{-1}{\alpha \log_e 2[\mathbf{\Lambda}_M^{-2}]_{kk}} - N_0\Gamma \right)^+, \qquad (4.48)$$

where $(x)^+ = \max(x, 0)$, and the constant α is chosen to satisfy

$$\sum_{k=0}^{M-1} \left(\frac{-1}{\alpha \log_e 2} - N_0\Gamma[\mathbf{\Lambda}_M^{-2}]_{kk} \right)^+ = P_0. \qquad (4.49)$$

The solution in (4.44) is the so-called "water-filling" solution. The number of bits allocated to the k-th subchannel is given by

$$b_k = \log_2 \left(\frac{|d_k^*|^2}{N_0\Gamma} + 1 \right). \qquad (4.50)$$

34

From (4.48), we see that for subchannels that correspond to larger singular values of the channel, $\{|d_k^*|^2\}$ is larger and more bits are assigned. Once the optimal $\{|d_k^*|^2\}$ is obtained by (4.48), the bit allocation in (4.50) can be determined. Using the choice of $\mathbf{Q}$ in (4.31), the matrix $\mathbf{A}$ becomes

$$\mathbf{A} = \begin{bmatrix} \boldsymbol{\Lambda}_M^{-1} \\ \mathbf{0} \end{bmatrix} \mathbf{D}. \tag{4.51}$$

Substituting (4.51) into (4.13), the optimal transmitter is given by

$$\mathbf{F} = \mathbf{V} \begin{bmatrix} \boldsymbol{\Lambda}_M^{-1} \\ \mathbf{0} \end{bmatrix} \mathbf{D}, \tag{4.52}$$

Using the optimal transmitter in (4.52), the optimal receiver in (2.13) becomes

$$\mathbf{G} = \tilde{\mathbf{D}}[\ \mathbf{I}_M \quad \mathbf{0}\ ]\mathbf{U}^\dagger, \tag{4.53}$$

where $\tilde{\mathbf{D}}$ is a diagonal matrix whose diagonal elements is

$$[\tilde{\mathbf{D}}]_{kk} = \frac{d_k^*}{1 + N_0^{-1}|d_k|^2}, \tag{4.54}$$

In the optimal solution, only d_k depends on the transmit power P_0 and the given error rate. The unitary matrices $\mathbf{V}$, $\mathbf{U}$ and the diagonal matrix $\boldsymbol{\Lambda}_M$ depend only on the channel matrix $\mathbf{H}$. When the optimal transceiver is applied, the output of the receiver is given by

$$\hat{\mathbf{s}} = \mathbf{GHFs} + \mathbf{Gq} \tag{4.55}$$

$$= \mathbf{Ts} + \mathbf{n}, \tag{4.56}$$

where $\mathbf{n} = \mathbf{Gq}$ and $\mathbf{T} = \mathbf{GHF}$. The autocorrelation of $\mathbf{n}$ is $N_0\tilde{\mathbf{D}}\tilde{\mathbf{D}}^\dagger$, which is a diagonal matrix. The overall transfer function $\mathbf{T}$ is diagonal and the diagonal element is

$$[\mathbf{T}]_{kk} = \frac{|d_k|^2}{1 + N_0^{-1}|d_k|^2}. \tag{4.57}$$

Let $\mathbf{T}_r$ be the $M_r \times M_r$ diagonal matrix obtained by removing the rows and columns of $\mathbf{T}$ that correspond to the zero diagonal elements. Then $\mathbf{T}_r$ is the overall transfer function of the reduced system, i.e.,

$$\hat{\mathbf{s}}_r = \mathbf{T}_r \mathbf{s}_r + \mathbf{G}_r \mathbf{q}. \tag{4.58}$$

Since $\mathbf{T}_r$ is diagonal, for the same transmitter $\mathbf{F}_r$ and signal autocorrelation matrix $\mathbf{\Lambda}_r$, we can a ZF receiver that achieve the same bit rate. Consider a ZF receiver given by $\mathbf{G}_{r,zf} = \mathbf{T}_r^{-1} \mathbf{G}_r$. We have $\mathbf{G}_{r,zf} \mathbf{H} \tilde{\mathbf{F}} = I_{M_r}$. The unbiased signal to noise ratio of the k-th subchannel for the new system is

$$\frac{1}{N_0 [\mathbf{G}_{r,zf} \mathbf{G}_{r,zf}^{\dagger}]_{kk}} = \frac{1}{N_0 [\mathbf{T}_r^{-1} \mathbf{G}_r \mathbf{G}_r \dagger \mathbf{T}_r^{-\dagger}]_{kk}} \tag{4.59}$$

$$= \frac{|[\mathbf{T}_r]_{kk}|^2}{N_0 [\mathbf{G}_r \mathbf{G}_r^{\dagger}]_{kk}}, \tag{4.60}$$

which is the same as the optimal solution. Thus the bit rate of the ZF system is the same as the optimal solution. This implies the solution of the MMSE transceiver is the same as the ZF transceiver.

In general, the bit allocation obtained in (4.50) is not integer. To obtain the solution with integer bit allocation, we can use the results of [34]. The results in [34] shows the greedy algorithm is optimal when a transceiver with diagonal structure is given. The algorithm is shown below:

Greedy algorithm for integer bit allocation:

Suppose the power constraint P_0 is given. Initially, set $b_0 = b_1 = \cdots = b_{M-1} = 0$. Define the power increase of k-th subchannel as $\triangle p_k = N_0 \Gamma [\mathbf{\Lambda}_M^{-2}]_{kk} (2^{b_k+1} - 2^{b_k})$.

1. Compute $\triangle p_k$ for $k = 0, 1, \cdots, M - 1$.

2. Find the index i such that $\triangle p_k$ is minimal. Set $b_i = b_i + 1$.

3. Computed the transmit power $P = \sum_{k=0}^{M-1} N_0 \Gamma (2^{b_k} - 1) [\mathbf{\Lambda}_M^{-2}]_{kk}$. If $P < P_0$, go to step 1. If $P = P_0$, the optimal bit allocation is $\{b_0, \cdots, b_{M-1}\}$. If $P > P_0$, then the optimal bit allocation is $\{b_0, \cdots, b_i - 1, \cdots, b_{M-1}\}$.

4.4 Simulations

In this section, we evaluate the performance of the proposed method. The number of subchannels M is 4. The channel used is a 4×4 MIMO channel ($P = N = 4$). The elements of $\mathbf{H}$ are complex Gaussian random variables with zero mean and unit variance. The noise vector $\mathbf{q}$ is assumed to be complex Gaussian with $\mathrm{E}[\mathbf{q}\mathbf{q}^\dagger] = \mathbf{I}_4$. QAM modulation is used for the input symbols. In the following examples, we will use the optimal transceiver in (4.52) and (4.53). In (4.50), the bits are not integers in general. For integer bit allocation, we will use the greedy algorithm to find the optimal solution.

Example 1. Fig. 4.1 shows the transmission rates for different transmit power to noise ratio (P_0/N_0). The symbol error rates are 10^{-2} for all the subchannels. The transmission rates are evaluated for 10^6 channel realizations. For comparison, we have also shown the results of five more systems: the bit rate maximizing zero-forcing transceiver in chapter 3, the zero-forcing transceiver in [18], the zero-forcing maximum signal to noise ratio (MSNR) transceiver in [8], and the zero-forcing unit noise variance (UNV) transceiver in [14], and also the results of the optimal transceivers in [13] that using a maximum mutual information (MMI) criterion. In the UNV [14] and MSNR [8] systems, as all the subchannels have the same signal to noise ratios, the same bits are assigned for all subchannels. For the MMI systems, we use the bit loading method mentioned in equation (46) of [13]. For the system in [18], we use the bit allocation in (24) of [13] and then truncate it to be integer. Fig. 4.1 shows that the proposed method can achieve a higher bit rate. For example, when $P_0/N_0 = 12$ dB, the number of bits that can be transmitted is 9 per block for the proposed system, 8 for the system in chapter 3, 7.8, 6, 3, and 2 respectively for MMI [13], [18], UNV [14] and MSNR [8] systems. The proposed method can achieve a higher bit rate the other five systems.
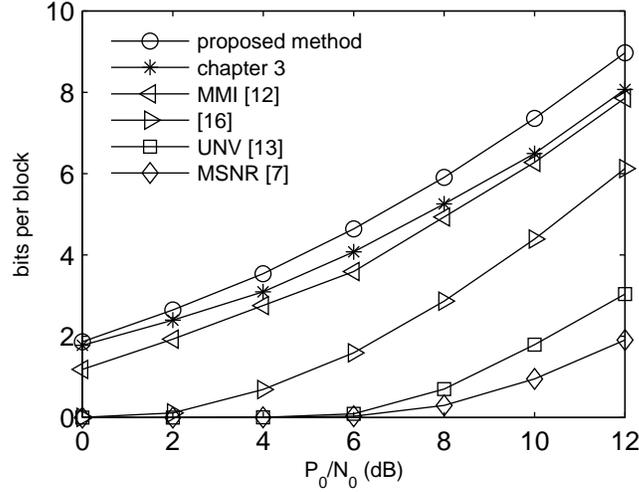
Figure 4.1: Transmission bit rates for a fixed error rate.

To better illustrate the advantage of the proposed method, we show in Fig. 4.2 the bit allocation when the data blocks are sent over a fixed channel for $P_0/N_0 = 20$ dB. The channel matrix $\mathbf{H}$ in this case is given by

$$\mathbf{H} = \begin{pmatrix} -0.31 - 0.36i & 0.28 - 0.43i & -0.91 - 0.73i & 0.05 + 0.64i \\ -0.12 + 0.17i & 0.53 - 0.86i & -1.65 + 0.94i & 0.03 - 0.21i \\ -0.15 + 0.17i & 1.26 + 0.22i & 0.64 - 0.30i & 1.57 + 0.73i \\ 0.38 + 0.05i & 0.86 - 0.95i & -1.30 - 0.10i & -0.05 - 0.24i \end{pmatrix}. \quad (4.61)$$

In the proposed method, the bits are allocated according to the subchannel signal to noise ratios, and 16 bits per block can be transmitted for this channel. 15 and 14 bits per block can be transmitted respectively for the transceiver in chapter 3 and MMI [13]. For UNV [14] and MSNR [8], all subchannels carry the same number of bits. The number of bits that can be transmitted in each block are eight and four respectively.

Example 2. In Fig. 4.3, we plot the bit error rates for a fixed transmission rate. The total number of bits per block is fixed to eight for the same five systems in example 1. For the proposed method, we compute the bit allocation that is obtained when $P_0/N_0 = 11$ dB (the corresponding bits per block is eight for the proposed system in Fig. 4.1) and the same bit allocation is used in generating
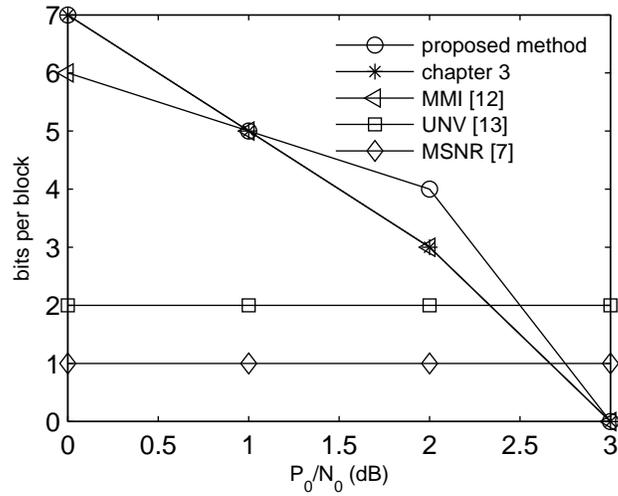
Figure 4.2: Bit allocation for the channel in (4.61) when $P_0/N_0 = 20$ dB.

the plot in Fig. 4.3. Similarly, we allocate bits for the other five system as in example 1 such that the number of total bits is eight. The bit error rates are evaluated for 10^6 channel realizations. In Fig. 4.3, we can see that the proposed method has the smallest bit error rate.
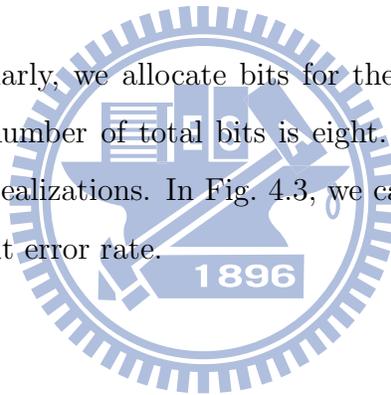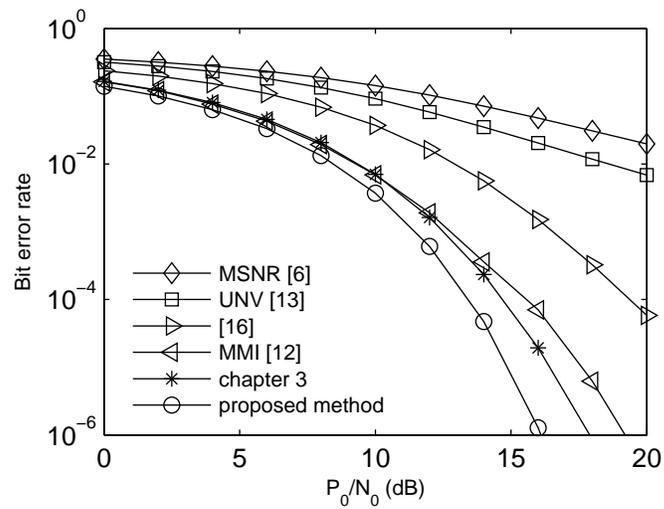
Figure 4.3: Bit error rate performance.

4.5 Summary

In this paper, we considered the problem of maximizing the bit rate over MIMO channels. The transceiver and bit allocation are jointly optimized without a high bit rate assumption. The optimal transceiver is obtained in a simple closed form. In the simulation, we have demonstrated that more bits can be transmitted when compared with earlier systems that use the same constellation size for all subchannels or systems that have a high bit rate assumption on bit allocation.

Chapter 5

On the Duality of Transceiver Designs for MIMO Channels

In chapter 3 and chapter 4, we optimized the transceiver to maximize the transmission rate. Another commonly used optimally criteria for MIMO transceiver design is transmission power. In the literature, bit rate maximization and power minimization problems are viewed as different problems. The solution of these two problems have been obtained independently when the bit allocation is not constrained to be integers. These two problems with integer bit allocation are still open. In this chapter, we will consider the connection between these two problems. We will first consider the case without integer constraint. We will show that if a transceiver is optimal for the power-minimizing problem, it is also optimal for the rate maximizing problem, and vice versa. For the case when the bit allocation is constrained to be non-negative integer, we will show the duality continue to hold with a modification in the rate maximization problem. Using the duality, we will develop an algorithm to find the optimal solutions of rate maximization problem when an integer bit constraint is imposed.

5.1 Power-minimizing and Rate-maximizing Transceiver design

Power-minimizing and rate-maximizing problems with real bit allocation have been considered in [14]-[22] and [23]-[24]. The aim of this section to establish the connection between these two and show that they are dual problems. For a given symbol error rate constraint ϵ and target bit rate B_0, the power-minimizing problem $\mathscr{A}_{pow}$ with real bit allocation can be formulated as [14]-[22]

$$
(\mathscr{A}_{pow}) \quad
\begin{array}{c}
\underset{\mathbf{F},\{\sigma_{s_k}^2\},\{b_k\}}{\text{minimize}} \\
\text{subject to}
\end{array}
\quad
\begin{array}{l}
P = \sum_{k=0}^{M-1}[\mathbf{F}^\dagger\mathbf{F}]_{kk}\sigma_{s_k}^2 \\
\left\{
\begin{array}{ll}
B = \sum_{k=0}^{M-1} b_k \geq B_0, \\
\epsilon_k \leq \epsilon, & k = 0, 1, \cdots, M-1,
\end{array}
\right.
\end{array}
\tag{5.1}
$$

where ϵ_k is the symbol error rate of the k-th subchannel. Given a symbol error rate ϵ and power constraint P_0, the rate-maximizing problem $\mathscr{A}_{rate}$ with real bit allocation is [23]-[24]

$$
(\mathscr{A}_{rate}) \quad
\begin{array}{c}
\underset{\mathbf{F},\{\sigma_{s_k}^2\},\{b_k\}}{\text{maximize}} \\
\text{subject to}
\end{array}
\quad
\begin{array}{l}
B = \sum_{k=0}^{M-1} b_k \\
\left\{
\begin{array}{ll}
P = \sum_{k=0}^{M-1}[\mathbf{F}^\dagger\mathbf{F}]_{kk}\sigma_{s_k}^2 \leq P_0, \\
\epsilon_k \leq \epsilon, & k = 0, 1, \cdots, M-1.
\end{array}
\right.
\end{array}
\tag{5.2}
$$

In either problem, we need to design the triplet $(\mathbf{F}, \{\sigma_{s_k}^2\}, \{b_k\})$, i.e., designing the transmit matrix $\mathbf{F}$, power allocation $\{\sigma_{s_k}^2\}$, and bit allocation $\{b_k\}$ jointly to maximizing bit rate or minimizing power. The following lemmas will be useful for subsequent discussion.

Lemma 5.1 *Given a channel matrix, consider a system with a fixed power allocation $\{\sigma_{s_k}^2\}$ and error rate constraint ϵ. Suppose the transmit power of the system is equal to $\alpha\mathbf{F}$, where $\mathbf{F}$ is some $N \times M$ matrix such that $\mathbf{HF} \neq \mathbf{0}$ and α is a positive real number. Then the transmit power and the achievable bit rate of the system are continuous and strictly increasing functions of α.*

Proof: See Appendix A.

Using Lemma 5.1 we know that for the same error rate, the power and bit rate are both continuous and increasing functions of α. This implies that if we increase the transmit power by choosing $\alpha > 1$, the bit rate will always be increased. Next we will show that if we decrease $\sigma_{s_k}^2$ for some subchannel k and keep the other symbol variances $\{\sigma_{s_l}^2\}_{l \neq k}$ unchanged, the error variance of all subchannels will be decreased.

Lemma 5.2 *Consider the MIMO transceiver in Fig. 2.1 with a given transmitter* **F***, channel matrix* **H***, and MMSE receive* **G***. Suppose $\sigma_{s_k}^2 > 0$ for some k. If we keep $\sigma_{s_0}^2, \cdots, \sigma_{s_{k-1}}^2, \sigma_{s_{k+1}}^2, \cdots, \sigma_{s_{M-1}}^2$ and vary only $\sigma_{s_k}^2$, then the error variances $\sigma_{e_l}^2$ are functions of $\sigma_{s_k}^2$ and*

$$\frac{\partial \sigma_{e_l}^2}{\partial \sigma_{s_k}^2} = \sigma_{s_k}^{-4} |[\mathbf{E}]_{lk}|^2, \ for \ l = 0, \cdots, M - 1. \tag{5.3}$$

Each error variance $\sigma_{e_l}^2$ is an increasing and continuous function of $\sigma_{s_k}^2$. More-over, $\sigma_{e_k}^2$ is a strictly increasing and concave function of $\sigma_{s_k}^2$.

Proof: See Appendix B.

In the following lemma, we will show that inequalities in the power-minimizing problem (5.1) and the rate-maximizing problem (5.2) become equalities when optimal designs are used.

Lemma 5.3 *If $(\mathbf{F}^*, \{\sigma_{s_k}^{*2}\}, \{b_k^*\})$ is optimal for the power-minimizing problem $\mathscr{A}_{pow}$ in (5.1), the transmission bit rate B is equal to the target bit rate B_0 and all the error rate ϵ_k are equal to ϵ. Similarly, for the rate-maximizing problem $\mathscr{A}_{rate}$ in (5.2), the transmit power P of the optimal solution is equal to P_0 and all the error rate ϵ_k are equal to ϵ.*

Proof: See Appendix C.

Using Lemma 5.1 and Lemma 5.3, we can show that, for the problem $\mathscr{A}_{rate}$ the maximal bit rate is a strictly increasing function of the power constraint. That is, $B^*(P_1) < B^*(P_2)$ whenever $P_1 < P_2$, where $B^*(x)$ denotes the maximal

bit rate for $\mathscr{A}_{rate}$ when the power constraint is x. To see this, let $P_1 < P_2$. It follows that $B^*(P_1) \le B^*(P_2)$. So we only need to show that $B^*(P_1) \neq B^*(P_2)$. Suppose $B^*(P_1) = B^*(P_2)$ for $P_1 < P_2$. By Lemma 5.3, the transmit power of the optimal solution that achieves $B^*(P_1)$ is equal to P_1. Using Lemma 5.1, we can always find a new system that achieves bit rate $\tilde{B} > B^*(P_1)$ using power $\tilde{P} = P_2$, which contracts the definition of $B^*(P_2)$. This completes the proof.

Remarks:

1. Lemma 5.3 shows that all the inequalities in constraints of $\mathscr{A}_{pow}$ and $\mathscr{A}_{rate}$ become equalities when the solutions are optimal. This means that when the optimal transmitter $\mathbf{F}^*$ and power allocation $\{\sigma_{s_k}^{*2}\}$ are given for $\mathscr{A}_{pow}$ or $\mathscr{A}_{rate}$, the bit allocation can be obtained directly using $\epsilon_k = \epsilon$ in (2.19). Therefore, we only need to design $\mathbf{F}$ and $\{\sigma_{s_k}^2\}$ directly but not bit allocation in these two problems.

2. When the error rate is constrained to be equal to ϵ for all subchannels, it has been shown that equality in the power and bit rate constraints will hold [22][24][27] using majorization theorem [42] and optimization theorem [77]. In Lemma 5.3 we consider error rate inequality constraint in addition to power and bit rate inequality constraints.

Using Lemma 5.1 and Lemma 5.3, we establish the duality between $\mathscr{A}_{pow}$ and $\mathscr{A}_{rate}$ in Theorem 5.1 and Theorem 5.2.

Theorem 5.1 *Given a target transmission rate B_0 and symbol error rate constraint ϵ, suppose the transmitter $\mathbf{F}^*$ and power allocation $\{\sigma_{s_k}^{*2}\}$ form an optimal solution for $\mathscr{A}_{pow}$, and the minimized power is P^*. Now, given transmit power constraint $P_0 = P^*$ and symbol error rate constraint ϵ, the same $\mathbf{F}^*$ and $\{\sigma_{s_k}^{*2}\}$ also maximize the bit rate for the problem in $\mathscr{A}_{rate}$. Furthermore, the maximized*

rate in this case is equal to B_0.

Proof: As $\mathbf{F}^*$ and $\{\sigma_{s_k}^{*2}\}$ are optimal for $\mathscr{A}_{pow}$, the minimized transmit power is

$$P^* = \sum_{k=0}^{M-1}[\mathbf{F}^{*\dagger}\mathbf{F}^*]_{kk}\sigma_{s_k}^{*2}. \tag{5.4}$$

By Lemma 5.3, the total bit rate is equal to the target rate B_0 and all the symbol error rates ϵ_k^* are equal to ϵ. Now, let us consider the problem in $\mathscr{A}_{rate}$ with power constraint $P_0 = P^*$ and error rate constraint ϵ. Suppose $\tilde{\mathbf{F}}$ and $\{\tilde{\sigma}_{s_k}^2\}$ are optimal for $\mathscr{A}_{rate}$. By Lemma 5.3, the transmit power used in this case is equal to P^* and symbol error rates are equal to ϵ. Since we already know $\mathbf{F}^*$ and $\{\sigma_{s_k}^{*2}\}$ can achieve bit rate B_0 with transmit power P^*, the maximal bit rate $\tilde{B}$ achieved in $\mathscr{A}_{rate}$ must be larger than or equal to B_0, i.e.,

$$\tilde{B} \geq B_0. \tag{5.5}$$

If $\tilde{B} = B_0$, we get the desired result that $\mathbf{F}^*$ and $\{\sigma_{s_k}^{*2}\}$ are also optimal for $\mathscr{A}_{rate}$. Suppose $\tilde{B} > B_0$, i.e., more than B_0 bits can be transmitted when P^* is given. Consider a new transceiver with transmitter $\mathbf{F}' = \alpha\tilde{\mathbf{F}}$, where $0 < \alpha < 1$, and power allocation $\{\tilde{\sigma}_{s_k}^2\}$ is unchanged. By Lemma 5.1 we know the bit rate of such a system is a strictly increasing function of α and is continuous on α. So we can always find $\alpha < 1$ such that $B' = B_0$. Since $\alpha < 1$, the required power is smaller than P^*. This is a contradiction to the assumption that P^* is the minimal transmit power when B_0 is given in the power-minimizing problem. Therefore, the maximal bit rate is B_0. Since $(\mathbf{F}^*, \{\sigma_{s_k}^{*2}\})$ can achieve bit rate rate B_0 with power P^*, it is an optimal solution for $\mathscr{A}_{rate}$. $\triangle\triangle\triangle$

Theorem 5.2 *Given a transmit power constraint P_0 and symbol error rate constraint ϵ, suppose the transmitter $\mathbf{F}^*$ and power allocation $\{\sigma_{s_k}^{*2}\}$ form an optimal solution for the rate-maximizing problem $\mathscr{A}_{rate}$, and the maximized rate is B^*. Then the same $\mathbf{F}^*$ and $\{\sigma_{s_k}^{*2}\}$ also minimize the transmit power for the problem*

$\mathscr{A}_{pow}$ when the target bit rate B_0 is equal to B^* and symbol error rate constraint is ϵ. Furthermore, the minimized power in this case is equal to P_0.

