The Dissertation Committee for Zukang Shen

certifies that this is the approved version of the following dissertation:

Multiuser Resource Allocation in Multichannel Wireless Communication Systems

Committee:

Jeffrey G. Andrews, Supervisor

Brian L. Evans, Supervisor

Melba M. Crawford

Gustavo de Veciana

Robert W. Heath, Jr.

Edward J. Powers

Multiuser Resource Allocation in Multichannel Wireless Communication Systems

by

Zukang Shen, B.S.E.E.; M.S.E.E.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2006

To my family

Acknowledgments

I would like to express my highest gratitude to my parents Haochuan Shen and Xi'an Chen. Without their continuous support, much of this dissertation would not have been possible. I would also like to thank my wife Tiefeng Qiang, who constantly encouraged me throughout the ups and downs in my Ph.D. studies. I am so lucky to have them in my life.

I would like to show special thanks to my supervisors, Prof. Jeffrey G. Andrews and Prof. Brian L. Evans, not only for their excellent academic advice, but also for all kinds of support in my life. I applied for admission at The University of Texas at Austin as an undergraduate in 2001. Without Prof. Evans' trust and help, I would not have had the chance to pursue my Ph.D. in the United States. I want to thank Prof. Andrews for introducing me to the area of wireless communications. I can still vividly remember the first time I walked into Prof. Andrews' office in 2002 and he showed me the problem of adaptive resource allocation, which became my dissertation topic. Further, their integrity and ethical behavior will greatly influence the rest of my life. I am blessed to have them as my Ph.D. supervisors.

I wish to thank my committee members, Prof. Melba M. Crawford, Prof. Gustavo de Veciana, Prof. Robert W. Heath, Jr., and Prof. Edward J. Powers, for their constructive feedback on my dissertation. I'm honored to have them on my committee. I would like to show my special thanks to Prof. Heath, not only for the courses (including wireless communications, advanced signal processing, and space-time communications) I took from him, but also for his critical comments on some of our co-authored technical papers. I also want to thank Prof. de Veciana for his excellent teaching in the course of probability and stochastic process. I have learned so much from them.

Last but not least, I would like to thank all current and former members in the Wireless Networking and Communications Group for the wonderful friendship, especially Runhua Chen, Aamir Hasan, Wei Wu, Taiwen Tang, Ming Ding, Ian Wong, and Daifeng Wang for all the joys we had together. I also wish to thank Runhua Chen for the joint work during my Ph.D. studies, some of which is included in this dissertation.

Zukang Shen

The University of Texas at Austin
May 2006

Multiuser Resource Allocation in Multichannel Wireless Communication Systems

Publication No. ____________________

Zukang Shen, Ph.D.
The University of Texas at Austin, 2006

Supervisors: Jeffrey G. Andrews and Brian L. Evans

A downlink wireless system features a centralized basestation communicating to a number of users physically scattered around the basestation. The purpose of resource allocation at the basestation is to intelligently allocate the limited resources, e.g. total transmit power and available frequency bandwidth, among users to meet users' service requirements. Channel-aware adaptive resource allocation has been shown to achieve higher system performance than static resource allocation, and is becoming more critical in current and future wireless communication systems as the user data rate requirements increase. Adaptive resource allocation in a multichannel downlink system is more challenging because of the additional degree of freedom for resources, but offers the potential to provide higher user data rates. Multiple channels can be created in the frequency domain using multiple carrier frequencies, a.k.a. multicarrier modulation (MCM), or in the spatial domain with multiple transmit and receive antennas, a.k.a. multiple-input multiple-output (MIMO) systems. This dissertation aims to study the system performance, e.g. total throughput and/or fairness, in multiuser multicarrier and multiuser MIMO systems with adap-

tive resource allocation, as well as low complexity algorithms that are suitable for cost-effective real-time implementations in practical systems.

The first contribution of this dissertation is a general framework for adaptive resource allocation in multiuser multicarrier systems that maximizes the total throughput subject to fairness constraints to enforce arbitrary proportional data rates among users. Whereas the global optimality is computationally intensive to obtain, a low complexity algorithm that decouples the subchannel and power allocation is proposed.

The second contribution concerns precoding using block diagonalization (BD) for single-carrier downlink multiuser MIMO systems. The contribution is twofold. First, it is shown that BD, as a practically realizable precoding technique, can achieve a significant part of the sum capacity achieved by dirty paper coding (DPC), which is optimal. Practical coding schemes that approach the DPC sum capacity, however, are still largely unknown. Second, an upper bound on the ergodic sum capacity gain of DPC over BD in Rayleigh fading channels is derived.

The third contribution concerns low-complexity BD precoding algorithms. Due to the zero inter-user interference requirement imposed by BD, the maximum number of simultaneously supportable users is limited. The brute-force search for the optimal user set, however, is computationally prohibitive for systems with a large number of users. The dissertation proposes two suboptimal user selection algorithms for BD that have linear complexity in the number of users, yet achieve total throughput close to the optimal.

A common characteristic of the resource allocations for multiuser multicarrier and multiuser MIMO systems is that the limited resources shall be allocated among multiple users as well as multiple parallel subchannels. As MCM and MIMO have been widely adopted in various standards, the research in this dissertation contributes to a better understanding of the system performance, and bridges the theory to practical implementations with the proposed low complexity algorithms.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 A Brief History of Cellular Systems

Powered by enabling technologies, such as advanced digital signal processing and very large scale integrated circuits, wireless communication has been experiencing an explosive growth in the last decades. Cellular systems are one of the most successful wireless applications, having billions of subscribers. It owes its birth to Bell Laboratories, where the cellular concept was conceived in the 1970s [52]. Due to the fact that radio signal strength weakens with distance, the limited frequency bandwidth can be spatially reused, rendering the possibility of wide coverage over a large population.

The first generation of cellular systems in the United States was advanced mobile phone systems (AMPS), which was deployed in 1980s. AMPS adopted analog FM technology with frequency division multiple access (FDMA). A similar analog cellular system, named The European Total Access Communication System (E-TACS), was deployed in Europe. Soon the first generation of cellular systems reached its capacity and was phased out by the second generation in the early 1990s. The 2G systems adopted digital technologies and provided much higher communication capacity at an even lower cost. Due to the debate on the spectrum access technologies, three major 2G standards were born, namely IS-136, IS-95 in the United States and Global System for Mobile (GSM) in Europe. The 2G standards have high data rate versions, e.g. General Packet Radio Service (GPRS) and Enhanced Data rates for GSM Evolution (EDGE) for GSM, IS-136 high speed (IS-136HS) for IS-136, and IS-95 high data rate (IS-95 HDR) for IS-95 [26]. These improved 2G cellular systems are generally referred to as 2.5G systems. The third and current generation of cellular systems includes wideband code division multiple access (WCDMA) and CDMA2000. The WCDMA frequency division duplex (FDD) and time division duplex (TDD) standards have been adopted in Europe and China, respectively, while CDMA2000 has been deployed in Korean and America. With different spreading factors and modulation methods, WCDMA and CDMA2000 can support transmission rate up to several mega-bits per second. The next generation of wireless cellular systems is envisioned to be multicarrier-based for its efficient bandwidth usage [20] [65] [83].

1.2 Spectrum Sharing Technologies

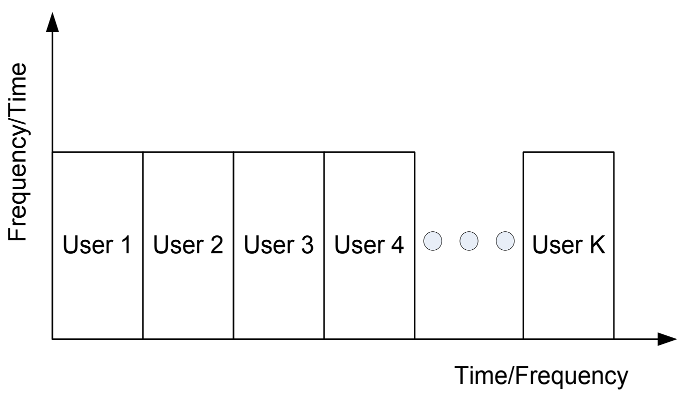

Figure 1.1: TDMA and FDMA

Wireless communication systems are essentially multiuser communication systems. The limited spectrum resources are shared among multiple users for successful communication. The typical spectrum sharing technologies include time division multiple access (TDMA), frequency division multiple access (FDMA), code division multiple access (CDMA), and spatial division multiple access (SDMA).

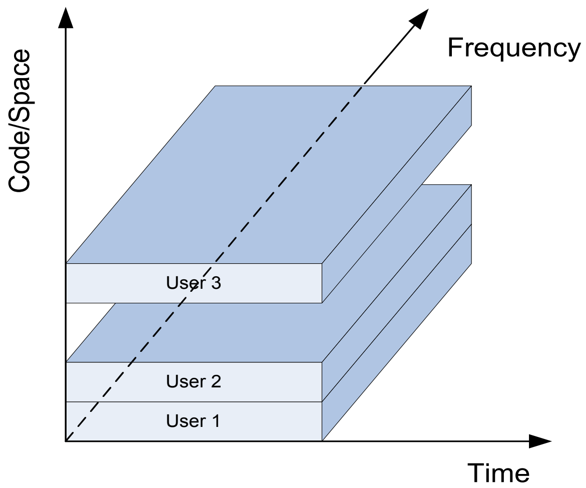

Figure 1.2: CDMA and SDMA

TDMA divides the transmit time into a serial number of time slots. One user is allowed to transmit in a time slot over the entire bandwidth. Similarly, FDMA creates multiple subbands in the frequency domain. A user may be able to occupy a subband throughout the whole transmission period. Fig. 1.1 shows the basic idea of TDMA and FDMA. TDMA and FDMA were widely employed in earlier generations of cellular systems.

CDMA, instead of separating users in either the time or frequency domain, distinguishes users in the code domain. Each active user is allocated a specific sequence, with which multiple users can enjoy the same bandwidth at the same time without causing significant interference to each other. SDMA utilizes multiple transmit and receive antennas to separate users in the spatial domain, also allowing users to access the same bandwidth simultaneously. The basic idea of CDMA and SDMA is shown in Fig. 1.2.

It is important to notice that those multiple access technologies are usually used in combination. For example, WCDMA TDD employs CDMA with TDMA, where the transmission time is divided into a number of time slots and within each time slot, multiple users employ CDMA to access the whole bandwidth. Further, FDMA is used in almost all cellular systems.

1.3 Resource Allocation in Wireless Communication Systems

In a downlink wireless system, a centralized basestation needs to communicate to multiple users, with limited resources, e.g. total transmit power and available frequency bandwidth. Given the freedom of separating users in the time, frequency, code, or spatial domain, how the basestation allocates the resources among users is critical to system performance. Earlier generations of wireless systems adopted static resource allocations such as time or frequency division multiple access, where the basestation takes turns to serve one user in a designated time slot or frequency band, irrespective of the user channels. The wireless channel is, however, time-varying and frequency selective. The channels experienced by different users are largely independent because of users' different locations. The basestation should allocate the limited resources among users by taking the user channel conditions into consideration and enhance the system performance. Further, adaptive resource allocation in a multichannel downlink system is more challenging because of the additional degree of freedom for resources. Multiple channels can be created in frequency domain using multiple carrier frequencies, a.k.a. multicarrier modulation (MCM) or in spatial domain with multiple transmit and receive antennas, a.k.a. multiple-input multiple-output (MIMO) systems. MCM and MIMO are two promising technologies that have been adopted in various standards. Adaptive resource allocation in multiuser multichannel wireless systems has drawn significant attention recently.

1.3.1 Multicarrier Modulation

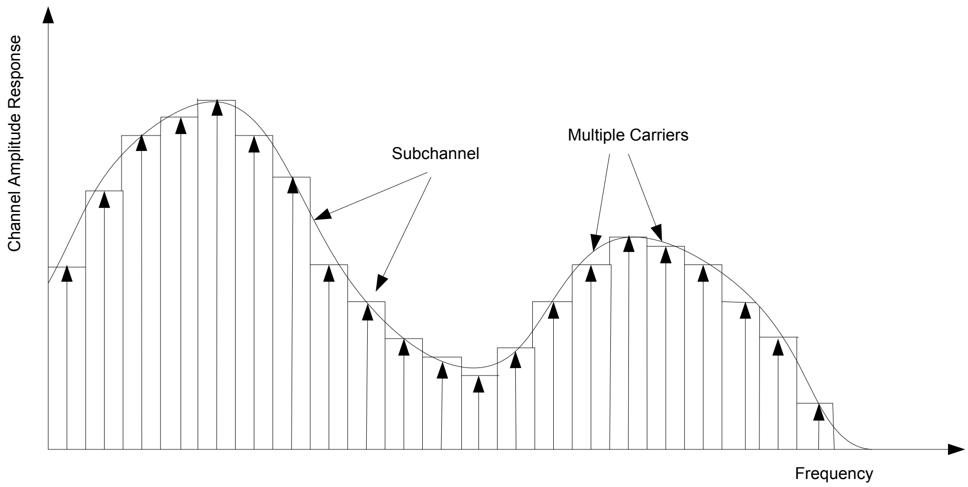

Figure 1.3: Multicarrier Modulation

Multicarrier modulation efficiently utilizes the bandwidth to enable high speed transmission for wireless [24] [65] and wireline [8] communication systems. As the data rate requirements get higher and higher, the transmission bandwidth increases significantly. Consequently, the wireless channel exhibits multipath property in the time domain, or equivalently selectivity [61] in the frequency domain. Successful transmission over a frequency selective channel is more challenging than a narrowband frequency flat channel, as inter-symbol interference degrades the system performance. Advanced signal processing techniques, such as equalization [30] [85], have been proposed to combat the channel dispersion. Multicarrier modulation divides the whole bandwidth into a number of parallel subchannels. As long as the number of subchannels is sufficiently large, the frequency response in each subchannel is close to be flat, as shown in Fig. 1.3. Hence equalization per subchannel is much easier to perform.

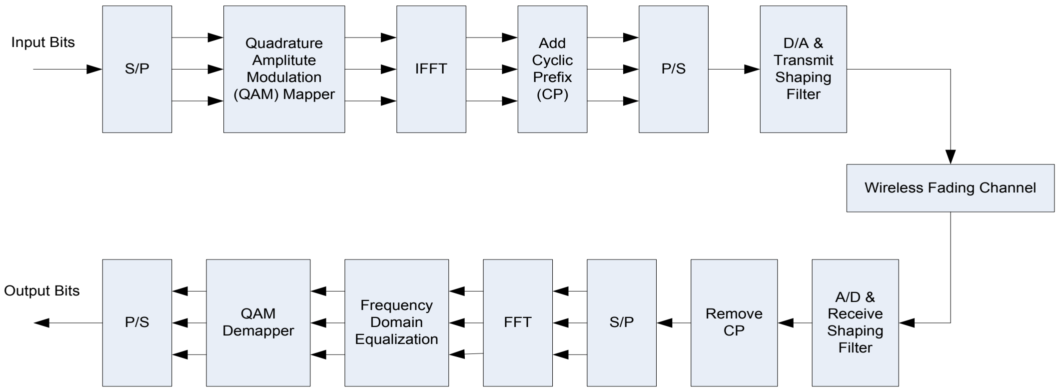

Figure 1.4: OFDM Transceiver Block Diagram

An OFDM transceiver block diagram is shown in Fig. 1.4. The serial input information bits are converted into a number of parallel streams. After quadrature amplitude modulation (QAM), the QAM symbols are fed into an inverse fast Fourier transform (IFFT) block. Parallel to serial conversion is performed subsequently to form an OFDM symbol in the time domain. The last ν IFFT samples are copied to the front of an OFDM symbol as a cyclic prefix (CP). As long as the CP length is greater than the wireless channel dispersion, only 1-tap frequency domain equalization per subchannel is required to retrieve the transmitted data, because the multi-path channel appears to be circular with CP. The rest of receiver blocks essentially invert the operations at the transmitter.

Since OFDM creates multiple parallel subchannels, an OFDM based multiple access technology, namely multiuser OFDM or orthogonal frequency division multiple access (OFDMA), has been proposed [47] [64] [96]. In a multiuser OFDM system, multiple users may be scheduled for transmission on different subchannels

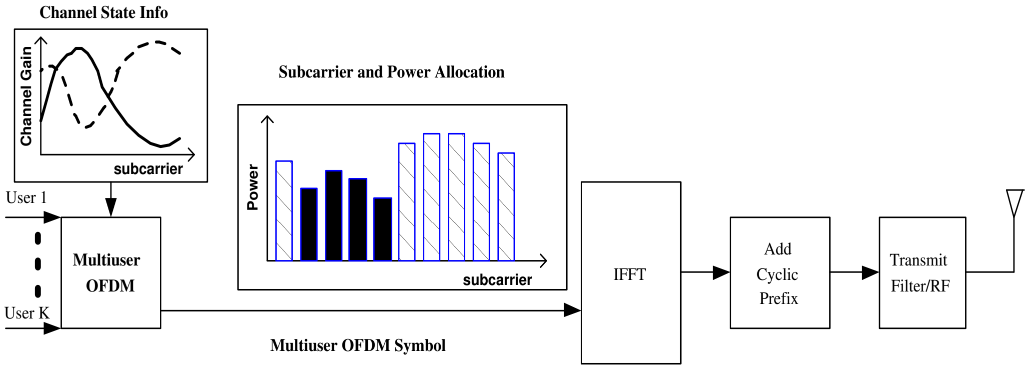

Figure 1.5: Multiuer OFDM Block Diagram

within an OFDM symbol. Further, due to users' different locations and their independent channel fading statistics, the subchannels can be allocated to the users who have good subchannel gains. By exploiting the multiuser diversity, channel-aware adaptive resource allocation outperforms the static resource allocation such as TDMA or FDMA in terms of system throughput. Fig. 1.5 shows a multiuser OFDM system.

Adaptive resource allocation in multiuser OFDM systems can be usually formulated as an optimization problem, e.g. minimizing the total transmit power with user data rate requirements [46] [96] or maximizing the total throughput with a transmit power constraint [36] [48] [56] [72] [108]. The formulated optimization problems are often very difficult to solve and low complexity algorithms have been proposed. Recently, an optimal spectrum management (OSM) algorithm was proposed in [10] in the context of Digital Subscriber Line (DSL) systems. DSL also employs a type of multicarrier modulation named Discrete Multi-Tone (DMT), which is the wireline version of OFDM. Optimal spectrum management aims to maximize a weighted sum capacity by optimally allocation the subchannel and power among users. While the objective, i.e. the weighted sum capacity, is not concave or convex, the algorithm in [10] converts the primal problem into its dual, which is much easier to solve. The algorithm in [10] was further improved into low complexity algorithms in [51] and [105].

Despite the huge amount of work on adaptive resource allocation in multiuser OFDM systems, there still lacks a general optimization framework with which the total throughput and fairness among user (in terms of data rate) can be balanced. For example, although the total throughput can be maximized as in [36] [48], the algorithms may allocate most of the resources to one user while leaving others little. With the weighted sum capacity [10], the fairness issue can be addressed to a certain extent by varying the weights in the objective function, but it is difficult to design the weights so that the fairness can be specifically controlled. A fair multiuser channel allocation algorithm based on Nash bargaining solutions and coalitions has been proposed in [31]. The user data rates, however, cannot be guaranteed to be proportional to each other with the algorithm in [31]. To that end, in this dissertation, I propose an optimization problem to maximize the total throughput while maintaining proportional user data rates exactly. Hence, the fairness among users can be easily determined and designed by a set of parameters.

1.3.2 Multiple Antenna Systems

Multiple-input-multiple-output (MIMO) antenna communication systems have been an intensive research area in the last decade. Equipped with multiple antennas at the transmitter and receiver, MIMO systems fully utilize the spatial dimension to improve the transmission reliability and/or the system throughput.

A point-to-point narrowband MIMO system is shown in Fig. 1.6. In contrast to conventional single antenna systems, the wireless MIMO channel between the communication pair can be represented as a matrix. In a rich-scattering environment without line-of-sight, each element in the MIMO channel matrix can be modeled as a complex Gaussian random variable, resulting from the Central Limit Theorem.