Proof: As $\mathbf{F}^*$ and $\{\sigma_{s_k}^{*2}\}$ are optimal for the problem $\mathscr{A}_{rate}$, by Lemma 5.3, the transmit power used in this case is equal to the constraint P_0 and the error rate is $\epsilon_k^* = \epsilon$ for $k = 0, \cdots, M-1$. Consider the problem $\mathscr{A}_{pow}$ with target bit rate $B_0 = B^*$ and error rate constraint ϵ. Suppose $(\tilde{\mathbf{F}}, \{\tilde{\sigma}_{s_k}^2\})$ is an optimal solution for $\mathscr{A}_{pow}$ and the minimized power is $\tilde{P}$. By Lemma 5.3, the transmitted bit rate is equal to the target B^* and all the error rates are ϵ. Also the minimal power $\tilde{P}$ in $\mathscr{A}_{pow}$ must be smaller than or equal to P_0 since we already know $\mathbf{F}^*$ and $\{\sigma_{s_k}^{*2}\}$ can achieve bit rate B^* with transmit power P_0, i.e.,

$$\tilde{P} \le P_0. \tag{5.6}$$

If the minimized transmit power $\tilde{P}$ is equal to P_0, we get the desired result that $\mathbf{F}^*$ and $\{\sigma_{s_k}^{*2}\}$ are also optimal for $\mathscr{A}_{pow}$. Suppose $\tilde{P} < P_0$, i.e., transmit power smaller than P_0 can be achieved when target rate B_0 is B^*. Consider a new system with transmitter $\mathbf{F}' = \alpha \tilde{\mathbf{F}}$ and power allocation $\{\tilde{\sigma}_{s_k}^2\}$, where $\alpha = \sqrt{P_0/\tilde{P}} > 1$. Then the transmit power of the new system is $P' = P_0$. Using Lemma 5.1 we know the bit rate of the new system will be larger than B^* for the same error rate constraint ϵ. This is a contradiction to the assumption that B^* is the maximal bit rate for $\mathscr{A}_{rate}$ when P_0 is given. Therefore, the minimal power $\tilde{P}$ is equal to P_0 and $(\mathbf{F}^*, \{\sigma_{s_k}^{*2}\})$ is an optimal solution for $\mathscr{A}_{pow}$. $\triangle\triangle\triangle$

Theorems 5.1 and 5.2 together show that if a transceiver is optimal in the power-minimizing problem, it is also optimal in the rate-maximizing problem, and vice versa. In the above discussion, the bits b_k assigned to the subchannel are not constrained to be integers. Such a duality also exists for the case when bit allocation is constrained to be integer. However, there are some subtle differences

46

as we will see in the next section.

5.2 Transceiver design with integer bit allocation

In this section, we consider the power-minimizing problem and rate-maximizing problem with integer bit allocation. With the constraint of integer bit allocation, the power-minimizing problem becomes

$$
(\mathscr{A}_{pow,int}) \quad
\begin{aligned}
&\underset{\mathbf{F},\{\sigma_{s_k}^2\},\{b_k\}}{\text{minimize}} \quad P = \sum_{k=0}^{M-1}[\mathbf{F}^\dagger\mathbf{F}]_{kk}\sigma_{s_k}^2 \\
&\text{subject to} \quad
\begin{cases}
\sum_{k=0}^{M-1} b_k \geq B_0, & \\
\epsilon_k \leq \epsilon, & k = 0, 1, \cdots, M-1, \\
b_k \in Z^+, & k = 0, 1, \cdots, M-1,
\end{cases}
\end{aligned}
\tag{5.7}
$$

where Z^+ denotes the set of nonnegative integers. The rate-maximizing problem with integer bit allocation is formulated as

$$
(\mathscr{A}_{rate,int}) \quad
\begin{aligned}
&\underset{\mathbf{F},\{\sigma_{s_k}^2\},\{b_k\}}{\text{maximize}} \quad B = \sum_{k=0}^{M-1} b_k \\
&\text{subject to} \quad
\begin{cases}
\sum_{k=0}^{M-1}[\mathbf{F}^\dagger\mathbf{F}]_{kk}\sigma_{s_k}^2 \leq P_0, & \\
\epsilon_k \leq \epsilon, & k = 0, 1, \cdots, M-1, \\
b_k \in Z^+, & k = 0, 1, \cdots, M-1.
\end{cases}
\end{aligned}
\tag{5.8}
$$

The following lemma shows that for the power-minimizing problem with integer bit constraint, the inequalities in the bit rate constraint and error rate constraint become equalities when the solution is optimal. This is similar to the power minimization problem without integer constraint. Such a property does not hold for the rate maximization problem with integer constraint as we will see later.

Lemma 5.4 *For the power-minimizing problem $\mathscr{A}_{pow,int}$ in (5.7), the bit rate of the optimal solution is equal to B_0 and the symbol error rates $\epsilon_k = \epsilon$ for all k.*
Proof: See Appendix D.

Lemma 5.4 leads to the following result that if a solution is optimal for $\mathscr{A}_{pow,int}$, it is also optimal for $\mathscr{A}_{rate,int}$.

47

Theorem 5.3 *Consider the power-minimizing problem $\mathscr{A}_{pow,int}$ with a target transmission rate B_0 and symbol error rate constraint ϵ. Suppose $(\mathbf{F}^*, \{\sigma_{s_k}^{*2}\}, \{b_k^*\})$ is optimal for $\mathscr{A}_{pow,int}$, and in this case the minimized power is P^*. Now for the problem $\mathscr{A}_{rate,int}$ with transmit power constraint $P_0 = P^*$ and error rate constraint ϵ, the same $(\mathbf{F}^*, \{\sigma_{s_k}^{*2}\}, \{b_k^*\})$ also maximizes the transmission rate and the maximized rate is equal to B_0.*

Proof: As $(\mathbf{F}^*, \{\sigma_{s_k}^{*2}\}, \{b_k^*\})$ is optimal for the problem $\mathscr{A}_{pow,int}$. By Lemma 5.4, the bit rate is $B^* = \sum_{k=0}^{M-1} b_k^* = B_0$, and all the symbol error rates satisfy $\epsilon_k^* = \epsilon$. Now, let us consider the problem $\mathscr{A}_{rate,int}$ with power constraint $P_0 = P^*$ and error rate constraint ϵ. Suppose $(\tilde{\mathbf{F}}, \{\tilde{\sigma}_{s_k}^2\}, \{\tilde{b}_k\})$ is optimal for the problem $\mathscr{A}_{rate,int}$ and the maximal bit rate is

$$\tilde{B} = \sum_{k=0}^{M-1} \tilde{b}_k. \tag{5.9}$$

All the corresponding error rates $\tilde{\epsilon}_k$ satisfy $\tilde{\epsilon}_k \leq \epsilon$ and the transmit power $\tilde{P}$ satisfies the power constraint, i.e., $\tilde{P} \leq P^*$. Since we already know the solution of $\mathscr{A}_{pow,int}$ can achieve bit rate B_0 with power P^*, the maximal bit rate $\tilde{B}$ in $\mathscr{A}_{rate,int}$ must be larger than or equal to B_0, i.e., $\tilde{B} \geq B_0$. We will prove the theorem by showing (i) the transmit power $\tilde{P}$ is equal exactly to P^*, and (ii) the maximized rate $\tilde{B}$ is in fact equal to B_0.

(i) $\tilde{P} = P^*$: Suppose $\tilde{P} < P^*$. This means $\tilde{\mathbf{F}}$, $\{\tilde{b}_k\}$, and $\tilde{\sigma}_{s_k}^2$ can achieve a smaller transmit power and still satisfy all the constraints in $\mathscr{A}_{pow,int}$. This contradicts the assumption that $\mathbf{F}^*$, $\{b_k^*\}$, and $\sigma_{s_k}^{*2}$ are optimal for $\mathscr{A}_{pow,int}$. So we have $\tilde{P} = P^*$.

(ii) $\tilde{B} = B_0$: If $\tilde{B} = B_0$, we get the desired result that $(\mathbf{F}^*, \sigma_{s_k}^{*2}, \{b_k^*\})$ is optimal for $\mathscr{A}_{rate,int}$. Suppose $\tilde{B} > B_0$. Similarly to the procedure in Lemma 5.4, we can find another system that achieves bit rate $B' = \tilde{B} - 1 \geq B_0$, with transmit power $P' < P^*$, and error rate $\epsilon_k' \leq \epsilon$. This contradicts the assumption that $(\mathbf{F}^*$,

$\{\sigma_{s_k}^{*2}\}$, $\{b_k^*\}$) is optimal for $\mathscr{A}_{pow,int}$. Therefore, we conclude that the maximized bit rate for the problem $\mathscr{A}_{rate,int}$ is B_0 and the power used is P^*. Therefore, the solution $(\mathbf{F}^*, \{\sigma_{s_k}^{*2}\}, \{b_k^*\})$ of $\mathscr{A}_{pow,int}$ is also an optimal solution for the problem $\mathscr{A}_{rate,int}$. $\triangle\triangle\triangle$

In Section 5.1, we saw that the transmit power of the optimal solution for the rate-maximizing problem is equal to the power constraint P_0 when the bit loading is not constrained to be integer. Such a property may not hold when there is integer bit constraint as we will see later. When the symbol error rate constraint ϵ is fixed, the maximal rate for $\mathscr{A}_{rate,int}$ is a function of the power constraint P_0. Similarly, for a fixed ϵ, the minimal power of $\mathscr{A}_{pow,int}$ is a function of target rate B_0. For convenience, we use $P_{int}^*(x)$ to denote the minimal transmit power for $\mathscr{A}_{pow,int}$ when the target bit rate x is given and $B_{int}^*(x)$ to denote as the maximal bit rate for $\mathscr{A}_{rate,int}$ when the power constraint is x.

The function $B_{int}^*(x)$ and $P_{int}^*(x)$. Using theorem 3, we will see that $B_{int}^*(x)$ is not continuous. It is a staircase-like function as shown in Fig. 5.1(a). This means a nonzero increase in the power constraint does not necessarily implies a nonzero increase in the maximized bit rate. This is different from the case without integer constraint in section 5.1. To explain this, consider the problem $\mathscr{A}_{pow,int}$ with two target bit rates B_1 and B_1+1. Let $P_1 = P_{int}^*(B_1)$ and $P_2 = P_{int}^*(B_1+1)$. We can plot the minimal transmit power as a function of target bit rate as in Fig. 5.1(b). By Theorem 5.3, we know $B_{int}^*(P_1) = B_1$ and $B_{int}^*(P_2) = B_1+1$. Now suppose the power constraint P_0 for $\mathscr{A}_{rate,int}$ is such that $P_1 < P_0 < P_2$. Then the maximal bit rate $B_{int}^*(P_0)$ for $\mathscr{A}_{rate,int}$ is equal to B_1 as we will see next. Since we already know that the maximal bit rate is B_1 when the power constraint is P_1, we have $B_{int}^*(P_0) \geq B_1$. Suppose $B_{int}^*(P_0) > B_1$. This contradicts the fact that P_2 is the minimal power for $\mathscr{A}_{pow,int}$ when the target bit rate is B_1+1. Hence we have $B_{int}^*(P_0) = B_1$. This implies that for any power constraint P that satisfies $P_1 \leq P_0 < P_2$, the maximal bit rate is $B_{int}^*(P_0) = B_1$. When the power constraint

$P_0 = P_2$, the maximal bit rate is increased to $B_1 + 1$. Therefore, $B_{int}^*(x)$ is the staircase like function in Fig. 5.1(a).

From the plot of $B_{int}^*(P_0)$ in Fig. 5.1(a) we can see that for $\mathscr{A}_{rate,int}$ there can be many solutions that achieve the same maximal bit rate, but with transmit power smaller than P_0. Hence for the problem $\mathscr{A}_{rate,int}$, the results in Lemma 5.3 is not true any more and the results of the real bit allocation case do not carry over to the the integer bit allocation case. To establish the duality with $\mathscr{A}_{pow,int}$, we will consider the solution that achieve the maximal rate B with the smallest transmit power among all possible solutions.
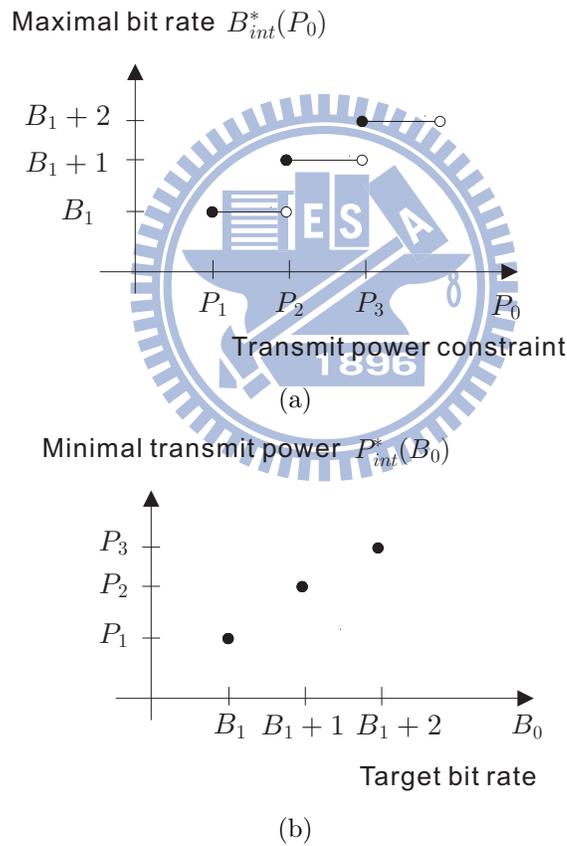
Figure 5.1: (a) Maximal bit rate as a function of power constraint for $\mathscr{A}_{rate,int}$. (b) Minimal transmit power as a function of target bit rate for $\mathscr{A}_{pow,int}$.

Theorem 5.4 *Consider the problem $\mathscr{A}_{rate,int}$ with power constraint P_0 and symbol error rate constraint ϵ. Suppose $(\mathbf{F}^*, \{\sigma_{s_k}^{*2}\}, \{b_k^*\})$ forms the solution that has the smallest transmit power P^* among all possible solutions. Let the maximized rate be B^*. Given target rate $B_0 = B^*$ and error rate constraint ϵ for the problem $\mathscr{A}_{pow,int}$, the same solution also minimizes the transmit power and the minimal power is P^*.*

Proof: As $(\mathbf{F}^*, \{b_k^*\}, \{\sigma_{s_k}^{*2}\})$ is optimal for $\mathscr{A}_{rate,int}$, the maximized rate is $B^* = \sum_{k=0}^{M-1} b_k^*$. The transmit power is $P^* \leq P_0$, and all the error rates satisfy $\epsilon_k^* \leq \epsilon$. Consider the power minimizing problem $\mathscr{A}_{pow,int}$ with target bit rate $B_0 = B^*$ and the same error rate constraint ϵ. Suppose $(\tilde{\mathbf{F}}, \{\tilde{b}_k\}, \{\tilde{\sigma}_{s_k}^2\})$ is optimal for $\mathscr{A}_{pow,int}$ and the minimized power is $\tilde{P}$. By Lemma 5.4, the bit rate $\sum_{k=0}^{M-1} \tilde{b}_k$ is equal to the target bit rate B^*. Since we already know $(\mathbf{F}^*, \{b_k^*\}, \{\sigma_{s_k}^{*2}\})$ can achieve bit rate B^* with transmit power P^*, the minimal power $\tilde{P}$ must be smaller than or equal to P^*, i.e., $\tilde{P} \leq P^*$. If $\tilde{P}$ is equal to P^*, we get the desired result that $(\mathbf{F}^*, \{b_k^*\}, \{\sigma_{s_k}^{*2}\})$ is an optimal solution for $\mathscr{A}_{pow,int}$. Assume $\tilde{P}$ is smaller, i.e., $\tilde{P} < P^*$. This means $(\tilde{\mathbf{F}}, \{\tilde{b}_k\}, \tilde{\sigma}_{s_k}^2)$ can achieve bit rate B^* with a smaller power $\tilde{P}$. It contradicts the assumption that $(\mathbf{F}^*, \{b_k^*\}, \sigma_{s_k}^{*2})$ is the optimal solution for the problem $\mathscr{A}_{rate,int}$ that has the smallest transmit power. Hence we have $\tilde{P} = P^*$ and the solution $(\mathbf{F}^*, \{b_k^*\}, \sigma_{s_k}^{*2})$ is optimal for $\mathscr{A}_{pow,int}$. $\triangle\triangle\triangle$

Theorem 5.3 shows that the optimal solution obtained in the power-minimizing problem is also an optimal solution in the rate-maximizing problem. Theorem 5.4 shows that the solution with the smallest transmit power in the rate-maximizing problem is also optimal in the power-minimizing problem.

Remark on ZF receiver: The derivations in Section 5.1 and Section 5.2 are considered for the MMSE receiver. Duality between the power minimization and rate maximization problems also hold for the ZF case. For the MMSE case, we have used the results in Lemmas 5.1, 5.3, and 5.4 to prove the main results

in Theorems 5.1-5.4. Lemma 5.2 is used in the proof of Lemmas 5.3 and 5.4. For the ZF case, Lemma 5.2 is not needed as the MSE matrix of the ZF receiver in (2.14) is independent of the power allocation. Using the methods of MMSE case, we can prove the results in Lemmas 5.1, 5.3, 5.4, and also Theorems 5.1-5.4 for the ZF case.

5.3 Optimal solution for transceiver design with bit allocation

Many optimal transceiver designs have been proposed to solve the power minimization problem $\mathscr{A}_{pow}$ [14]-[22] and bit rate maximization problem $\mathscr{A}_{rate}$ [23]-[27]. For the power minimization problem with integer bit allocation, the solution has been found in [24]. There is no solution yet for the rate maximization problem with integer bit allocation. In Section 5.3.1, we will review the solution of $\mathscr{A}_{pow}$ and $\mathscr{A}_{rate}$ (no integer constraint on bit allocation). In Section 5.3.2, we will review the solution of $\mathscr{A}_{pow,int}$ and show how to find the solutions of $\mathscr{A}_{rate,int}$ using the solution of $\mathscr{A}_{pow,int}$.

5.3.1 Optimal solution of $\mathscr{A}_{pow}$ and $\mathscr{A}_{rate}$

Let the singular value decomposition of the $P \times N$ channel matrix $\mathbf{H}$ be

$$\mathbf{H} = \mathbf{U} \begin{bmatrix} \boldsymbol{\Lambda} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{V}^{\dagger}, \tag{5.10}$$

where $\boldsymbol{\Lambda}$ is diagonal that contains the nonzero singular values of $\mathbf{H}$. The elements of $\boldsymbol{\Lambda}$ are in nonincreasing order. The $P \times P$ matrix $\mathbf{U}$ and the $N \times N$ matrix $\mathbf{V}$ are unitary. For the power-minimizing problem $\mathscr{A}_{pow}$ with target bit rate B_0 and error rate constraint ϵ, the solution is given by [22, 26]

$$\mathbf{F} = \mathbf{V}_1 \mathbf{D}^{1/2}, \tag{5.11}$$

where $\mathbf{V}_1$ contains the first M columns of $\mathbf{V}$ and $\mathbf{D}$ is a diagonal matrix with diagonal element $[\mathbf{D}]_{kk} = (\alpha - [\boldsymbol{\Lambda}]_{kk}^{-2} N_0 \Gamma)^+$, where $(x)^+ = \max(x, 0)$. The con-

stant α is chosen such that $\sum_{k=0}^{M-1} b_k = B_0$. The power allocation $\sigma_{s_k}^2$ is equal to one. The solution of rate-maximizing problem $\mathscr{A}_{rate}$ with power constraint P_0 and error rate constraint ϵ is given by [24, 26]

$$\mathbf{F} = \mathbf{V}_1 \mathbf{D}^{1/2}, \tag{5.12}$$

where $\mathbf{D}$ is a diagonal matrix with diagonal element $[\mathbf{D}]_{kk} = (\beta - [\mathbf{\Lambda}]_{kk}^{-2} N_0 \Gamma)^+$. The constant β is chosen such that $\sum_{k=0}^{M-1}[\mathbf{D}]_{kk} = P_0$. The power allocation $\sigma_{s_k}^2 = 1$.

Note that the overall transfer function $\mathbf{T} = \mathbf{GHF}$ of the optimal solutions of $\mathscr{A}_{pow}$ and $\mathscr{A}_{rate}$ is a diagonal matrix that can be singular. Let $\mathbf{T}_r$ be the $M_r \times M_r$ nonsingular diagonal matrix obtained by removing the rows and columns of $\mathbf{T}$ that correspond to the zero diagonal elements. Consider the reduced system $(\mathbf{F}_r, \mathbf{\Lambda}_r)$ with the ZF receiver given by $\mathbf{G}_{r,zf} = \mathbf{T}_r^{-1} \mathbf{G}_r$, where $\mathbf{G}_r$ is the MMSE receiver for the system $(\mathbf{F}_r, \mathbf{\Lambda}_r)$. It can be verified that the unbiased signal to noise ratio of the ZF receiver is the same as that of the optimal MMSE solution. Thus the bit rate of the ZF receiver is the same as the optimal MMSE receiver. Therefore, for $\mathscr{A}_{pow}$ and $\mathscr{A}_{rate}$ the optimal solution of the MMSE transceiver is the same as the ZF transceiver.

5.3.2   Optimal solution of $\mathscr{A}_{pow,int}$ and $\mathscr{A}_{rate,int}$

First, we will review the optimal solution for $\mathscr{A}_{pow,int}$.

Optimal solution for $\mathscr{A}_{pow,int}$[24]

The optimal power-minimizing transceiver can be found using [15] if the optimal integer bit allocation is given to us. However, we do not know optimal bit allocation beforehand. Nonetheless, for a given target bit rate B_0, there are only a finite number of possible integer bit allocation. In particular, $\{b_k\}$ is such that $b_k \in Z^+$ and

$$b_0 + b_1 + \cdots + b_{M-1} = B_0. \tag{5.13}$$

For each integer bit allocation $\{b_k\}$ that satisfies the condition in (5.13), we remove the subchannels that correspond to $b_k = 0$. Then we can use the result in [15] to find the transceiver that minimizes the transmit power. The optimal solution of $\mathscr{A}_{pow,int}$ can be obtained by choosing the integer bit allocation and transceiver that have the minimal transmit power among all the possible integer bit allocations. From [24], we know that the optimal solution of $\mathscr{A}_{pow,int}$ for the MMSE receiver is different from that for the ZF receiver. When B_0 and M become large, the number of possible bit allocation L becomes large and the computational cost for solving $\mathscr{A}_{pow,int}$ will be large.

Optimal solution for $\mathscr{A}_{rate,int}$

For the rate-maximizing problem $\mathscr{A}_{rate,int}$ with power constraint P_0, if the maximal rate $B^*_{int}(P_0)$ is known, we can solve it using the solution of $\mathscr{A}_{pow,int}$ based on Theorem 5.3. We can find $B^*_{int}(P_0)$ using an iterative search. For example, starting from $B_0 = 1$ we compute $P^*_{int}(B_0)$. If $P^*_{int}(B_0) \leq P_0$, we increase B_0 by one and compute $P^*_{int}(B_0)$ again until $P^*_{int}(B_0) > P_0$. Then $B^*_{int}(P_0) = B_0 - 1$. To reduce the number of iterations we note that $B^*_{int}(P_0) \leq B^*(P_0)$, where $B^*(P_0)$ is the maximal bit rate of the rate maximization problem $\mathscr{A}_{rate}$ without integer bit constraint. As a result, $B^*_{int}(P_0) \leq \lfloor B^*(P_0) \rfloor$, where the notation $\lfloor x \rfloor$ denotes the largest integer that is less than or equal to x. Using this property and Theorem 5.3 we have the following algorithm.

Algorithm for finding the solution of $\mathscr{A}_{rate,int}$:

1. Initially, given the power constraint P_0, compute the maximal bit rate $B^*(P_0)$ for $\mathscr{A}_{rate}$. Then set $B_0 = \lfloor B^*(P_0) \rfloor$.

2. Given the target bit rate B_0, find the optimal bit allocation and transceiver for minimizing transmit power in $\mathscr{A}_{pow,int}$. Compute the minimal power $P^*_{int}(B_0)$.

3. If $P^*_{int}(B_0) > P_0$, set $B_0 = B_0 - 1$ and go to step 2. If $P^*_{int}(B_0) \leq P_0$, then

the maximal bit rate $B_{int}^*(P_0) = B_0$.

In this algorithm, the number of iterations is equal to $\lfloor B^*(P_0) \rfloor - B_{int}^*(P_0)$. This number is in fact less than M as we explain below. Suppose $\lfloor B^*(P_0) \rfloor - B_{int}^*(P_0) \geq M$. Let $\{b_k^*\}$ be the optimal real-valued bit allocation of $\mathscr{A}_{rate}$, i.e., $B^*(P_0) = \sum_{k=0}^{M-1} b_k^*$. Then $\{\lfloor b_k^* \rfloor\}$ is also a valid integer bit allocation that satisfies the error rate constraint. Since $\{b_k^*\}$ is real, we have $\lfloor B^*(P_0) \rfloor - \sum_{k=0}^{M-1} \lfloor b_k^* \rfloor \leq B^*(P_0) - \sum_{k=0}^{M-1} \lfloor b_k^* \rfloor < M$. This implies $\sum_{k=0}^{M-1} \lfloor b_k^* \rfloor > B_{int}^*(P_0)$, which contradicts the definition of $B_{int}^*(P_0)$. Therefore we have $\lfloor B^*(P_0) \rfloor - B_{int}^*(P_0) < M$. Note the number M is an upper bound of the number of iterations. As the optimal solution for $\mathscr{A}_{rate,int}$ is obtained using the solution of $\mathscr{A}_{pow,int}$, the optimal solution of the MMSE receiver is different from that of the ZF receiver.

5.4 Simulation

In the simulations, we will demonstrate the duality between power-minimizing problem and rate-maximizing problem. In the following examples, the number of subchannels M is 4. The noise vector $\mathbf{q}$ is assumed to be complex white Gaussian with $\mathrm{E}[\mathbf{q}\mathbf{q}^\dagger] = \mathbf{I}_4$. The symbol error rate constraint ϵ is assumed to be 10^{-4}. In examples 1-2, we use a fixed 4×4 MIMO channel as shown in example 1. In examples 3-4, the results are averaged over random channels. For the problems $\mathscr{A}_{pow}$ and $\mathscr{A}_{rate}$, we use the solutions in Section 5.3.1. For $\mathscr{A}_{pow,int}$ and $\mathscr{A}_{rate,int}$, we use the solutions in Section 5.3.2.

Example 1. Duality between $\mathscr{A}_{pow}$ and $\mathscr{A}_{rate}$. In this example, we will demonstrate the results in Theorem 5.1 and Theorem 5.2. Consider a 4×4 channel $\mathbf{H}$ that is given by

$$\begin{pmatrix} -0.5 + 0.6i & -0.5 - 1.1i & 0.2 - 0.2i & 0.4 - 0.5i \\ -0.3 + 0.6i & -0.2 + 1.4i & -0.4 + 0.9i & 0.8 - 0.5i \\ -0.1 + 0.5i & -0.4 - 0.3i & 0.9 + 0.3i & 0.1 + 0.2i \\ 1.1 + 0.6i & -0.5 + 0.4i & 0.0 - 0.2i & -0.3 + 1.4i \end{pmatrix}. \tag{5.14}$$

	(a)
B_0 (bits)	$P^*(B_0)$ (dB)
2	3.2833
4	8.0001
6	11.1790
8	13.7891
10	16.1370

	(b)
$P_0 = P^*(B_0)$ (dB)	$B^*(P_0)$ (bits)
3.2833	2
8.0001	4
11.1790	6
13.7891	8
16.1370	10

Table 5.1: (a) Minimal power $P(B_0)$ for $\mathscr{A}_{pow}$ when $B_0 = 2, 4, 6, 8, 10$ bits. (b) Maximal bit rate for $\mathscr{A}_{rate}$ when the power constraint $P_0 = P(B_0)$.

Given target bit rate B_0, we use (5.11) to find the optimal transceiver, and (2.1) to compute the corresponding transmit power $P^*(B_0)$ for the problem $\mathscr{A}_{pow}$. Table 5.1(a) shows the minimal transmit power $P^*(B_0)$ when the target bit rates are $B_0 = 2, 4, 6, 8, 10$ bits. Using the minimized power in Table 5.1(a) as power constraint, Table 5.1(b) shows the maximal bit rate for the rate maximizing problem $\mathscr{A}_{rate}$. The rates are computed using (2.20) for the optimal transceiver in (5.12). We can see that $B^*(P^*(B_0)) = B_0$ and the solution of the power-minimizing problem is also optimal for the rate-maximizing problem as we have shown in Theorem 5.1.

Table 5.2(a) shows the maximal bit rate $B^*(P_0)$ for $\mathscr{A}_{rate}$ when the power constraints are $P_0 = 2, 4, 8, 16, 32$ dB. Table 5.2(b) shows the minimal power $P^*(B_0)$ for the problem $\mathscr{A}_{pow}$ when the target bit rates are equal to the maximized rate in Table 5.2(a). We can see that $P^*(B^*(P_0)) = P_0$ and the solution of rate-maximizing problem is also a solution of the power-minimizing problem as we have shown in Theorem 5.2.

Example 2. **ZF and MMSE receivers for** $\mathscr{A}_{pow,int}$. When there is no integer bit constraint, the optimal solution with ZF receiver and the optimal solution with MMSE receiver are the same. When there is integer constraint, the solutions are in general different as we demonstrated in this example. We use the same channel as in example 1. Let us compute the optimal integer bit

56

(a)	
P_0 (dB)	$B^*(P_0)$ (bits)
2	1.6108
4	2.2447
8	3.9999
16	9.8792
32	28.7068

(b)	
$B_0 = B^*(P_0)$ (bits)	$P^*(B_0)$ (dB)
1.6108	2
2.2447	4
3.9999	8
9.8792	16
28.7068	32

Table 5.2: (a) Maximal bit rate for $\mathscr{A}_{rate}$ when $P_0 = 2, 4, 8, 16, 32$ dB. (b) Minimal power $P(B_0)$ for $\mathscr{A}_{pow}$ when $B_0 = B(P_0)$.

allocation and transceiver of $\mathscr{A}_{pow,int}$ when MMSE reception is considered. We use the method mentioned in section 5.3.2. The target bit rate B_0 is set to be 8 bits. In this case the optimal integer bit allocation is

$$\mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ 2 \\ 0 \end{bmatrix},$$

and the minimal transmit power is 13.891 dB. The optimal transmitter is given by

$$\mathbf{F} = \begin{bmatrix} -1.35 - 0.36i & 0.83 - 1.19i & -0.68 + 0.82i \\ -1.63 + 1.61i & 0.22 + 0.51i & 0.28 - 0.45i \\ -0.39 + 0.52i & 1.15 + 0.28i & -1.93 - 0.15i \\ 0.05 & -1.95 & -2.03 \end{bmatrix}, \tag{5.15}$$

and the overall transfer function $\mathbf{T}$ is

$$\mathbf{T} = \mathbf{GHF} = \begin{bmatrix} 0.98 & 0.0021 & 0 \\ 0.0021 & 0.97 & 0 \\ 0 & 0 & 0.94 \end{bmatrix}. \tag{5.16}$$

We can see that the overall transfer function for the MMSE receiver is not diagonal. For the ZF case. For the ZF case, the minimized power is 13.8915 dB, which is very close to the MMSE case.

Example 3. Duality between $\mathscr{A}_{pow}$ and $\mathscr{A}_{rate}$. In this example, we use random channels to demonstrate the connections between power minimization and rate maximization problems. The channel is of size 4×4 and the elements

are complex Gaussian random variables whose real and imaginary parts are independent with zero mean and variance $1/2$. Monte Carlo simulation using 10^6 channel realizations is used to generate the following results. For each channel, we compute the optimal solutions of $\mathscr{A}_{pow}$ and $\mathscr{A}_{rate}$ using (5.11) and (5.12) in Section 5.3.1. Fig. 5.2 shows the maximal transmission rates $B^*(P_0)$ of $\mathscr{A}_{rate}$ as a function of power constraint. Fig. 5.3 shows the minimal transmit power $P^*(B_0)$ of $\mathscr{A}_{pow}$ as a function of target bit rate. We can observe the duality between that the power-minimizing and rate-maximizing problems from Fig. 5.2 and Fig. 5.3. For example, the minimal power of $\mathscr{A}_{pow}$ is 9 dB when the target bit rate is 5 bits. When we set the power constraint in $\mathscr{A}_{rate}$ to be 9 dB, the maximal bit rate is 5 bits. On the other hand, the maximal bit rate of $\mathscr{A}_{rate}$ is 9 bits when the power constraint is 15 dB. When we set the target bit rate in $\mathscr{A}_{pow}$ to be 9 bits, the minimal power is 15 dB.
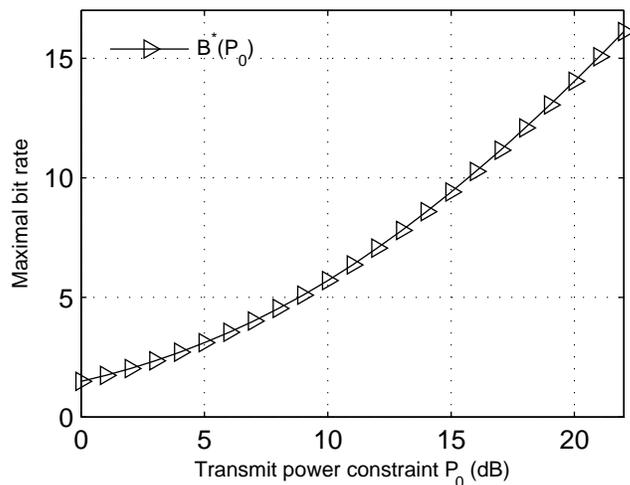
Figure 5.2: Maximal bit rate $B^*(P_0)$ for $\mathscr{A}_{rate}$ as a function of power constraint P_0 without integer constraint.

Example 4. **Minimal power for $\mathscr{A}_{pow,int}$ and maximal bit rate for $\mathscr{A}_{rate,int}$**. We use the same random channel as in example 3. In Table 5.3, we compute the minimal transmit power of $\mathscr{A}_{pow,int}$ with MMSE and ZF receivers.
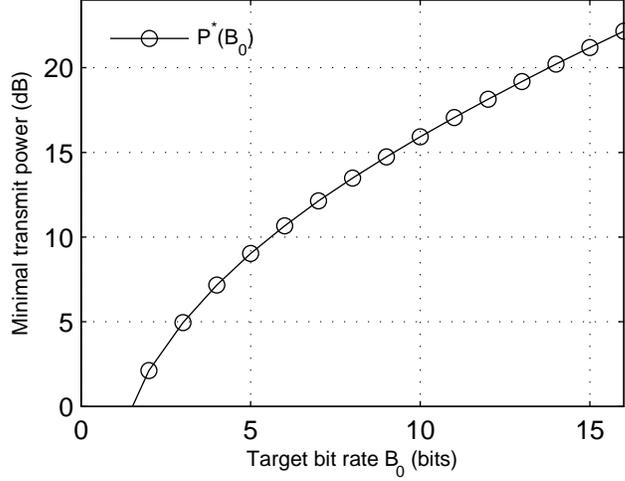
Figure 5.3: Minimal transmit power $P^*(B_0)$ for $\mathscr{A}_{pow}$ as a function of target bit rate B_0 without integer constraint.

When the target bit rate is B_0, the minimal transmit powers of the MMSE case and the ZF case are denoted by $P^*_{int,mmse}(B_0)$ and $P^*_{int,zf}(B_0)$, respectively. For comparison, we also show the transmit power $P^*(B_0)$ of $\mathscr{A}_{pow}$ (without integer constraint). We can see that the gap between $P^*_{int,mmse}(B_0)$ and $P^*_{int,zf}(B_0)$ is small. Also, the difference between $P^*_{int,mmse}(B_0)$ and $P^*(B_0)$ is smaller than 0.21 dB. In Table 5.4, we compute the maximal bit rate of $\mathscr{A}_{rate,int}$ for the MMSE and ZF receivers. The maximal bit rate for the MMSE and ZF cases are denoted respectively by $B^*_{int,mmse}(P_0)$ and $B^*_{int,zf}(P_0)$. Also shown in Table 5.4 is the maximal bit rate $B^*(P_0)$ of $\mathscr{A}_{rate}$ (without integer constraint). We can see that $B^*_{int,zf}(P_0)$ is close to $B^*_{int,mmse}(P_0)$. The difference between $B^*_{int,mmse}(P_0)$ and $B^*(P_0)$ is smaller than 0.6 bits. This gap is less than 0.15 bits per symbols.

5.5 Summary

In this chapter, we consider two commonly used transceiver design criteria: power minimization criterion and rate maximization criterion. The duality are derived

B_0 (bits)	$P^*(B_0)$ (dB)	$P^*_{int,zf}(B_0)$ (dB)	$P^*_{int,mmse}(B_0)$ (dB)
2	2.1167	2.3322	2.3254
4	7.1634	7.3107	7.3078
6	10.6614	10.7832	10.7826
8	13.4708	13.5796	13.5788
10	15.9114	16.0175	16.0170
12	18.1311	18.2314	18.2312
14	20.1976	20.2972	20.2970

Table 5.3: Transmit power of $\mathscr{A}_{pow}$ (without integer bit allocation), $\mathscr{A}_{pow,int}$ (ZF), and $\mathscr{A}_{pow,int}$ (MMSE) when the target bit rate is $B_0 = 2, 4, 6, 8, 10, 12$ bits.

P_0 (dB)	$B^*(P_0)$ (dB)	$B^*_{int,zf}(P_0)$ (bits)	$B^*_{int,mmse}(P_0)$ (bits)
2	2.0305	1.4549	1.4572
4	2.7096	2.1549	2.1557
6	3.5429	2.9858	2.9865
8	4.5391	3.9689	3.9700
10	5.7103	5.1326	5.1333
12	7.0629	6.4788	6.4794
14	8.5888	8.0033	8.0037
16	10.2714	9.6811	9.6815

Table 5.4: Bit rate of $\mathscr{A}_{rate}$ (without integer bit allocation), $\mathscr{A}_{rate,int}$ (ZF), and $\mathscr{A}_{rate,int}$ (MMSE) when the target bit rate is $P_0 = 2, 4, 6, 8, 10, 12, 14, 16$ dB.

for these two problems. If there is no integer bit constraint, the optimal solution in either one solution is also optimal in the other problem. When there is an integer bit constraint, we have shown the rate-maximizing problem is equivalent to the power-minimizing problem with power modifications. Using the duality, the optimal solution of the rate maximization problem with integer bit constraint can be found using the solution of the power minimization problem. Simulation results have been shown to demonstrate the duality of these two problems.

Chapter 6

Overview of Multicarrier Systems

Multicarrier systems have found many applications in DMT systems and OFDM systems. For the multicarrier systems, the frequency band of the channel is divided into a number of subchannels and information is transmitted on each of the subchannel. In this chapter, we will introduce the multicarrier systems and find the filterbank representation of the multicarrier systems.

6.1 DFT Based Multicarrier System
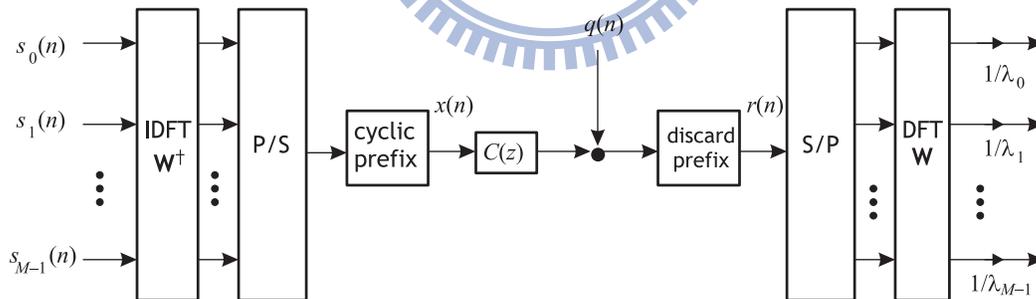
Figure 6.1: Block diagram of the DFT based multicarrier system.

The block diagram of the DFT based multicarrier system is as shown in Fig. 6.1. The input of the transmitter $\mathbf{s}$ is an $M \times 1$ vector of modulation symbols. The symbol vector $\mathbf{s}$ are assumed to be zero-mean and uncorrelated.

The autocorrelation matrix of the input vector s is assumed to be

$$\mathbf{R}_s = \varepsilon_s \mathbf{I}_M. \tag{6.1}$$

The channel is modeled as an FIR filter of order L, i.e.,

$$C(z) = \sum_{n=0}^{L} c(n) z^{-n}. \tag{6.2}$$

The channel noise $q(n)$ is assumed to be a circularly symmetric complex Gaussian random process with zero mean and variance N_0. The channel noise $q(n)$ is assumed to be uncorrelated with the symbols $s_k(n)$. At the transmitter, IDFT is applied to the input symbol vector $\mathbf{s}$ and the output vector $\mathbf{x}$ is

$$\mathbf{x} = \mathbf{W}^\dagger \mathbf{s}, \tag{6.3}$$

where $\mathbf{W}$ denotes the $M \times M$ normalized DFT matrix given by

$$[\mathbf{W}]_{mn} = \frac{1}{\sqrt{M}} e^{-j\frac{2\pi mn}{M}}, \text{ for } 0 \leq m, n \leq M - 1.$$

The outputs are converted to a block of M serial samples by the parallel to serial operation (P/S). Then a cyclic prefix of length ν is inserted by copying the last ν samples of the block to the beginning. The length of the cyclic prefix ν is chosen so that $\nu \geq L$, which ensures that inter-block-interference (IBI) can be removed easily by discarding the prefix at the receiver.

At the receiver, after prefix removal the samples are blocked into M by 1 vectors $\mathbf{r}$ by the serial to parallel operation (S/P). When there is no channel noise, it can be shown that the transfer matrix from $\mathbf{x}$ to $\mathbf{r}$ is the $M \times M$ circulant matrix given by

$$\mathbf{C}_{circ} = \begin{bmatrix} c(0) & 0 & \cdots & \cdots & c(L) & \cdots & c(1) \\ c(1) & c(0) & & & & \ddots & \vdots \\ \vdots & & \ddots & & & & c(L) \\ c(L) & & & \ddots & & & 0 \\ 0 & \ddots & & & c(0) & & \vdots \\ \vdots & \ddots & \ddots & & & \ddots & 0 \\ 0 & \cdots & 0 & c(L) & \cdots & c(1) & c(0) \end{bmatrix}. \tag{6.4}$$

In the presence of channel noise, the received vector $\mathbf{r}$ is

$$\mathbf{r} = \mathbf{C}_{circ}\mathbf{x} + \mathbf{q}, \tag{6.5}$$

where $\mathbf{q}$ is the blocked channel noise vector of size M. Then the DFT matrix is applied, i.e.,

$$\mathbf{y} = \mathbf{W}\mathbf{r} \tag{6.6}$$

$$= \mathbf{W}\mathbf{C}_{circ}\mathbf{W}^{\dagger}\mathbf{s} + \mathbf{W}\mathbf{q}. \tag{6.7}$$

From [41], we know that circulant matrices $\mathbf{C}_{circ}$ can be diagonalized using DFT and IDFT matrices,

$$\mathbf{C}_{circ} = \mathbf{W}^{\dagger}\boldsymbol{\Lambda}\mathbf{W}, \tag{6.8}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix. The diagonal element λ_k of $\boldsymbol{\Lambda}$ corresponding to the M-point DFT of the channel impulse response, i.e.,

$$\lambda_k = [\boldsymbol{\Lambda}]_{kk} = C(z)|_{z=e^{-j2\pi k/M}}. \tag{6.9}$$

The DFT output vector becomes

$$\mathbf{y} = \boldsymbol{\Lambda}\mathbf{s} + \mathbf{W}\mathbf{q}. \tag{6.10}$$

Then the scalar multipliers $1/\lambda_k$, which are called frequency domain equalizers (FEQ), are applied to $\mathbf{y}$. The transceiver is ISI free and the receiver is a zero-forcing receiver. The receiver outputs are identical to the inputs of the transmitter in the absence of channel noise. From (6.10), the signal to noise ratio (SNR) of the k-th subchannel is given by

$$\beta_k = \frac{|\lambda_k|^2 \varepsilon_s}{N_0}. \tag{6.11}$$

Transmission rate: For the QAM modulation, suppose the target symbol error rate ϵ_k are given. Then the number of bits loaded on the k-th subchannel can be computed by (2.19), i.e.,

$$b_k = \log_2\left(1 + \frac{\beta_k}{\Gamma_k}\right). \tag{6.12}$$

63

Suppose the sampling time of the system is T_s. Then the transmission rate is given by

$$R = \frac{\sum_{k=0}^{M-1} b_k}{NT_s}.$$ (6.13)

Complexity:

The main computations of the transceiver are those of the IDFT and DFT matrices, for which fast algorithms can be applied. The complexity of the transmitter is simply that of an IDFT matrix and the complexity of the receiver is that of a DFT matrix plus M multiplications for FEQs. Moreover, except for the FEQs, the computations are channel independent.

6.2   Filterbank Representation

In this section, we will look at the multicarrier system from the viewpoint of filterbanks, which will be useful for later discussion. In Fig. 6.1, the operation 'P/S' followed by the insertion of cyclic prefix can be viewed as the interconnection of the matrix

$$\begin{bmatrix} \mathbf{0} & \mathbf{I}_\nu \\ \mathbf{I}_M \end{bmatrix}$$

followed by 'P/S' for every $N = M + \nu$ parallel samples as shown in Fig. 6.2(a). The 'P/S' operation is represented using decimators and a delay chain in the figure. On the other hand, the operation 'discard prefix' followed by 'serial to parallel' and M-point DFT for every M samples in Fig. 6.1 can be viewed as 'serial to parallel' for every N samples followed by the matrix

$$\begin{bmatrix} \mathbf{0} & \mathbf{W} \end{bmatrix}.$$

as shown in Fig. 6.2(a). Thus the transmitter and receiver can be redrawn as in Fig. 6.2(a), where we have combined the two matrices at the transmitter as one matrix $\mathbf{G}$.