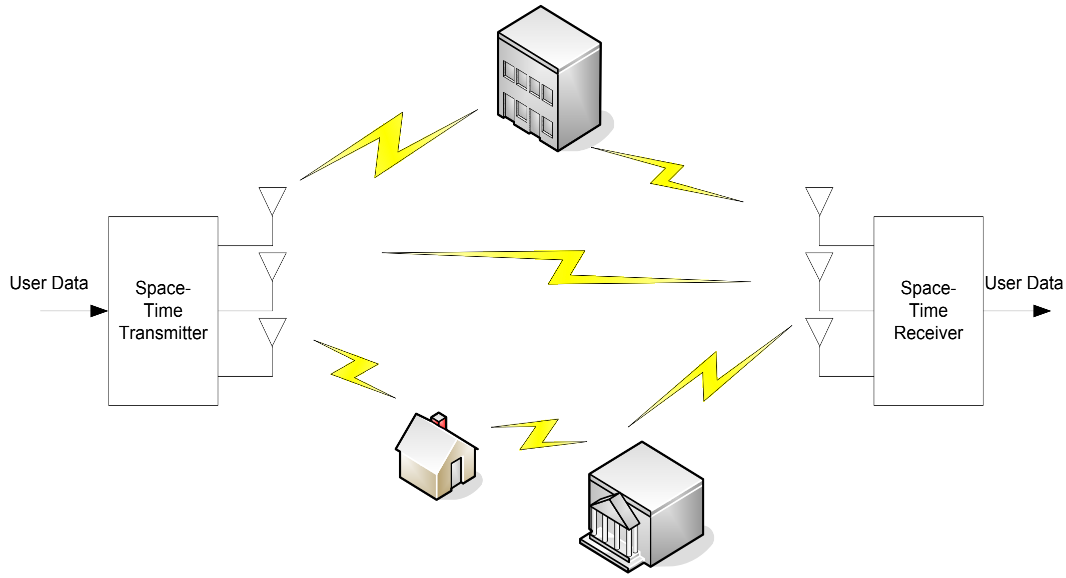

Figure 1.6: A Point to Point MIMO System

This MIMO channel model, the Rayleigh MIMO fading channel, is widely adopted in the literature for system performance evaluations. Other physical and non-physical models can be found in [102].

Due to the time-varying nature of the wireless channel, the signal reception is likely to be very poor when the channel is in deep fading. A common means to combat channel fading is to employ diversity in the communication link. The idea of using multiple receiver antennas to exploit the spatial diversity was proposed decades ago [35]. With optimal combining of the received signals from multiple antennas, the transmission reliability can be significantly improved. Further, with the additional degree of freedom in the spatial domain, multi-antenna systems can even suppress co-channel interference [63] [94]. Later, researchers found that if multiple antennas are both equipped at the transmit and receiver, then a number of parallel channels can be established to increase the spectral efficiency [22] [84] [95]. It was proven in [84] that for point-to-point Rayleigh fading channels, the

MIMO channel capacity scales linearly with the minimum number of transmit and receive antennas in high SNR regime. The results in [84] theoretically show the potential of MIMO systems in spectral efficiency enhancement. With experimental results, the researchers in Bell Laboratories showed that the V-BLAST (Vertical Bell Laboratories Layered Space-Time) architecture [21] [25] can provide a spectral efficiency of tens of bits per second per Hertz. In summary, MIMO technologies provide the diversity and multiplexing opportunities to improve the communication reliability and spectral efficiency [69]. A theoretical study on the tradeoff between diversity and multiplexing of MIMO systems was presented in [110], and a practical algorithm on the switching between diversity and multiplexing was proposed in [42].

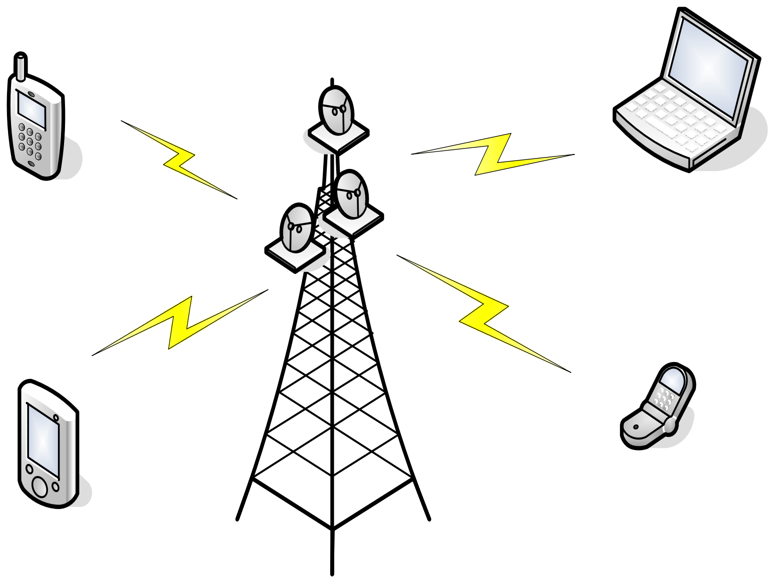

Figure 1.7: A Downlink Multiuser MIMO System

In a downlink multiuser MIMO system, as shown in Fig. 1.7, a basestation is communicating simultaneously to multiple users. Both the basestation and the users are equipped with multiple antennas. The basestation needs to allocate the limited transmit power among users such that an objective function, e.g. the sum capacity, can be achieved. From the information theory point of view, the downlink multiuser MIMO system is named as the MIMO broadcast channel (BC). Although the capacity results for point-to-point MIMO systems are well understood, only recently has the capacity region of the multiuser MIMO Gaussian broadcast channels been discovered. The MIMO BC capacity region is very difficult to obtain since the channels are usually non-degraded. It was conjectured that the MIMO BC capacity region is achieved with dirty paper coding (DPC) [17] and subsequently proven in [93]. Several researchers [9] [88] [89] [103] have established the duality relationship between the MIMO BC capacity region and the MIMO Multiple Access Channel (MAC) capacity region. The sum capacity, which is defined as the maximum aggregation of all users' data rates, can be obtained by iterative water-filling algorithms [40] [106].

Although the sum capacity of a Gaussian MIMO BC channel is achievable with DPC, a practical coding scheme that approaches the DPC sum capacity is still unavailable. Recently, significant efforts have been made, e.g. [2] [82] [107], in designing implementable algorithms to achieve the DPC sum capacity. The proposed algorithms, however, are typically too complicated for cost-effective implementations. An alternative linear precoding technique for downlink multiuser MIMO systems, generally named Block Diagonalization (BD), was proposed in [14] [59] [80] [99] . With BD, each user's data is multiplied by a linear precoding matrix before transmission. The precoding matrix for every user lies in the null space of all other users' channels. Consequently, if the channel matrices of all users are perfectly known at the transmitter, then there is no interference at every receiver, rendering a simple receiver structure. Hence, BD is a potentially realizable precoding method for a MIMO broadcast channel.

Since the user signal covariance matrices of BD are generally not optimal for the sum capacity, it deserves a thorough study on how good BD is relative to the optimal DPC sum capacity. In this dissertation, I analytically compare BD to DPC, for a given set of channels and in Rayleigh fading channels. Further, due to the zero inter-user interference requirement imposed by BD, the maximum number of simultaneously supportable users is limited. I propose two low complexity user selection algorithms with BD to avoid the computationally expensive complete search for the optimal user set.

1.4 Nomenclature

AMPS	Advanced Mobile Phone Systems
AWGN	Additive White Gaussian Noise
BC	Broadcast Channel
BD	Block Diagonalization
BER	Bit Error Rate
CDMA	Code Division Multiple Access
CP	Cyclic Prefix
CSI	Channel State Information
DMT	Discrete Multi-Tone
DPC	Dirty Paper Coding
DSL	Digital Subscriber Line
EDGE	Enhanced Data rates for GSM Evolution
E-TACS	European Total Access Communication System
FDD	Frequency Division Duplex
FDMA	Frequency Division Multiple Access
GPRS	General Packet Radio Service
GSM	Global System for Mobile
GSO	Gram-Schmidt Orthogonalization
HDR	High Data Rate
IFFT	Inverse Fast Fourier Transform

LAN	Local Area Network
MA	Margin Adaptive
MAC	Multiple Access Channel
MCM	Multicarrier Modulation
MIMO	Multiple-input multiple-output
MISO	Multiple-input single-output
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
OSM	Optimal Spectrum Management
QAM	Quadrature Amplitude Modulation
RA	Rate Adaptive
RHS	Right Hand Side
RxAS	Receive Antenna Selection
SDMA	Spatial Division Multiple Access
SNR	Signal-to-noise Ratio
SVD	Singular Value Decomposition
TDD	Time Division Duplex
TDMA	Time Division Multiple Access
V-BLAST	Vertical Bell Laboratories Layered Space-Time
WCDMA	Wideband Code Division Multiple Access
ZF	Zero Forcing

1.5 Assumptions in the Dissertation

- *Perfect channel station information of all users available at the basestation*

  User channel state information is crucial for exploiting multiuser diversity in multiuser wireless communication systems. In this dissertation, I assume users perfectly estimate and feedback their channel information to the basestation. The amount of feedback information increases the system overhead, especially for multi-antenna systems as each user's channel is represented by a matrix. Limited feedback technique [13] [50] or channel prediction [71] [97] can be used to reduced the amount of feedback overhead. The throughput of multiuser systems with imperfect channel state information is still an intensive on-going research area [38].

- *Continuous Shannon channel capacity formula as user throughput measure*

  The Shannon capacity, which is a continuous function, is used as the user throughput in this dissertation. In practical systems, user data rates assume discrete values due to different modulation and coding schemes. The continuous Shannon capacity formula, however, simplifies the analysis of adaptive resource allocation and provides an upper bound on the achievable throughput. A signal-to-noise ratio gap can be included in the Shannon capacity formula to model the signal-to-noise ratio degradation [15] [16]. This gap is widely used in digital subscriber line standards, e.g. [3] [4].

- *Single cell environment*

  In this dissertation, only resource allocation in a single cell is considered. Hence, other-cell interference is not modeled. For users at the cell edges, other-cell interference is not negligible as it greatly impacts the user channel-to-interference-plus-noise ratio. To schedule users in cell edges or in soft handover, either basestation coordination or static frequency planning is required. Several researchers have already discussed resource allocation in multi-cell environment or with inter-user interference, e.g. [104] [108]. Generally, resource allocation in a multi-cell scenario is much more complicated than single cell. The resource allocation algorithms discussed in this dissertation can be applied to users for whom other-cell interference does not dominate the amount of additive white Gaussian noise.

- *Flat power spectrum density mask*

  The total transmit power available in a basestation is usually limited by a power spectrum density mask. Due to the limited frequency bandwidth, multiple standards may co-exist in the same frequency range. To reduce the interference to other systems, the transmit power of every communication system is usually limited by a power spectrum density mask defined in the standards. In this dissertation, we assume a flat power spectrum density mask to simplify the analysis. A non-flat power spectrum density mask can be incorporated into problem formulations by adding different power constraints on different subchannels.

- *Infinitely backlogged user queues*

  The goal of resource allocation discussed in this dissertation is to maximize the throughput given various constraints. The user queues are assumed to be infinitely backlogged. In other words, when one user is scheduled for transmission, he/she always has some information data to transmit. Although the amount of user data is limited in practice, there is always a subset of users who require an opportunity to communicate. Hence, the resource allocation algorithms presented in this dissertation can be applied to those active users.

1.6 Contributions and Organization of the Dissertation

Chapter 2 presents the first contribution of this dissertation: an optimization framework for adaptive resource allocation in multiuser OFDM systems. I impose a set of proportional fairness constraints to assure that each user can achieve a required data rate, as in a system with quality of service guarantees. With the proposed framework, the sum capacity can be distributed fairly and flexibly among users. Since the optimal solution to the constrained fairness problem is extremely computationally complex to obtain, I propose a low-complexity suboptimal algorithm that separates subchannel allocation and power allocation. In the proposed algorithm, subchannel allocation is first performed by assuming an equal power distribution. An optimal power allocation algorithm then maximizes the sum capacity while maintaining proportional fairness. The proposed algorithm is shown to achieve about 95% of the optimal capacity in a two-user system, while reducing the complexity from exponential to linear in the number of subchannels.

The second contribution, presented in Chapter 3, is on the sum capacity of Block Diagonalization with and without receive antenna selection in downlink multiuser MIMO systems. I analytically compare BD to dirty paper coding (DPC), which is optimal for the sum capacity. For a set of given channels, it is shown that 1) if the user channels are orthogonal to each other, then BD achieves the same sum capacity as DPC; 2) if the user channels lie in the same subspace, then the gain of DPC over BD can be up-bounded by the minimum of the number of transmit and receive antennas. I also study the ergodic sum capacity of BD with and without receive antenna selection in a Rayleigh fading channel. Simulations show that BD can achieve a significant part of the total throughput of DPC. An upper bound on the ergodic sum capacity gain of DPC over BD is proposed for easy estimation of the gap between the sum capacity of DPC and BD without receive antenna selection.

The third contribution of this dissertation includes two low complexity user selection algorithms for BD in downlink multiuser MIMO systems, which are presented in Chapter 4. Due to the zero inter-user interference requirement, the number of simultaneously supportable users with BD is limited. In a downlink MIMO system with a large number of users, the basestation may select a subset of users to serve in order to maximize the total throughput. The brute-force search for the optimal user set, however, is computationally prohibitive. Both of the two proposed algorithms aim to select a subset of users such that the total throughput is nearly maximized.

The first user selection algorithm greedily maximizes the total throughput, whereas the criterion of the second algorithm is based on the channel energy. I show that both algorithms have linear complexity in the number of users and achieve around 95% of the total throughput of the complete search method in simulations.

In Chapter 5, I summarize the contributions of this dissertation. Future research topics are also discussed in Chapter 5.

Chapter 2

Adaptive Resource Allocation in Multiuser OFDM Systems with Proportional Rate Constraints

2.1 Introduction

Orthogonal frequency division multiplexing (OFDM) is a promising technique for the next generation of wireless communication systems [62] [65]. OFDM divides the available bandwidth into N orthogonal subchannels. By adding a cyclic prefix (CP) to each OFDM symbol, the channel appears to be circular if the CP length is longer than the channel length. Each subchannel thus can be modeled as a time-varying gain plus additive white Gaussian noise (AWGN). Besides the improved immunity to fast fading [8] brought by the multicarrier property of OFDM systems, multiple access is also possible because the subchannels are orthogonal to each other.

Multiuser OFDM adds multiple access to OFDM by allowing a number of users to share an OFDM symbol. Two classes of resource allocation schemes exist: fixed resource allocation [47] and dynamic resource allocation [36] [45] [64] [96].

Fixed resource allocation schemes, such as time division multiple access (TDMA) and frequency division multiple access (FDMA), assign an independent dimension, e.g. time slot or subchannel, to each user. A fixed resource allocation scheme is not optimal since the scheme is fixed regardless of the current channel condition. On the other hand, dynamic resource allocation allocates a dimension adaptively to the users based on their channel gains. Due to the time-varying nature of the wireless channel, dynamic resource allocation makes full use of multiuser diversity to achieve higher performance.

Two classes of optimization techniques have been proposed in the dynamic multiuser OFDM literature: margin adaptive (MA) [96] and rate adaptive (RA) [36], [64]. The margin adaptive objective is to achieve the minimum overall transmit power given the constraints on the users' data rate or bit error rate (BER). The rate adaptive objective is to maximize each user's error-free capacity with a total transmit power constraint. These optimization problems are nonlinear and hence computationally intensive to solve. In [45], the nonlinear optimization problems were transformed into a linear optimization problem with integer variables. The optimal solution can be achieved by integer programming. However, even with integer programming, the complexity increases exponentially with the number of constraints and variables.

Two rate adaptive optimization problems have been proposed by researchers. Recently, Jang and Lee proposed the rate maximization problem [36]. In [36], they proved that the sum capacity is maximized when each subchannel is assigned to the user with the best subchannel gain and power is then distributed by the water-filling algorithm. However, fairness is not considered in [36]. When the path loss differences among users are large, it is possible that the users with higher average channel gains will be allocated most of the resources, i.e. subchannels and power, for a significant portion of time. The users with lower average channel gains may be unable to receive any data, since most of the time the subchannels will be assigned to users with higher channel gains. In [64], Rhee and Cioffi studied the *max-min* problem, where by maximizing the worst user's capacity, it is assured that all users achieve a similar data rate. However, the *max-min* optimization problem can only provide maximum fairness among the users. In most wireless systems of interest, different users require different data rates, which may be accommodated by allowing users to subscribe to different levels of service.

In [90], Viswanath, Tse, and Laroia discussed long-term proportional fairness resource allocation with "dumb" antennas. They pointed out that in multiuser systems, channel fading can be exploited as a source of randomness, i.e. multiuser diversity. However, in some scenarios, due to the limited scatters in the environment and slow channel variation, the dynamic range of channel fluctuation in the time scale of interest may be small.

Proportionally fair resource allocation has been well-studied in the networking literature, e.g. [43] [44] [55] [66] [81] [86] [111]. In networking literature, the resource allocation is usually formulated as an optimization problem to maximize a certain utility function given the constraints on resources [44]. A vector of rates is said to be proportionally fair if it is feasible and the aggregate of proportional changes between it and any other vector of feasible rates is non-positive [44].

In this chapter, I formulate a new optimization problem that balances the tradeoff between capacity and fairness. The objective function is still the sum capacity, but proportional user data rates are assured by imposing a set of nonlinear constraints into the optimization problem. In contrast to the definition of proportional fairness in [44], which compares the aggregate of proportional changes of two vectors of feasible rates, I incorporate a set of system parameters in the problem formulation such that the ratio of the user data rates strictly follows the set of system parameters after resource allocation. Hence the proportionality in this chapter compares the user data rates to the set of system parameters instead of another feasible set of user data rates as in the networking area. The set of system parameters can be determined in various ways, e.g. users' service applications. Hence, by varying the set of proportional parameters, different service privileges and pricing can be achieved. Further, while large channel fluctuations are intentionally created with "dumb" antennas for long-term proportional fairness resource allocation in [90], the proposed algorithm in this chapter maintains proportional rates among users for each channel realization, which ensures the rates of different users to be proportional in any time scale of interest.

2.2 System Model

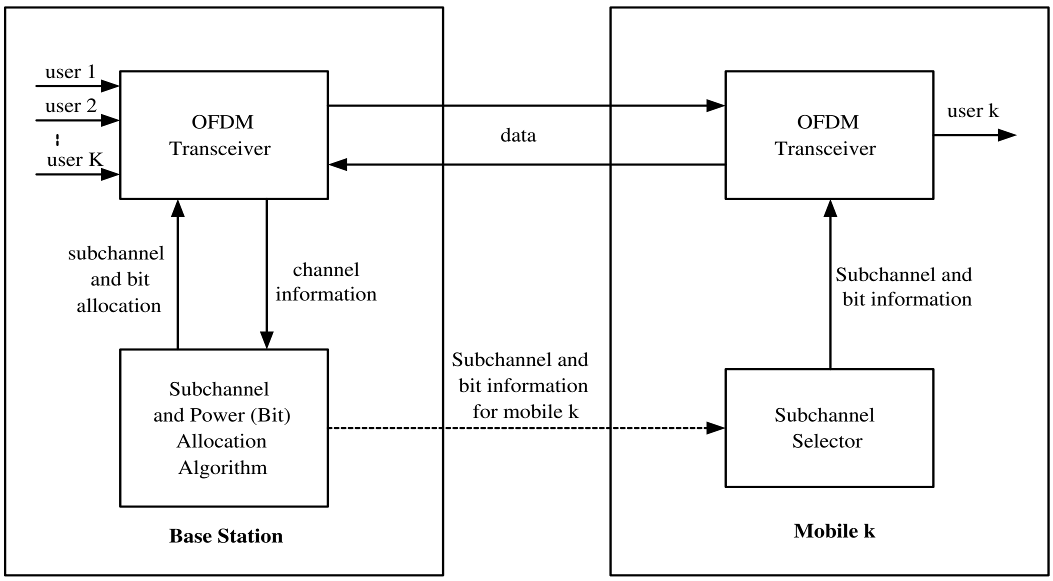

Figure 2.1: Multiuser OFDM System Block Diagram

A multiuser OFDM system is shown in Fig. 2.1. In the basestation, all channel information is sent to the subchannel and power allocation algorithm through feedback channels from all mobile users. The resource allocation scheme made by the algorithm is forwarded to the OFDM transmitter. The transmitter then selects different numbers of bits from different users to form an OFDM symbol. The resource allocation scheme is updated as fast as the channel information is collected. In this chapter, perfect instantaneous channel information is assumed to be available at the basestation and only the broadcast scenario is studied. It is also assumed that the subchannel and bit allocation information is sent to each user by a separate channel.

Throughout this chapter, it is assumed a total of K users in the system sharing N subchannels, with total transmit power constraint P_{total}. The objective is to optimize the subchannel and power allocation in order to achieve the highest sum error-free capacity under the total power constraint. The equally weighted sum capacity is adopted as the objective function, but the idea of proportional fairness is introduced into the system by adding a set of nonlinear constraints. The benefit of the proportional fairness is that the capacity ratios among users can be explicitly controlled to meet each user's target data rate, given sufficient total available transmit power.

Mathematically, the optimization problem considered in this chapter is formulated as

$$\max_{p_{k,n},\rho_{k,n}} \sum_{k=1}^{K}\sum_{n=1}^{N} \frac{\rho_{k,n}}{N} \log_2\left(1+\frac{p_{k,n}h_{k,n}^2}{N_0\frac{B}{N}}\right) \tag{2.1}$$

$$\begin{aligned}
\text{subject to} \quad & \sum_{k=1}^{K}\sum_{n=1}^{N} p_{k,n} \le P_{total} \\
& p_{k,n} \ge 0 \text{ for all } k,n \\
& \rho_{k,n} = \{0,1\} \text{ for all } k,n \\
& \sum_{k=1}^{K} \rho_{k,n} = 1 \text{ for all } n \\
& R_1 : R_2 : \ldots : R_K = \gamma_1 : \gamma_2 : \ldots : \gamma_K
\end{aligned}$$

where K is the total number of users; N is the total number of subchannels; N_0 is the power spectral density of additive white Gaussian noise; B and P_{total} are the total available bandwidth and power, respectively; $p_{k,n}$ is the power allocated for user k in the subchannel n; $h_{k,n}$ is the channel gain for user k in subchannel n; $\rho_{k,n}$ can only be the value of either 1 or 0, indicating whether subchannel n is used by user k or not. The fourth constraint shows that each subchannel can only be used by one user. The capacity for user k, denoted as R_k, is defined as

$$R_k = \sum_{n=1}^{N} \frac{\rho_{k,n}}{N} \log_2 \left(1 + \frac{p_{k,n} h_{k,n}^2}{N_0 \frac{B}{N}} \right). \tag{2.2}$$

Finally, $\{\gamma_i\}_{i=1}^{K}$ is a set of predetermined values which are used to ensure proportional fairness among users.

The fairness index is defined as

$$\mathcal{F} = \frac{\left(\sum\limits_{k=1}^{K} \gamma_k \right)^2}{K \sum\limits_{k=1}^{K} \gamma_k^2} \tag{2.3}$$

with the maximum value of 1 to be the greatest fairness case in which all users would achieve the same data rate. When all γ_i terms are equal, the objective function in (2.1) is similar to the objective function of the *max-min* problem [64], since maximizing the sum capacity while making all R_k terms equal is equivalent to maximizing the worst user's capacity. Hence, [64] is a special case of the proposed constrained-fairness problem.

2.3 Optimal Subchannel Allocation and Power Distribution

The optimization problem in(2.1) is generally very hard to solve. It involves both continuous variables $p_{k,n}$ and binary variables $\rho_{k,n}$. Such an optimization problem is called a mixed binary integer programming problem. Furthermore, the nonlinear constraints in (2.1) increase the difficulty in finding the optimal solution because the feasible set is not convex.

In a system with K users and N subchannels, there are K^N possible subchannel allocations, since it is assumed that no subchannel can be used by more than one user. For a certain subchannel allocation, an optimal power distribution can be used to maximize the sum capacity, while maintaining proportional fairness. The optimal power distribution method is derived in the next section. The maximum capacity over all K^N subchannel allocation schemes is the global maximum and the corresponding subchannel allocation and power distribution is the optimal resource allocation scheme. However, it is prohibitive to find the global optimizer in terms of computational complexity. A suboptimal algorithm is derived in this chapter to reduce the complexity significantly while still delivering performance close to the global optimum.

An alternative approach [36] [64] [96] to make the optimization problem in (2.1) easier to solve is to relax the constraint that subchannels can only be used by one user. Thus $\rho_{k,n}$ is reinterpreted as the sharing factor of user k to subchannel n, which can be any value on the half-open interval of $(0,1]$. The optimization in (2.1) can be transformed into

$$\min_{p_{k,n},\rho_{k,n}} -\sum_{k=1}^{K}\sum_{n=1}^{N}\frac{\rho_{k,n}}{N}\log_2\left(1+\frac{p_{k,n}h_{k,n}^2}{\rho_{k,n}N_0\frac{B}{N}}\right) \tag{2.4}$$

$$\begin{aligned}
\text{subject to}\quad & \sum_{k=1}^{K}\sum_{n=1}^{N} p_{k,n} \leq P_{total} \\
& p_{k,n} \geq 0 \text{ for all } k,n \\
& \rho_{k,n} \in (0,1] \text{ for all } k,n \\
& \sum_{k=1}^{K} \rho_{k,n} = 1 \text{ for all } n \\
& R_1 : R_2 : \ldots : R_K = \gamma_1 : \gamma_2 : \ldots : \gamma_K.
\end{aligned}$$

That is, the original maximization problem is transformed into a minimization problem. In the third constraint in (2.4), $\rho_{k,n}$ is not allowed to be zero since the objective function is not defined for $\rho_{k,n}=0$. However, when $\rho_{k,n}$ is arbitrarily close to 0,

$\frac{\rho_{k,n}}{N} \log_2 \left(1 + \frac{p_{k,n} h_{k,n}^2}{\rho_{k,n} N_0 \frac{B}{N}}\right)$ also approaches 0. Thus, the nature of the objective function remains unchanged by excluding the case $\rho_{k,n} = 0$.

A desirable property of the objective function in (2.4) is that it is convex on the set defined by the first two constraints. The convexity is shown in Appendix A. However, the nonlinear equality constraints make the feasible set non-convex. In general, such optimization problems require linearization of the nonlinear constraints. The linearization procedure may lead the solution slightly off the feasible set defined by the nonlinear constraints. There is always a tradeoff between satisfaction of the constraints and improvement of the objective. Furthermore, it is still computationally complex to find the optimal solution. For these reasons, I propose a suboptimal technique in the next section.

2.4 Suboptimal Subchannel Allocation and Power Distribution

Ideally, subchannels and power should be allocated jointly to achieve the optimal solution in (2.1). However, this poses a prohibitive computational burden at the basestation in order to reach the optimal allocation. Furthermore, the basestation has to rapidly compute the optimal subchannel and power allocation as the wireless channel changes. Hence low-complexity suboptimal algorithms are preferred for cost-effective and delay-sensitive implementations. Separating the subchannel and power allocation is a way to reduce the complexity because the number of variables in the objective function is almost reduced by half. Section 2.4.1 discusses a subchannel allocation scheme. Section 2.4.2 presents the optimal power distribution given a certain subchannel allocation.

2.4.1 Suboptimal Subchannel Allocation

In this section, a suboptimal subchannel algorithm based on [64] is proposed. In the suboptimal subchannel allocation algorithm, equal power distribution is assumed across all subchannels. The channel-to-noise ratio for user k in subchannel n is defined as $H_{k,n} = \frac{h_{k,n}^2}{N_0 \frac{B}{N}}$ and Ω_k is the set of subchannels assigned to user k. The algorithm can be described as

1. Initialization
   set $R_k = 0$, $\Omega_k = \emptyset$ for $k = 1, 2, ..., K$ and $A = \{1, 2, ..., N\}$

2. For $k = 1$ to K

   (a) find n satisfying $| H_{k,n} | \geq | H_{k,j} |$ for all $j \in A$

   (b) let $\Omega_k = \Omega_k \cup \{n\}$, $A = A - \{n\}$ and update R_k according to (2.2)

3. While $A \neq \emptyset$

   (a) find k satisfying $R_k/\gamma_k \leq R_i/\gamma_i$ for all i, $1 \leq i \leq K$

   (b) for the found k, find n satisfying $| H_{k,n} | \geq | H_{k,j} |$ for all $j \in A$

   (c) for the found k and n, let $\Omega_k = \Omega_k \cup \{n\}$, $A = A - \{n\}$ and update R_k according to (2.2)

The principle of the suboptimal subchannel algorithm is for each user to use the subchannels with high channel-to-noise ratio as much as possible. At each iteration, the user with the lowest proportional capacity has the option to pick which subchannel to use. The subchannel allocation algorithm is suboptimal because equal power distribution in all subchannels is assumed. After subchannel allocation, only coarse proportional fairness is achieved. The goal of maximizing the sum capacity while maintaining proportional fairness is achieved by the power allocation in the next section.

2.4.2 Optimal Power Distribution for a Fixed Subchannel Allocation

To a certain determined subchannel allocation, the optimization problem is formulated as

$$\max_{p_{k,n}} \sum_{k=1}^{K} \sum_{n \in \Omega_k} \frac{1}{N} \log_2 \left(1 + \frac{p_{k,n} h_{k,n}^2}{N_0 \frac{B}{N}}\right) \tag{2.5}$$

$$\begin{aligned} \text{subject to} \quad & \sum_{k=1}^{K} \sum_{n \in \Omega_k} p_{k,n} \leq P_{total} \\ & p_{k,n} \geq 0 \text{ for all } k, n \\ & \Omega_k \text{ are disjoint for all } k \\ & \Omega_1 \cup \Omega_2 \cup ... \cup \Omega_K \subseteq \{1, 2, ..., N\} \\ & R_1 : R_2 : ... : R_K = \gamma_1 : \gamma_2 : ... : \gamma_K \end{aligned}$$

where Ω_k is the set of subchannels for user k, and Ω_k and Ω_l are mutually exclusive when $k \neq l$.

The optimization problem in (2.5) is equivalent to finding the maximum of the following cost function

$$\begin{aligned} L \;=\; & \sum_{k=1}^{K} \sum_{n \in \Omega_k} \frac{1}{N} \log_2 (1 + p_{k,n} H_{k,n}) + \lambda_1 \left(\sum_{k=1}^{K} \sum_{n \in \Omega_k} p_{k,n} - P_{total} \right) \\ & + \sum_{k=2}^{K} \lambda_k \left(\sum_{n \in \Omega_1} \frac{1}{N} \log_2 (1 + p_{1,n} H_{1,n}) - \frac{\gamma_1}{\gamma_k} \sum_{n \in \Omega_k} \frac{1}{N} \log_2 (1 + p_{k,n} H_{k,n}) \right) \end{aligned} \tag{2.6}$$

where $\{\lambda_i\}_{i=1}^{K}$ are the Lagrangian multipliers. After differentiating (2.6) with respect to $p_{k,n}$ and setting each derivative to 0, it can be obtained that

$$\frac{\partial L}{\partial p_{1,n}} = \frac{1}{N \ln 2} \frac{H_{1,n}}{1 + H_{1,n} p_{1,n}} + \lambda_1 + \sum_{k=2}^{K} \lambda_k \frac{1}{N \ln 2} \frac{H_{1,n}}{1 + H_{1,n} p_{1,n}} = 0 \tag{2.7}$$

$$\frac{\partial L}{\partial p_{k,n}} = \frac{1}{N \ln 2} \frac{H_{k,n}}{1 + H_{k,n} p_{k,n}} + \lambda_1 - \lambda_k \frac{\gamma_1}{\gamma_k} \frac{1}{N \ln 2} \frac{H_{k,n}}{1 + H_{k,n} p_{k,n}} = 0 \tag{2.8}$$

for $k = 2, 3, ..., K$ and $n \in \Omega_k$.

Power Distribution for a Single User

In this section, the optimal power distribution strategy for a single user k is derived.

From either (2.7) or (2.8), it can obtained that

$$\frac{H_{k,m}}{1 + H_{k,m}p_{k,m}} = \frac{H_{k,n}}{1 + H_{k,n}p_{k,n}} \tag{2.9}$$

for $m, n \in \Omega_k$ and $k = 1, 2, ..., K$. Without loss of generality, I assume that $H_{k,1} \leq H_{k,2} \leq ... \leq H_{k,N_k}$ for $k = 1, 2, ..., K$ and N_k is number of subchannels in Ω_k. Thus, (2.9) can be rewritten as

$$p_{k,n} = p_{k,1} + \frac{H_{k,n} - H_{k,1}}{H_{k,n}H_{k,1}} \tag{2.10}$$

for $n = 1, 2, ..., N_k$ and $k = 1, 2, ..., K$. Equation (2.10) shows that the power distribution for a single user k on subchannel n. More power will be put into the subchannels with higher channel-to-noise ratio. This is the water-filling algorithm [18] in frequency domain.

By defining $P_{k,tot}$ as the total power allocated for user k and using (2.10), $P_{k,tot}$ can be expressed as

$$P_{k,tot} = \sum_{n=1}^{N_k} p_{k,n} = N_k p_{k,1} + \sum_{n=2}^{N_k} \frac{H_{k,n} - H_{k,1}}{H_{k,n}H_{k,1}} \tag{2.11}$$

for $k = 1, 2, ..., K$.

Power Distribution among Users

Once the set $\{P_{k,tot}\}_{k=1}^{K}$ is known, power allocation can be determined by (2.10) and (2.11). The total power constraint and capacity ratio constraints in (2.5) are used to obtain $\{P_{k,tot}\}_{k=1}^{K}$. With (2.9) and (2.11), the capacity ratio constraints can be expressed as

$$\begin{aligned} &\frac{1}{\gamma_1} \cdot \frac{N_1}{N} \left(\log_2 \left(1 + H_{1,1} \frac{P_{1,tot} - V_1}{N_1} \right) + \log_2 W_1 \right) \\ = &\frac{1}{\gamma_k} \cdot \frac{N_k}{N} \left(\log_2 \left(1 + H_{k,1} \frac{P_{k,tot} - V_k}{N_k} \right) + \log_2 W_k \right) \end{aligned} \tag{2.12}$$

for $k = 2, 3, ..., K$, where V_k and W_k are defined as

$$V_k = \sum_{n=2}^{N_k} \frac{H_{k,n} - H_{k,1}}{H_{k,n} H_{k,1}} \tag{2.13}$$

and

$$W_k = \left(\prod_{n=2}^{N_k} \frac{H_{k,n}}{H_{k,1}} \right)^{\frac{1}{N_k}} \tag{2.14}$$

for $k = 1, 2, ..., K$.

Adding the total power constraints

$$\sum_{k=1}^{K} P_{k,tot} = P_{total} \tag{2.15}$$

there are K variables $\{P_{k,tot}\}_{k=1}^{K}$ in the set of K equations in (2.12) and (2.15). Solving the set of functions provides the optimal power allocation scheme. The equations are, in general, nonlinear. Iterative methods, such as the Newton-Raphson or Quasi-Newton methods [5], can be used to obtain the solution, with a certain amount of computational effort. In the Newton-Raphson method, the computational complexity primarily comes from finding the update direction. In Appendix B, the computational complexity of each iteration is shown to be $\mathcal{O}(K)$. Under certain conditions, the optimal or near-optimal solution to the set of nonlinear equations can be found in one iteration. Two special cases are analyzed below.

- Linear Case

  If $N_1 : N_2 : ... : N_K = \gamma_1 : \gamma_2 : ... : \gamma_K$, then the set of equations, i.e. (2.12) and (2.15), can be transformed into a set of linear equations with the following expression

$$\begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & a_{2,2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \dots & a_{K,K} \end{bmatrix} \begin{bmatrix} P_{1,tot} \\ P_{2,tot} \\ \vdots \\ P_{K,tot} \end{bmatrix} = \begin{bmatrix} P_{total} \\ b_2 \\ \vdots \\ b_K \end{bmatrix} \tag{2.16}$$

where

$$a_{k,k} = -\frac{N_1}{N_k}\frac{H_{k,1}W_k}{H_{1,1}W_1} \tag{2.17}$$

$$b_k = \frac{N_1}{H_{1,1}W_1}(W_k - W_1 + \frac{H_{1,1}V_1W_1}{N1} - \frac{H_{k,1}V_kW_k}{N_k}) \tag{2.18}$$

for $k = 2, 3, ..., K$. The matrix of $\{a_{i,i}\}_{i=2}^{K}$ in (2.16) has nonzero elements only on the first row, the first column and the main diagonal. By substitution, the solution to (2.16) can be obtained with a computational complexity of $\mathcal{O}(K)$.

- High Channel-to-Noise Ratio Case

  In adaptive modulation, the linear condition rarely happens and the set of equations remains nonlinear, which requires considerably more computation to solve. However, if the channel-to-noise ratio is high, approximations can be made to simplify the problem.

First consider (2.13), in which V_k could be relatively small compared to $P_{k,tot}$ if the channel-to-noise ratios are high. Furthermore, if adaptive subchannel allocation is used, the best subchannels will be chosen and they have relatively small channel gain differences among them. Thus, the first approximation is $V_k = 0$.

Second, assuming that the basestation could provide a large amount of power and the channel-to-noise ratio is high, the term $H_{k,1}P_{k,tot}/N_k$ is much larger than 1.

With the above two approximations, (2.12) can be rearranged and simplified to be

$$\left(\frac{H_{1,1}W_1}{N_1}\right)^{\frac{N_1}{\gamma_1}} (P_{1,tot})^{\frac{N_1}{\gamma_1}} = \left(\frac{H_{k,1}W_k}{N_k}\right)^{\frac{N_k}{\gamma_k}} (P_{k,tot})^{\frac{N_k}{\gamma_k}} \tag{2.19}$$

where $k = 2, 3, ..., K$.

Substituting (2.19) into (2.15), a single equation with the variable $P_{1,tot}$ can be derived as

$$\sum_{k=1}^{K} c_k (P_{1,tot})^{d_k} - P_{total} = 0 \tag{2.20}$$

where

$$c_k = \begin{cases} 1 & \text{if } k = 1 \\ \dfrac{\left(\frac{H_{1,1}W_1}{N_1}\right)^{\frac{N_1\gamma_k}{N_k\gamma_1}}}{\frac{H_{k,1}W_k}{N_k}} & \text{if } k = 2, 3, ..., K \end{cases} \tag{2.21}$$

and

$$d_k = \begin{cases} 1 & \text{if } k = 1 \\ \frac{N_1\gamma_k}{N_k\gamma_1} & \text{if } k = 2, 3, ..., K. \end{cases} \tag{2.22}$$

Numerical algorithms, such as Newton's root-finding method [1] or the false position method [1], can be applied to find the zero of (2.20).

2.4.3 Existence of Power Allocation Scheme

Solution to Single User Power Allocation

For a certain user k, there is no power allocation if $V_k > P_{k,tot}$. This situation could happen when a subchannel is allocated to a user who does not have a high channel gain in that subchannel. The greedy water-filling algorithm would rather stop using this subchannel. In case this situation happens, the set of Ω_k, as well as the corresponding values of N_k, V_k and W_k, need to be updated and the power allocation algorithm presented in 2.4.2 should be executed again, as shown in Fig. 2.2.

Solution to Multiuser Power Allocation

In case that the channel-to-noise ratio is high, there is one and only one solution to (2.20) since every item in the summation monotonically increases and (2.20) achieves different signs at $P_{1,tot} = 0$ and $P_{1,tot} = P_{total}$. A numerical algorithm can be used to find the solution to (2.20). The complexity of finding the solution will

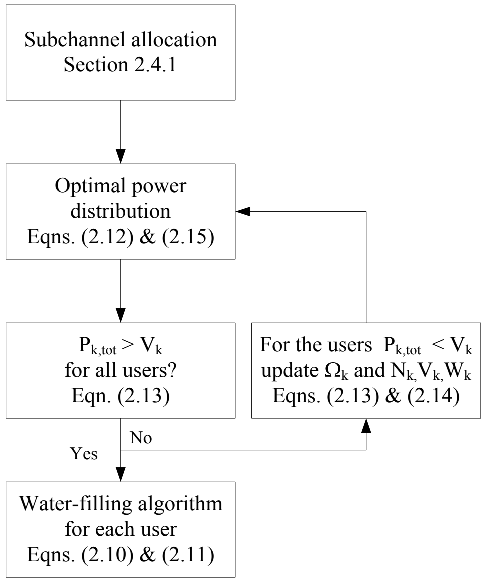

Figure 2.2: Proposed Resource Allocation Algorithm

primarily rely on the choice of the numerical algorithm and the precision required in the results. After $P_{1,tot}$ is found, $\{P_{k,tot}\}_{k=2}^{K}$ can be calculated using (2.19). Then the overall power allocation scheme can be determined by (2.10) and (2.11).

In general, it can be proved that there must be an optimal subchannel and power allocation scheme that satisfies the proportional fairness constraints and the total power constraint. Furthermore, the optimal scheme must utilize all available power. Several facts lead to the above conclusion. First, to a certain user, the capacity of the user is maximized if water-filling algorithm is adopted. Furthermore, the capacity function is continuous with respect to the total available power to that user. In other words, $R_k(P_{k,tot})$ is continuous with $P_{k,tot}$. Second, if the optimal allocation scheme does not use all available transmit power, there is always a way to redistribute the unused power among users while maintaining the capacity ratio constraints, since $R_k(P_{k,tot})$ is continuous with $P_{k,tot}$ for all k. Thus, the sum capacity is further increased. The Newton-Raphson method descried in Appendix B finds $P_{k,tot}$, without considering the constraints on $P_{k,tot}$, i.e. $P_{k,tot} > V_k$ for $k = 1, 2, ..., K$. If the Newton-Raphson method returns a non-feasible $P_{k,tot}$, the set Ω_k and the associated N_k, V_k, and W_k would need to be updated. The Newton-Raphson method should be performed until all $P_{k,tot} > V_k$.

2.4.4 Complexity Analysis

The best subchannel allocation scheme can be found by exhaustive search; i.e., for each subchannel allocation, one would run the optimal power allocation algorithm in Fig. 2.2, which has the computational complexity of $\mathcal{O}(K)$. The subchannel allocation that gives the highest sum capacity is the optimum. In a K-user N-subchannel system, it is prohibitive to find the global optimum since there are K^N possible subchannel allocations. The complexity of the proposed algorithm consists of two parts: subchannel allocation with the complexity of $\mathcal{O}(KN)$ and power allocation of $\mathcal{O}(K)$. Hence the complexity of the proposed method is approximately on the order of K^N times less than that of the optimal, because the power allocation is only executed once. The proposed method is described by Fig. 2.2.