As $\mathbf{G}$ is a constant matrix, we can exchange $\mathbf{G}$ and the expanders; the resulting transmitter is as shown in Fig. 6.2(b). Similarly, we can exchange $\begin{bmatrix} \mathbf{0} & \mathbf{W} \end{bmatrix}$.
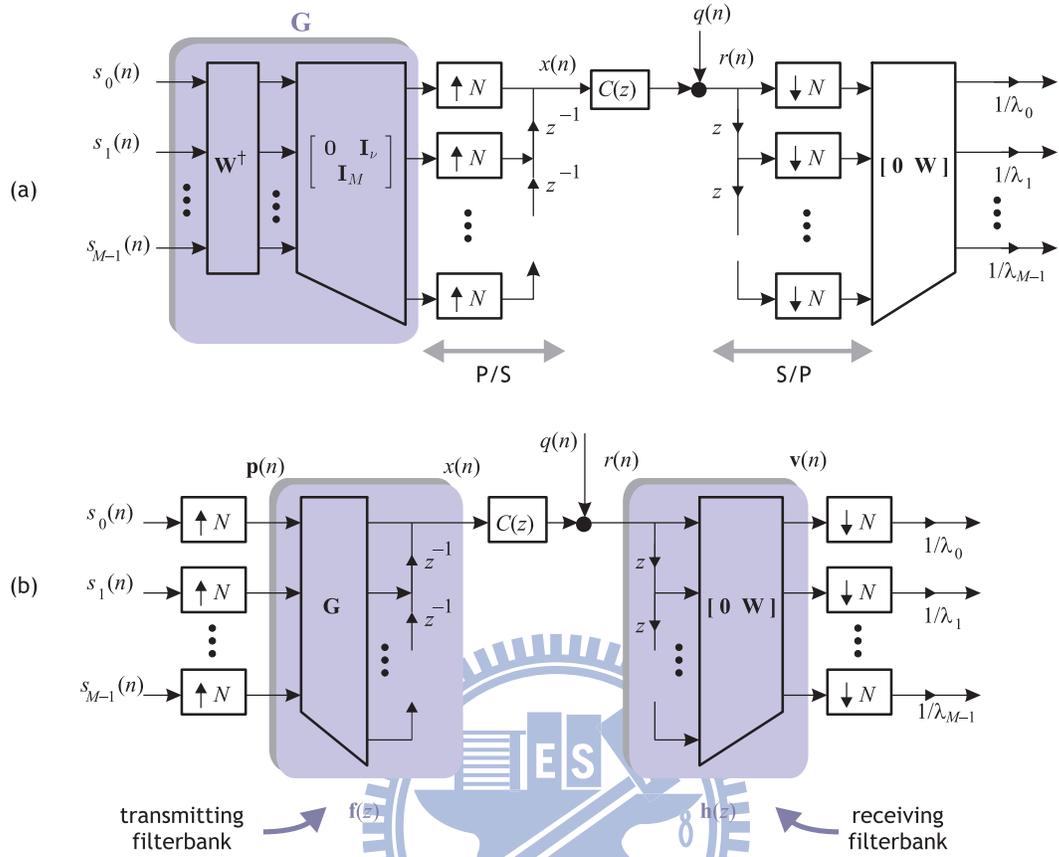
Figure 6.2: Matrix forms of the transmitter and receiver for the DMT system.

and the decimators to yield the receiver shown in Fig. 6.2(b). Note that the $1 \times M$ system from $\mathbf{p}(n)$ to $x(n)$ is LTI. Let's call the $1 \times M$ transmitting bank $\mathbf{f}(z)$, then $\mathbf{f}(z)$ is a row vector given by $[\; 1 \quad z^{-1} \quad \cdots \quad z^{-(N-1)} \;]\mathbf{G}$. Each element of the row vector can be obtained by multiplying out the above expression. Suppose the k-th element is $F_k(z)$ (k-th transmitting filter), we have

$$F_k(z) = \frac{1}{\sqrt{M}} \sum_{i=0}^{N-1} W^{-(i-\nu)k} z^{-i}, \tag{6.14}$$

where $W = e^{-j2\pi/M}$. Then the transmitter in Fig. 6.2(b) can be redrawn as in Fig. 6.3. Now consider the receiver side. Denote the $M \times 1$ system from $r(n)$ to
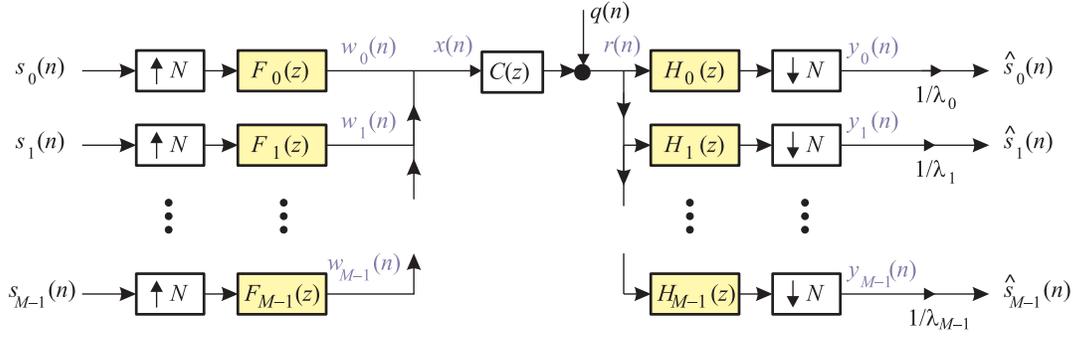
Figure 6.3: Filterbank representation of the DMT system.

$\mathbf{v}(n)$ in Fig. 6.2(b) as $\mathbf{h}(z)$. We can write $\mathbf{h}(z)$ as

$$\mathbf{h}(z) = \begin{bmatrix} \mathbf{0} & \mathbf{W} \end{bmatrix} \begin{bmatrix} 1 \\ z \\ \vdots \\ z^{N-1} \end{bmatrix}. \tag{6.15}$$

Suppose the k-th element is $H_k(z)$ (the k-th receiving filter), we have

$$H_k(z) = \frac{z^\nu}{\sqrt{M}} \sum_{i=0}^{M-1} W^{ik} z^i. \tag{6.16}$$

We can redraw the receiver as the receiving bank structure in Fig. 6.3. Note the first transmitting filter $F_0(z)$ is a rectangular window of length N. All the other transmitting filters are scaled and frequency-shifted versions of the first transmitting filter (prototype filter),

$$F_k(z) = W^{\nu k} F_0(zW^k). \tag{6.17}$$

Similarly, the first receiving filter is also a rectangular window, but of length M. All the other receiving filters are scaled and frequency-shifted versions of the first receiving filter,

$$H_k(z) = W^{-\nu k} H_0(zW^k). \tag{6.18}$$

This means both the transmitting and receiving filters form the DFT bank structure [76].

66
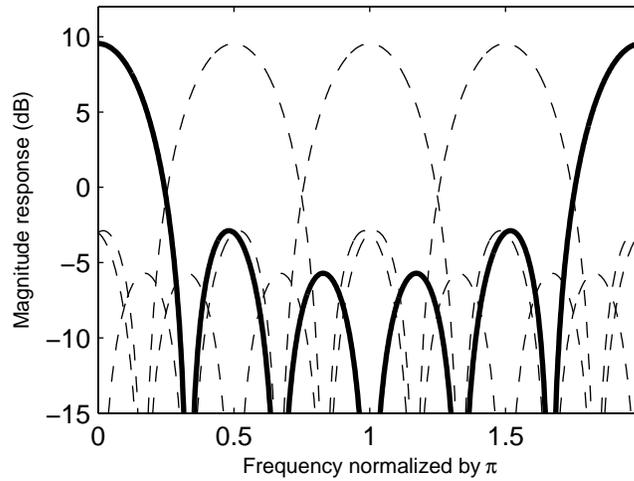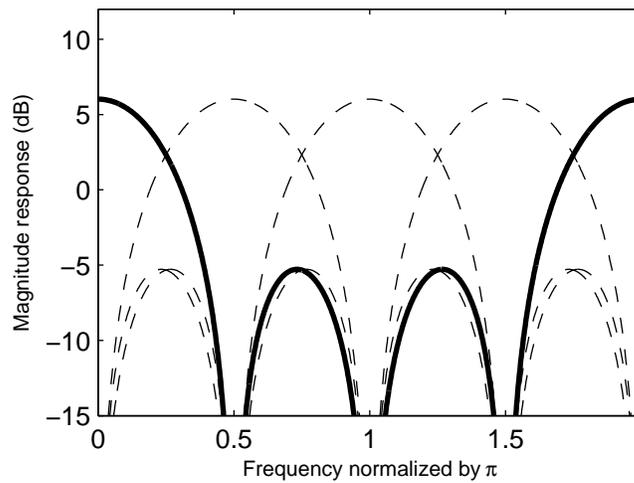
Figure 6.4: The magnitude response of the transmitting and receiving filters for $M = 4$ and $\nu = 2$.(a) the transmitting filters, and (b) the receiving filters.

A numerical example of the transmitting and receiving filters for $M = 4$ and $\nu = 2$ is shown in Fig. 6.4. The magnitude response of the two prototype filters $F_0(z)$ and $H_0(z)$ are drawn with a solid line. The magnitude response of all the other filters, which are shifted versions of the corresponding prototype filters, are drawn with a dotted line. We can see that the first sidelobe has an attenuation of about 13 dB only and the stopband decays slowly. The attenuation is not

adequate in many applications. In chapter 7 and chapter 8, we will design the transmitting and receiving windows to improve the frequency characteristics at the transmitter and receiver side.

6.3   Transmitted Power Spectrum

The filterbank representation allows us to express the power spectrum of the transmitted signal $x(n)$ in terms of the transmitting filters. For OFDM systems in wireless applications, the inputs $s_k(n)$ can be assumed to be uncorrelated and the transmitted power spectrum has been derived in [60]. The assumption of uncorrelated input symbols is not valid for DMT systems in wired applications. This is because the DMT system uses baseband transmission and the signal to be transmitted is real. This requires that the inputs of the IDFT matrix have the conjugate symmetric property, $s_k(n) = s_{M-k}^*(n)$, $k = 1, 2, \cdots, M-1$, and $s_0(n)$ is real. For even M, usually the case in practice, $s_{M/2}(n)$ is also real. This conjugate symmetric property means that the symbols assigned to the second half and the first half of the subchannels are related. Therefore for DMT systems, we can no longer assume that the inputs are uncorrelated.

For those inputs $s_k(n)$ that are in conjugate pairs, let the real part be $s_k^{(r)}(n)$ and the imaginary part be $s_k^{(i)}(n)$. We can treat these real parts and imaginary parts as random processes and assume, reasonably, that these random processes are white, uncorrelated, jointly wide-sense stationary with zero mean and variance $\mathcal{E}_{s,k}/2$. The scalar $1/2$ is included so that the variance of $s_k(n)$ is $\mathcal{E}_{s,k}$. For the k-th and $(M-k)$-th subchannels, the inputs are a complex conjugate pair. When the transmitting filters are shifted versions of the prototype filter as in (6.17) and the prototype has real coefficients, the coefficients of the transmitting filters are also in conjugate pairs, $f_{M-k}(n) = f_k^*(n)$. As a result, the outputs of each pair are also conjugates of each other. Now instead of considering the output of individual subchannel, let us consider the sum of outputs of each pair. Let the

output of the k-th transmitting filter be $w_k(n)$ as indicated in Fig. 6.3 and define $w'_k(n) = w_k(n) + w_{M-k}(n)$ Then $w'_k(n) = 2\text{Real}\{w_k(n)\}$ and it can be written as

$$w'_k(n) = 2 \sum_\ell \left(s_k^{(r)}(\ell) f_k^{(r)}(n - N\ell) - s_k^{(i)}(\ell) f_k^{(i)}(n - N\ell) \right),$$

where $f_k^{(r)}(n)$ and $f_k^{(i)}(n)$ are respectively the real and imaginary part of $f_k(n)$. As the real and imaginary parts of the transmitter inputs are uncorrelated, the power spectrum of $w'_k(n)$ is

$$S_{w'_k}(e^{j\omega}) = \frac{2\mathcal{E}_{s,k}}{N}(|F_k^{(r)}(e^{j\omega})|^2 + |F_k^{(i)}(e^{j\omega})|^2).$$

It turns out the summation of the two terms on the right hand side is equal to the $\frac{\mathcal{E}_{s,k}}{N}(|F_k(e^{j\omega})|^2 + |F_{M-k}(e^{j\omega})|^2)$. We can obtain the transmitted power spectrum by summing up contributions from $w'_k(n)$, plus $w_0(n)$ and $w_{M/2}(n)$ (if M is even). We arrive at the following simple expression for the transmitted spectrum

$$S_x(e^{j\omega}) = \frac{1}{N} \sum_{k=0}^{M-1} \mathcal{E}_{s,k}|F_k(e^{j\omega})|^2. \tag{6.19}$$

We can further observe that if an equal power allocation is used, the inputs of all the subchannels have the same variance $\mathcal{E}_s$ and the transmitted power spectrum becomes the same as that of the OFDM system derived in [60]. In some applications of the DMT system such as VDSL and ADSL, and the OFDM system such as wireless local area networks [81], only a subset of the subchannels are actually used for data transmission. Thus the transmitted spectrum becomes

$$S_x(e^{j\omega}) = \frac{1}{N} \sum_{k \in U} \mathcal{E}_{s,k}|F_k(e^{j\omega})|^2, \tag{6.20}$$

where U is the set of tones that are used for the current transmission. As $F_k(e^{j\omega})$ is not ideal filter, the spectrum is nonzero not only in the frequency bins of the subchannels that are used but also in other frequency bands as well. This is referred to as **spectral leakage**. The sidelobes of the transmitting filters directly affect the amount of spectral leakage.

6.4 Radio Frequency Interference

DFT based multicarrier system in Fig. 6.1 have been found applications in DMT systems, e.g., ADSL and VDSL [49][50]. In ADSL and VDSL environment, radio frequency signals such as amateur radio and AM radio may interfere the received signal at the receiver. This kind of noise is called the radio frequency interference (RFI). The radio interference is known to be of a narrowband nature but has a large amplitude in frequency. For the duration of one DMT symbol, it can be considered as a sum of sinusoids. We assume that RFI interference occurs at frequency ω_l with amplitude α_l and phase θ_l, $l = 0, \cdots, R-1$. Thus we can model the interference as

$$v(n) = \sum_{l=0}^{R-1} \alpha_l \cos(\omega_l n + \theta_l). \tag{6.21}$$

To analyze the effect of $v(n)$, we apply the interference-only signal $v(n)$ to the DMT receiver as shown in Fig. 6.5.
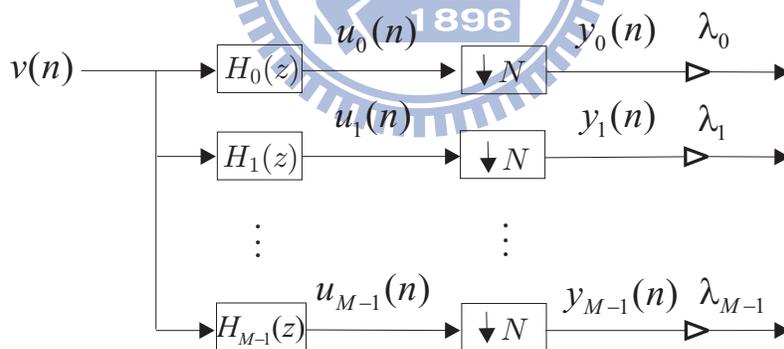
Figure 6.5: Filter bank representation of the receiver with windowing.

The output of the i-th receiving filter is

$$u_i(n) = \frac{1}{2} \sum_{l=0}^{R-1} \alpha_l [H_i(e^{j\omega_l})e^{j(\omega_l n + \theta_l)} + H_i(e^{-j\omega_l})e^{-j(\omega_l n + \theta_l)}]. \tag{6.22}$$

70

Using (6.18), we have

$$H_i(e^{j\omega_l}) = e^{j2\pi\nu i/M} H_0(e^{j(\omega_l - 2\pi i/M)}), \text{ and } H_i(e^{-j\omega_l}) = e^{j2\pi\nu i/M} H_0(e^{j(-\omega_l - 2\pi i/M)}).$$
(6.23)

The output $u_i(n)$ becomes

$$u_i(n) = \frac{1}{2} \sum_{l=0}^{R-1} \alpha_l e^{j2\pi\nu i/M} [H_0(e^{j(\omega_l - 2\pi i/M)}) e^{j(\omega_l n + \theta_l)} + H_0(e^{-j(\omega_l + 2\pi i/M)}) e^{-j(\omega_l n + \theta_l)}].$$
(6.24)

We can see that the amplitude of $u_i(n)$ is scaled by the prototype filter $H_0(e^{j\omega})$ at frequencies $\{\pm\omega_l - 2\pi i/M\}_{l=0}^{R-1}$. Suppose there is one RFI source occurs in the frequency band of the 0-th subchannel, i.e., $0 < \omega_l < 2\pi/M$ for some l. Then $u_1(n), u_2(n), \cdots, u_{M-1}(n)$ will be large if the sidelobes of $H_0(e^{j\omega})$ are large. Thus the spectral roll-off of the $H_0(e^{j\omega})$ determines how neighboring tones are affected by RFI. For the DMT system, since the prototype filter of the receiving filter is a rectangular window, the sidelobe is large and results in poor RFI suppression. In the next section, we will introduce the conventional windowing technique to improve the frequency characteristics.

6.5 Summary

In this section, we introduced the multicarrier systems. The filterbank representation of the multicarrier systems was also derived. Using the filterbank representation, we have shown the power spectrum of the the multicarrier systems and the spectral leakage. We have also introduced the RFI interference in the DSL applications.

Chapter 7

Receiver Window Designs for Radio Frequency Interference Suppression for Multicarrier Systems

In chapter 6, we have introduced the multicarrier system and its filterbank representation. We have also studied the RFI interference in the multicarrier system. The spectral roll-off of the receiving filters determine how neighboring tones are affected by RFI. To improve the frequency characteristics of the receiving filters, windowing technique is often used at the receiver. In this chapter, first we will introduce the conventional windowing technique. Then we will design the receiving windows to mitigate the RFI interference by minimizing the total interference. We will consider both the informed receiver (RFI information available to the receiver) and uninformed receiver (RFI information unavailable to the receiver). In either case, the proposed window is channel independent and can be obtained in a closed form.

7.1   Receiver Windowing in Multicarrier System

For the conventional DMT system, the sidelobes of the receiving filters are too large to have a good RFI suppression. As a result, windowing technique is often applied at the receiver to suppress the RFI suppression. To apply windows, the receiver takes the last $M + \beta$ samples, multiplies the first β samples by the coefficients w_n, $n = 0, 1, \cdots, \beta - 1$, and multiplies the last β samples by $1 - w_n$, where w_n are free parameters [71]. In other words, the $M + \beta$ samples are applied by a window of the following form.

$$\mathbf{g} = \begin{bmatrix} \mathbf{b} \\ \mathbf{1}_{M-\beta} \\ \mathbf{1}_\beta - \mathbf{b} \end{bmatrix}, \tag{7.1}$$

where $\mathbf{b} = [b_0 \cdots b_{\beta-1}]^{\mathrm{T}}$, and the notation $\mathbf{1}_n$ denotes an $n \times 1$ column vector whose elements are equal to one. In this chapter, we assume the channel order is smaller than or equal to $\nu - \beta$. This implies only the first $\nu - \beta$ samples contain the interference from the previous block. Since the first $\nu - \beta$ samples are discarded, there will be no IBI. This operation is shown in Fig. 7.1(a). After applying the window $\mathbf{g}$, the receiver folds the first β samples and adds to the last β samples, which is shown in Fig. 7.1(b). Combining the operation in Fig. 7.1(a) and Fig. 7.1(b), the implementation of the windowing operation is shown in Fig. 7.2.

The windowing operation in Fig. 7.2 can be represented by an $M \times N$ matrix $\mathbf{B}$. The matrix $\mathbf{B}$ is given by

$$\mathbf{B} = \begin{bmatrix} \mathbf{0} & \mathbf{I}_M \\ \mathbf{I}_\beta & \end{bmatrix} \mathrm{diag}(\mathbf{g}) \begin{bmatrix} \mathbf{0} & \mathbf{I}_{M+\beta} \end{bmatrix}, \tag{7.2}$$

where $\mathrm{diag}(\mathbf{g})$ is a diagonal matrix with the elements of $\mathbf{g}$ on its diagonal.
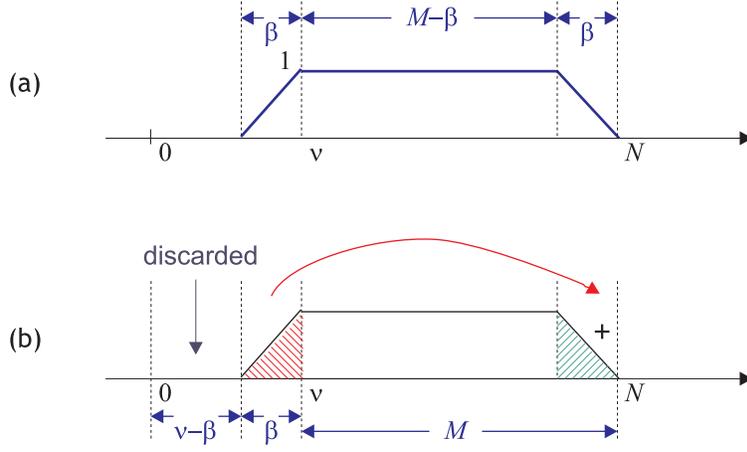
Figure 7.1: (a) An example of receiver window; (b) receiver windowing.
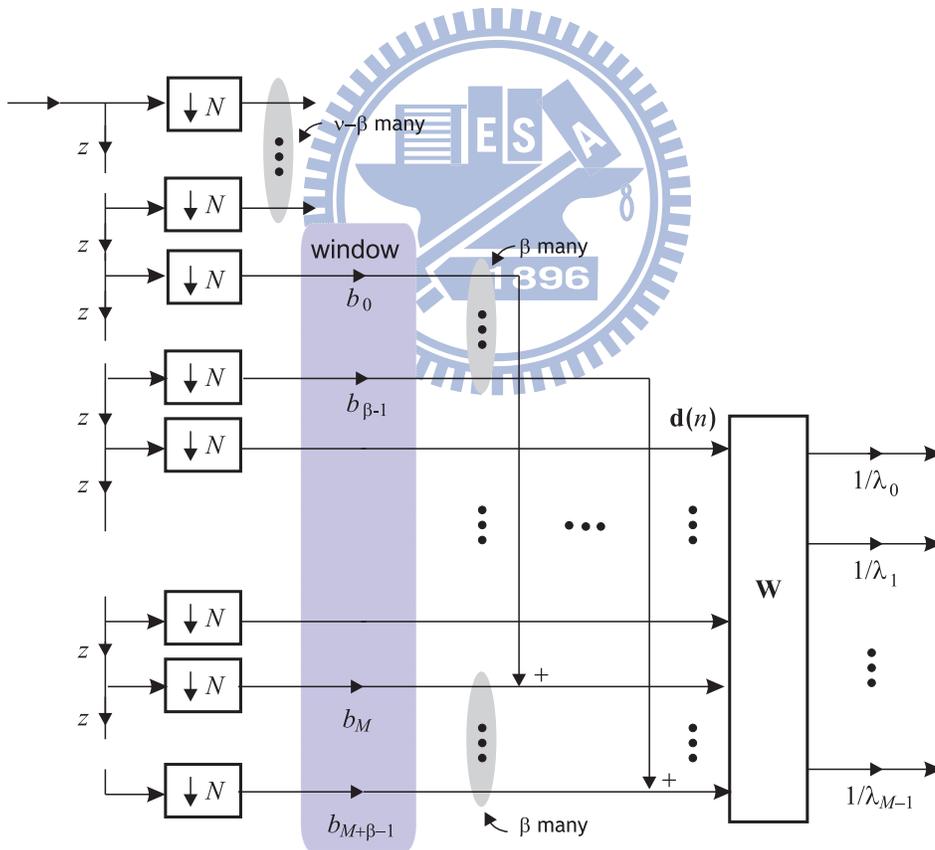
Figure 7.2: Receiver with windowing in the multicarrier system.
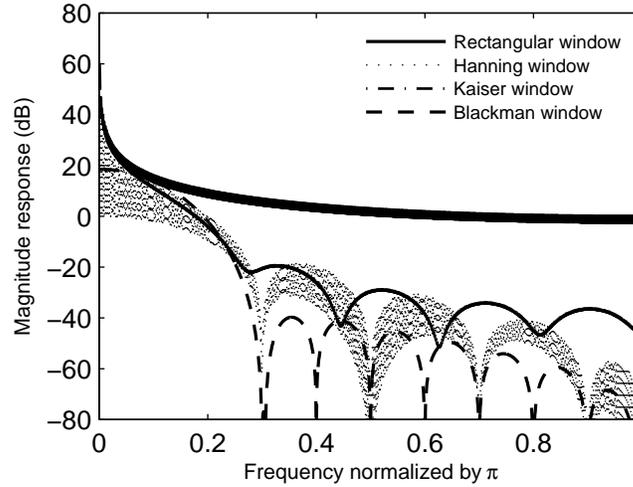
Figure 7.3: Frequency response of receiving windows.