2.5 Numerical Results

In this section, simulation results are presented to show the performance of the proposed resource allocation algorithm. The tradeoff between sum capacity and the fairness constraints is also illustrated.

In all simulations presented in this section, the wireless channel is modeled as a frequency-selective channel consisting of six independent Rayleigh multipaths. Each multipath component is modeled by Clarke's flat fading model [61]. It is assumed that the power delay profile is exponentially decaying with e^{-2l}, where l is the multipath index. Hence, the relative power of the six multipath components are [0, −8.69, −17.37, −26.06, −34.74, −43.43] dB. The total available bandwidth and transmit power are 1 MHz and 1 W, respectively.

2.5.1 A System with Two Users and Ten Subchannels

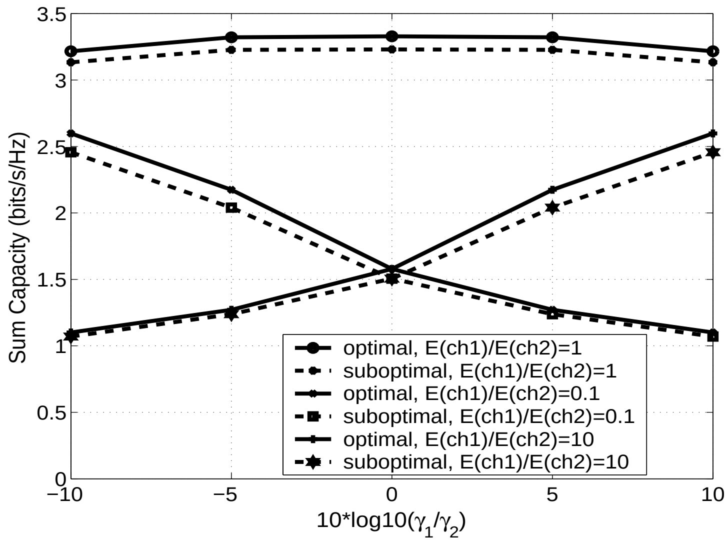

Figure 2.3: Optimal vs. suboptimal adaptive resource allocation in a two-user ten-subchannel system.

Fig. 2.3 shows the sum capacity of a two-user ten-subchannel system vs. γ_1/γ_2, from which the fairness index as defined in (2.3) can be calculated. The AWGN power spectrum density is -70 dBW/Hz. Both the suboptimal results and the optimal results are plotted. A small number of users and subchannels are used in order to reduce the time to find the optimal solution. The sum capacities shown in Fig. 2.3 are averaged over 200 channel realizations. Fig. 2.3 shows that the sum capacity is not very sensitive to the fairness constraint ratio γ_1/γ_2 when there is no path loss difference between the two users. However, when there exists path loss difference, e.g. 10 dB, the sum capacity varies greatly with the fairness constraint ratio. For example, when the averaged channel power of user 1, denoted as $E(\text{ch}_1)$, is 10 dB higher than average channel power of user 2, denoted as $E(\mathrm{ch}_2)$, the sum capacity reduces as γ_1/γ_2 decreases. The reason is as γ_1/γ_2 decreases, more priority is assigned to user 2. Hence user 2 will be assigned most of the available resources, i.e. power and bandwidth, which consequently lowers the sum capacity since the average channel power of user 2 is 10 dB lower than user 1.

From Fig. 2.3, the proposed method achieves about 95% of the optimal performance in a two-user ten-subchannel system. Although in a real cellular or wireless LAN system, the number of users and subchannels is much larger, it is still expected the proposed method to perform close to the optimum because the subchannel allocation algorithm is designed to utilize the subchannels with large channel-to-noise ratio as much as possible, and the power distribution is always optimal for any determined subchannel allocation.

2.5.2 Comparison with Maximum Fairness

The objective in [64] is to maximize the minimum user's capacity. By setting $\gamma_1 : \gamma_2 : ... : \gamma_K = 1 : 1 : ... : 1$, the objective of the optimization problem in (2.1) is identical to the one in [64], since the worst user's capacity is maximized when all users have the same capacity and the sum capacity is maximized. Hence, the problem in [64] is a special case of the framework presented in this chapter. In this section of simulations, the worst user's capacity is compared. In [64], a suboptimal algorithm is proposed to achieve near-optimal capacity using adaptive subchannel allocation, but an equal power distribution is assumed. When the number of users increases, the equal power distribution does not equalize every user's capacity. By transferring power from the users with high capacity to the users with low capacity, the worst user's capacity could be even increased. For the purpose of comparison, I use the suboptimal algorithm in [64], which is a special case of the subchannel allocation algorithm in 2.4.1, to allocate the subchannels first and then apply the optimal power allocation scheme proposed in 2.4.2. Both of these adaptive schemes are compared with the fixed time division multiple access (TDMA) resource allocation scheme.

The wireless channel is modeled as before, and the total transmit power available at the basestation is 1 W. The power spectral density of additive white Gaussian noise is -80 dBW/Hz, and the total bandwidth is 1 MHz, which is divided into 64 subchannels. The maximum path loss difference is 40 dB, and the user locations are assumed to be uniformly distributed. In this part of the simulation, the subchannel SNR is high, hence the power allocation algorithm can be reduced to the high channel-to-noise case discussed in section 2.4.2.

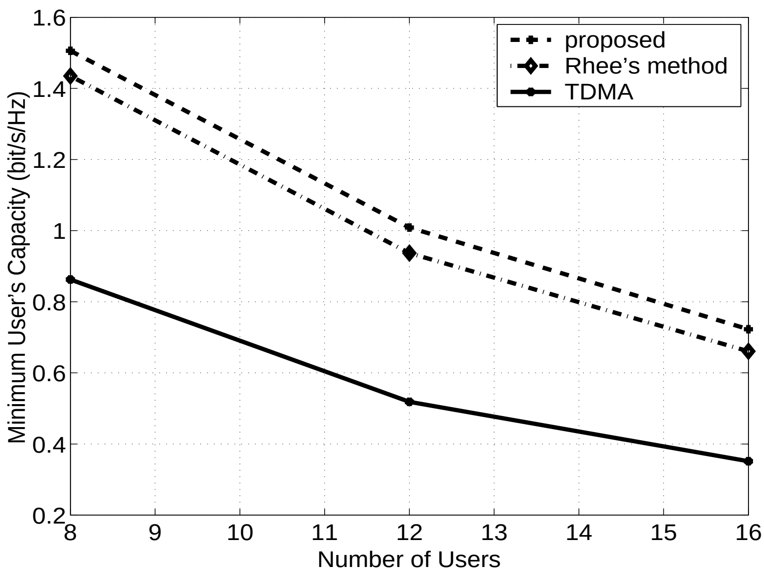

Figure 2.4: Minimum user's capacity vs. number of users.

Fig. 2.4 shows the capacity vs. number of users in the OFDM system. From Fig. 2.4, adaptive resource allocation can achieve significant capacity gain over non-adaptive TDMA. Also the adaptive scheme with optimal power allocation achieves even higher capacity than the scheme with equal power distribution. Notice that this capacity gain is purely from the optimal power allocation algorithm, since both adaptive resource allocation algorithms adopt the same subchannel allocation. Further, Fig. 2.4 shows that the capacity gain over TDMA increases when the number of users increases. This can be explained by multiuser diversity: the more users in the system, the lower the probability that a given subchannel is in a deep fade for all users. In a system of 16 users, the adaptive scheme with the proposed optimal power allocation achieves 17% more capacity gain than the scheme with equal power distribution, when compared to fixed TDMA.

2.5.3 Comparison with Maximum Total Throughput

In this section, I compare the sum capacity achieved by the proposed algorithm with the method in [36]. The simulation parameters are the same as the previous section, i.e. the total available bandwidth is 1 MHz, the total transmit power at basestation is 1W, the AWGN power density is -80 dBW/Hz, and the number of subchannels is $N = 64$.

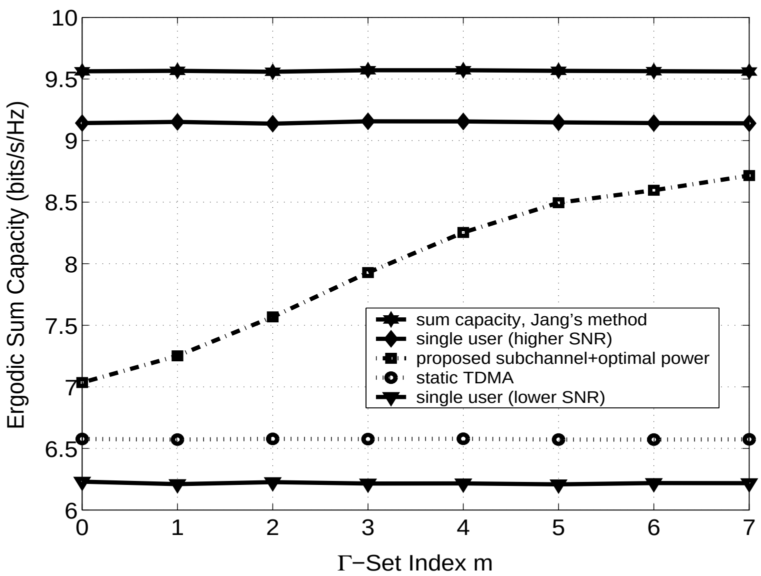

Figure 2.5: Ergodic sum capacity of an 8-user OFDM system vs. various Γ-sets.

Fig. 2.5 shows that the sum capacity of the proposed resource allocation algorithm in an eight-user OFDM system vs. different fairness constraints, which are defined in Table 2.1. The average channel power of user 2 to user 8 are the same, while the average channel power of user 1 is 10 dB higher than the other seven users. In Fig. 2.5, I also include

- the sum capacity achieved by the method in [36],
- the capacity achieved by a static TDMA system, in which each user is allocated an equal share of time slots and equal transmit power, and
- the capacities of two types of single user systems, one for the user with high average channel power, and the other for those with low average channel power.

All sum capacities shown in Fig. 2.5 are ergodic capacities averaged over 5×10^4 channel realizations. It can be seen that the sum capacity maximization method in [36] achieves the maximum sum capacity, because all resources are allocated to the users with the best channel gains. The capacity achieved by the proposed algorithm varies as the rate constraint changes. As more priority is allocated to user 1, i.e. as the Γ-set index increases, higher sum capacity is achieved. This is reasonable since user 1 has higher average channel gain and hence can more efficiently utilize the resources.

Fig. 2.6 shows the normalized ergodic capacity distribution among users for Γ-set index 3 in Table 2.1, where $\gamma_1 = 8$ and $\gamma_2 = \gamma_2 = \cdots = \gamma_8 = 1$. With the proposed subchannel and power allocation algorithm, the capacity is distributed very well among users according to the rate constraints. However, for the capacity maximization method in [36], user 1 gets most of the resources and hence achieves a significant part of the sum capacity. Static TDMA tends to allocate similar capacity to each user, since all users get the same opportunity to transmit. Notice that the capacity distribution of the method in [36] and static TDMA cannot be changed

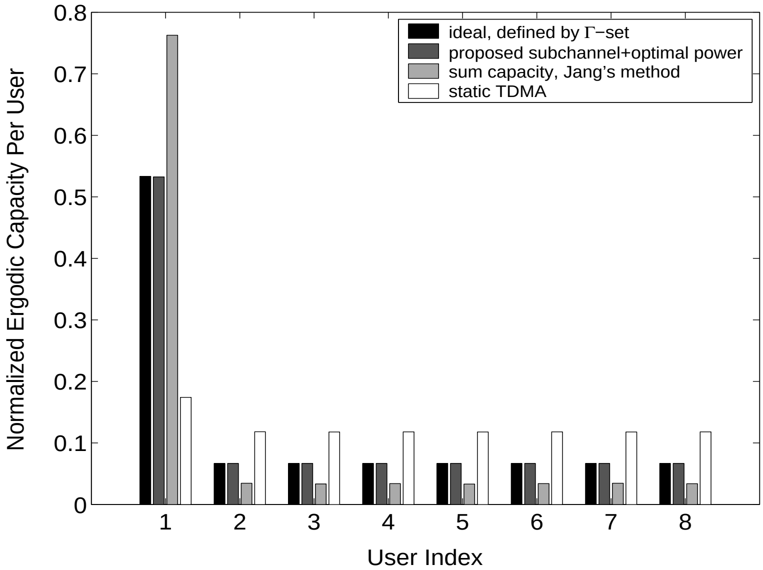

Figure 2.6: Normalized Ergodic sum capacity distribution among 8 users.

by varying the Γ-set values, because there is no fairness control mechanism in these systems.

Since the problem formulation in (2.1) is to allocate resources to satisfy the rate constraints strictly for each channel realization, I define a quantity to measure how well the rate constraints are satisfied. Let $R_{k,i}$ be the capacity of user k for a certain channel realization i, $\widetilde{R}_{k,i} = \frac{R_{k,i}}{\sum_{k=1}^{K} R_{k,i}}$ be the normalized capacity for user k,

Table 2.1: Rate Constraints (Γ-sets) and Rate Constraint Deviations for Figs. 2.5 & 2.6, $K = 8$

Γ-Set Index m	0	1	2	3	4	5	6	7
$\gamma_1 = 2^m$	2^0	2^1	2^2	2^3	2^4	2^5	2^6	2^7
$\gamma_2 = \cdots = \gamma_8$	1	1	1	1	1	1	1	1
$\overline{\mathcal{D}}$, proposed algorithm	0.0026	0.0024	0.0020	0.0015	0.0012	0.0010	0.0013	0.0012
$\overline{\mathcal{D}}$, sum capacity [36]	0.8848	0.7825	0.6441	0.5004	0.3878	0.3216	0.2902	0.2751
$\overline{\mathcal{D}}$, static TDMA	0.1118	0.1114	0.2247	0.3867	0.5453	0.6633	0.7377	0.7799

Table 2.2: Rate Constraints (Γ-sets) and Rate Constraint Deviations for Figs. 2.7 & 2.8, $K = 16$

Γ-Set Index m	0	1	2	3	4
$\gamma_1 = \cdots = \gamma_4 = 2^m$	2^0	2^1	2^2	2^3	2^4
$\gamma_5 = \cdots = \gamma_{16}$	1	1	1	1	1
$\overline{\mathcal{D}}$, proposed algorithm	0.0015	0.0015	0.0013	0.0012	0.0018
$\overline{\mathcal{D}}$, sum capacity [36]	0.9238	0.8361	0.7438	0.6662	0.6133
$\overline{\mathcal{D}}$,static TDMA	0.1150	0.1093	0.2548	0.4071	0.5193

and $\widetilde{\gamma}_k = \frac{\gamma_k}{\sum_{k=1}^{K}\gamma_k}$ be the normalized rate constraint. The normalized rate constraint deviation measure for channel realization i is defined as

$$\mathcal{D}_i = \frac{\sum\limits_{k=1}^{K} |\widetilde{R}_{k,i} - \widetilde{\gamma}_k|}{\max\limits_{\widetilde{R}_{k,i}} \sum\limits_{k=1}^{K} |\widetilde{R}_{k,i} - \widetilde{\gamma}_k|}. \tag{2.23}$$

Notice that the denominator in (2.23) refers to the maximum deviation over all possible $\widetilde{R}_{k,i}$ values. It is shown in Appendix C that

$$\max_{\widetilde{R}_{k,i}} \sum_{k=1}^{K} |\widetilde{R}_{k,i} - \widetilde{\gamma}_k| = 2 - 2\min_k \widetilde{\gamma}_k. \tag{2.24}$$

Table 2.1 shows the averaged rate constraint deviations, denoted as $\overline{\mathcal{D}} = \sum_{i=1}^{I} \mathcal{D}_i / I$, where I is the total number of channel realizations of the eight-user OFDM system. The rate constraint deviation of the proposed subchannel and power allocation is orders of magnitude smaller than those achieved by the method in [36] and the static TDMA. In other words, the price of maximizing ergodic capacity is that the short-term data rates vary widely and users may have poor performance over a certain block of time.

Fig. 2.7 shows the ergodic sum capacities in a multiuser OFDM systems with 16 users. The simulation parameters are the same as those in the previous 8-user system. The average channel power of the first four user are 10 dB higher than the rest of twelve users. The Γ-set index and the rate constraint deviations are shown in

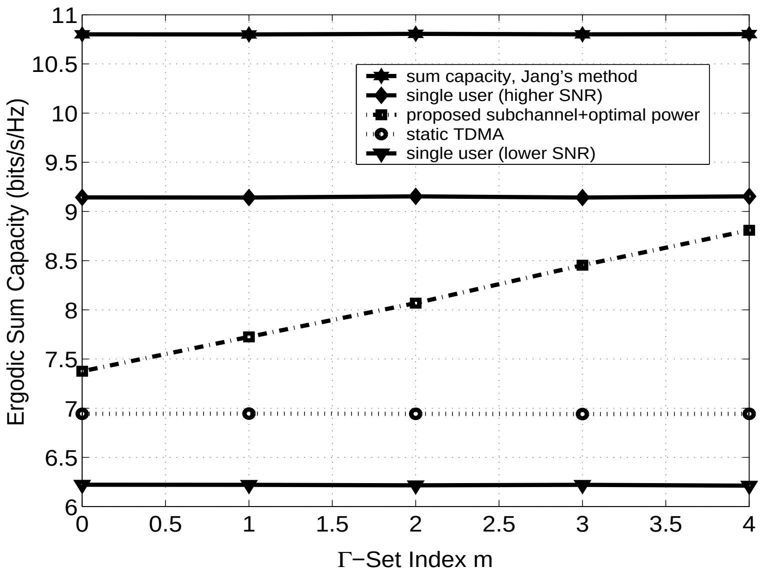

Figure 2.7: Ergodic sum capacity of a 16-user OFDM system vs. various Γ-sets.

Table 2.2. Fig. 2.8 shows the normalized average sum capacity distribution among users with $\gamma_1 = \cdots = \gamma_4 = 8$ and $\gamma_5 = \cdots = \gamma_{16} = 1$. It should be noted that higher sum capacity is achieved by the method in [36] in this 16-user OFDM system since more users with high channel power are in this system, hence more multiuser diversity. However, it can be seen from Fig. 2.8 that the users with lower average channel power, i.e. users 5-16, get very small portions of the sum capacity, since in most channel realizations, the subchannels and power are allocated to the users with larger subchannel gains.

2.6 Conclusion

In this chapter, I presented a resource allocation framework in multiuser OFDM systems to achieve variable proportional rate constraints. For different rate constraints, i.e. $\{\gamma_k\}_{k=1}^{K}$, different proportional rates can be achieved among users.

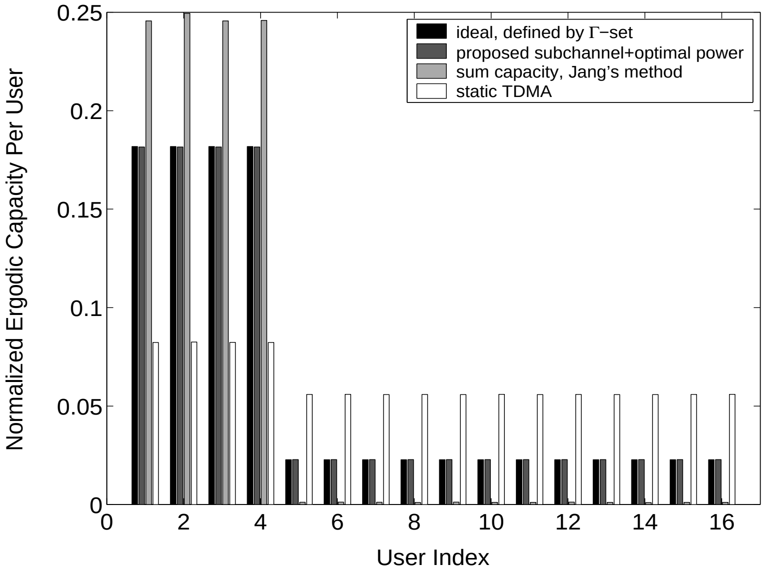

Figure 2.8: Normalized Ergodic sum capacity distribution among 16 users.

The term "variable" refers to the facts that the rate constraints can be configured at the basestation and hence rate allocation among users is flexible.

The proposed optimization problem considers maximizing the sum capacity while maintaining proportional fairness among users for each channel realization. The algorithm to find the optimal solution is discussed, and a low complexity suboptimal algorithm, which reduces complexity from $\mathcal{O}(K^N)$ to $\mathcal{O}(KN)$, is also proposed. In the suboptimal algorithm, subchannel and power allocation are carried out separately. The optimal power allocation to a determined subchannel scheme is developed. A two-step procedure may be taken to get the optimal power distribution. First, a set of nonlinear equations has to be solved in order to get the power distribution among users. Then to a particular user, the greedy water-filling algorithm is adopted to maximize the capacity. The existence of power allocation is also discussed.