Some commonly used window are Hanning window, Blackman window, and Kaiser window. These windows are computed using (7.1), where b_n is defined by the following equations [75]:

Hanning window.

$$b_n = 0.5 - 0.5\cos(\pi n/(\beta+1)),\ 0 \le n \le \beta. \tag{7.3}$$

Blackman window.

$$b_n = 0.54 - 0.46\cos(\pi n/(\beta+1)),\ 0 \le n \le \beta. \tag{7.4}$$

Kaiser window.

$$b_n = \frac{I_0[\gamma(1-[(n-\alpha)/\alpha]^2)^{1/2}]}{I_0(\gamma)},\ 0 \le n \le \beta, \tag{7.5}$$

where $\alpha = \beta+1$, and $I_0(\cdot)$ represents the zeroth-order modified Bessel function of the first kind. γ is the shape parameter. Fig. 7.3 shows the frequency response of the rectangular window, Hanning window, Blackman window, and Kaiser window with shape parameter $\beta = 5$. We can see that the rectangular window has larger sidelobes than the other three windows.

75

7.2 Filterbank Representation of Receiver Windowing in Multicarrier System

In this section, we derive the filterbank representation of the receiver with windowing. The representation will be useful in analyzing the interference of individual tones. Similar to (6.15), the M receiving filters $H_i'(z)$ for $i = 0, 1, \cdots, M-1$ are related to $\mathbf{B}$ and $\mathbf{W}$ by

$$
\begin{bmatrix} H_0'(z) \\ H_1'(z) \\ \vdots \\ H_{M-1}'(z) \end{bmatrix} = \mathbf{W}\mathbf{B} \begin{bmatrix} 1 \\ z \\ \vdots \\ z^{N-1} \end{bmatrix}.
\tag{7.6}
$$

The equivalent filter bank representation is shown in Fig. 6.5. Using the expression of $\mathbf{B}$ in (7.2), we can verify that the coefficients of the first receiving filter $h_0(n)$ are given by

$$
h_0'(n) = \frac{1}{\sqrt{M}} \begin{cases} b_{-n-\nu+\beta}, & -(\nu-1) \leq n \leq -(\nu-\beta), \\ 1, & -(N-\beta-1) \leq n \leq -\nu, \\ 1 - b_{-n-N+\beta}, & -(N-1) \leq n \leq -(N-\beta), \\ 0, & \text{otherwise.} \end{cases}
\tag{7.7}
$$

Comparing (7.7) with (7.1), we have

$$
\sqrt{M}\mathbf{g} = \begin{bmatrix} h_0'(-\nu+\beta) \\ \vdots \\ h_0'(-N+1) \end{bmatrix}.
\tag{7.8}
$$

Thus the magnitude response of $\sqrt{M}\mathbf{g}$ is equal to the magnitude response of $h_0'(n)$. We can further verify that all the receiving filters are shifted versions of the first receiving filter except for some scalars,

$$
H_i'(z) = W^{-i\nu} H_0'(zW^i).
\tag{7.9}
$$

We can see that the new receiving filters $\{H_i'(z)\}_{i=0}^{M-1}$ still have the DFT bank structure.

7.3 Informed Window

For the DMT system, the RFI interference is modeled as in (6.21). When we apply the interference-only signal $v(n)$ to the receiver, the output of the i-th receiving filter $H_i'(e^{j\omega})$ is

$$u_i(n) = \frac{1}{2} \sum_{l=0}^{R-1} \alpha_l [c_{l,i} e^{j(\omega_l n + \theta_l)} + c_{l,i}' e^{-j(\omega_l n + \theta_l)}], \tag{7.10}$$

where $c_{l,i} = H_i'(e^{j\omega_l})$ and $c_{l,i}' = H_i'(e^{-j\omega_l})$. The interference at the i-th receiver output is $y_i(n) = u_i(Nn)$, which has the same amplitude as $u_i(n)$. Note that the RFI interference due to the l-th source will be small if $\alpha_l^2(|c_{l,i}|^2 + |c_{l,i}'|^2)$ is small. Hence the RFI interference of the i-th individual tone will be small if

$$J_i = \sum_{l=0}^{R-1} \alpha_l^2(|c_{l,i}|^2 + |c_{l,i}'|^2) \tag{7.11}$$

is small. The total RFI interference can be mitigated by minimizing

$$
\begin{aligned}
J &= \sum_{i \in U} J_i \\
&= \sum_{i \in U} \sum_{l=0}^{R-1} \alpha_l^2 \quad (|H_0'(e^{j(\omega_l - 2\pi i/M)})|^2 + |H_0'(e^{-j(\omega_l + 2\pi i/M)})|^2)
\end{aligned} \tag{7.12}
$$

where we have used

$$c_{l,i} = W^{-iP} H_0'(e^{j(\omega_l - 2\pi i/M)}), \text{ and } \quad c_{l,i}' = W^{-iP} H_0'(e^{-j(\omega_l + 2\pi i/M)}). \tag{7.13}$$

U is the set of tones that are used for the current transmission. From (7.7) we can verify that $H_0'(e^{j(\omega_l - 2\pi i/M)})$ can be given in terms of $\mathbf{b}$ as

$$H_0'(e^{j(\omega_l - 2\pi i/M)}) = t_{l,i} + \mathbf{a}_{l,i}^\dagger \mathbf{b}, \tag{7.14}$$

where the notation '†' denotes Hermitian, $b_{l,i}$ is a scalar and $\mathbf{a}_{l,i}$ is an $\beta \times 1$ column vector given respectively by

$$
\begin{aligned}
t_{l,i} &= \sum_{k=\nu}^{\nu+M-1} e^{j(\omega_l - 2\pi i/M)k}, \\
[\mathbf{a}_{l,i}]_m &= e^{j(\omega_l - 2\pi i/M)(\nu - \beta + m)} - e^{j(\omega_l - 2\pi i/M)(N - \beta + m)}.
\end{aligned} \tag{7.15}
$$

Similarly, we can verify that $H_0'(e^{-j(\omega_l + 2\pi i/M)})$ can be expressed by

$$H_0'(e^{-j(\omega_l + 2\pi i/M)}) = t_{l,i}' + \mathbf{a'}_{l,i}^\dagger \mathbf{b}, \qquad (7.16)$$

where $b_{l,i}'$ and $\mathbf{a'}_{l,i}$ are respectively

$$
\begin{aligned}
t_{l,i}' &= \sum_{k=\nu}^{\nu+M-1} e^{-j(\omega_l + 2\pi i/M)k}, \\
[\mathbf{a'}_{l,i}]_m &= e^{-j(\omega_l + 2\pi i/M)(\nu-\beta+m)} - e^{-j(\omega_l + 2\pi i/M)(N-\beta+m)}.
\end{aligned}
\qquad (7.17)
$$

Using (7.14)-(7.17), the objective function can be written in terms of $\mathbf{b}$ as

$$J = \mathbf{b}^{\mathrm{T}} \mathbf{A} \mathbf{b} + \mathbf{b}^{\mathrm{T}} \mathbf{t} + \mathbf{t}^\dagger \mathbf{b} + c, \qquad (7.18)$$

where $\mathbf{A}$ is an $\beta \times \beta$ matrix, $\mathbf{t}$ is an $\beta \times 1$ vector, and c is a scalar given respectively by

$$
\begin{aligned}
\mathbf{A} &= \sum_{i \in U} \sum_{l=0}^{R-1} \alpha_l^2 [\mathbf{a}_{l,i} \mathbf{a}_{l,i}^\dagger + \mathbf{a'}_{l,i} \mathbf{a'}_{l,i}^\dagger], \\
\mathbf{t} &= \sum_{i \in U} \sum_{l=0}^{R-1} \alpha_l^2 [\mathbf{a}_{l,i} t_{l,i} + \mathbf{a'}_{l,i} t'_{l,i}], \\
c &= \sum_{i \in U} \sum_{l=0}^{R-1} \alpha_l^2 [|t_{l,i}|^2 + |t'_{l,i}|^2].
\end{aligned}
\qquad (7.19)
$$

To minimize the objective function in (7.18), we can use the method of optimization in [77] to obtain a closed-form solution. In particular, when the objective function J in (7.18) is minimal, the optimal $\mathbf{b}$ must satisfy $\partial J/\partial \mathbf{b} = 0$. The optimal solution can be written as follows

$$\mathbf{b} = -[\Re(\mathbf{A})]^{-\mathrm{T}} \Re(\mathbf{t}), \qquad (7.20)$$

where the notation $\Re(X)$ denote the real part of X. In the above solution, channel information is not required; only the statistics of the RFI interference are needed.

7.4   Uninformed Window

We now consider the case when the statistics of RFI interference are not available to the receiver (uninformed receiver). In this case, the frequency and amplitude

of RFI are not known. We can minimize the total interference by minimizing the stopband energy of $H_0'(e^{j\omega})$

$$\phi_h = \int_{\omega_s}^{2\pi - \omega_s} |H_0'(e^{j\omega})|^2 \frac{d\omega}{2\pi}, \tag{7.21}$$

where ω_s is the stopband bandedge. From (7.7) we can write $H_0'(e^{j\omega})$ as

$$H_0'(e^{j\omega}) = \mathbf{s}^\dagger \mathbf{g}, \tag{7.22}$$

where $\mathbf{g}$ is the window vector and $\mathbf{s}$ is an $(M + \beta) \times 1$ column vector given by

$$\mathbf{s} = \begin{bmatrix} e^{j\omega(\nu - \beta)} \\ e^{j\omega(\nu - \beta + 1)} \\ \vdots \\ e^{j\omega(\nu + M - 1)} \end{bmatrix} \tag{7.23}$$

Then the stopband energy ϕ_h can be rewritten as

$$\phi_h = \int_{\omega_s}^{2\pi - \omega_s} (\mathbf{g}^\dagger \mathbf{s}\mathbf{s}^\dagger \mathbf{g}) \frac{d\omega}{2\pi} = \mathbf{g}^\dagger \mathbf{Q} \mathbf{g}, \tag{7.24}$$

where

$$\mathbf{Q} = \int_{\omega_s}^{2\pi - \omega_s} \mathbf{s}\mathbf{s}^\dagger \frac{d\omega}{2\pi}. \tag{7.25}$$

The elements of Q are given by

$$[\mathbf{Q}]_{mn} = \begin{cases} -\frac{\sin(m-n)\omega_s}{\pi(m-n)}, & m \neq n, \\ 1 - \frac{\omega_s}{\pi}, & m = n. \end{cases} \tag{7.26}$$

The window vector $\mathbf{g}$ can be written as

$$\mathbf{g} = \mathbf{d} + \mathbf{E}\mathbf{b}, \tag{7.27}$$

where $\mathbf{d}^\mathrm{T} = [\, \mathbf{0} \quad \mathbf{1}_M^\mathrm{T} \,]$, and $\mathbf{E}^\mathrm{T} = [\, \mathbf{I}_\beta \quad \mathbf{0} \quad -\mathbf{I}_\beta \,]$.

As a result, the objection function can be given in terms of $\mathbf{b}$,

$$\phi_h = (\mathbf{d} + \mathbf{E}\mathbf{b})^\mathrm{T} \mathbf{Q} (\mathbf{d} + \mathbf{E}\mathbf{b}), \tag{7.28}$$

Similarly to the informed window, using the method of optimization in [77], we can obtain the following optimal uninformed solution $\mathbf{b}$ that minimizes the stopband energy

$$\mathbf{b} = -(\mathbf{E}^{\mathrm{T}}\mathbf{Q}\mathbf{E})^{-\mathrm{T}}(\mathbf{E}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}\mathbf{d}). \tag{7.29}$$

In this case, neither the channel nor the RFI information is needed for obtaining the window.

7.5 Simulations

In this section, we will evaluate the proposed window design technique. The channels used for our evaluations are seven VDSL loops [49]. The DFT size $M = 1024$, cyclic prefix $\nu = 80$, and window length $\beta = 10$. The channel noise consists of AWGN of -140 dBm, FEXT and NEXT crosstalk as described in [49]. The time domain equalizer of length 20 is used to shorten the channel to length less than 70 [79]. The RFI interference is of differential mode with strength -55dBm [49]. Three RFI sources with frequencies at 1.44, 1.9, and 2.0MHz are considered. In this simulation, the RFI signal is generated as in [49]. We will first use VDSL loop1 of length 4500ft as an example to examine the frequency response of the proposed window and demonstrate the effect on subchannel interference and SINR.

Frequency response: Suppose the statistics of RFI is available to the receiver. We compute $\mathbf{w}$ using (7.20) and obtain the informed window form (7.1). Fig. 7.4 shows the frequency response of the informed window $\mathbf{g}$. For comparison, we have also shown the frequency responses of the Hanning window, Blackman window, and Kaiser window with shape parameter $\beta = 5$ [75]. We can see that the informed window has a faster roll-off in low frequency while the other three windows have much smaller sidelobes in high frequency. However, the roll-off in high frequency will not be important when the sidelobes are so small that RFI

80

is not the dominating noise. As the proposed window has the characteristics of fast roll-off in low frequency, fewer tones will be dominated by RFI as we will see next.
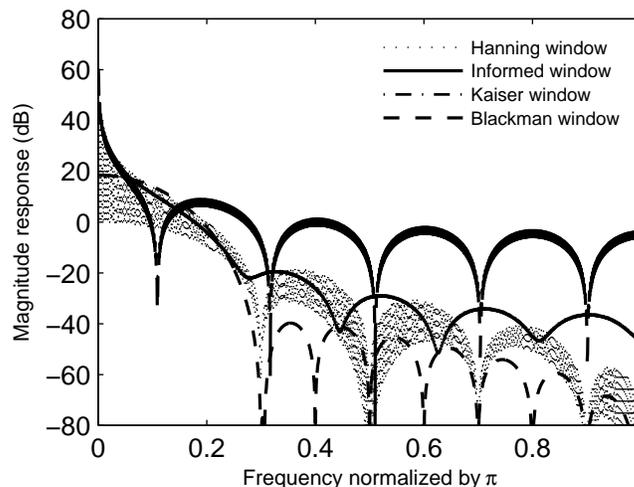
Figure 7.4: Frequency response of receiving windows.

Subchannel Interference: We compute the interference power at the receiver outputs for the receiving windows. Fig. 7.5 shows the RFI interference power of individual tones for the informed window, uninformed window, window in [71], Hanning window, Blackman window, and Kaiser window with shape parameter $\beta = 5$. In Fig. 7.5(a), we compare with the window in [71] and Hanning window. In Fig. 7.5(b), we compare with Blackman window and Kaiser window. We can see that the informed window and uninformed windows have lower RFI power than the other four windows near the RFI source frequencies. Also shown in Fig. 7.5(a) and Fig. 7.5(b) are the combined effects of channel noise (AWGN, FEXT, and NEXT) and the residual ISI for the informed window, uninformed window, window in [71], Hanning window, Blackman window, and Kaiser window, which are labeled as "other noise (informed)", "other noise (uninformed)", "other noise (Window [71])", "other noise (Hanning)", "other noise

(Blackman)", and "other noise (Kaiser)". In both Fig. 7.5(a) and Fig. 7.5(b), the curves of "other noise" overlap with each other and are indistinguishable in the figure. From Fig. 7.5, we can see that RFI is dominating in the tones around the RFI frequencies. For the tones away from the interference sources, other noise is dominating. As a result, higher attenuation of the window in high frequency is of little significance. In this case, the commonly used Hanning window and Blackman window are over designed in high frequency region. The proposed windows, due to their faster roll-off in low frequency, has fewer RFI dominating tones.

Subchannel SINRs: Fig. 7.6 shows the SINRs of the individual tones for both informed and uninformed windows. For comparison, in Fig. 7.6(a)(b), we have also shown the SINRs of the window in [71], Hanning window, Blackman window and Kaiser window with shape parameter $\beta = 5$. From Fig. 7.6(a)(b) we see that the SINRs of the informed and uninformed window are higher than those of the other windows near the RFI source frequency, i.e., in the tones where RFI interference is dominating. This is due to the fact that the proposed windows achieve a better trade-off in low frequency and high frequency. Therefore, we can transmit more bits in the neighboring tones by using the proposed windows. The two curves corresponding to the two proposed windows almost overlap with each other. This shows that the use of uninformed window leads to only a minor performance degradation.

Table 1 shows the bit rates for seven VDSL loops [1] with window length $\beta = 10$, where VDSL loop 1 to 4 are of length 4500 ft. The sampling frequency is $f_s = 4.416$ MHz. For comparison purpose, we have also included the bit rates of the rectangular window, Hanning window, Blackman window, Kaiser window, and the window in [71]. In addition, the bit rates for the case when there is no RFI interference are also shown in the table. From the table, we can see that the proposed windows have better performance for all the test loops.

82
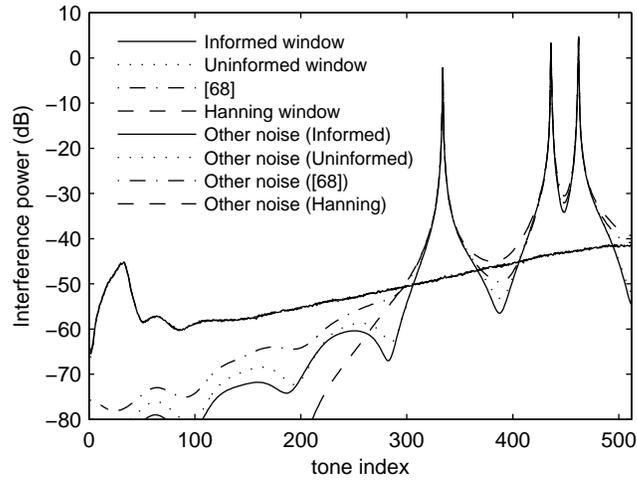
(a)
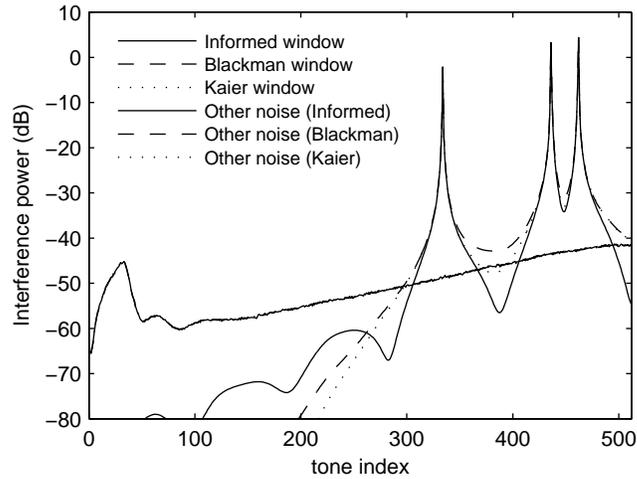
(b)

Figure 7.5: Subchannel interference power of the DMT system with windowing.
(a) Informed window, uninformed window, window in [71], and Hanning win-
dow. (b) Informed window, Blackman window, and Kaiser window with shape
parameter $\beta = 5$.

7.6 Summary

We have proposed a window design method for RFI suppression in DMT sys-
tems. The proposed windows strike a balance between low frequency and high
frequency response. Thus, fewer tones are dominated by RFI and better bit rates
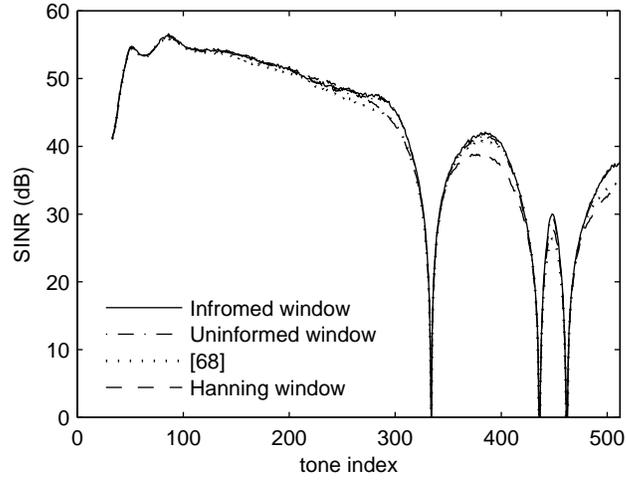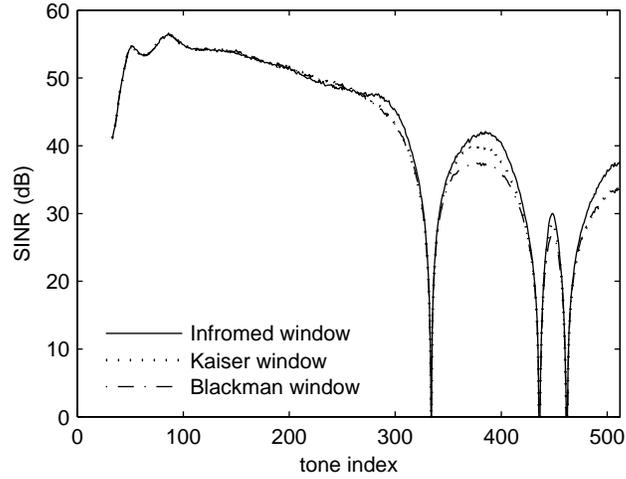
Figure 7.6: Subchannel SINRs of the DMT system with windowing. (a) Informed window, uninformed window, window in [71], and Hanning window. (b) Informed window, Blackman window, and Kaiser window with shape parameter $\beta = 5$.

is achieved. We consider both the case when the receiver knows the statistics of the interference (informed receiver) and the case when the statistics are not available to the receiver (uninformed receiver). In both cases the windows are channel independent and can be obtained in a closed form. Windows designed for uninformed receiver (interference-independent window) has the advantage that

Loop	1	2	3	4	5	6	7
informed	20.74	20.42	18.94	11.25	26.60	22.75	17.97
uninformed	20.60	20.40	18.90	11.22	26.58	22.70	17.94
rectangular	19.72	19.49	17.95	10.40	26.38	21.92	16.94
Hanning	20.23	19.96	18.59	10.90	26.48	22.33	17.42
Blackman	20.14	19.86	18.48	10.80	26.39	22.32	17.46
Kaiser $\beta = 5$	20.38	20.02	18.78	10.97	26.46	22.33	17.68
window [71]	20.24	20.23	18.82	11.06	26.52	22.60	17.79
No RFI	23.34	22.78	21.49	13.45	27.59	22.39	20.57

Table 7.1: Bit rate (Mbits/sec) on VDSL loops.

the window coefficients need not be updated when the statistics of the RFI interference changes. We also shows not knowing the statistics of the RFI source leads to only a minor performance degradation.
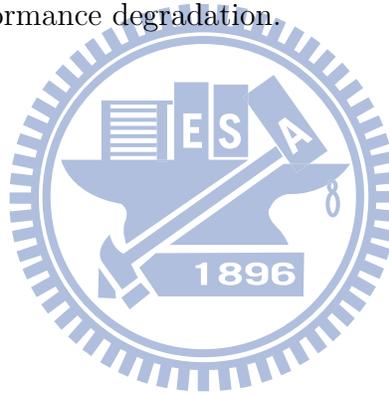
Chapter 8

A Filterbank Approach to Window Designs for Multicarrier Systems

In chapter 7, we have designed the receiving windows for RFI suppression at the receiver. At the transmitting side, spectral leakage is also an important issue in the multicarrier system, and transmitting windows have been used to mitigate the out of band spectral leakage. Better frequency separation among the transmitting filters leads to a smaller out-of-band spectral leakage and also less interference to radio frequency transmission. In this chapter, we will propose a unified filterbank approach to the design of transmitting/receiving windows for multicarrier systems. The approach used here will be more general. We will introduce the so-called subfilters. The use of subfilters will enhance the frequency selectivity of the transmitting and receiving filters. It can be shown that the receiving windows in chapter 7 are special cases of this filterbank approach. The filterbank viewpoint provides an additional insight into the transmitter design for spectral leakage reduction as well as to the receiver design for interference suppression.

8.1 System Model

From section 6.3 and section 6.4, we know that the spectral leakage at the transmitter and the number of subchannels affected by RFI at the receiver depend on the sidelobes of the transmitting and receiving filters. To have a better frequency selectivity, we will design the transmitter and receivers using the filterbank representation in Fig. 6.3. Employing the polyphase identity [76], we observe that the transfer function $T_{k,i}(z)$ from the i-th transmitter input $s_i(n)$ to the k-th signal $y_k(n)$ at the receiver is given by

$$T_{ki}(z) = [H_k(z)C(z)F_i(z)]_{\downarrow N},\tag{8.1}$$

where the notation $[A(z)]_{\downarrow N}$ denotes the N-fold decimated version of $A(z)$. Note that the DMT system has zero inter-block and inter-subchannel ISI, and the transmitter inputs are the same as the receiver outputs $s_k(n) = \hat{s}_k(n)$ when there is no channel noise. As $y_k(n) = \lambda_k \hat{s}_k(n)$, we have

$$T_{ki}(z) = \lambda_k \delta(k - i).\tag{8.2}$$

Summarizing, we can obtain the following lemma.