Simulation results show that the suboptimal algorithm can achieve above

95% of the optimal performance in a two-user system. Simulation results also shows that in a system of 16 users, the proposed optimal power allocation achieves 17% more capacity over fixed TDMA than the *max-min* method in [64]. It is also shown that with the proposed resource allocation algorithm, the sum capacity is distributed more fairly among users than the sum capacity maximization algorithm in [36].

Chapter 3

Sum Capacity of Multiuser MIMO Broadcast Channels with Block Diagonalization

3.1 Introduction

Multiple-input multiple-output (MIMO) systems have drawn a lot of attention in the last decade. The spatial dimension, in addition to the time and frequency dimensions, can be exploited with multiple antennas at the transmitter and receiver. A pioneering paper on point-to-point MIMO channel capacity is due to Telatar [84], followed by other papers, e.g. [7] [12] [27] [37] [49] [60] [78] [87]. In [84], Telatar showed that for Rayleigh fading channels, the MIMO channel capacity scales linearly with the minimum number of transmit and receive antennas.

The sum capacity of a Gaussian MIMO BC channel is achievable with DPC [88]. A practical coding scheme that approaches the DPC sum capacity, however, is still unavailable. Several nonlinear or linear algorithms have been proposed in [2] [82] [107]. These algorithms, however, are typically too complicated for cost-effective implementations. An alternative linear precoding technique for downlink multiuser MIMO systems is Block Diagonalization (BD) [14] [59] [80] [99]. With BD, each user's data is multiplied by a linear precoding matrix before transmission. The precoding matrix for every user lies in the null space of all other users' channels. Consequently, with perfect channel state information at the basestation, each user sees no inter-user interference, rendering a simple receiver structure. Hence, BD is a potentially realizable precoding method for a MIMO broadcast channel, although it is suboptimal as far as the sum capacity.

The sum capacity gain of DPC vs. TDMA for MIMO broadcast channels has been studied in [39] [67]. In TDMA, the transmitter only sends data to the user with the largest channel capacity. It has been proven in [39] that for a set of given channels, for any number of users, any number of transmit and receive antennas, and any SNR, the ratio of the DPC sum capacity over TDMA can be up-bounded by the minimum of the number of users and the number of transmit antennas. In this chapter, I focus on the sum capacity gain of DPC over BD. BD's sum capacity is defined to be the maximum total throughput over all possible user sets. Hence the TDMA sum capacity, where no precoding is employed, is automatically incorporated in BD's sum capacity definition. Therefore, the general bound on the gain of DPC vs. TDMA applies to the gain of DPC vs. BD.

In this chapter, I analyze the sum capacity of BD with receive antenna selection. The sum capacity gain of DPC over BD without receive antenna selection for a set of given channels is presented. I show that 1) if user channels are orthogonal to each other, then BD achieves the same sum capacity as DPC; 2) if user channels lie in the same subspace, then the gain of DPC over BD can be reduced to the minimum of the number of transmit and receive antennas. These statements also hold for BD with receive antenna selection. Further, the ergodic sum capacity of DPC is compared to that of BD in a Rayleigh fading channel. I propose an upper bound on the ergodic sum capacity gain of DPC over BD. The proposed upper bound on the gain can be evaluated with a few numerical integrations.

3.2 System Model and Background on Block Diagonalization

In this section, I introduce the system model and briefly describe the generalized block diagonalization method for multiuser MIMO systems presented in [59].

Consider a downlink multiuser MIMO system with K users. Let us denote the number of transmit antennas at the basestation as N_t and the number of receive antennas for the jth user as $N_{r,j}$. It is assumed that $N_{r,j} \leq N_t$ for all $j = 1, 2, \cdots, K$ in this chapter. The transmitted symbol of user j is denoted as a N_j-dimensional vector $\mathbf{x}_j$, which is multiplied by a $N_t \times N_j$ precoding matrix $\mathbf{T}_j$. At receiver j, a $M_j \times N_{r,j}$ $(M_j \leq N_{r,j})$ matrix $\mathbf{R}_j$ is applied to the received signals from all receive antennas. The purpose of the post-processing matrices $\{\mathbf{R}_j\}_{j=1}^{K}$ is to form a set of better effective user channels to improve the sum capacity. Hence, the post-processed received signal $\mathbf{y}_j$ for user j can be represented as

$$\begin{aligned} \mathbf{y}_j &= \mathbf{R}_j \left(\mathbf{H}_j \mathbf{T}_j \mathbf{x}_j + \sum_{k=1, k\neq j}^{K} \mathbf{H}_j \mathbf{T}_k \mathbf{x}_k + \mathbf{v}_j \right) \\ &= \mathbf{R}_j \mathbf{H}_j \mathbf{T}_j \mathbf{x}_j + \mathbf{R}_j \sum_{k=1, k\neq j}^{K} \mathbf{H}_j \mathbf{T}_k \mathbf{x}_k + \mathbf{R}_j \mathbf{v}_j \end{aligned} \tag{3.1}$$

where the first item in the right-hand-side (RHS) of (3.1) is the desired signal for user j; the second item in the RHS of (3.1) is the interference seen by user j from the other users' signals; and $\mathbf{v}_j$ denotes the additive white Gaussian noise (AWGN) vector for user j with variance $E[\mathbf{v}_j \mathbf{v}_j^*] = \sigma^2 \mathbf{I}$, where $()^*$ denotes the matrix conjugate transpose. Matrix $\mathbf{H}_j \in \mathbb{C}^{N_{r,j} \times N_t}$ denotes the channel transfer matrix from the basestation to the jth user, with each entry following an i.i.d. complex Gaussian distribution $\mathcal{CN}(0, 1)$. For analytical simplicity, I assume that $\text{rank}(\mathbf{H}_j) = \min(N_{r,j}, N_t)$ for all users. It is also assumed that the channels $\mathbf{H}_j$ experienced by different users are statistically independent due to the users' different locations.

The key idea of block diagonalization is to design $\mathbf{T}_j$ and $\mathbf{R}_j$, such that

$$\begin{aligned}
&\mathbf{T}_j \in \mathbb{U}(N_t, N_j) \\
&\mathbf{R}_j^T \in \mathbb{U}(N_{r,j}, M_j) \\
&\mathbf{R}_i \mathbf{H}_i \mathbf{T}_j = 0 \quad \text{for all } i \neq j \text{ and } 1 \leq i, j \leq K,
\end{aligned} \tag{3.2}$$

where $()^T$ denotes the matrix transpose and $\mathbb{U}(n, k)$ represents the set of $n \times k$ $(n \geq k)$ matrices with orthonormal columns, i.e. for any $\mathbf{U} \in \mathbb{U}(n, k)$, it holds that $\mathbf{U}^* \mathbf{U} = \mathbf{I}_k$ where $\mathbf{I}_k$ is the identity matrix of size $k \times k$. The constraint that $\mathbf{T}_j \in \mathbb{U}(N_t, N_j)$ ensures that the total transmit power is not changed; the constraint that $\mathbf{R}_j \in \mathbb{U}(N_{r,j}, M_j)$ avoids noise enhancement after applying $\mathbf{R}_j$ to the received signals; and the last constraint in (3.2) eliminates inter-user interference.

Hence with precoding matrices $\mathbf{T}_j$ and post-processing matrix $\mathbf{R}_j$, the post-processed received signal for user j can be simplified to

$$\begin{aligned}
\mathbf{y}_j &= \mathbf{R}_j \mathbf{H}_j \mathbf{T}_j \mathbf{x}_j + \mathbf{R}_j \sum_{k=1, k \neq j}^{K} \mathbf{H}_j \mathbf{T}_k \mathbf{x}_k + \mathbf{R}_j \mathbf{v}_j \\
&= \mathbf{R}_j \mathbf{H}_j \mathbf{T}_j \mathbf{x}_j + \mathbf{R}_j \mathbf{v}_j.
\end{aligned} \tag{3.3}$$

For a fixed set of $\{\mathbf{R}_j\}_{j=1}^K$, let

$$\widetilde{\mathbf{H}}_j = [(\mathbf{R}_1 \mathbf{H}_1)^T \ \cdots \ (\mathbf{R}_{j-1} \mathbf{H}_{j-1})^T \ (\mathbf{R}_{j+1} \mathbf{H}_{j+1})^T \ \cdots \ (\mathbf{R}_K \mathbf{H}_K)^T]^T. \tag{3.4}$$

To satisfy the constraint in (3.2), $\mathbf{T}_j$ shall be in the null space of $\widetilde{\mathbf{H}}_j$. Let $\widetilde{N}_j$ denote the rank of $\widetilde{\mathbf{H}}_j$. Let the singular value decomposition (SVD) of $\widetilde{\mathbf{H}}_j$ be

$$\widetilde{\mathbf{H}}_j = \widetilde{\mathbf{U}}_j \widetilde{\mathbf{\Lambda}}_j [\widetilde{\mathbf{V}}_{j,1} \ \widetilde{\mathbf{V}}_{j,0}]^* \tag{3.5}$$

where $\widetilde{\mathbf{U}}_j \in \mathbb{U}(\widetilde{N}_j, \widetilde{N}_j)$, $\widetilde{\mathbf{\Lambda}}_j$ is a diagonal matrix of size $\widetilde{N}_j \times N_t$; $\widetilde{\mathbf{V}}_{j,1} \in \mathbb{U}(N_t, \widetilde{N}_j)$ contains the first $\widetilde{N}_j$ right singular vectors and $\widetilde{\mathbf{V}}_{j,0} \in \mathbb{U}(N_t, N_t - \widetilde{N}_j)$ contains the last $N_t - \widetilde{N}_j$ right singular vectors of $\widetilde{\mathbf{H}}_j$. The columns in $\widetilde{\mathbf{V}}_{j,0}$ form a set of basis in the null space of $\widetilde{\mathbf{H}}_j$, and hence the columns in $\mathbf{T}_j$ are linear combinations of those in $\widetilde{\mathbf{V}}_{j,0}$. In fact, $\mathbf{T}_j$ can be any set of basis in the null space of $\widetilde{\mathbf{H}}_j$.

Due to the zero inter-user interference requirement, the number of data streams for each user is limited in a block diagonalization system. The following Lemma shows the maximum number of possible data streams for a user.

Lemma 1: For a fixed set of $\{\mathbf{R}_j\}_{j=1}^K$, let $\mathbf{H} = [(\mathbf{R}_1\mathbf{H}_1)^T \ \cdots \ (\mathbf{R}_K\mathbf{H}_K)^T]^T$ and N be the rank of $\mathbf{H}$. To satisfy the zero-interference constraint (3.2), the number of possible independent data streams of user j, denoted $\overline{N}_j$ ($\overline{N}_j \leq N_j$), must satisfy

$$\overline{N}_j \leq \min\{N - \widetilde{N}_j, M_j\}. \tag{3.6}$$

Proof: A detailed proof can be found in [11] [59] [80]. □

In the rest of the chapter, for notational and analytic simplicity, it is assumed that every user has the same number of receive antennas, i.e. $N_{r,k} = N_r$ for $k = 1, 2, \cdots, K$. The results in this chapter can be easily extended to the case where different users have different numbers of receive antennas.

3.3 Sum Capacity of Block Diagonalization with Receiver Antenna Selection

Consider a given set of channel realizations for a multiuser MIMO system, where $\mathbf{H}_j$ denotes the channel for user j. Notice that the precoding matrices $\{\mathbf{T}_j\}_{j=1}^K$ can be determined based on $\{\mathbf{H}_j\}_{j=1}^K$ and $\{\mathbf{R}_j\}_{j=1}^K$, i.e. $\mathbf{T}_j$ can be any set of basis in the null space of $\widetilde{\mathbf{H}}_j = [(\mathbf{R}_1\mathbf{H}_1)^T \ \cdots \ (\mathbf{R}_{j-1}\mathbf{H}_{j-1})^T \ (\mathbf{R}_{j+1}\mathbf{H}_{j+1})^T \ \cdots \ (\mathbf{R}_K\mathbf{H}_K)^T]^T$. Ideally, the sum capacity can be obtained by jointly optimizing $\{\mathbf{R}_j\}_{j=1}^K$ and the users' transmit signal covariance matrices $\{\mathbf{Q}_j\}_{j=1}^K$ in the following problem

$$\max_{M_j, \mathbf{R}_j, \mathbf{Q}_j} \sum_{j=1}^{K} \log \left| \mathbf{I} + \frac{1}{\sigma^2} \mathbf{R}_j \mathbf{H}_j \mathbf{T}_j \mathbf{Q}_j \mathbf{T}_j^* \mathbf{H}_j^* \mathbf{R}_j^* \right| \tag{3.7}$$

$$\begin{aligned} \text{subject to} \quad & \mathbf{R}_j^T \in \mathbb{U}(N_r, M_j) \quad \text{for all } j \\ & \mathbf{R}_i \mathbf{H}_i \mathbf{T}_j = 0 \quad \text{for all } i \neq j \\ & 0 \leq M_j \leq N_r \quad \text{for all } j \\ & \sum_{j=1}^{K} \mathrm{Tr}(\mathbf{Q}_j) \leq P \\ & \mathbf{Q}_j \geq 0 \quad \text{for all } j \end{aligned}$$

where $\mathbf{R}_j \mathbf{H}_j \mathbf{T}_j$ denotes the effective channel for user j, $\mathbf{Q}_j = E[\mathbf{x}_j \mathbf{x}_j^*]$ is user j's $N_j \times N_j$ input covariance matrix, N is the rank of $\mathbf{H} = [(\mathbf{R}_1 \mathbf{H}_1)^T \ \cdots \ (\mathbf{R}_K \mathbf{H}_K)^T]^T$, $\widetilde{N}_j$ is the rank of $\widetilde{\mathbf{H}}_j = [(\mathbf{R}_1 \mathbf{H}_1)^T \ \cdots \ (\mathbf{R}_{j-1} \mathbf{H}_{j-1})^T \ (\mathbf{R}_{j+1} \mathbf{H}_{j+1})^T \ \cdots \ (\mathbf{R}_K \mathbf{H}_K)^T]^T$, and P denotes the total transmit power available at the basestation. The optimization over $\mathbf{Q}_j$ ensures the best signal covariance for user j. The maximization over $\mathbf{R}_j$, as well as its dimension M_j, ensures that the total throughput can be maximized. Notice that for a single user j, the user's throughput may be decreased by choosing $M_j < N_r$. The total throughput, however, can be increased because user j saves additional dimension for other users. Notice that due to the zero-forcing requirement, i.e. the third constraint in (3.7), not all K users can be simultaneously supported with block diagonalization, i.e. $M_j = 0$ for those users that are not scheduled for transmission.

The optimization problem in (3.7) is difficult to solve, especially the maximization over $\{\mathbf{R}_j\}_{j=1}^K$. The difficulty primarily comes from the zero inter-user interference requirement, i.e. the third constraint in (3.7). In [59], an iterative algorithm was proposed to optimize $\{\mathbf{R}_j\}_{j=1}^K$ and $\{\mathbf{T}_j\}_{j=1}^K$ so that the aggregate effective channel energy is maximized. The sum capacity, however, is not directly optimized in [59]. In this chapter, I consider a set of special $M_j \times N_r$ matrices $\mathbf{R}_j$ (for $j = 1, 2, \cdots, K$) that are formed by taking M_j rows from $\mathbf{I}_{N_r}$ [41]. For example, if $M_j = 2$ and $N_r = 3$, then $\mathbf{R}_j$ must be in the following set:

$$\mathcal{R}^{(2,3)} = \left\{ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right\} \tag{3.8}$$

where $\mathcal{R}^{(m,n)}$ denotes the set of matrices formed by taking m rows from $\mathbf{I}_n$. The motivation of studying this special $\mathbf{R}_j$ are:

1. Since the matrices $\{\mathbf{R}_j\}_{j=1}^K$ and $\{\mathbf{T}_j\}_{j=1}^K$ are designed at the basestation, the post-processing matrices $\{\mathbf{R}_j\}_{j=1}^K$ need to be conveyed to the users, which is a system overhead that should be kept low. To successfully convey the post-processing matrices to the users, much less overhead (in the number of bits) is required for this specially formed $\mathbf{R}_j$ than a general $M_j \times N_r$ matrix. For example, $\log_2 \left(\sum_{M_j=0}^{N_r} |\mathcal{R}^{(M_j,N_r)}| \right) = N_r$ bits are sufficient to convey $\mathbf{R}_j$ to user j, where $|\mathcal{R}^{(M_j,N_r)}|$ denotes the cardinality of set $\mathcal{R}^{(M_j,N_r)}$.

2. With this special $\mathbf{R}_j$, user j can select M_j receive antennas to use. Hence, user selection and receive antenna selection can be combined to optimize the total throughput of all users. If $\mathbf{R}_j = \mathbf{I}_{N_r}$ for those users scheduled for transmission and $\mathbf{R}_j = \emptyset$ (i.e. $M_j = 0$) for those unscheduled users, then the generalized block diagonalization [59] reduces to the BD algorithm without post-processing presented in [14] [80].

3. With the additional constraint that $\mathbf{R}_j \in \mathcal{R}^{(M_j,N_r)}$ for $j = 1, 2, \cdots, K$, the optimization problem in (3.7) is solvable by exhaustively searching over all possible sets of $\{\mathbf{R}_j\}_{j=1}^K$. For each set of $\{\mathbf{R}_j\}_{j=1}^K$, the corresponding $\{\mathbf{T}_j\}_{j=1}^K$ can be found according to the SVD outlined in Section II. Due to the zero-forcing condition, the effective channels for all users do not interfere each other. Hence, the optimal $\{\mathbf{Q}_j\}_{j=1}^K$ can be obtained by the water-filling algorithm with an overall transmit power constraint [80].

3.4 BD vs. DPC: Sum Capacity for a Given Set of Channels

In this section, I compare the sum capacity achieved by block diagonalization without receive antenna selection with the sum capacity achieved by dirty paper coding. The BD sum capacity is defined based on the algorithm in [80], i.e. $\mathbf{R}_j = \mathbf{I}_{N_r}$ for all scheduled users. With this assumption, each user will utilize all N_r receive antennas provided he/she is scheduled for transmission. The simulation results in section VI show that the sum capacity of BD without receive antenna selection is very close to that with receive antenna selection in Rayleigh fading channels.

Consider a given set of channel realizations for a multiuser MIMO system, where $\mathbf{H}_j$ denotes the channel for user j. Let $\mathcal{K} = \{1, 2, \cdots, K\}$ denote the set of user indices. Assume all user sets are ordered and let $\mathcal{A}_i \in \mathcal{K}$ be the ith set. Let $\overline{\mathbf{H}}_j = \mathbf{H}_j\mathbf{T}_j$ denote the effective channel after precoding for user $j \in \mathcal{A}_i$, then the total throughput achieved with BD applied to the user set $\mathcal{A}_i$ with total power P can be expressed as

$$C_{BD|\mathcal{A}_i}(\mathbf{H}_{\mathcal{A}_i}, P, \sigma^2) = \max_{\{\mathbf{Q}_j :\, \mathbf{Q}_j \geq 0,\, \sum_{j\in\mathcal{A}_i} \mathrm{Tr}(\mathbf{Q}_j) \leq P\}} \sum_{j\in\mathcal{A}_i} \log \left| \mathbf{I} + \frac{1}{\sigma^2}\overline{\mathbf{H}}_j \mathbf{Q}_j \overline{\mathbf{H}}_j^* \right| \qquad (3.9)$$

where $\mathbf{Q}_j = E[\mathbf{x}_j\mathbf{x}_j^*]$ is user j's input covariance matrix of size $N_j \times N_j$. Let $\mathcal{A}$ be the set containing all possible user sets, i.e. $\mathcal{A} = \{\mathcal{A}_1, \mathcal{A}_2, \cdots\}$. The sum capacity of BD is defined as the maximum total throughput of BD as

$$C_{BD}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2) = \max_{\mathcal{A}_i \in \mathcal{A}} C_{BD|\mathcal{A}_i}(\mathbf{H}_{\mathcal{A}_i}, P, \sigma^2). \qquad (3.10)$$

It has been proven that the sum capacity of a multiuser Gaussian broadcast channel is achieved with dirty paper coding [88]. With the duality results in [88], the sum capacity can be expressed as

$$C_{DPC}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2) = \max_{\{\mathbf{S}_j :\, \mathbf{S}_j \geq 0,\, \sum_{j=1}^{K} \mathrm{Tr}(\mathbf{S}_j) \leq P\}} \log \left| \mathbf{I} + \frac{1}{\sigma^2} \sum_{j=1}^{K} \mathbf{H}_j^* \mathbf{S}_j \mathbf{H}_j \right| \qquad (3.11)$$

where $\mathbf{S}_j$ of size $N_r \times N_r$ is the signal covariance matrix for user j in the dual multiple access channel.

In this section, I am interested in the gain of DPC over BD in terms of sum capacity. Analogous to [39], I define the ratio of DPC to BD as

$$G(\mathbf{H}_{1,\cdots,K}, P, \sigma^2) \triangleq \frac{C_{DPC}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2)}{C_{BD}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2)}. \tag{3.12}$$

The gain is obviously dependent on the channel realizations $\{\mathbf{H}_k\}_{k=1}^K$, the total power, and noise variance. The next theorem gives a bound on $G(\mathbf{H}_{1,\cdots,K}, P, \sigma^2)$ that is valid for any $\{\mathbf{H}_k\}_{k=1}^K$, P, and σ^2.

Theorem 1: The sum capacity gain of DPC over BD is lower bounded by 1 and upper bounded by the minimum of N_t and K, i.e.

$$1 \leq G(\mathbf{H}_{1,\cdots,K}, P, \sigma^2) \leq \min\{N_t, K\} \tag{3.13}$$

Proof: Theorem 3 in [39] states that

$$\frac{C_{DPC}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2)}{C_{TDMA}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2)} \leq \min\{N_t, K\} \tag{3.14}$$

where

$$\begin{aligned} C_{TDMA}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2) &\triangleq \max_{k\in\mathcal{K}} C(\mathbf{H}_k, P, \sigma^2) && (3.15) \\ &= \max_{k\in\mathcal{K}} \max_{\{\mathbf{Q}_k:\, \mathbf{Q}_k \geq 0, \mathrm{Tr}(\mathbf{Q}_k)\leq P\}} \log \left| \mathbf{I} + \frac{1}{\sigma^2}\mathbf{H}_k\mathbf{Q}_k\mathbf{H}_k^* \right| && (3.16) \end{aligned}$$

It is easy to show that

$$C_{BD}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2) \geq C_{TDMA}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2). \tag{3.17}$$

Hence, following Theorem 3 in [39], it can be immediately concluded that

$$\frac{C_{DPC}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2)}{C_{BD}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2)} \leq \frac{C_{DPC}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2)}{C_{TDMA}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2)} \leq \min\{N_t, K\}. \tag{3.18}$$

Furthermore, it is true that

$$C_{DPC}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2) \geq C_{BD}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2). \tag{3.19}$$

Combining (3.18) and (3.19) completes the proof. □

Although the bound in Theorem 1 holds for any N_t, N_r, K, $\{\mathbf{H}_i\}_{i=1}^K$, P, and σ^2, it is generally a loose bound. In the following, several cases are presented where the bound on $G(\mathbf{H}_{1,\cdots,K}, P, \sigma^2)$ can be tightened. I first show a sufficient condition where $C_{DPC}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2) = C_{BD}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2)$.

Lemma 2: Assume $N_r \leq N_t$ and $K \leq \lfloor \frac{N_t}{N_r} \rfloor$. If $\{\mathbf{H}_k\}_{k=1}^K$ are mutually orthogonal, i.e. $\mathbf{H}_i\mathbf{H}_j^* = 0$ for $i \neq j$, then

$$C_{DPC}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2) = C_{BD}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2). \tag{3.20}$$

Proof: Please see Appendix D. □

Lemma 2 shows that when the user channels are mutually orthogonal to each other, BD achieves the same capacity as DPC. This is very different from [39] where DPC vs. TDMA is compared. For TDMA, even though the users are mutually orthogonal, it is not possible to achieve the same sum capacity as DPC. In fact, the gain of DPC over TDMA can still be at the maximum, i.e. $\min\{N_t, K\}$, when the user channels are mutually orthogonal, e.g. when each user has one receiver antenna and the user channels have the same energy.

While Lemma 2 shows DPC and BD are essentially the same when the user channels are orthogonal to each other, in the next Lemma, I show that when all user channels are in the same vector subspace, the bound on the gain of DPC over BD in Theorem 1 can be tightened to $\min\{N_r, K\}$ for $N_r \leq N_t$.

Lemma 3: Assume $N_r \leq N_t$. If the row vector spaces of all user channels are the same, i.e. $\text{span}(\mathbf{H}_1) = \text{span}(\mathbf{H}_2) = \cdots = \text{span}(\mathbf{H}_K)$, which is denoted as $\mathcal{W}$, then

$$G(\mathbf{H}_{1,\cdots,K}, P, \sigma^2) \leq \min\{N_r, K\}. \tag{3.21}$$

Proof: Please see Appendix E. □

Notice Lemma 2 and Lemma 3 are two extreme cases, where the channels $\{\mathbf{H}_i\}_{i=1}^K$ are either mutually orthogonal or in the same subspace. When the user channels are mutually orthogonal, BD is the same as DPC. On the other hand, when the user channels are in the same subspace, BD is the same as TDMA. For the general case where the user channels are partially overlapped to each other, BD may be superior to TDMA since multiple users may be supported at the same time. However, a good bound on the sum capacity gain of DPC over BD for the general case is very difficult to obtain.

For BD with receive antenna selection, if the optimal receive antenna set is obtained for each user through complete search as outlined at the end of Section III, then the sum capacity may be increased compared to BD without receive antenna selection. Since BD without receive antenna selection is a special case of BD with receive antenna selection, i.e. $\mathbf{R}_j = \mathbf{I}_{N_r}$ for $j = 1, 2, \cdots, K$, the results in Theorem 1, Lemma 2 and 3 also hold for BD with receive antenna selection.

3.5 BD vs. DPC: Ergodic Sum Capacity in Rayleigh Fading Channels

In this section, I analyze the ergodic capacity of a multiuser MIMO system in Rayleigh fading channels for block diagonalization without receive antenna selection vs. DPC. Let $\overline{\mathbf{H}}_j = \mathbf{H}_j\mathbf{T}_j$ be the equivalent channel for user j after precoding. With the assumptions that $\{\mathbf{H}_j\}_{j=1}^{K}$ are statistically independent for different j and the elements in $\mathbf{H}_j$ are i.i.d. complex Gaussian random variables, the following theorem on the probability density function of $\overline{\mathbf{H}}_j$ holds.

Theorem 2: In a downlink MIMO system with block diagonalization applied to a fixed set of users, if the MIMO channel for each user is modeled as i.i.d. complex Gaussian, then the effective channel after precoding is also an i.i.d. complex Gaussian matrix.

Proof: Since $\overline{\mathbf{H}}_j = \mathbf{H}_j\mathbf{T}_j$ and $\mathbf{H}_j$ is i.i.d. complex Gaussian, then $\overline{\mathbf{H}}_j$ conditioned on $\mathbf{T}_j$ is also complex Gaussian. Recall that for BD without receive antenna selection, $\mathbf{T}_j$ is a set of basis in the null space of $\widetilde{\mathbf{H}}_j = [\mathbf{H}_1^T \ \cdots \ \mathbf{H}_{j-1}^T \ \mathbf{H}_{j+1}^T \ \cdots \ \mathbf{H}_K^T]^T$, hence $\overline{\mathbf{H}}_j$ is independent of $\mathbf{T}_j$. Therefore, the theorem is proved. $\square$

Theorem 2 indicates that if BD is applied to a fixed set of users, the effective channel for each user still follows the i.i.d. complex Gaussian distribution if the original channels are i.i.d. complex Gaussian. Hence the ergodic capacity of user j can be easily evaluated with the eigenvalue distribution of $\overline{\mathbf{H}}_j\overline{\mathbf{H}}_j^*$, where $\overline{\mathbf{H}}_j\overline{\mathbf{H}}_j^*$ follows a Wishart distribution [19] [58] [84].

3.5.1 A Lower Bound on Ergodic Sum Capacity with BD

Let $\mathcal{A}_i = \{1, 2, \cdots, i\}$ be the set of the first i users, for $i = 1, 2, \cdots, I$ where $I = \min\left\{K, \left\lfloor \frac{N_t}{N_r} \right\rfloor\right\}$. Notice that for $i \leq \min\left\{K, \left\lfloor \frac{N_t}{N_r} \right\rfloor\right\}$, when the elements in $\{\mathbf{H}\}_{k=1}^K$ are generated according to an i.i.d. complex Gaussian distribution, a lower bound on the ergodic sum capacity of BD can be obtained as

$$
\begin{aligned}
E\left[C_{BD}(\mathbf{H}_{\mathcal{A}_i}, P, \sigma^2)\right] &\overset{(a)}{\geq} E\left[\sum_{j=1}^{i} \log\left|I + \frac{1}{\sigma^2}\overline{\mathbf{H}}_j\mathbf{Q}_j\overline{\mathbf{H}}_j^*\right|\right] && (3.22)\\
&= \sum_{j=1}^{i} E\left[\sum_{n=1}^{N_r} \log\left|1 + \frac{P_{j,n}}{\sigma^2}\overline{\lambda}_{j,n}^2\right|\right] && (3.23)\\
&\overset{(b)}{\geq} \sum_{j=1}^{i} E\left[\sum_{n=1}^{N_r} \log\left|1 + \frac{P}{iN_r\sigma^2}\overline{\lambda}_{j,n}^2\right|\right] && (3.24)\\
&= \sum_{j=1}^{i} N_r E\left[\log\left|1 + \frac{P}{iN_r\sigma^2}\overline{\lambda}_{j,1}^2\right|\right] && (3.25)\\
&\overset{(c)}{=} i\ N_r E\left[\log\left|1 + \frac{P}{iN_r\sigma^2}\overline{\lambda}_{i,1}^2\right|\right] && (3.26)\\
&\triangleq \overline{C}_{BD}(\mathbf{H}_{\mathcal{A}_i}, P, \sigma^2) && (3.27)
\end{aligned}
$$

where $\overline{\lambda}_{j,n}^2$ are nth unordered eigenvalues of $\overline{\mathbf{H}}_j\overline{\mathbf{H}}_j^*$ and $\overline{\mathbf{H}}_j$ is of size $N_r \times (N_t - (i-1)N_r)$. Inequality (a) holds because the RHS of (3.22) assumes all i users are simultaneously transmitting for all channel realizations. Inequality (b) holds because the RHS of

(3.24) assumes equal power is allocated to every non-zero eigenmode. Equality (c) holds because $\overline{\lambda}_{j,1}$ has the same distribution for $j = 1, 2, \cdots, i$.

For notational simplicity, I denote $\alpha_i = \overline{\lambda}_{i,1}^2$ and $\overline{N}_i = N_t - (i-1)N_r$. With Theorem 2 and [84], the distribution of α_i can be expressed as

$$p_{\overline{N}_i, N_r}(\alpha_i) = \frac{1}{N_r} \sum_{m=1}^{N_r} \varphi_m(\alpha_i)^2 \alpha_i^{\overline{N}_i - N_r} e^{-\alpha_i} \tag{3.28}$$

where

$$\varphi_{k+1}(\alpha_i) = \left[\frac{k!}{(k + \overline{N}_i - N_r)!} \right]^{1/2} L_k^{\overline{N}_i - N_r}(\alpha_i) \tag{3.29}$$

for $k = 0, 1, \cdots, m-1$, and

$$L_k^{n-m}(x) = \frac{1}{k!} e^x x^{m-n} \frac{d^k}{dx^k} \left(e^{-x} x^{n-m+k} \right). \tag{3.30}$$

Hence (3.26) can be easily evaluated with a numerical integration.

Hence, the ergodic sum capacity with BD can be lower bounded by

$$E\left[C_{BD}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2)\right] \quad \geq \max_{i \in \{1,2,\cdots,I\}} \overline{C}_{BD}(\mathbf{H}_{\mathcal{A}_i}, P, \sigma^2). \tag{3.31}$$

It is important to note that to evaluate the lower bound, up to $I = \min\left\{K, \left\lfloor \frac{N_t}{N_r} \right\rfloor\right\}$ numerical integrations need to be carried out because of the maximization in the RHS of (3.31).

3.5.2 An Upper Bound on the Ergodic Sum Capacity of DPC

It is well known that the sum capacity of a K-user broadcast channel with DPC is upper bounded if the receivers are allowed to cooperate [88]. Let $\mathbf{H} = [\mathbf{H}_1^T \ \mathbf{H}_2^T \ \cdots \ \mathbf{H}_K^T]^T$, and $N = \max\{N_t, KN_r\}$ and $M = \min\{N_t, KN_r\}$, then

$$
\begin{aligned}
E\left[C_{DPC}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2)\right] &\leq E\left[\log\left|\mathbf{I} + \frac{1}{\sigma^2}\mathbf{HQH}^*\right|\right] &(3.32)\\
&= \sum_{m=1}^{M} E\left[\log\left(1 + \frac{P_m}{\sigma^2}\lambda_m^2\right)\right] &(3.33)\\
&= ME\left[\log\left(1 + \frac{P_1}{\sigma^2}\alpha_1\right)\right] &(3.34)\\
&\leq M \int_{\sigma^2/\Gamma_0}^{\infty} \log\left(\frac{\Gamma_0\alpha_1}{\sigma^2}\right) p_{N,M}(\alpha_1)d\alpha_1 &(3.35)\\
&\triangleq \overline{C}_{coop}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2) &(3.36)
\end{aligned}
$$

where λ_m^2 is mth unordered eigenvalue of $\mathbf{H}^*\mathbf{H}$ and $\alpha_1 \triangleq \lambda_1^2$; $p_{N,M}(\alpha_1)$ is the distribution for α_1, which is given by (3.28) with N_r and $\overline{N}_i$ replaced by M and N respectively. The parameter Γ_0 is optimized so that the ergodic sum capacity is maximized with the average power constraint, i.e. $M \int_{\sigma^2/\Gamma_0}^{\infty} \left(\Gamma_0 - \frac{\sigma^2}{\alpha}\right) p_{N,M}(\alpha)d\alpha = P$. Details on the inequality (3.35) can be found in [78].

3.5.3 An Upper Bound on the Ergodic Capacity of DPC vs. BD

From the above two sections, an upper bound on the ergodic sum capacity gain of DPC over BD can be derived as

$$
\frac{E\left[C_{DPC}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2)\right]}{E\left[C_{BD}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2)\right]} \leq \frac{\overline{C}_{coop}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2)}{\max\limits_{i\in\{1,2,\cdots,I\}} \overline{C}_{BD}(\mathbf{H}_{\mathcal{A}_i}, P, \sigma^2)}. \tag{3.37}
$$

Notice that the upper bound in (3.37) is a function of N_t, N_r, K, P, and σ^2. Furthermore, $\min\left\{K, \left\lfloor \frac{N_t}{N_r} \right\rfloor\right\} + 1$ numerical integrations are necessary to evaluate the bound in (3.37). The tightness of this bound is shown in the numerical results section.

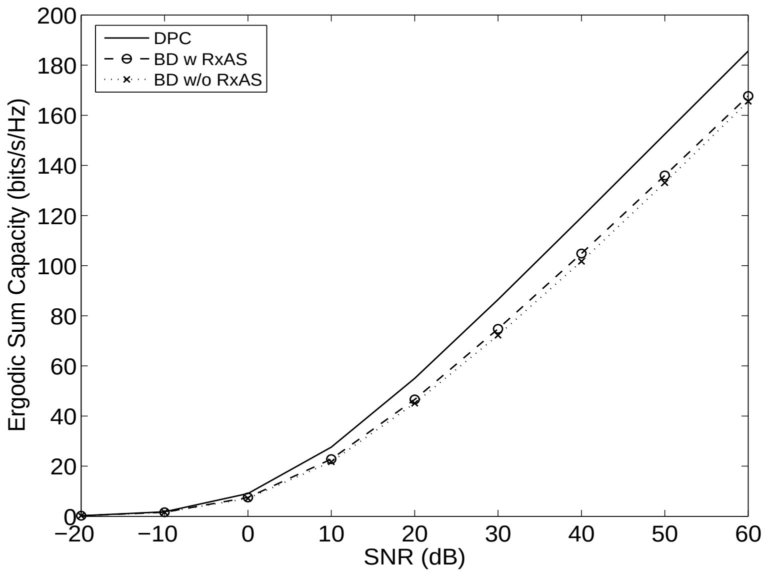

Figure 3.1: Ergodic sum capacity of DPC vs. BD in Rayleigh fading channels. $N_t = 10$, $N_r = 2$, $K = 5$.

3.6 Numerical Results

In this section, I provide some numerical demonstrations of the gain of DPC over BD. I compare the sum capacity achieved by

- dirty paper coding (DPC) implemented with the iterative water-filling algorithm [40],
- block diagonalization with receive antenna selection (BD w RxAS),
- block diagonalization without receive antenna selection (BD w/o RxAS).

I show the ergodic sum capacity of DPC and BD by Monte Carlo simulations and compare the gain with the bound in (3.37) for various system parameters. In Rayleigh fading channels, BD achieves a significant part of the sum capacity of DPC in most cases. And the bound in (3.37) is tight for medium to high SNRs or when $K \leq \lfloor \frac{N_t}{N_r} \rfloor$.

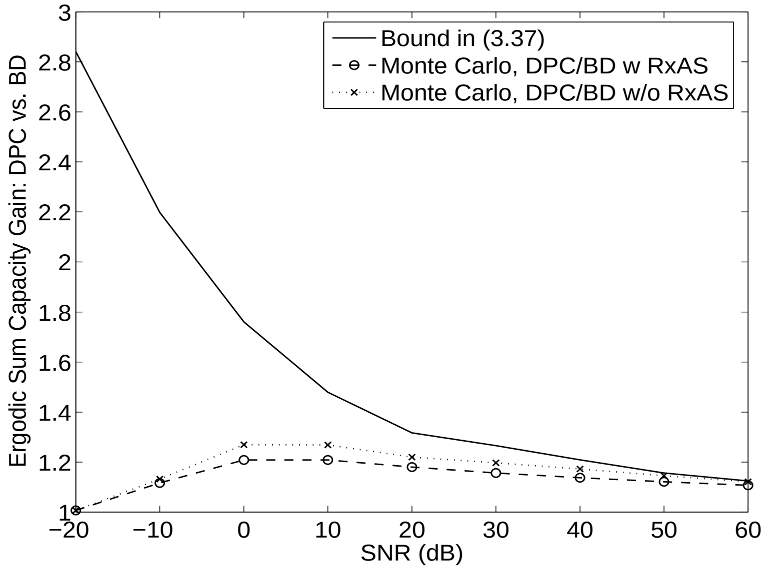

Figure 3.2: Ergodic sum capacity gain of DPC over BD in Rayleigh fading channels. $N_t = 10$, $N_r = 2$, $K = 5$.

Fig. 3.1 shows the ergodic sum capacity of DPC vs. BD under different SNRs, with $N_t = 10$, $N_r = 2$, and $K = 5$. In the low SNR regime, BD achieves almost the same sum capacity as DPC. As SNR goes to infinity, the sum capacity of both DPC and BD increase with the same slope. Essentially, the ratio of the sum capacity of BD and DPC equals one in asymptotically low and high SNR regimes [39]. Fig. 3.2 shows the gain of DPC over BD from the curves in Fig. 3.1, as well as the bound on the gain evaluated from (3.37). As the SNR increases, the bound in (3.37) gets tighter. For low SNR, the bound in (3.37) is loose mainly because 1) the lower bound on BD assumes an equal power allocation to all non-zero eigenvalues; 2) the cooperative upper bound on DPC is also loose in low SNR. The bound in the low SNR regime is, however, less interesting because it has been proven in [39] that the sum capacity of BD equal that of DPC for asymptotically low SNRs.

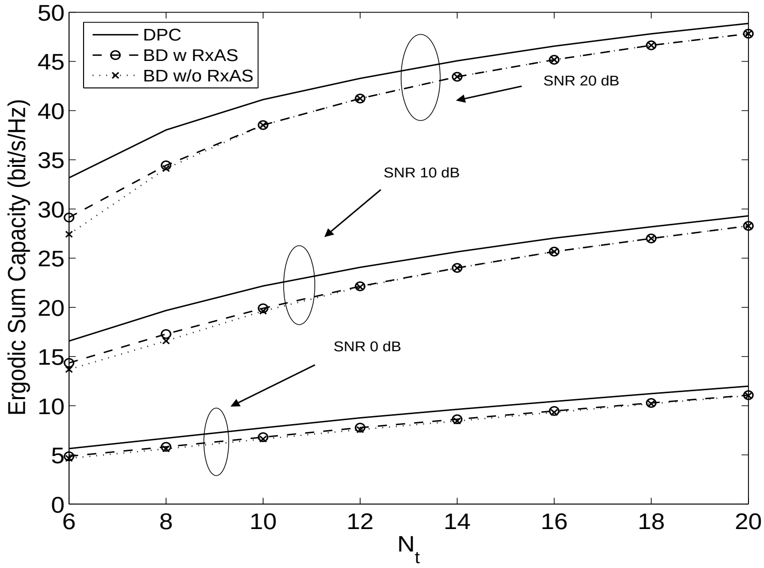

Figure 3.3: Ergodic sum capacity of DPC vs. BD in Rayleigh fading channels. $N_r = 2$, $K = 3$.

Fig. 3.3 shows the ergodic sum capacity of DPC vs. BD for different N_t, with $N_r = 2$ and $K = 3$. As the number of transmit antenna increases, the sum capacity of BD gets closer to the sum capacity of DPC. Fig. 3.4 shows the gain of DPC over BD from the curves in Fig. 3.3, with $SNR = 20$ dB. It is observed that the bound from (3.37) is fairly tight for $N_t > KN_r$.