Lemma 8.1 *Consider the system in Fig. 6.3. The transfer function $T_{k,i}(z)$ from the i-th transmitter input $s_i(n)$ to the k-th signal $y_k(n)$ at the receiver is given by*

$$T_{ki}(z) = \lambda_k \delta(k - i), \quad 0 \le k, i \le M - 1.\tag{8.3}$$

The result holds for any FIR filter $C(z)$ of order $L \le \nu$. The constant λ_k are the M-point DFT of $c(n)$.

So long as the order of $C(z)$ is not larger than ν, the system is free from inter-block interference and inter-subchannel interference. This means that, if we cascade another filter before or after the channel, as long as the product of this extra filter and $C(z)$ has order no larger than ν the overall system remains ISI free. We will use this observation later to design transmitters and receivers in the following sections.

8.2 Receivers with Subfilters

To improve the frequency selectivity of the receiving filters, we introduce the subfilters $Q_k(z)$ to the receiving bank, as shown in Fig. 8.1. With the subfilters, the k-th effective receiving filter becomes $H'_k(z) = H_k(z)Q_k(z)$; the frequency responses of the receiving filters are further shaped by the subfilters. The transfer function from the i-th transmitter input $s_i(n)$ to the k-th signal $y_k(n)$ at the receiver becomes

$$T_{ki}(z) = [H_k(z)(Q_k(z)C(z))F_i(z)]_{\downarrow N}, \tag{8.4}$$

which is the same expression as (8.1) except that the channel is replaced by the composite channel $Q_k(z)C(z)$. From the result in Lemma 8.1, we know the system is free from ISI as long as the order of the composite channel is not larger than ν. In particular, $T_{ki}(z)$ is the same as in (8.3) except that the coefficients λ_k are now the M-point DFT of the composite channel.
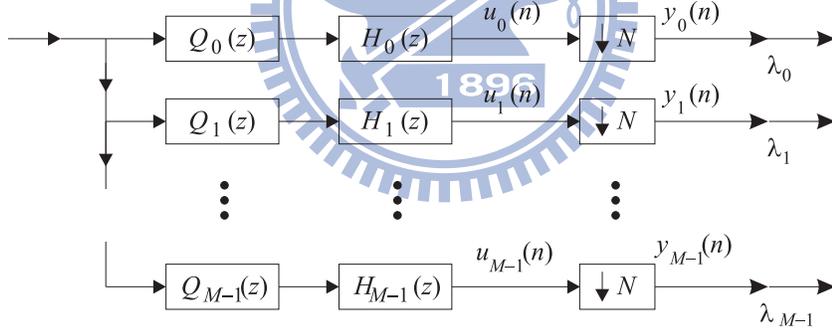
Figure 8.1: The receiving bank with subfilters.

We can choose the subfilters coefficients so that λ_k remain the same after the subfilters are included. To have this property, we need

$$Q_k(e^{j2\pi k/M}) = 1, \tag{8.5}$$

i.e., the k-th DFT coefficient of $Q_k(z)$ normalized to one. In the special case that the subfilters are chosen as shifted versions of the first subfilter, $Q_k(z) =$

$Q_0(zW^k)$, then $Q_k(e^{j2\pi k/M})$ is equal to $Q_0(e^{j0})$. This means that the FEQ coefficients remain the same if the DC value of the first subfilter is one. This translates to the time-domain condition that the sum of the coefficients is one. Suppose that $Q_0(z)$ is a causal FIR filter of order β, then the condition is

$$\sum_{n=0}^{\beta} q_0(n) = 1. \tag{8.6}$$

This condition can be easily satisfied by a simple normalization. Therefore we can design the first subfilter without constraint and then normalize the coefficients to one. The normalization in (8.6) will be assumed in the following discussions. Furthermore when the subfilters are shifted versions of the first subfilter, the new receiving filter becomes $H'_k(z) = W^{-\nu k} H'_0(zW^k)$. They are also shifted versions of the new prototype filter $H'_0(z)$ except for some scalars. We will see below that in the special case $Q_k(z) = Q_0(zW^k)$, these receiving filters form a DFT bank and thus can be implemented efficiently as we discussed in chapter 7. The complexity is almost the same as the conventional DMT system without subfilters.

8.3 Implementation of Receiving Bank with Sub-filters

The new prototype filter is the product of $Q_0(z)$ and the rectangular window $H_0(z)$ given in (6.16). Let the coefficients of $H'_0(z)$ be $b_i/\sqrt{M}$ and we write it as

$$H'_0(z) = \frac{z^{\nu-\beta}}{\sqrt{M}} \sum_{i=0}^{M+\beta-1} b_i z^i. \tag{8.7}$$

We will call b_i receiver window coefficients for reasons that will become clear later. Using the relation $H'_k(z) = W^{-\nu k} H'_0(zW^k)$, we can write the new k-th

receiving filter as

$$H_k'(z) = \frac{z^{\nu-\beta}}{\sqrt{M}} \sum_{i=0}^{M+\beta-1} b_i W^{k(i-\beta)} z^i \tag{8.8}$$

$$= \frac{z^{\nu-\beta}}{\sqrt{M}} \begin{pmatrix} 1 & W^k & \cdots W^{k(M-1)} \end{pmatrix} \mathbf{g}(z), \tag{8.9}$$

where

$$\mathbf{g}(z) = \begin{pmatrix} \mathbf{0} & \mathbf{I}_M \\ \mathbf{I}_\beta & \end{pmatrix} \operatorname{diag} \begin{pmatrix} b_0 & b_1 & \cdots & b_{M+\beta-1} \end{pmatrix} \begin{pmatrix} 1 \\ z \\ \vdots \\ z^{M+\beta-1} \end{pmatrix}. \tag{8.10}$$

Using the above equation, the new receiving bank $\mathbf{h}'(z)$ as indicated in Fig. 8.1 can be written as $\mathbf{h}'(z) = z^{\nu-\beta}\mathbf{W}\mathbf{g}(z)$. Using this expression and the Noble identity for decimators [76], we obtain the same implementation as in Fig. 7.2. This means when the subfilters are chosen as shifted versions of the first subfilter, the receiver with subfilters is the same as the usual receiver windowing described in chapter 7. Thus the windowing technique in chapter 7 can be viewed as a special case of the subfilter problems in section 8.2.

8.4   Window coefficients b_k

The new prototype filter $H_0'(z)$ is the convolution of $h_0(n)$ and a much shorter $q_0(n)$. As $h_0(z)$ is a rectangular window, each window coefficient b_k is a partial sum of the coefficients of $q_0(n)$. With the normalization in (8.6), most of the window coefficients are equal to one, except for those on the two ends. The middle $M - \beta$ coefficients are equal to one, the remaining coefficients, β coefficients on each side, have non-unity values. Fig. 7.1(a) gives an example of window coefficients. Furthermore, we can verify that the time shifts of b_k add up to one, in particular

$$\sum_{\ell=-\infty}^{\infty} b_{k-\ell M} = 1. \tag{8.11}$$

This is known as the time-domain Nyquist property [67, 68]. The subfilter view-point has the advantage that the time-domain Nyquist property is satisfied inherently and the expression can be easily incorporated in the receiving window design.

8.5   Design of Receiver Subfilters

The frequency selectivity of the receiving filters are important for RFI suppression. The radio interference is known to be of a narrowband nature. For the duration of one DMT symbol, it can be considered as a sum of sinusoids. To analyze the effect of interference, we can apply an interference-only signal $v(n)$ to the receiver in Fig. 8.1. Suppose there are J interference sources, and the interference is modeled as

$$v(n) = \sum_{l=0}^{J-1} \mu_l \cos(\omega_l n + \theta_l). \tag{8.12}$$

The interference term at the output of the k-th receiving filter $H'_k(z)$ is

$$u_k(n) = \frac{1}{2} \sum_{l=0}^{J-1} \mu_l \left[H'_k(e^{j\omega_l}) e^{j(\omega_l n + \theta_l)} + H'_k(e^{-j\omega_l}) e^{-j(\omega_l n + \theta_l)} \right]. \tag{8.13}$$

Minimization of interference terms requires the knowledge of μ_l, ω_l and θ_l.

First let us consider the case when the information of the interference is not available. In this case, we can alleviate the effect of interference in the k-th subchannel by minimizing the stopband of the receiving filters. When the receiving filters are frequency shifted versions of the prototype, we use the same objective function as in section 7.4 to design $Q_0(z)$.

$$\phi_h = \int_{\omega_s}^{2\pi - \omega_s} |H'_0(e^{j\omega})|^2 d\omega. \tag{8.14}$$

To consider the optimization of the above objective function, we note that $H'_0(z)$ is the product of $Q_0(z)$ and $H_0(z)$. We can write its Fourier transform as $H'_0(e^{j\omega}) =$

91

$H_0(e^{j\omega})\boldsymbol{\tau}_\beta(\omega)\mathbf{q}_0$, where $\mathbf{q}_0$ is an $\beta \times 1$ vector consisting of the coefficients of $q_0(n)$ and $\boldsymbol{\tau}_\beta(\omega)$ is the $1 \times \beta$ row vector $\begin{pmatrix} 1 & e^{-j\omega} & \cdots & e^{-j\beta\omega} \end{pmatrix}$. Therefore, we can write the stopband energy as

$$\phi_h = \mathbf{q}_0^\dagger \mathbf{B} \mathbf{q}_0, \tag{8.15}$$

where

$$\mathbf{B} = \int_{\omega \in O_h} |H_0(e^{j\omega})|^2 \boldsymbol{\tau}_\beta^\dagger(\omega)\boldsymbol{\tau}_\beta(\omega) d\omega. \tag{8.16}$$

To avoid a trivial solution, we can fix the energy of the first subfilter to be one, $\mathbf{q}_0^\dagger \mathbf{q}_0 = 1$. The matrix $\mathbf{B}$ is always positive definite because the objective function represents the stopband energy of the prototype filter, which is always positive. The minimization of the objective function becomes the optimization of the first subfilter such that the quadratic form in (8.15) is minimized. To minimize ϕ_h, we can choose $\mathbf{q}_0$ as the eigen vector associated with the smallest eigen value of $\mathbf{B}$. Such an approach does not depend on the RFI statistics or the channel; it has the advantage that the subfilters need to be designed only once. The subfilters need not be redesigned when the interference changes. When $Q_k(z)$ is not constrained to be frequency shifted version of $Q_0(z)$, we can design $Q_k(z)$ to minimize the stopband energy of $H_k'(z)$, i.e.,

$$\tilde{\phi}_{k,h} = \int_{\omega_s'}^{2\pi-\omega_s'} |H_k'(e^{j\omega})|^2 d\omega, \tag{8.17}$$

where $\omega_s' = \omega_s + 2\pi k/M$ is the stopband bandedge of the k-the receiving filter. Let $\omega' = \omega - 2\pi k/M$. Using $H_k'(e^{j\omega}) = H_k(e^{j\omega})Q_k(e^{j\omega})$ and (6.18), we have

$$\tilde{\phi}_{k,h} = \int_{\omega_s}^{2\pi-\omega_s} |H_0(e^{j\omega'})|^2 |Q_k(e^{j(\omega'+2\pi k/M)})|^2 d\omega'. \tag{8.18}$$

Thus the optimal subfilter $Q_k(z)$ will satisfy $Q_k(z) = Q_0(zW^k)$. Therefore, for the case that the information of the interference sources is not available to the receiver, the solution of minimizing $\tilde{\phi}_{k,h}$ individually is the same as that of minimizing ϕ_h.

If the information of the interference sources is available to the receiver, the subfilters can be individually optimized. The amplitude of the k-th interference signal $u_k(n)$ is a nonlinear function of the k-th subfilter coefficients. To simplify the problem, note that the interference due to the l-th source will be small if $\mu_l^2(|H_k'(e^{j\omega_l})|^2 + |H_k'(e^{-j\omega_l})|^2)$ is small. The k-th subchannel interference can be mitigated by designing $Q_k(z)$ to minimize $\phi_{k,h}$,

$$\phi_{k,h} = \sum_{l=0}^{J-1} \mu_l^2(|H_k'(e^{j\omega_l})|^2 + |H_k'(e^{-j\omega_l})|^2). \tag{8.19}$$

We can write $\phi_{k,h}$ in a quadratic form similar to that in (8.15) and find the optimal subfilters. Such an optimization requires only the amplitudes and frequencies, but not the phases, of the interference sources. When the subfilters are so designed, the receiving bank does not have the DFT bank structure in Fig. 7.2. Nonetheless, the receiver can be implemented with a much reduced complexity using the sliding window approach in [76]. When the subfilters $Q_k(z)$ are shifted versions of $Q_0(z)$, we can design $Q_0(z)$ to minimize the total interference $\sum_k \phi_{k,h}$ (as shown in section 7.3).

8.6　Transmitter with Subfilters

Similar to the case of the receiving end, we can also introduce subfilters to the transmitter side to improve the frequency selectivity of the transmitting filters. Fig. 8.2 shows the transmitting bank with subfilters. Suppose the subfilters are FIR filters $P_k(z)$ with order α. The k-th new transmitting filter is

$$F_k'(z) = F_k(z)P_k(z). \tag{8.20}$$

The new transmitting filters are of length $N + \alpha$, as $F_k(z)$ are of length N. Now the transfer function from the i-th transmitter input $s_i(n)$ to the k-th signal $y_k(n)$ at the receiver (Fig. 6.3) becomes

$$T_{ki}(z) = [H_k(z)(P_i(z)C(z))F_i(z)]_{\downarrow N}. \tag{8.21}$$

We can also apply the result in Lemma 8.1 here. The overall system remains ISI free as long as the order of the subfilters α satisfy $\alpha + L \leq \nu$. The transfer function $T_{ki}(z)$ is the same as in (8.3), except that now the coefficients λ_k are the M-point DFT of $p_k(n) * c(n)$. As in the case of receiver windowing, we can choose the subfilters to be shifted versions of the first subfilter, i.e., $P_k(z) = P_0(zW^k)$. In this case we can have λ_k remain the same after subfilters are included by normalizing the DC value of $P_0(z)$ like that in (8.6) (Without loss of generality, such a normalization will be assumed in the following discussion.) Furthermore, as we will derive next, the resulting transmitting filters form a DFT bank, which can be implemented very efficiently.
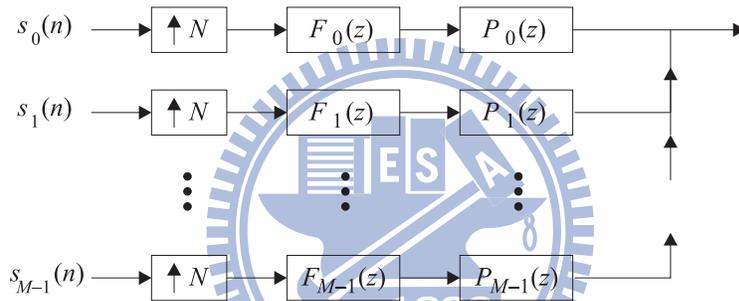
Figure 8.2: The transmitting bank with subfilters.

8.7 Implementation of Transmitting Bank with Subfilters

Similar to the receiver case, when the subfilters are frequency shifted versions of the first subfilter, the new transmitting filters are also frequency shifted versions of the new prototype except for some scalars. In particular, $F'_k(z) = W^{\nu k} F'_0(z W^k)$. Let the coefficients of the prototype be $a_i / \sqrt{M}$ and

$$F'_0(z) = \frac{1}{\sqrt{M}} \sum_{i=0}^{N+\alpha-1} a_i z^{-i}. \tag{8.22}$$

Like the case of receiver windowing, we call these a_i window coefficients. As there is a frequency shifting relation among the transmitting filters, we can obtain the coefficients of all the other transmitting filters given the coefficients of the prototype. Arranging all the transmitting as a row vector, we have the new transmitting bank $\mathbf{f}'(z) = \begin{pmatrix} F'_0(z) & F'_1(z) & \cdots & F'_{M-1}(z) \end{pmatrix}$ as indiciated in Fig. 8.2. The new transmitting bank can be expressed as

$$\mathbf{f}'(z) = \begin{pmatrix} 1 & z^{-1} & \cdots & z^{-N+1} \end{pmatrix} \mathbf{G}(z^N), \tag{8.23}$$

where

$$\mathbf{G}(z) = \begin{pmatrix} \mathbf{D}_0 & \mathbf{D}_1 z^{-1} \\ & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{0} & \mathbf{I}_\nu \\ \mathbf{I}_M & \\ \mathbf{I}_\alpha & \mathbf{0} \end{pmatrix} \mathbf{W}^\dagger. \tag{8.24}$$

The matrices $\mathbf{D}_0$ and $\mathbf{D}_1$ are diagonal matrices given respectively by diag $\begin{pmatrix} a_0 & a_1 & \cdots & a_{N-1} \end{pmatrix}$, and diag $\begin{pmatrix} a_N & a_{N+1} & \cdots & a_{N+\alpha-1} \end{pmatrix}$.

Such an expression of the transmitting bank gives rise to the implementation in Fig. 8.3, where we have used the Noble identity for exchanging LTI filters and expanders [76] to move $\mathbf{G}(z^N)$ to the left of the expanders. The coefficients a_i come from convolution of an N-point rectangular window with a much shorter $p_0(n)$ of length α. When the sum of the coefficients of $p_0(n)$ is normalized to one, most of the coefficients a_i are equal to one. Only the remaining 2α coefficients can have non-unity values and only for these coefficients multiplications are needed.
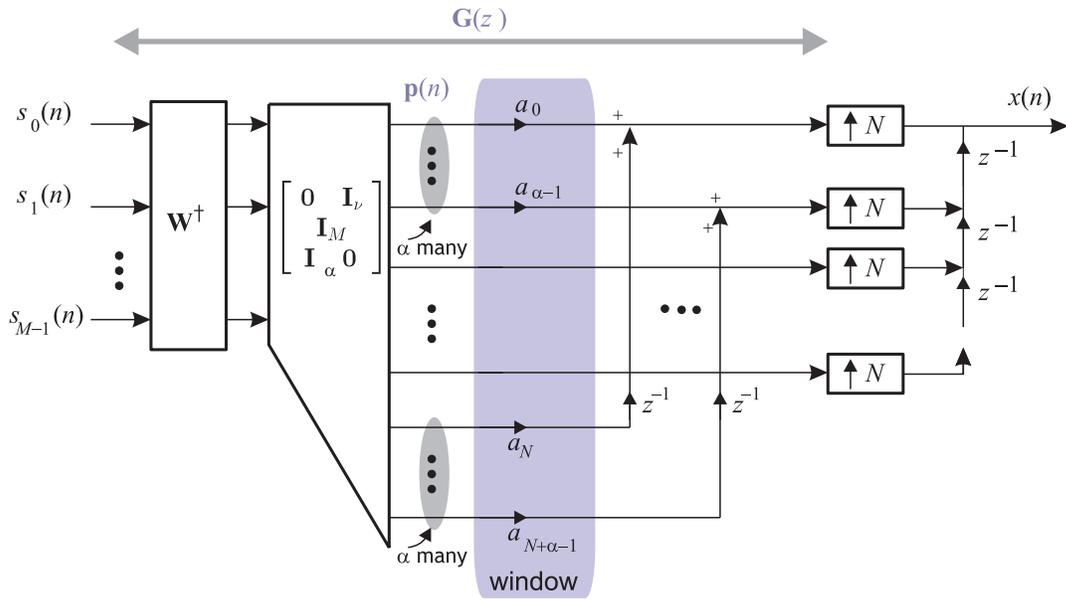
Figure 8.3: Efficient DFT implementation of the transmitting bank.

Connection with the usual transmitter windowing. Observing the DFT bank implementation in Fig. 8.3, we see that for each input block, M-point IDFT is performed, followed by the insertion of cyclic prefix of length ν and also the insertion of suffix of length α. The resulting vector $\mathbf{p}(n)$, as shown in Fig. 8.3, is of size $N+\alpha$. The window coefficients are applied to each vector. Then the last α samples of the previous block are added to the first α samples of the current block, as shown in Fig. 8.4. This is the same as the usual transmitter windowing [49].
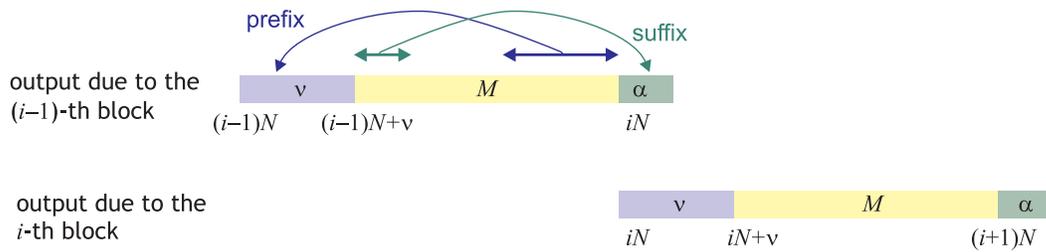
Figure 8.4: Time-domain illustration of transmitter windowing.

8.8 Design of Transmitter Subfilters

For the transmitter side, let us first consider the case when the transmitting filters are constrained to be shifted versions of one prototype. From the expression in (6.19), we see that spectral leakage can be minimized by minimizing the stopband energy of the prototype filter $F_0'(z)$. Following a procedure similar to the design of receiver subfilters, we can write the stopband energy ϕ_f of the prototype $F_0'(z)$ as

$$\phi_f = \mathbf{p}_0^\dagger \mathbf{A} \mathbf{p}_0, \tag{8.25}$$

where

$$\mathbf{A} = \int_{\omega \in O_f} |F_0(e^{j\omega})|^2 \boldsymbol{\tau}_\alpha^\dagger(\omega) \boldsymbol{\tau}_\alpha(\omega) d\omega. \tag{8.26}$$

where O_f denotes the stopband of the prototype filter. We can see that ϕ_f can be minimized by choosing $\mathbf{p}_0$ to be the eigenvector associated with the minimum eigenvalue of $\mathbf{A}$.

Now consider the case when the subfilters are not constrained. The total spectral leakage is

$$\int_{\omega \in O_u} S_x(j\omega) d\omega, \tag{8.27}$$

where $S_x(j\omega)$ is the transmitted spectrum given in (6.19) and O_u denotes the band in which leakage is undesired. The total leakage can be minimized if we can minimize the individual contribution $\phi_{k,f}$ from each subchannel,

$$\phi_{k,f} = \int_{\omega \in O_u} |F_k'(e^{j\omega})|^2 d\omega, \tag{8.28}$$

where O_u denotes the bands that are not used. We can write $\phi_{k,f}$ in a quadratic form like that in (8.25) and find the optimal subfilters. In this case the subfilters do not form a DFT bank, and neither do the new transmitting filters. An efficient implementation of the resulting transmitting bank can be found in [66].

8.9 Simulations

Example 1. Receiver Subfiltering–RFI reduction. In this example, we design the subfilters for RFI reduction at the receiver. The DFT size is $M = 512$ and cyclic prefix length is $\nu = 40$. The order of the subfilters is $\beta = 10$. The channel used in this example is VDSL loop#1 (4500 ft) [49]. and the channel noise is AWGN of -140 dBm. Model 1 differential mode RFI interference is considered [49]. Four RFI sources are assumed in the simulations, at respectively 660, 710, 770 and 1050 KHz, of strength -60, -40, -70, and -55 dBm, respectively. The sampling frequency is $f_s = 2.208$ MHz.

We will consider two different subfilter designs. In the first design, the sub-filters $Q_k(z)$ are shifted versions of $Q_0(z)$ and only $Q_0(z)$ needs to be designed. The subfilter $Q_0(z)$ is the solution to the minimization problem in (8.15). In this case the receiving filters form a DFT bank and the solution is the same as that in section 7.4. In the second design, the RFI source is known to the receiver and the subfilters $Q_k(z)$ are individually optimized by minimizing the objective function $\phi_{k,h}$ in (8.19). The SINRs (signal-to-noise-interference ratio) of the subchannels are as shown in Fig. 8.5. The first case is labelled 'DFT bank (chapter 7)' while the second case 'Subfilters (RFI known)'. For comparison, we have also shown the subchannel SINRs for the cases of rectangular, Hanning windows, and also the window from [71]. The receivers with subfilters enjoy higher SINRs for the tones that are close to the RFI frequencies, especially when the statistics of the RFI source is known and the subfilters are optimized individually. As a result, higher transmission rates can be achieved. The transmission rate of the first case is 7.44 Mbits/sec, and that of the second case is 8.54 Mbits/sec. The transmission rates for the cases of rectangular, Hanning windows, and [71] are 6.84, 7.16, and 7.27 Mbits/sec, respectively.

Example 2. Transmitter Subfiltering–spectral leakage suppression. The
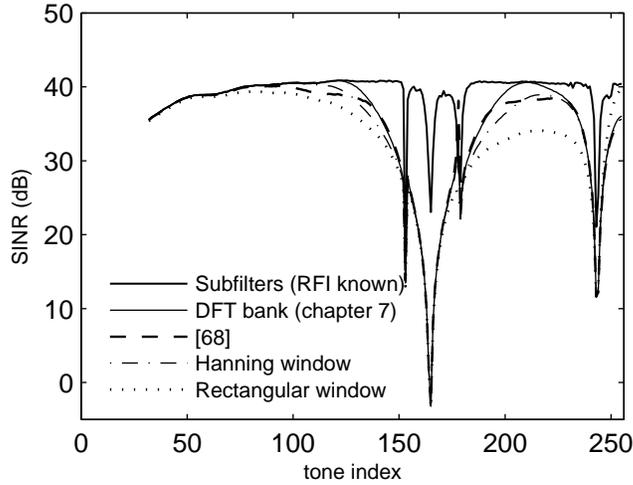
Figure 8.5: The subchannel SINRs.

block size $M = 512$ and prefix length $\nu = 40$. The order α of the subfilters is 20. First we consider the case when the subfilters are shifted versions of the first subfilter $P_0(z)$ and thus the transmitting filters form a DFT bank. We form the positive definite matrix $\mathbf{A}$ and compute the eigenvector corresponding to the smallest eigenvalue to obtain $\mathbf{p}_0$. Second we design the subfilters by minimizing the individual $\phi_{k,f}$ in (8.28) for each subchannel. The first case is labelled 'Subfilters (DFT bank)' while the second case 'Subfilters'. Fig. 8.6 shows the spectrum of the transmitter output. The subcarriers used are 38 to 90 and 111 to 255. The subcarriers with indices smaller than 38 are reserved for voice band and upstream transmission, and those with indices between 91 and 110 are for egress (interference of DMT signals to wireless radio frequency transmission) control. Also shown in the figure are the output spectrums when the rectangular window and transmitter window of [64] is used. The transmitter window in [64] requires no extra cyclic prefix but additional post-processing is needed at the receiver. We see that the spectrum with the subfilters has a much smaller spectral leakage in unused bands.
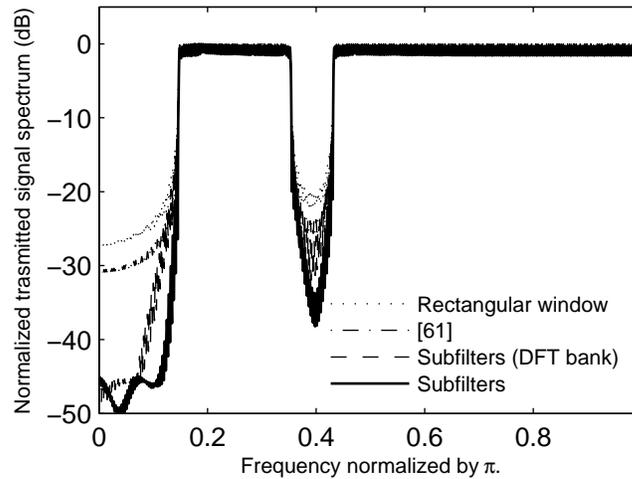
Figure 8.6: The power spectrum of the transmitted signal.

8.10 Summary

In this chapter, we have presented a filterbank approach to the design of transmitter/receiver by introducing subfilters. The frequency separation among the subchannels can be considerably improved. Better separation among the transmitting filters translates to less spectral leakage in the transmitted spectrum while better separation among the receiving filters leads to improved RFI suppression. As these are frequency based characteristics, the filterbank transceiver representation provides a natural and useful framework for formulating the problem. The transmitter/receiver designs are converted to simple eigen-problems and closed form solutions have be obtained.

Chapter 9

Conclusion

In the first half of the thesis, we considered the problem of designing the transceiver with bit allocation. In the earlier works, the transceiver were designed for a given constellation or designed with real-valued bit allocation. In chapter 3, we designed the zero-forcing transceiver with bit allocation for maximizing bit rate under the high bit rate assumption. The optimal transceiver and bit allocation can be obtained in a closed form using simple Hadamard inequality and the Poincaré separation theorem. In chapter 4, we designed the MMSE transceiver with bit allocation for maximizing bit rate. In this approach, we did not use the high bit rate assumption. We have shown that the optimal solution diagonalizes the channel matrix and optimal solution can be obtained by the water-filling solution. For the rate maximizing problem. In chapter 5, we derived the dualities between the power-minimizing problem and the rate-maximizing problem with bit allocation. We considered both the case without the integer bit constraint and the case when the integer bit constraint is imposed. We have shown that whether the bit allocation is integer-constrained or not, if a transceiver is optimal for the power-minimizing problem, it is also optimal for the rate maximizing problem and the converse is true. We also presented an algorithm to find the optimal solution for the power-minimizing problem and the rate-maximizing problem with integer bit constraint.

In the second half of this thesis, we considered the problems of designing the transmitting and receiving windows for the multicarrier systems. For the multicarrier system, better separation among the transmitting filters translates to less spectral leakage in the transmitted spectrum while better separation among the receiving filters leads to improved RFI suppression. In chapter 7, we designed the receiving wondows for RFI suppression in the multicarrier system. We consider both the case when the receiver knows the statistics of the interference and the case when the statistics are not available to the receiver. In both cases the windows are channel independent and can be obtained in a closed form. In chapter 8, we proposed a filterbank approach to the design of transmitter and receiver by introducing subfilters. The filterbank transceiver representation provides a natural and useful framework. At the receiver side, we design the subfilters to mitigate RFI interference. The proposed filterbank approach here is more general. The design in chapter 7 is the special case of the filterbank approach when the subfilters are constrained to be the frequency shifted version of the first subfilter. At the transmitter side, we design the subfilters to minimize the spectral leakage. The designs of the transmitting and receiving subfilters are converted to simple eigen-problems and closed form solutions can be obtained.

Future work:

- For the transceiver design with integer bit constraint, an exhaustive search is used to find the optimal solutions. In the future work, it is interesting to solve the problems with integer bit constraint in a closed form.

- The MIMO channel considered in this thesis is memoryless, in the future we will consider the case when the MIMO channel has a memory.

- The duality between the power minimization problem and rate maximization problem has been proposed in this thesis. In the future we will try to find the connection between other design criteria, for example, BER minimization or capacity maximization.
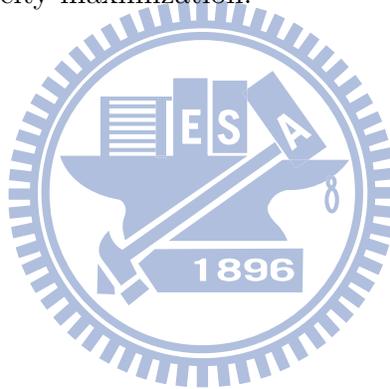
Appendix
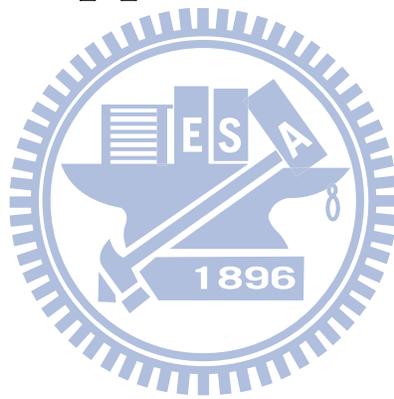
Appendix A: Proof of Lemma 5.1

Let us consider the system $(\mathbf{F}_r, \mathbf{\Lambda}_r)$, where $\mathbf{F}_r$ is the $N \times M_r$ matrix obtained by deleting the columns of $\mathbf{F}$ that correspond to the subchannels assigned with zero power and $\mathbf{\Lambda}_r$ is the $M_r \times M_r$ diagonal matrix obtained by deleting the columns and rows of $\mathbf{\Lambda}_s$ with zero power. The transmit power and bit rate of $(\mathbf{F}_r, \mathbf{\Lambda}_r)$ is the same as $(\mathbf{F}, \mathbf{\Lambda}_s)$. Define the new transmitter as

$$\tilde{\mathbf{F}} = \alpha \mathbf{F}_r, \tag{A-1}$$

where $\alpha > 0$ is a scalar. Let us consider the new system $(\tilde{\mathbf{F}}, \mathbf{\Lambda}_r)$ and fix the target error rate to be ϵ. The new transmit power is given by

$$\text{Tr}(\tilde{\mathbf{F}} \mathbf{\Lambda}_r \tilde{\mathbf{F}}'^{\dagger}) = \alpha^2 \text{Tr}(\mathbf{F}_r \mathbf{\Lambda}_r \mathbf{F}_r^{\dagger}). \tag{A-2}$$

So the new transmit power is a continuous and strictly increasing function of α. Next we will show the new achievable bit rate is also a continuous and strictly increasing function in terms of α. The MSE matrix of the new system becomes

$$\tilde{\mathbf{E}} = [\alpha^2 N_0^{-1} \mathbf{F}_r^{\dagger} \mathbf{H}^{\dagger} \mathbf{H} \mathbf{F}_r + \mathbf{\Lambda}_r^{-1}]^{-1}. \tag{A-3}$$

Let $\sigma_{s_{r,k}}^2 = [\mathbf{\Lambda}_r]_{kk}$ and $\tilde{\sigma}_{e_k}^2 = [\tilde{\mathbf{E}}]_{kk}$. The achievable transmission bit rate is

$$\tilde{B} = \sum_{k=0}^{M_r-1} \log_2 \left(1 + \left(\frac{\sigma_{s_{r,k}}^2}{\tilde{\sigma}_{e_k}^2} - 1 \right) / \Gamma \right). \tag{A-4}$$

The derivative of $\tilde{B}$ with respect to α is

$$\frac{\partial \tilde{B}}{\partial \alpha} = \frac{1}{\log_e 2} \sum_{k=0}^{M_r-1} \frac{-\sigma_{s_{r,k}}^2 \tilde{\sigma}_{e_k}^{-4}}{\sigma_{s_{r,k}}^2 \tilde{\sigma}_{e_k}^{-2} + \Gamma - 1} \cdot \frac{\partial \tilde{\sigma}_{e_k}^2}{\partial \alpha}. \tag{A-5}$$

To derive $\partial \tilde{\sigma}_{e_k}^2 / \partial \alpha$, we compute the derivative of $\tilde{\mathbf{E}}$ with respect to α [40]

$$\frac{\partial \tilde{\mathbf{E}}}{\partial \alpha} = -\tilde{\mathbf{E}} \frac{\partial \tilde{\mathbf{E}}^{-1}}{\partial \alpha} \tilde{\mathbf{E}}, \tag{A-6}$$

$$= -2\alpha N_0^{-1} \tilde{\mathbf{E}} \mathbf{F}_r^{\dagger} \mathbf{H}^{\dagger} \mathbf{H} \mathbf{F}_r \tilde{\mathbf{E}}. \tag{A-7}$$

We have

$$\frac{\partial \tilde{\sigma}_{e_k}^2}{\partial \alpha} = \left[\frac{\partial \tilde{\mathbf{E}}}{\partial \alpha}\right]_{kk} = -2\alpha N_0^{-1}[\mathbf{A}]_{kk}, \tag{A-8}$$

where $\mathbf{A} = \tilde{\mathbf{E}}\mathbf{F}_r^\dagger\mathbf{H}^\dagger\mathbf{H}\mathbf{F}_r\tilde{\mathbf{E}}$ is positive semi-definite and thus $[\mathbf{A}]_{kk} \geq 0$. We now show that the diagonal elements of $\mathbf{A}$ can not all be zeros. Suppose $[\mathbf{A}]_{kk} = 0$ for $k = 0, \cdots, M_r - 1$. This means all the norms of the columns of $\mathbf{H}\mathbf{F}_r\tilde{\mathbf{E}}$ are zeros, i.e., $\mathbf{H}\mathbf{F}_r\tilde{\mathbf{E}} = \mathbf{0}$. As $\tilde{\mathbf{E}}$ is invertible, $\mathbf{H}\mathbf{F}_r\tilde{\mathbf{E}} = \mathbf{0}$ implies $\mathbf{H}\mathbf{F}_r = \mathbf{0}$. In this case, no signal is transmitted and only the noise $\mathbf{q}$ is received by the receiver. Therefore, the diagonal elements of $\mathbf{A}$ cannot be all zeros. Substituting (A-8) into (A-5), we have

$$\frac{\partial \tilde{B}}{\partial \alpha} > 0. \tag{A-9}$$

Hence the bit rate $\tilde{B}$ of the new system is a continuous and strictly increasing function of α. $\triangle\triangle\triangle$

Appendix B: Proof of Lemma 5.2

Define the set

$$U = \{i : \sigma_{s_i}^2 > 0, \ 0 \le i \le M-1\}. \tag{B-1}$$

From (2.14), we have $[\mathbf{E}]_{lk} = 0$ for $l \notin U$ or $k \notin U$. So (5.3) holds when $l \notin U$ or $k \notin U$. We only need to consider (5.3) for $l, k \in U$. Let $\mathbf{F}_r$ and $\mathbf{\Lambda}_r$ be the reduced transmit matrix and reduced symbol autocorrelation matrix as we defined in Section2.2. Suppose the k-th symbol s_k corresponds to the m_k-th symbol in the system with transmitter $\mathbf{F}_r$ and autocorrelation matrix $\mathbf{\Lambda}_r$. For $l \in U$, the noise variance $\sigma_{e_l}^2$ is given by

$$\sigma_{e_l}^2 = [\mathbf{E}]_{ll} = [\mathbf{E}_r]_{m_l,m_l} = [(N_0^{-1}\mathbf{F}_r^{\dagger}\mathbf{H}^{\dagger}\mathbf{H}\mathbf{F}_r + \mathbf{\Lambda}_r^{-1})^{-1}]_{m_l,m_l}. \tag{B-2}$$

The derivative of $\sigma_{e_l}^2$ with respect to $\sigma_{s_k}^2$ is

$$\frac{\partial \sigma_{e_l}^2}{\partial \sigma_{s_k}^2} = \left[\frac{\partial \mathbf{E}}{\partial \sigma_{s_k}^2}\right]_{ll} = \left[\frac{\partial \mathbf{E}_r}{\partial \sigma_{s_k}^2}\right]_{m_l,m_l}. \tag{B-3}$$

The derivative of $\mathbf{E}_r$ with respect to $\sigma_{s_k}^2$ is [40].

$$\frac{\partial \mathbf{E}_r}{\partial \sigma_{s_k}^2} = -\mathbf{E}_r \frac{\partial \mathbf{E}_r^{-1}}{\partial \sigma_{s_k}^2} \mathbf{E}_r. \tag{B-4}$$

Using (2.16), $\mathbf{E}_r = [N_0^{-1}\mathbf{F}_r^{\dagger}\mathbf{H}^{\dagger}\mathbf{H}\mathbf{F}_r + \mathbf{\Lambda}_r^{-1}]^{-1}$, we have

$$\frac{\partial \mathbf{E}_r^{-1}}{\partial \sigma_{s_k}^2} = \begin{bmatrix} 0 & & \cdots & & 0 \\ & \ddots & & & \\ \vdots & & -\sigma_{s_k}^{-4} & & \vdots \\ & & & \ddots & \\ 0 & & \cdots & & 0 \end{bmatrix}. \tag{B-5}$$

So we can obtain

$$\frac{\partial \mathbf{E}_r}{\partial \sigma_{s_k}^2} = \sigma_{s_k}^{-4} \mathbf{e}_{m_k} \mathbf{e}_{m_k}^{\dagger}, \tag{B-6}$$

where $\mathbf{e}_{m_k}$ is the m_k-th column of $\mathbf{E}_r$. Using (B-3) and (B-6), for $l \in U$ we have

$$\frac{\partial \sigma_{e_l}^2}{\partial \sigma_{s_k}^2} = \left[\sigma_{s_k}^{-4} \mathbf{e}_{m_k} \mathbf{e}_{m_k}^{\dagger} \right]_{m_l, m_l} \tag{B-7}$$

$$= \sigma_{s_k}^{-4} |[\mathbf{E}_r]_{m_l, m_k}|^2, \tag{B-8}$$

$$= \sigma_{s_k}^{-4} |[\mathbf{E}]_{l,k}|^2. \tag{B-9}$$

As $\sigma_{s_k}^{-4}|[\mathbf{E}]_{lk}|^2 \geq 0$, we conclude that $\sigma_{e_i}^2$ is an increasing and continuous function of $\sigma_{s_k}^2$. Using (B-9), we have

$$\frac{\partial \sigma_{e_k}^2}{\partial \sigma_{s_k}^2} = (\sigma_{s_k}^2 / \sigma_{e_k}^2)^{-2} > 0, \tag{B-10}$$

which means $\sigma_{e_k}^2$ is strictly increasing on $\sigma_{s_k}^2$. The second order derivative of $\sigma_{e_k}^2$ with respect to $\sigma_{s_k}^2$ is

$$\frac{\partial^2 \sigma_{e_k}^2}{\partial \sigma_{s_k}^4} = -2\sigma_{s_k}^{-6} \sigma_{e_k}^4 + \sigma_{s_k}^{-4} \cdot 2\sigma_{e_k}^2 \cdot \frac{\partial \sigma_{e_k}^2}{\partial \sigma_{s_k}^2} \tag{B-11}$$

$$= 2\sigma_{s_k}^{-6} \sigma_{e_k}^4 \left(\frac{\sigma_{e_k}^2}{\sigma_{s_k}^2} - 1 \right). \tag{B-12}$$

As $\sigma_{e_k}^2 < \sigma_{s_k}^2$ for the MMSE receiver, we have

$$\frac{\partial^2 \sigma_{e_k}^2}{\partial \sigma_{s_k}^4} < 0, \tag{B-13}$$

which implies $\sigma_{e_k}^2$ is a strictly concave function of $\sigma_{s_k}^2$. $\triangle\triangle\triangle$

Appendix C: Proof of Lemma 5.3

Equalities hold in the power-minimizing problem.

Suppose $(\mathbf{F}^*, \mathbf{\Lambda}_s^*, \{b_k^*\})$ is optimal for $\mathscr{A}_{pow}$ and the minimized power is P^*. Let $\{\epsilon_k^*\}$ and B^* be the symbol error rates and bit rate achieved by the optimal solution. Then we have $\epsilon_k^* \leq \epsilon$ and $B^* \geq B_0$. First we show that $\epsilon_k^* = \epsilon$ for all k. Suppose the k_0-th subchannel is assigned with nonzero power $\sigma_{s_{k_0}}^{*2} > 0$ and the error rate is $\epsilon_{k_0}^* < \epsilon$. Consider a new system with the same transmitter $\mathbf{F}^*$ and bit allocation $\{b_k^*\}$, but power allocation is changed to

$$\tilde{\sigma}_{s_k}^2 = \begin{cases} \alpha \sigma_{s_{k_0}}^{*2}, & k = k_0, \\ \sigma_{s_k}^{*2}, & \text{otherwise.} \end{cases} \tag{C-1}$$

where $0 < \alpha < 1$, to be chosen later. Using (2.1), the new transmit power $\tilde{P}$ is

$$\tilde{P} = \sum_{k=0}^{M-1} [\mathbf{F}^{*\dagger}\mathbf{F}^*]_{kk} \tilde{\sigma}_{s_k}^2 < P^*. \tag{C-2}$$

The bit rate of the new system is still B^* as bit allocation is not changed. Next we will show that there always exists $\alpha < 1$ such that the same error rate requirement will be satisfied, i.e.,

$$\tilde{\epsilon}_k = 4\left(1 - \frac{1}{2^{b_k^*/2}}\right) Q\left(\sqrt{\frac{3\tilde{\sigma}_{s_k}^2}{(2^{b_k^*} - 1)\tilde{\sigma}_{e_k}^2}}\right) \leq \epsilon. \tag{C-3}$$

Using Lemma 5.2, when $\alpha < 1$ we have

$$\tilde{\sigma}_{e_k}^2 \leq \sigma_{e_k}^{*2}, \text{ for } k = 0, \cdots, M - 1. \tag{C-4}$$

Using (C-1) and (C-4) we have

$$\frac{\tilde{\sigma}_{s_k}^2}{\tilde{\sigma}_{e_k}^2} \geq \frac{\sigma_{s_k}^{*2}}{\sigma_{e_k}^{*2}} \text{ for } k \neq k_0, \tag{C-5}$$

which implies $\tilde{\epsilon}_k \leq \epsilon_k^* \leq \epsilon$ for $k \neq k_0$. For $k = k_0$, we rearrange the inequality in (C-3), and the error rate constraint for the k_0-th subchannel can be rewritten as

$$\frac{\tilde{\sigma}_{s_{k_0}}^2}{\tilde{\sigma}_{e_{k_0}}^2} \geq \gamma, \tag{C-6}$$

109

where

$$\gamma = \left(\frac{2^{b_{k_0}^*} - 1}{3}\right) \left[Q^{-1}\left(\frac{\epsilon}{4(1 - 2^{b_{k_0}^*/2})}\right)\right]^2.$$

Choose $\alpha = \gamma \sigma_{e_{k_0}}^{*2} / \sigma_{s_{k_0}}^{*2}$. Since $\epsilon_{k_0}^* < \epsilon$, we have $\sigma_{s_{k_0}}^{*2} / \sigma_{e_{k_0}}^{*2} > \gamma$, which implies that $\alpha < 1$. Using (C-4) we have

$$\frac{\tilde{\sigma}_{s_{k_0}}^2}{\tilde{\sigma}_{e_{k_0}}^2} = \frac{\gamma \sigma_{e_{k_0}}^{*2}}{\tilde{\sigma}_{e_{k_0}}^2} \geq \gamma, \tag{C-7}$$

i.e., $\tilde{\epsilon}_{k_0} \leq \epsilon$. With this choice of α, the new system $(\mathbf{F}^*, \{b_k^*\}, \{\tilde{\sigma}_{s_k}^2\})$ can achieve a smaller transmit power $\tilde{P} < P^*$ and still satisfy all the constraints in $\mathscr{A}_{pow}$. This contradicts the assumption that P^* is the minimal power when B_0 is given. Hence we have that $\epsilon_k^* = \epsilon$ for all k. Next we prove that the bit rate B^* is equal to B_0. Assume the bit rate is

$$B^* > B_0. \tag{C-8}$$

Consider a new system with transmitter $\tilde{\mathbf{F}} = \alpha \mathbf{F}^*$ and $\tilde{\mathbf{\Lambda}}_s = \mathbf{\Lambda}_s^*$, where $\alpha > 0$ is a scalar. For the target error rate ϵ, we know from Lemma 5.1 that the bit rate of the new system is a strictly increasing and continuous function of α. So we can properly choose $\alpha < 1$ such that the new bit rate $\tilde{B} = B_0$. In this case, the required power is smaller than P^*. This contradicts the assumption that P^* is the minimal power when B_0 is given. Hence the equality in the bit rate constraint will hold when the design is optimal.

Equality holds in the rate-maximizing problem.