Fig. 3.5 shows the ergodic sum capacity of DPC vs. BD for different numbers of users, with $N_t = 10$ and $N_r = 2$. For small numbers of users, BD achieves almost the same sum capacity as DPC. As the number of users increases, DPC exhibits higher performance than BD. Fig. 3.6 shows the gain of DPC over BD from the curves in Fig. 3.5, with $SNR = 20$ dB. For small numbers of users, the bound from (3.37) is very tight. For larger numbers of users, the bound from (3.37) loosens. The main reason is that for $N_t = 10$ and $N_r = 2$, BD without receive antenna selection can support at most 5 users simultaneously. The increase in the ergodic sum ca-

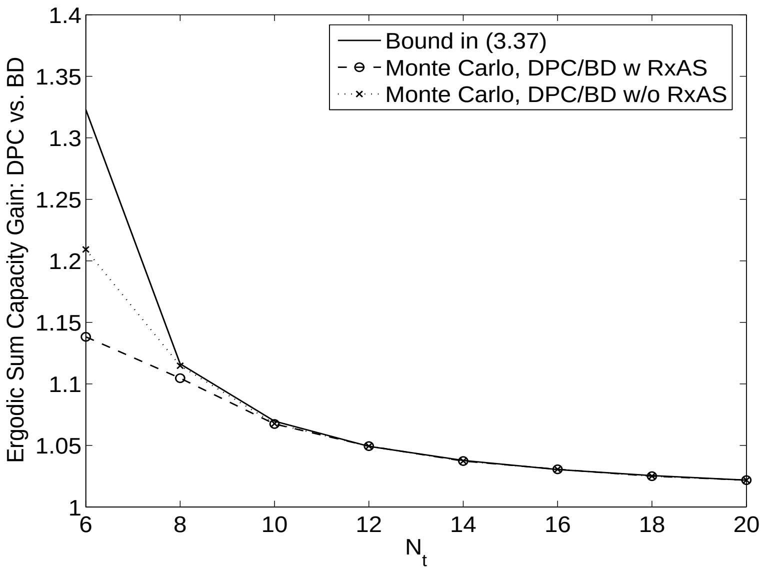

Figure 3.4: Ergodic sum capacity gain of DPC over BD in Rayleigh fading channels. $N_r = 2$, $K = 3$, SNR = 20 dB.

pacity of BD as K increases is mainly from multiuser diversity. However, the lower bound on the sum capacity of BD in (3.31) does not take multiuser diversity into consideration, which means the lower bound on BD is the same for $k = 5, 6, \cdots, 10$.

Notice that throughout the simulations, BD with receive antenna selection (RxAS) achieves higher sum capacity than BD without receive antenna selection. On the other hand, to find the optimal user and/or receive antenna set for the sum capacity, the search space for BD with RxAS is much higher that that of BD without RxAS. Some low complexity user selection algorithms for BD or zero-forcing (ZF) can be found in [74] [101].

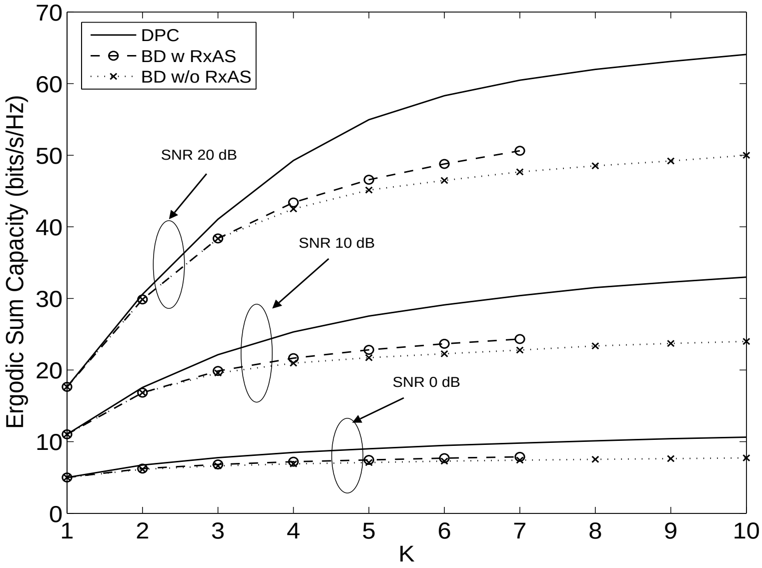

Figure 3.5: Ergodic sum capacity of DPC vs. BD in Rayleigh fading channels. $N_t = 10$, $N_r = 2$.

3.7 Conclusion

In this chapter, I compare the sum capacity of BD with and without receive antenna selection to that of DPC. For a given set of channel realizations, the sum capacity gain of DPC over BD can be generally bounded by $\min\{N_t, K\}$, where N_t and K are the number of transmit antennas and the number of users, respectively. The gain can be tightened in two special cases: 1) if the user channels are orthogonal to each other, BD achieves the same sum capacity with DPC; 2) if the user channels are in the same vector space, the gain can be reduced to $\min\{N_r, K\}$, where N_r is the number of receive antennas at each user with $N_r \leq N_t$. The ergodic sum capacity gain of DPC over BD is also studied in a Rayleigh fading channel. Simulations show that BD can achieve a significant part of the total throughput of DPC. An upper bound on the ergodic sum capacity gain of DPC over BD is proposed. The

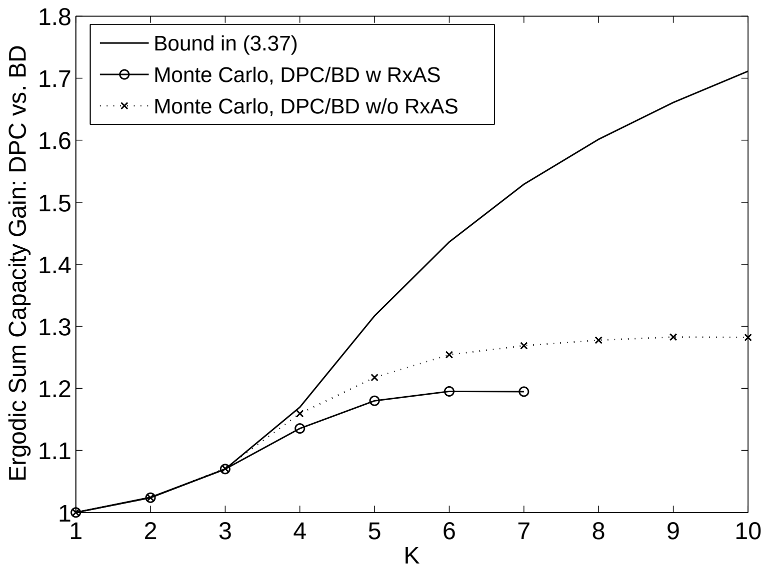

Figure 3.6: Ergodic sum capacity gain of DPC over BD in Rayleigh fading channels. $N_t = 10$, $N_r = 2$, SNR = 20 dB.

bound is very tight for medium to high SNRs or when $K \leq \lfloor \frac{N_t}{N_r} \rfloor$, which is useful in estimating how far away BD is from being optimal in terms of ergodic sum capacity.

Chapter 4

Low Complexity User Selection Algorithms for Multiuser MIMO Systems with Block Diagonalization

4.1 Introduction

As discussed in the previous chapter, Block Diagonalization is a practical precoding technique for downlink broadcast MIMO channels [14] [59] [80] [99] . With BD, each user's signal is multiplied by a precoding matrix before transmission. Every user's precoding matrix is restricted to be in the null space of all other users' channels. Hence if the channel matrices of all users are perfectly known at the transmitter, each user perceives an interference-free channel.

Due to the rank condition imposed by the fact that each user's precoding matrix must lie in the null space of all other users' channels, the number of users that can be simultaneously supported with BD is limited by the number of transmit antennas, the number of receive antennas, and the richness of the channels [80]. For systems with a large number of users, a subset of users can be selected to maximize the total throughput (defined as the aggregate error-free capacity). In this chapter, I assume that every user utilizes all its receive antennas. A brute-force complete search over all possible user sets guarantees that the total throughput is maximized. The complexity, however, is prohibitive if the number of users in the system is large. For example, if $\hat{K}$ is the maximum number of users that can be simultaneously supported by BD and K is the total number of users, then the complete search for the optimal user set has combinatoric complexity because every i ($1 \leq i \leq \hat{K}$) out of K users must searched.

A user selection algorithm for downlink multiuser MISO systems has been proposed in [101], where the users are equipped with one receive antenna and zero-forcing beamforming is performed at the transmitter, which is equivalent to BD. The algorithm in [101] constructs a set of semi-orthogonal users whose total throughput is close the sum capacity achieved by DPC. Analogous to the user selection problem is the antenna selection problem where the transmitter and receiver select a subset of antennas to transmit and receive signals [23] [29] [57], e.g. a low complexity antenna selection algorithm is proposed in [23] that achieves almost the same outage capacity as the optimal selection method. Antenna selection has also been considered in downlink multiuser MIMO systems with BD [11], where it has been shown that a significant reduction in symbol error rate can be achieved even with one extra transmit antenna. Space division multiple access (SDMA) with scheduling for multimedia services has been studied in [100]. It was shown in [100] that the system throughput-delay characteristics can be improved by scheduling the users with nearly orthogonal spatial signatures at each time slot. Several other scheduling as well as admission control algorithms for downlink SDMA systems can also be found in [6].

In this chapter, I propose two suboptimal user selection algorithms for BD with the aim of maximizing the total throughput while keeping the complexity low. Both algorithms iteratively select users until the maximum number of simultaneously supportable users are reached. The first user selection algorithm greedily maximizes the total throughput. In each user selection step, the algorithm selects a user who provides the maximum total throughput with those already selected users. While the first algorithm requires frequent singular value decomposition (SVD) of the channel matrices, the second proposed algorithm selects the users based on the channel energy, thus reducing the computational complexity. I show that the proposed algorithms achieve around 95% of the total throughput of the optimal user set, and the complexity of the proposed algorithms is linear in the total number of users K.

4.2 System Model and Background

The system model and the BD algorithm has been presented in Chapter 3. In a downlink multiuser MIMO system with K users, I denote the number of transmit antennas at the basestation as N_t and the number of receive antennas for the jth user as $N_{r,j}$. In the rest of the chapter, it is assumed that every user has and uses the same number of receive antennas, i.e. $N_{r,j} = N_r$ for $j = 1, 2, \cdots, K$ for simplicity, where K is the total number of users in the system. With the assumption that each element in the user MIMO channel matrix $\mathbf{H}_j$ is generated by an i.i.d. complex Gaussian distribution, it can be inferred from the rank condition in [80] that the maximum number of simultaneous users is $\left\lceil \frac{N_t}{N_r} \right\rceil$, where $\lceil \cdot \rceil$ is the ceiling operation.

4.3 Low Complexity User Selection Algorithms

In this section, the sum capacity (i.e. the maximum total throughput) of BD defined in Section 3.4 is repeated. Two suboptimal user selection algorithms are then proposed to reduce the complexity of finding the optimal user set.

Consider a set of channels $\{\mathbf{H}_j\}_{j=1}^K$ for a multiuser MIMO system. Let $\mathcal{K} = \{1, 2, \cdots, K\}$ denote the set of all users, and $\mathcal{A}_i$ be a subset of $\mathcal{K}$, where the cardinality of $\mathcal{A}_i$ is less than or equal to the maximum number of simultaneous users $\hat{K}$. Let $\overline{\mathbf{H}}_j = \mathbf{H}_j\mathbf{T}_j$ denote the effective channel after precoding for user $j \in \mathcal{A}_i$, then the total throughput achieved with BD applied to the user set $\mathcal{A}_i$ with total power P can be expressed as

$$C_{BD|\mathcal{A}_i}(\mathbf{H}_{\mathcal{A}_i}, P, \sigma^2) = \max_{\{\mathbf{Q}_j:\, \mathbf{Q}_j \geq 0,\, \sum_{j\in\mathcal{A}_i} \mathrm{Tr}(\mathbf{Q}_j) \leq P\}} \sum_{j\in\mathcal{A}_i} \log \left| \mathbf{I} + \frac{1}{\sigma^2}\overline{\mathbf{H}}_j\mathbf{Q}_j\overline{\mathbf{H}}_j^* \right| \tag{4.1}$$

where $\mathbf{Q}_j = E[\mathbf{x}_j\mathbf{x}_j^*]$ is user j's input covariance matrix of size $N_j \times N_j$ and $\mathbf{H}_{\mathcal{A}_i}$ denotes the set of channels for those users in $\mathcal{A}_i$. Notice that the solution to the RHS of (4.1) can be obtained by the water-filling algorithm over the eigenvalues of $\{\overline{\mathbf{H}}_j\overline{\mathbf{H}}_j^*\}_{j\in\mathcal{A}_i}$ with total power constraint P, as discussed in [80].

Let $\mathcal{A}$ be the set containing all possible $\mathcal{A}_i$, i.e. $\mathcal{A} = \{\mathcal{A}_1, \mathcal{A}_2, \cdots\}$, then the sum capacity (maximum total throughput) with BD can be defined as

$$C_{BD}(\mathbf{H}_1, \mathbf{H}_2, \cdots, \mathbf{H}_K, P, \sigma^2) = \max_{\mathcal{A}_i\in\mathcal{A}} C_{BD|\mathcal{A}_i}(\mathbf{H}_{\mathcal{A}_i}, P, \sigma^2). \tag{4.2}$$

Denote $\hat{K} = \left\lceil \frac{N_t}{N_r} \right\rceil$ as the maximum number of simultaneous users, and the Cardinality of $\mathcal{A}$ is $|\mathcal{A}| = \sum_{i=1}^{\hat{K}} {}_KC_i$, where ${}_nC_m$ denotes the combination of n choosing m. Hence, it is clear that a brute-force exhaustive search over $\mathcal{A}$ is computationally prohibitive if $K \gg \hat{K}$.

4.3.1 Capacity-Based Suboptimal User Selection Algorithm

The exhaustive search method needs to consider roughly $\mathcal{O}(K^{\hat{K}})$ possible user sets. In this section, I present a suboptimal algorithm whose complexity is $\mathcal{O}(\hat{K}K)$.

Let s_i denote the user index selected in the ith iteration, i.e. $s_i \in \{1, 2, \cdots, K\}$ and $1 \leq i \leq \hat{K}$. Let Ω denote the set of unselected users and Υ denote the set of

Table 4.1: Capacity-Based Suboptimal User Selection Algorithm

1. Initially, let $\Omega = \{1, 2, \cdots, K\}$ and $\Upsilon = \emptyset$. Let $s_1 = \arg\max_{k \in \Omega} \log \left| \mathbf{I} + \frac{1}{\sigma^2} \mathbf{H}_k \mathbf{Q}_k \mathbf{H}_k^* \right|$ where $\mathrm{Tr}(\mathbf{Q}_k) \leq P$ and $\mathbf{Q}_k$ is semi-positive definite. Let $\Omega = \Omega - \{s_1\}$ and $\Upsilon = \Upsilon + \{s_1\}$. Let $C_{temp} = \max_{k \in \Omega} \log \left| \mathbf{I} + \frac{1}{\sigma^2} \mathbf{H}_k \mathbf{Q}_k \mathbf{H}_k^* \right|$.

2. for $i = 2 : \hat{K}$

   (a) For every $k \in \Omega$,

      i. Let $\overline{\Upsilon}_k = \Upsilon + \{k\}$.

      ii. Find the precoding matrix $\mathbf{T}_j$ for each $j \in \overline{\Upsilon}_k$, and obtain the effective channel $\overline{\mathbf{H}}_j = \mathbf{H}_j \mathbf{T}_j$ for each $j \in \overline{\Upsilon}_k$.

      iii. Perform a singular value decomposition (SVD) on $\overline{\mathbf{H}}_j$, and obtain the M singular values $\{\lambda_{j,m}\}_{i=1}^{M}$.

      iv. Water-fill over $\lambda_{j,m}^2$ for $j \in \overline{\Upsilon}_k$ and $1 \leq m \leq M$. Find the total throughput to the user set $\overline{\Upsilon}_k$, denoted as C_k.

   (b) Let $s_i = \arg\max_{k \in \Omega} C_k$.

   (c) If $\max_{k \in \Omega} C_k < C_{temp}$

      Algorithm terminated. The selected user set is Υ.

      else

      Let $\Omega = \Omega - \{s_i\}$ and $\Upsilon = \Upsilon + \{s_i\}$. And let $C_{temp} = \max_{k \in \Omega} C_k$.

selected users. The capacity-based user selection algorithm is described in Table 4.1. In words, the algorithm first selects the single user with the highest capacity. Then, from the remaining unselected users, it finds the user that provides the highest total throughput together with those selected users. The algorithm terminates when $\hat{K}$ users are selected or the total throughput drops if more users are selected (the total throughput may decrease with an additional user because the size of the null space for every user reduces in order to meet the zero inter-user interference requirement). Clearly, the proposed algorithm needs to search over no more than $\hat{K}K$ user sets, which greatly reduces the complexity compared to the exhaustive search method.

Since the user selection criterion is based on the sum capacity, I name the above algorithm the capacity-based suboptimal user selection algorithm. Its throughput performance will be shown in Section 4.5.

4.3.2 Frobenius Norm-Based Suboptimal User Selection Algorithm

Although the capacity-based suboptimal user selection algorithm greatly reduces the size of the search set, the algorithm still may not be cost-effective for real-time implementation because singular value decomposition, which is computationally intensive, is required for each user in each iteration to find the total throughput. In this section, I propose another suboptimal user selection algorithm which is based on channel Frobenius norm. The motivation is that the capacity is closely related to eigenvalues of the effective channel after precoding. Although the channel Frobenius norm cannot characterize the capacity completely, it is related to the capacity because the Frobenius norm indicates the overall energy of the channel, i.e. the sum of the eigenvalues of $\mathbf{HH}^*$ equals $||\mathbf{H}||_F^2$.

Let s_i denotes the user index selected in the ith iteration, i.e. $s_i \in \{1, 2, \cdots, K\}$ and $1 \leq i \leq \hat{K}$. Let Ω denote the set of unselected users and Υ denote the set of selected users. Let $\mathbf{V}_k$ be the basis for the row vector space of $\mathbf{H}_k$ after applying the Gram-Schmidt orthogonalization procedure to the rows of $\mathbf{H}_k$. The Frobenius norm-based user selection algorithm is described in Table 4.2. The idea of the norm-based user selection algorithm is to select the set of users such that the sum of the effective channel energy of those selected users is as large as possible. Notice that steps 1 and 2 in the norm-based algorithm are independent with SNR, i.e. P. Once the $\hat{K}$ users are selected, step 3 makes the final user selection (possibly a subset of the $\hat{K}$ users chosen by steps 1 and 2) with the capacity-based algorithm, where the SNR is taken into consideration. Clearly, the norm-based algorithm requires fewer SVD operations than the capacity-based algorithm. Detailed computational complexity will be analyzed in Section 4.4.

Table 4.2: Frobenius Norm-Based Suboptimal User Selection Algorithm

1. Initially, let $\Omega = \{1, 2, \cdots, K\}$ and $\Upsilon = \emptyset$. Let $s_1 = \arg\max_{k\in\Omega} ||\mathbf{H}_k||_F^2$. Let $\mathbf{V} = \mathbf{V}_{s_1}$. Let $\Omega = \Omega - \{s_1\}$ and $\Upsilon = \Upsilon + \{s_1\}$.

2. for $i = 2 : \hat{K}$

   (a) For each $k \in \Omega$, let $\widetilde{\mathbf{H}}_k = \mathbf{H}_k - \mathbf{H}_k\mathbf{V}^*\mathbf{V}$. Then $\widetilde{\mathbf{H}}_k$ is in the null space of $\mathbf{V}$.

   for $j = 1 : i-1$

   i. Let
   $$\hat{\mathbf{H}}_{s_j,k} = [\mathbf{H}_{s_1}^T \ \cdots \ \mathbf{H}_{s_{j-1}}^T \ \mathbf{H}_{s_{j+1}}^T \ \cdots \ \mathbf{H}_{s_{i-1}}^T \ \mathbf{H}_k^T]^T.$$

   ii. Let $\mathbf{W}_{s_j,k}$ be the row basis for $\hat{\mathbf{H}}_{s_j,k}$ after Gram-Schmidt orthogonalization.

   (b) For each $s \in \Upsilon$, let $\widetilde{\mathbf{H}}_s = \mathbf{H}_s - \mathbf{H}_s\mathbf{W}_{s,k}^*\mathbf{W}_{s,k}$. Then $\widetilde{\mathbf{H}}_s$ is in the null space of $\hat{\mathbf{H}}_{s,k}$. Let
   $$s_i = \arg\max_{k\in\Omega}\left(\sum_{s\in\Upsilon} ||\widetilde{\mathbf{H}}_s||_F^2 + ||\widetilde{\mathbf{H}}_k||_F^2\right).$$

   (c) Let $\Omega = \Omega - \{s_i\}$ and $\Upsilon = \Upsilon + \{s_i\}$. Apply the Gram-Schmidt orthogonalization procedure to $\widetilde{\mathbf{H}}_{s_i}$ and get $\widetilde{\mathbf{V}}_{s_i}$. Let $\mathbf{V} = [\mathbf{V}^T \ \widetilde{\mathbf{V}}_{s_i}^T]^T$.