Suppose $(\mathbf{F}^*, \mathbf{\Lambda}_s^*, \{b_k^*\})$ is optimal for $\mathscr{A}_{rate}$ and the maximized bit rate is B^*. Let P^* and $\{\epsilon_k^*\}$ be the transmit power and error rates of the optimal solution. Then we have $P^* \leq P_0$ and $\epsilon_k^* \leq \epsilon$. First we show that $\epsilon_k^* = \epsilon$ for all k. Suppose for the k_0-th subchannel, $\sigma_{s_{k_0}}^{*2} > 0$ and $\epsilon_{k_0}^* < \epsilon$. From (2.17) we know that the error rate ϵ_k is a continuous and increasing function of the number of bits allocated when SNR β_k is fixed. For the same $\mathbf{F}^*$ and $\mathbf{\Lambda}_s^*$, we can increase the number of

bits allocated to the k_0-th subchannel such that the new error rate $\tilde{\epsilon}_{k_0}$ satisfies

$$\epsilon_{k_0}^* < \tilde{\epsilon}_{k_0} \leq \epsilon. \qquad \text{(C-9)}$$

The error rates of other subchannels are not affected while a higher bit rate be achieved. This contradicts the assumption that B^* is the maximal bit rate when the power constraint P_0 is given. Hence we have that $\epsilon_k^* = \epsilon$ for all k. Now let us show $P^* = P_0$. Assume the transmit power is

$$P^* < P_0. \qquad \text{(C-10)}$$

Consider the new system with transmitter $\tilde{\mathbf{F}} = \alpha \mathbf{F}^*$ and $\tilde{\mathbf{\Lambda}}_s = \mathbf{\Lambda}_s^*$, where $\alpha = \sqrt{P_0/P^*} > 1$. The power of the new system is $\tilde{P} = P_0$. From Lemma 5.1, the bit rate of the new system is a strictly increasing function of α, and we have $\tilde{B} > B^*$. This contradicts the assumption that B^* is the maximal bit rate when the power constraint P_0 is given. Hence the power constraint will hold when a solution is optimal for $\mathscr{A}_{rate}$. $\qquad \triangle\triangle\triangle$
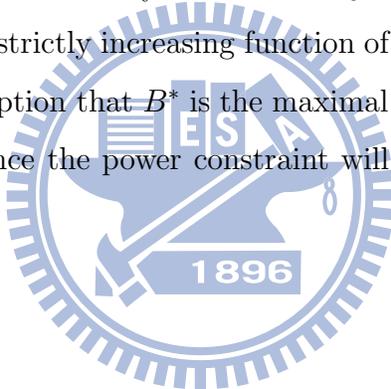
Appendix D: Proof of Lemma 5.4

Suppose $(\mathbf{F}^*, \{\sigma_{s_k}^{*2}\}, \{b_k^*\})$ is optimal for $\mathscr{A}_{pow,int}$. Let ϵ_k^* be the error rate on the k-th subchannel of the optimal system. Then ϵ_k^* is given by

$$\epsilon_k^* = 4\left(1 - \frac{1}{2^{b_k^*/2}}\right)Q\left(\sqrt{\frac{3\beta_k^*}{(2^{b_k^*} - 1)}}\right), \tag{D-1}$$

where the k-th SNR $\beta_k^* = \sigma_{s_k}^{*2}/\sigma_{e_k}^{*2} - 1$ as the receiver is MMSE. The minimized power P^* is given by

$$P^* = \sum_{k=0}^{M-1}[\mathbf{F}^{*\dagger}\mathbf{F}^*]_{kk}\sigma_{s_k}^{*2}, \tag{D-2}$$

where $\sigma_{s_k}^{*2} = [\mathbf{\Lambda}_s^*]_{kk}$. The bit rate B^* is $B^* = \sum_{k=0}^{M-1} b_k^*$. Using the technique in the proof of Lemma 5.3, we can show that $\epsilon_k^* = \epsilon$ for all k. But the technique does not work for property $B^* = B_0$ and a different proof is needed. Suppose

$$B^* > B_0, \tag{D-3}$$

Suppose $\sigma_{s_{k_0}}^{*2} > 0$ and $b_{k_0}^* > 0$ for some k_0-th subchannel. Consider a new system with the same transmitter $\mathbf{F}^*$, but the bit allocation is changed to

$$\tilde{b}_k = \begin{cases} b_{k_0}^* - 1, & k = k_0, \\ b_k^*, & \text{otherwise}, \end{cases} \tag{D-4}$$

and power allocation is changed to

$$\tilde{\sigma}_{s_k}^2 = \begin{cases} \alpha\sigma_{s_{k_0}}^{*2}, & k = k_0, \\ \sigma_{s_k}^{*2}, & \text{otherwise}. \end{cases} \tag{D-5}$$

where $0 < \alpha < 1$ and α will be chosen later. The bit rate of the new system is $\tilde{B} = B^* - 1 \geq B_0$. The transmit power $\tilde{P}$ of the new system is smaller than P^* because

$$\tilde{P} = \sum_{k=0}^{M-1}[\mathbf{F}^{*\dagger}\mathbf{F}^*]_{kk}\tilde{\sigma}_{s_k}^2 < \sum_{k=0}^{M-1}[\mathbf{F}^{*\dagger}\mathbf{F}^*]_{kk}\sigma_{s_k}^{*2} = P^*. \tag{D-6}$$

Next, we will show that with appropriate choice of α, the error rate $\tilde{\epsilon}_k$ of the new system still satisfies the error rate constraint in $\mathscr{A}_{pow,int}$. Using (2.17), $\tilde{\epsilon}_k$ is can

112

be expressed as

$$\tilde{\epsilon}_k = 4\left(1 - \frac{1}{2^{\tilde{b}_k/2}}\right) Q\left(\sqrt{\frac{3\tilde{\beta}_k}{(2^{\tilde{b}_k} - 1)}}\right) \tag{D-7}$$

$$< 4\left(1 - \frac{1}{2^{b_k^*/2}}\right) Q\left(\sqrt{\frac{3\tilde{\beta}_k}{(2^{\tilde{b}_k} - 1)}}\right), \tag{D-8}$$

where $\tilde{\beta}_k = \tilde{\sigma}_{s_k}^2 / \tilde{\sigma}_{e_k}^2 - 1$. Observe that the symbol error rate $\tilde{\epsilon}_k$ of the new system will be smaller than ϵ_k^* if the quantity in the Q function of (D-8) is larger than or equal to that in the Q function of (D-1), i.e.,

$$\frac{1}{2^{\tilde{b}_k} - 1}\left(\frac{\tilde{\sigma}_{s_k}^2}{\tilde{\sigma}_{e_k}^2} - 1\right) \geq \frac{1}{2^{b_k^*} - 1}\left(\frac{\sigma_{s_k}^{*2}}{\sigma_{e_k}^{*2}} - 1\right), \forall k. \tag{D-9}$$

When $\alpha < 1$, using Lemma 5.2 we have

$$\tilde{\sigma}_{e_k}^2 \leq \sigma_{e_k}^{*2}, \text{ for } k = 0, \cdots, M - 1. \tag{D-10}$$

For $k \neq k_0$, using (D-4), (D-5), and (D-10), we have

$$\frac{1}{2^{\tilde{b}_k} - 1}\left(\frac{\tilde{\sigma}_{s_k}^2}{\tilde{\sigma}_{e_k}^2} - 1\right) \geq \frac{1}{2^{b_k^*} - 1}\left(\frac{\sigma_{s_k}^{*2}}{\sigma_{e_k}^{*2}} - 1\right), \tag{D-11}$$

which implies $\tilde{\epsilon}_k < \epsilon_k^* = \epsilon$ for $k \neq k_0$. For $k = k_0$, we can always find $\alpha < 1$ such that (D-9) is satisfied. For example, we can choose

$$\alpha = \frac{1}{\beta_{k_0}^* + 1}\left(1 + \frac{2^{\tilde{b}_{k_0}} - 1}{2^{b_{k_0}^*} - 1}\beta_{k_0}^*\right). \tag{D-12}$$

It can be verified that $1 - \alpha = \frac{\beta_{k_0}^*}{\beta_{k_0}^* + 1}\left(1 - \frac{2^{\tilde{b}_{k_0}} - 1}{2^{b_{k_0}^*} - 1}\right) > 0$. So we have $\alpha < 1$. In this case, we have

$$\frac{\tilde{\sigma}_{s_{k_0}}^2}{\tilde{\sigma}_{e_{k_0}}^2} = \frac{\alpha\sigma_{s_{k_0}}^{*2}}{\tilde{\sigma}_{e_{k_0}}^2} \geq \frac{\alpha\sigma_{s_{k_0}}^{*2}}{\sigma_{e_{k_0}}^{*2}} = 1 + \frac{2^{\tilde{b}_{k_0}} - 1}{2^{b_{k_0}^*} - 1}\beta_{k_0}^*. \tag{D-13}$$

Rearranging (D-13), we can see that (D-9) is satisfied for $k = k_0$. Therefore, we have $\tilde{\epsilon}_k < \epsilon_k^* = \epsilon$ for all k. This means $(\mathbf{F}^*, \{\tilde{\sigma}_{s_k}^2\}, \{\tilde{b}_k\})$ can achieve a smaller transmit power and still satisfies all the constraints in $\mathscr{A}_{pow,int}$. This contradicts the assumption that $(\mathbf{F}^*, \{b_k^*\}, \{\sigma_{s_k}^{*2}\})$ is optimal for $\mathscr{A}_{pow,int}$. Hence the total bit rate B^* of the optimal MMSE system must be equal to B_0. $\triangle\triangle\triangle$

Bibliography

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, New York, NY, USA: Wiley,1991.

[2] G. G. Raleigh, and J. M. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Transactions on Communications*, vol. 46, pp. 357-366, March 1998.

[3] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on telecommunications*, vol. 10, pp. 585-595, November/December 1999.

[4] A. Scaglione, G. B. Giannakis, and S. Barbarossa, "Filterbank transceivers optimizing information rate in Block Transmissions over Dispersive Channels," *IEEE Transactions on Information Theory*, vol. 45, pp. 1019-1032, April 1999.

[5] W. Yu, T. Lan, "Transmitter Optimization for the Multi-Antenna Downlink With Per-Antenna Power Constraints," *IEEE Transactions on Signal Processing*, vol. 55, no. 6, pp. 2646-2660, June 2007.

[6] K. H. Lee and D. P. Petersen, "Optimal linear coding for vecotr channels," *IEEE Transactions on Communications*, Vol. COM24, pp. 1283-1290, December 1976.

[7] J. Yang, and S. Roy, "On Joint transmitter and receiver optimization for multiple-input-multiple-Output (MIMO) transmission systems," *IEEE Transactions on Communications*, vol. 42, pp. 3221-3231 , December 1994.

[8] A. Scaglione, G. B. Giannakis, and S. Barbarossa, "Redundant filterbank precoders and equalizers, Part I: unification and optimal designs," *IEEE Transactions on Signal Processing*, vol. 47, pp. 1988-2006, July 1999.

[9] H. Sampath, P. Stoica, and A. Paulraj, "Generalized linear precoder and decoder design for MIMO channelsusing the weighted MMSE criterion," *IEEE Transactions on Communications*, vol. 49, pp. 2198-2206, December 2001.

[10] Y. Ding, T. N. Davidson, Z.-Q. Luo, and K.-M. Wong, "Minimum BER block precoders for zero-forcing equalization," *IEEE Transactions on Signal Processing*, vol. 51, pp. 2410-2423, Sep. 2003.

[11] Y.-P. Lin and S.-M. Phoong, "BER minimized OFDM systems with channel independent precoders," *IEEE Transactions on Signal Processing*, vol. 51, pp. 2369-2380, Sept. 2003.

[12] D. P. Palomar, M. Bengtsson, B. Ottersten, "Minimum BER linear transceivers for MIMO channels via primal decomposition," *IEEE Transactions on Signal Processing*, vol.53, pp. 2866-2882, August 2005.

[13] A. Scaglione, P. Stoica, S. Barbarossa, G. B. Giannakis, and H. Sampath, "Optimal designs for space-time linear precoders and decoders," *IEEE Transactions on Signal Processing*, vol. 50, pp. 1051-1064, May 2002.

[14] A. Yasotharan, "Multirate zero-forcing Tx-Rx design for MIMO channel under BER constraints," *IEEE Transactions on Signal Processing*, vol. 54, pp. 2288 - 2301, June 2006.

[15] D. P. Palomar, M. A. Lagunas, and J. M. Cioffi, "Optimum linear joint transmit-receive processing for MIMO channels with QoS constraints," *IEEE Transactions on Signal Processing*, vol. 52, pp. 1179- 1197, May 2004.

[16] C.-H. F. Fung, W. Yu, T. J. Lim, "Precoding for the Multiantenna Downlink: Multiuser SNR Gap and Optimal User Ordering," , *IEEE Transactions on Communications*, vol. 55, no. 1, pp. 188-197, January 2007.

[17] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: a unified framework for convex optimization," *IEEE Transactions on Signal Processing*, vol. 51, pp. 2381-2401, Sept. 2003.

[18] S. Dasgupta and A. Pandharipande "Optimum multiflow biorthogonal DMT with unequal subchannel assignment," *IEEE Transactions on Signal Processing*, vol. 53, pp. 3572-3582, September 2005.

[19] X. Song and S. Dasgupta, "A separation principle for optimum biorthogonal DMT without a high bit rate assumption," *Proceedings of ISSPA*, Sydney, Australia, vol. 1, pp. 21-24, August 2005.

[20] Y.-P. Lin and S.-M. Phoong, "Perfect discrete multitone modulation with optimal transceivers," *IEEE Transactions on Signal Processing*, vol. 48, pp. 1702-1711, June 2000.

[21] Y.-P. Lin and S.-M. Phoong, "Optimal ISI-free DMT transceivers for distorted channels with colored noise," *IEEE Transactions on Signal Processing*, vol. 49, pp. 2702-2712, November 2001.

[22] D. P. Palomar, A. Pascual-Iserte, J. M. Cioffi, and M. A. Lagunas, "Convex optimization theory applied to joint transmitter-receiver design in MIMO

channels," *Space-Time Processing for MIMO Communications*, A. B. Gershman and N. D. Sidiropoulos, Eds. New York: Wiley, ch. 8, pp. 269-318.

[23] C.-C. Li, Y.-P. Lin, S.-H. Tsai, and P. P. Vaidyanathan, "Optimization of Transceivers with Bit Allocation to Maximize Bit Rate for MIMO Transmission," *IEEE Transactions on Communications*, vol. 57, no. 11, pp. 3556-3560, Dec. 2009.

[24] D. P. Palomar, S. Barbarossa, "Designing MIMO communication systems: constellation choice and linear transceiver design," *IEEE Transactions on Signal Processing*, vol. 53, pp. 3804-3818, October 2005.

[25] P. P. Vaidyanathan, Yuan-Pei Lin, Sony Akkarakaran, See-May Phoong, "Discrete multitone modulation with principal component filter banks," *IEEE Transactions on circuits and systems I: fundamental theory and applications* , vol. 49, pp. 1397-1412, october 2002.

[26] D. P. Palomar and Y. Jiang, "MIMO Transceiver Design via Majorization Theory," *Foundations and Trends in Communications and Information Theory*, vol. 3, pp. 331-551, 2007.

[27] P. P. Vaidyanathan, See-May Phoong, and Yuan-Pei Lin, *Signal Processing and Optimization for Transceiver Systems.* Cambridge University Press, 2010.

[28] A. Yasotharan, "Lower bound on the BER of a pre- and postequalized vector communication system," *IEEE Transactions on Communications,* Vol. 53, pp. 453-462, March 2005.

[29] J. Campello, "Optimal discrete bit loading for multicarrier modulation systems," *Proceedings, IEEE International Symposium on Information Theory*, pp. 193, Aug 1998.

[30] S. Kourtis, "Optimum bit allocation algorithm for DMT-based systems under minimum transmitted power constraint," *Electronics Letters* , vol. 35, pp. 2181-2182, Dec. 1999.

[31] N. Papandreou and T. Antonakopoulos, "A new computationally efficient discrete bit-loading algorithm for DMT applications," *IEEE Transactions on Communications*, vol. 53, pp. 785-789, May 2005.

[32] M. Zwingelstein-Colin, M. Gazalet, and M. Gharbi, "Non-iterative bit-loading algorithm for ADSL-type DMT applications," *IEE Proceedings on Communications*, vol. 150, no.6, pp. 414-418, Dec. 2003.

[33] P. S. Chow, J. M. Cioffi, and J. A. C. Bingham, "A practical discrete multitone transceiver loading algorithm for data transmission over spectrally shaped channels," *IEEE Transactions on Communications*, vol. 43, pp. 773-775, Feb/Mar/Apr 1995.

[34] B. Fox, "Discrete optimization via marginal analysis," *Management Science,* vol. 13, pp. 210-216, Nov. 1966.

[35] M. Vemulapalli, S. Dasgupta, and A. Pandharipande, "A new algorithm for optimum bit loading with a general cost," *IEEE Proceedings on International Symposium on Circuits and Systems*, pp. 21-24, May 2006.

[36] V. M. K. Chan, W. Yu, "Multiuser Spectrum Optimization for Discrete Multitone Systems With Asynchronous Crosstalk," *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5425-5435, November 2007.

[37] B. Farber and K. Zeger, "Quantization of Multiple Sources Using Nonnegative Integer Bit Allocation," *IEEE Transactions on Information Theory*, vol. 52, pp. 4945-4964, Nov. 2006.

[38] J. S. Chow, J. C. Tu, and J. M. Cioffi, "A Discrete Multitone Transceiver System for HDSL Applications," *IEEE Journal on Selected Areas in Communications.*, pp. 895-908, Aug 1991.

[39] G. D. Dudevoir, J. S. Chow, S. Kasturia, and J. M. Cioffi, "Vector coding for T1 transmission in the ISDN digital subscriber loop," *IEEE International Conference on Communications*, pp. 536-540, June 1989.

[40] T. K. Moon, and W. C. Stirling , *Mathematical Methods and Algorithms for Signal Processing.* Prentice Hall, 2000.

[41] R. A. Horn, and C. R. Johnson , *Matrix analysis.* Cambridge University Press, 1985.

[42] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applicaions.* New York: Academic, 1979.

[43] E. K. P. Chong, and S. H. Zak, *An introduction to optimization.* New York: John Wiley & Sons, 1996.

[44] A. Papoulis and S.U. Pillai, *Probability, Random Variables and Stochastic Processes* fourth edition New York: McGraw-Hill, 2002.

[45] J. G. Proakis, *Digital Communications*, third edition New York: McGraw-Hill, 1995.

[46] R. Steele, *Mobile Radio Communications.* New York: IEEE Press, 1992.

[47] J. A. C. Bingham, "Multicarrier modulation for data transmission: An idea whose time has come," *IEEE Communication Magazine*, vol. 28, pp. 5-14, May 1990.

[48] "Asymmetric Digital Subscriber Lines (ADSL)-Metallic Interface," ANSI T1.413, 1998.

[49] "Very-high Speed Digital Subscriber Lines (VDSL)-Metallic Interface," ANSI T1.424, 2002.

[50] "Very high speed digital subscriber line transceivers 2 (VDSL2)", ITU-T Recommendation G.993.2, 2006.

[51] ISO/IEC, IEEE Std. 802.11a, 1999.

[52] ETSI, "Digital Video Broadcasting; Framing, Structure, Channel Coding and Modulation for Digital Terrestrial Television (DVB-T)," ETS 300 744, 1997.

[53] L. De Clercq, M. Peeters, S. Schelstraete, and T. Pollet, "Mitigation of radio interference in xDSL transmission," *IEEE Communications Magazine*, vol. 38, pp. 168-173, March 2000.

[54] A. Vahlin and N. Holte, "Optimal Finite Duration Pulses for OFDM," *IEEE Transactions on Communications*, vol. 44, no. 1, pp. 10-14, Jan. 1996.

[55] H. Nikookar and R. Prasad, "Optimal waveform design for multicarrier transmission through a multipath channel," *IEEE Proceedings on Vehicular Technical Conference*, vol. 3, pp. 1812-1816, May 1997.

[56] K. Matheus, and K.-D. Kammeyer, "Optimal design of a multicarrier systems with soft impulse shaping including equalization in time or frequency direction," *IEEE Proceedings on Global Telecommunications Conference*, vol. 1, pp. 310-314, Nov. 1997.

[57] N. Laurenti and L. Vangelista, "Filter design for the conjugate OFDM-OQAM system," *presented at First Int'l Workshop on Image and Signal Processing and Analysis*, June 2000.

[58] S. B. Slimane, "Performance of OFDM systems with time-limited waveforms over multipath radio channels," *presented at Global Telecommunications Conference*, 1998.

[59] M. Pauli and P. Kuchenbecker, "On the reduction of the out-of-band radiation of OFDM-signals," *IEEE Proceedings on International Conference on Communications*, vol. 3, pp. 1304-1308, June 1998.

[60] Yuan-Pei Lin and See-May Phoong, "OFDM Transmitters: Analog Representation and DFT Based Implementation," *IEEE Transactions on Signal Processing*, Sepetember 2003.

[61] R. W. Lowdermilk, "Design and performance of fading insensitive orthogonal frequency division multiplexing (OFDM) using polyphase filtering techniques," *presented in Conference Record of the Thirtieth Asilomar Conference on Signals, Systems and Computers*, November 1996.

[62] H. Boelcskei, P. Duhamel, and R. Hleiss, "Design of pulse shaping OFDM/OQAM systems for high data-rate transmission over wireless channels," *IEEE Proceedings on International Conference on Communications*, 1999.

[63] P. Siohan, C. Siclet, and N. Lacaille, "Analysis and Design of OFDM/OQAM Systems Based on Filterbank Theory," *IEEE Transactions on Signal Processing*, May 2002.

[64] G. Cuypers, K. Vanbleu, G. Ysebaert, M. Moonen "Egress Reduction By Intra-symbol Windowing in DMT-based Transmissions", *IEEE Proceedings on International Conference on Acoustics, Speech and Signal Processing*, 2003.

[65] Yuan-Pei Lin and See-May Phoong, "Window Designs for DFT based Multicarrier System," *IEEE Transactions on Signal Processing*, March 2005.

[66] Chun-Yang Chen, and See-May Phoong, "Per tone Shaping filters for DMT Transmitters," *IEEE Proceedings on International Conference on Acoustic, Speech, and Signal Processing*, May 2004.

[67] C. Muschallik, "Improving an OFDM reception using an adaptive Nyquist windowing," *IEEE Transactions on Consumer Electronics*, vol. 42, no. 3, pp. 259-269, Aug. 1996.

[68] S. H. Müller-Weinfurtner, "Optimal Nyquist Windowing in OFDM Receivers," *IEEE Transactions on Communications*, vol. 49, no. 3, March 2001.

[69] P. Spruyt, P. Reusens, and S. Braet, "Performance of Improved DMT Transceiver for VDSL," ANSI T1E1.4, Doc. 96-104, April 1996.

[70] S. Kapoor and S. Nedic, "Interference suppression in DMT receivers using windowing," *IEEE International Conference on Communications*, vol. 2, pp. 778-782, June 2000.

[71] A. J. Redfern, "Receiver Window Design for Multicarrier Communication Systems," *IEEE Journal on Selected Areas in Communications*, vol. 20, pp. 1029-1036, June 2002.

[72] G. Cuypers, K. Vanbleu, G. Ysebaert, and M. Moonen, P. Vandaele, "Combining raised cosine windowing and per tone equalization for RFI mitigation in DMT receivers," *IEEE International Conference on Communications*, vol. 4, pp. 2852-2856, May 2003.

[73] B. Borna and T. N. Davidson, "Efficient filter bank design for filtered multitone modulation," *IEEE International Conference on Communications*, vol. 1, pp. 38-42, June 2004.

[74] G. Ysebaert, K. Vanbleu, G. Cuypers, and M. Moonen, "Joint Window and Time Domain Equalizer Design for Bit Rate Maximization in DMT-Receivers," *IEEE Transactions on Signal Processing*, vol. 53, pp. 1132-1146, March 2005.

[75] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing*, Prentice Hall, 1999.

[76] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*, Englewood Cliffs,Prentice-Hall, 1993.

[77] E. K. P. Chong, and S. H. Zak, *An introduction to optimization*, New York, John Wiley & Sons, 1996.

[78] K. H. Rosen, *Discrete mathematics and its applications,* 5th edition, McGraw Hill.

[79] R. K. Martin, J. Balakrishnan, W. A. Sethares, and C. R. Johnson Jr., "A blind Adaptive TEQ for Multicarrier Systems," *IEEE Signal Processing Letters*, vol. 9, no. 11, pp. 341-343, November 2002.

[80] K. Van Acker, G. Leus, M. Moonen, O. van de Wiel, and T. Pollet, "Per tone equalization for DMT-based systemss," *IEEE Transactions on Communications*, vol. 49, no. 1, Jan. 2001.

[81] ISO/IEC, IEEE Std. 802.11a, 1999.