3. Apply the capacity-based suboptimal user selection algorithm to the set Υ, and get the final selected user set and the total throughput.

4.4 Computational Complexity Analysis

Since the primary motivation for the two proposed suboptimal algorithm is their reduced computational complexity, in the section I quantify their complexity and compare with the brute-force approach. The complexity is counted as the number of flops, denoted as ψ. A flop is defined to be a real floating point operation [28]. A real addition, multiplication, or division is counted as one flop. A complex addition and multiplication have two flops and six flops, respectively. Although flop counting cannot characterize the true computational complexity, it captures the order of the computation load, so suffices for the purpose of the complexity analysis in this chapter.

4.4.1 Complexity of Typical Matrix Operations

For an $m \times n$ complex-valued matrix $\mathbf{H} \in \mathbb{C}^{m\times n}$, I first provide the flop count of several matrix operations that are frequently used in the suboptimal user selection algorithm. It is assumed that $K \gg \hat{K}$, $\hat{K}N_r \approx N_t$, and $m \leq n$ in this section.

- Frobenius norm $\|\mathbf{H}\|_F^2$ takes $2mn$ real multiplications and $2mn$ real additions, hence the flop count is $4mn$.

- Gram-Schmidt orthogonalization GSO($\mathbf{H}$) takes $4m^2n - 2mn$ real multiplications; $4m^2n - 2mn$ real additions; and $2mn$ real divisions. The flop count for GSO is $8m^2n - 2mn$.

- Water-filling over n eigenmodes takes up to $\frac{1}{2}(n^2 + 3n)$ real multiplication; $n^2 + 3n$ real additions; and $\frac{1}{2}(n^2 + 3n)$ real divisions. The flop count for water-filling is $2n^2 + 6n$.

- The flop count for SVD of real-valued $m \times n$ ($m \geq n$) matrices is $4m^2n + 8mn^2 + 9n^3$ [28]. For complex-valued $m \times n$ ($m \leq n$) matrices, the flop count is approximated as $24mn^2 + 48m^2n + 54m^3$ by treating every operation as complex multiplication.

4.4.2 Suboptimal User Selection Algorithm I: Capacity-Based Approach

1. $i = 1$: SVD of $\mathbf{H}_k$ has $48N_r^2N_t + 24N_rN_t^2 + 54N_r^3$ flops, water-filling needs $2N_r^2 + 6N_r$ flops, and the calculation of total throughput requires $2N_r$ flops. In total, step 1 has computational complexity $K\left(48N_r^2N_t + 24N_rN_t^2 + 54N_r^3 + 2N_r^2 + 8N_r\right)$.

2. $i \geq 2$:

   For each $k \in \Omega$, to get $\mathbf{T}_k$ by SVD needs $48(i-1)^2N_r^2N_t + 24(i-1)N_rN_t^2 + 54(i-1)^3N_r^3$ flops. To compute $\overline{\mathbf{H}}_k = \mathbf{H}_k\mathbf{T}_k$, the complexity of this multiplication is $8N_tN_r(N_t-(i-1)N_r)$. SVD of $\overline{\mathbf{H}}_k$ introduces $48N_r^2\left(N_t-(i-1)N_r\right) + 24N_r\left(N_t-(i-1)N_r\right)^2 + 54N_r^3$ flops. Water-filling needs $2iN_r(iN_r+3)$ flops, whereas the total throughput calculation has complexity $2iN_r$.

Hence, the flop count of the capacity-based user selection algorithm is

$$
\begin{aligned}
\psi_c \;\overset{(a)}{<}\; & \sum_{i=2}^{\left\lceil \frac{N_t}{N_r} \right\rceil} \Big\{ \left[48i(i-1)^2 + 48i\right] N_r^2 N_t \\
& + \left[24i(i-1) + 32i\right] N_r N_t^2 + \left(54i(i-1)^3 + 54i\right) N_r^3 \\
& + 2i^2N_r^2 + 8iN_r \Big\} \times (K-i+1) \\
& + K\left(48N_r^2N_t + 24N_rN_t^2 + 54N_r^3 + 2N_r^2 + 8N_r\right) \\
\approx\; & \mathcal{O}\left(K\left\lceil \frac{N_t}{N_r}\right\rceil^5 N_r^3\right) \approx \mathcal{O}\left(K\left\lceil \frac{N_t}{N_r}\right\rceil^2 N_t^3\right),
\end{aligned}
\tag{4.3}
$$

where the inequality in (a) is due to the upper bound of $(N_t-(i-1)N_r)$ by N_t in the calculation of $\overline{\mathbf{H}}_k$ and the SVD of $\overline{\mathbf{H}}_k$.

4.4.3 Suboptimal User Selection Algorithm II: Frobenius Norm Approach

1. $i = 1$: The Frobenius norm of K users needs $4KN_rN_t$ flop counts.

2. $i \geq 2$.

   For each $k \in \Omega$, $18(i-1)N_r^2N_t$ flops are needed for $\widetilde{\mathbf{H}}_k = \mathbf{H}_k - \mathbf{H}_k\mathbf{V}^*\mathbf{V}$, which include the flops for both matrix multiplications and additions; $8(i-1)^2N_r^2N_t - 2(i-1)N_rN_t$ flops for $\mathbf{W}_{s_j,k}$; $18(i-1)N_r^2N_t + 4(i-1)N_rN_t$ flops for $\|\mathbf{H}_s - \mathbf{H}_s\mathbf{W}_{s,k}^*\mathbf{W}_{s,k}\|_F^2$; and $4(i-1)N_rN_t$ flops for $\|\widetilde{\mathbf{H}}_k\|_F^2$.

3. The complexity of applying the capacity-based algorithm to the selected $\hat{K} = \left\lceil \frac{N_t}{N_r} \right\rceil$ users is $\mathcal{O}\left\{ \left\lceil \frac{N_t}{N_r} \right\rceil^3 N_t^3 \right\}$, which is not dependent on K, and hence negligible compared to the complexity of steps 1 and 2.

Therefore, the total flops of the norm-based user selection algorithm is

$$\begin{aligned}
\psi_n &\approx \sum_{i=2}^{\left\lceil \frac{N_t}{N_r} \right\rceil} \left\{ \left[8(i-1)^3 + 18(i-1)^2 + 18(i-1) \right] N_r^2N_t + \right. \\
&\quad \left. [2(i-1)^2 + 4(i-1)]N_rN_t \right\} \times (K - i + 1) + 4KN_rN_t \\
&\approx \mathcal{O}\left(K \left\lceil \frac{N_t}{N_r} \right\rceil^4 N_r^2N_t \right) \approx \mathcal{O}\left(K \left\lceil \frac{N_t}{N_r} \right\rceil^2 N_t^3 \right). \qquad (4.4)
\end{aligned}$$

4.4.4 Optimal User Selection Algorithm: Complete Search

In the optimal user selection algorithm, the basestation conducts an exhaustive search over the $\sum_{i=1}^{\left\lceil \frac{N_t}{N_r} \right\rceil} {}_K C_i$ possible user sets. The complexity of this complete search method is

$$
\begin{aligned}
\psi_{cs} &\overset{(a)}{\geq} {}_K C_{\left\lceil \frac{N_t}{N_r} \right\rceil} \left\lceil \frac{N_t}{N_r} \right\rceil \left[\left(48 \left(\left\lceil \frac{N_t}{N_r} \right\rceil - 1 \right)^2 + 8 \right) N_r^2 N_t + 24 \left(\left\lceil \frac{N_t}{N_r} \right\rceil - 1 \right) N_r N_t^2 \right. \\
&\quad \left. + \left(54 \left(\left\lceil \frac{N_t}{N_r} \right\rceil - 1 \right)^3 + 2 \left\lceil \frac{N_t}{N_r} \right\rceil^2 + 126 \right) N_r^3 + 8 \left\lceil \frac{N_t}{N_r} \right\rceil N_r \right] \\
&\approx \mathcal{O} \left({}_K C_{\left\lceil \frac{N_t}{N_r} \right\rceil} \left\lceil \frac{N_t}{N_r} \right\rceil N_t^3 \right) \qquad (4.5)
\end{aligned}
$$

where the inequality in (a) holds because only the case of picking $\hat{K} = \left\lceil \frac{N_t}{N_r} \right\rceil$ out of K users is considered to simplify the complexity analysis.

In summary, the proposed two suboptimal user selection algorithms have only a fraction of the complexity of the complete search method approximately equal to

$$
\eta \approx \frac{K \left\lceil \frac{N_t}{N_r} \right\rceil}{{}_K C_{\left\lceil \frac{N_t}{N_r} \right\rceil}}. \qquad (4.6)
$$

Both the capacity-based and the Frobenius norm-based algorithms have linear complexity with K, because no more than $\hat{K}K$ user sets need to be searched over. The norm-based algorithm has slightly lower complexity than the capacity-based one because SVD is less frequently used in the norm-based algorithm. In my Matlab 7.0 implementation of the two proposed suboptimal algorithms, I observed that both algorithms take tens to hundreds of milliseconds (on a Pentium M 1.6 GHz PC) to select a user set, and the CPU run time is linear in the number of users. Further, the norm-based algorithm runs roughly two times faster than the capacity-based algorithm, for systems with a large number of users.

4.5 Simulation Results

In this section, I compare the performance of the following algorithms:

- iterative water-filling for dirty paper coding [40] (DPC),

- optimal user selection by complete search (BD Optimal),
- capacity-based user selection algorithm (BD c-algorithm),
- Frobenius norm-based user selection algorithm (BD n-algorithm),
- round-robin algorithm for $\hat{K}$ simultaneous users (BD no selection).

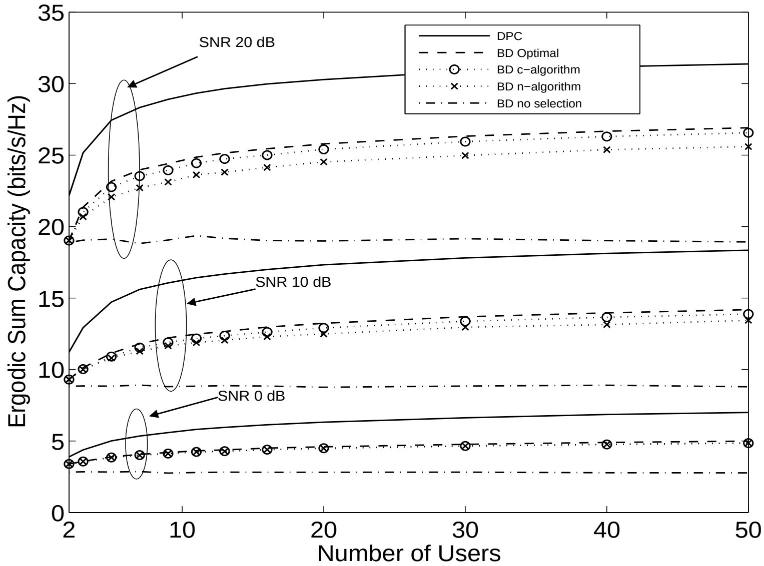

Figure 4.1: Ergodic sum capacity vs. the number of users. $N_t = 4$ and $N_r = 2$.

Figs. 4.1–4.3 show the ergodic sum capacity (averaged over 1000 channel realizations) vs. the number of users for $(N_t = 4, N_r = 2)$, $(N_t = 12, N_r = 4)$, and $(N_t = 8, N_r = 1)$ MIMO systems, where $\hat{K} = 2, 3, 8$, respectively. Fig. 4.3 shows only up to 16 users in the system due to the complexity of the exhaustive search method. In all simulations, the capacity-based and the norm-based user selection algorithms achieve around 95% of the total throughput of the complete search method. The capacity-based algorithm performs slightly better than the norm-based algorithm because its user selection criterion is directly based on the

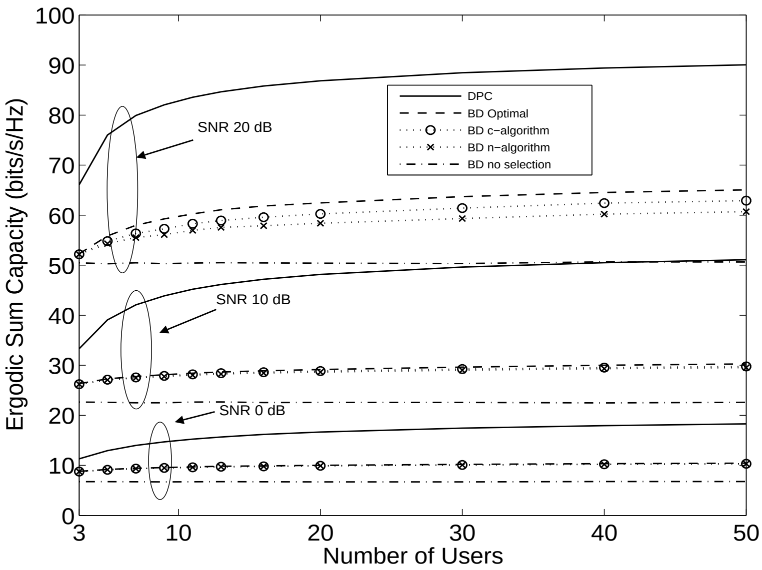

Figure 4.2: Ergodic sum capacity vs. the number of users. $N_t = 12$ and $N_r = 4$.

sum capacity. For low SNRs, e.g. $SNR = 0$ dB, the proposed algorithms achieve almost the same sum capacity as the exhaustive search method. This is true because beamforming to the user with the highest capacity, which is the first step in the capacity-based user selection algorithm, is asymptotically optimal for sum capacity of BD in the low SNR regime. For high SNRs, although the proposed algorithms may not always find the optimal user set due to their reduced search range, they can still achieve a significant part of the ergodic sum capacity of the exhaustive search method because both algorithms greedily try to maximize the total throughput. The sum capacity achieved by dirty paper coding (DPC) is also plotted in Figs. 4.1–4.3. In general, DPC achieves higher sum capacity than BD because DPC is optimal for the sum capacity of MIMO broadcast channels [88][93]. BD, however, still achieves a significant part of the DPC sum capacity. Further, the low complexity property of the BD algorithm (e.g. without the requirement for successively encoding and decoding user signals) makes it more suitable for practical implementations.

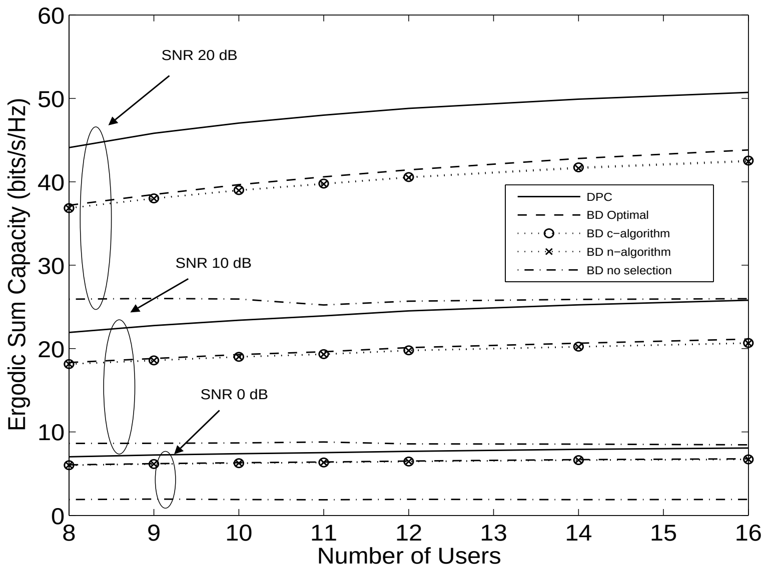

Figure 4.3: Ergodic sum capacity vs. the number of users. $N_t = 8$ and $N_r = 1$.

It is also noticeable that the throughput does not increase significantly when the number of total users is already large, for dirty paper coding, block diagonalization, and the proposed user selection algorithms. In fact, it has been proven that the sum capacity of dirty paper coding scales as $\log \log K$ for large K [67], where K denotes the total number of users in the system. Hence, most of the multiuser diversity can be achieved with a medium number of users. This motivates user grouping in practical systems to reduce the computational complexity with small throughput degradation compared to the optimal solution.

4.6 Conclusion

Two suboptimal user selection algorithms for multiuser MIMO systems with block diagonalization are proposed in this chapter. The goal is to select a subset of users to maximize the total throughput while keeping the complexity low. The brute-force complete search method yields the optimal user set, but the complexity is roughly $\mathcal{O}\{K^{\hat{K}}\}$, where K is the total number of users and $\hat{K}$ is the maximum number of simultaneous users. Simulations show that the proposed capacity-based and norm-based user selection algorithms achieve about 95% of the sum capacity whereas their complexity is $\mathcal{O}\{K\}$. Although the proposed user selection algorithms are greedy in nature, they can be easily extended to incorporate fairness, e.g. the rate proportional fairness in [72].

Chapter 5

Conclusion

As the data rate requirements increase for media-rich communications, channel-aware adaptive resource allocation is becoming more critical to system performance. Enabled by multicarrier modulation and multi-antenna technologies, multiple parallel channels can be created in either the frequency or spatial domain. Compared to single channel systems, resource allocation in multiuser multichannel systems is more challenging because of the additional degree of freedom for resources. In this dissertation, I study the performance of adaptive resource allocation in multiuser multichannel wireless communication systems [70] [72] [76] [77]. Adaptive resource allocation can usually be formulated as an optimization problem. The optimal solution is typically very difficult to obtain due to the large number of variables. Further, the wireless channel is time-varying, so adaptive resource allocation should be performed to match the channel variations. Due to the need for real-time channel adaptation, low complexity algorithms are mandatory for any practical implementation of resource allocation. This dissertation presents low complexity resource allocation algorithms for both multiuser OFDM systems [70] [72] and multiuser MIMO systems [74] [75].

5.1 Summary of Contributions

The first contribution of this dissertation is an optimization framework for multiuser OFDM systems, in which the tradeoff between total throughput and user fairness can be easily evaluated. In downlink multiuser OFDM systems, data streams from multiple users are multiplexed into each OFDM symbol. Hence, the basestation can serve multiple users simultaneously. While the channel conditions of different users are largely independent due to users' different locations, multiuser OFDM can exploit the multiuser diversity to improve the system performance. Previous works either maximize the total system throughput without consideration of user data fairness [36] [48] or provide maximum fairness among users with the sacrifice of system throughput [64]. In this dissertation, I propose to maximize the total throughput while maintaining user data rates proportional [70] [72]. With the proportional rate constraints, the data rate fairness among users can be flexibly controlled by a set of parameters. Further, the total system throughput is also adjustable by varying the proportional fairness parameters.

The formulated optimization problem for adaptive resource allocation in multiuser OFDM systems includes both continuous and binary variables and, hence, is difficult to solve. To lower the computational complexity, I propose a suboptimal algorithm that separates the subchannel and power allocation among users. By doing so, the number of variables each step has to optimized is almost reduced by half. First, the subchannels are allocated among users assuming equal power is distributed in each subchannel. Second, transmit power is optimally allocated among users and within each individual user according to the subchannel allocation scheme. In general, the optimal power allocation is the solution to a set of nonlinear equations, which can be found iteratively with the Newton-Raphson method. Due to the special characteristics of the nonlinear equations, it is shown in Appendix B that a computational complexity of $\mathcal{O}(K)$ is necessary to obtain the optimal power allocation, where K is the total number of users in the system. Two specially cases where the iterative Newton-Raphson method can be avoided are also analyzed in the dissertation. The Matlab implementation of the proposed algorithm is available at http://www.ece.utexas.edu/~bevans/papers/2005/multiuserOFDM/.

Simulations show that the proposed suboptimal algorithm can achieve about 95% of the optimal total throughput in a two-user ten-subchannel OFDM system, while reducing the complexity from exponential to linear in the number of subchannels. It is also shown that in a system of 16 users, the proposed optimal power allocation achieves 17% more capacity over fixed TDMA than the *max-min* method with equal power allocation in [64]. Further, with the proposed resource allocation algorithm, the sum capacity is distributed more fairly and flexibly among users than the sum capacity maximization algorithm in [36]. Since the proposed adaptive resource allocation applies to each channel realization, proportional data rates can be assured among users for any time scale of interest. Further, the proposed optimization framework allows different users request variable priorities of their services with different prices, which is suitable for systems with heterogenous user services.

The second contribution is an analysis on the sum capacity of multiuser MIMO broadcast channels with block diagonalization [76] [77]. Block diagonalization [14] [80] is a linear precoding method in downlink multiuser MIMO systems. Each user's signal is pre-multiplied by a precoding matrix before transmission. The precoding matrix of each user lies in the null space of all other users' channels, hence inter-user interference is completely eliminated if the channel state information of all users is available at the basestation. The effective channel for every user, therefore, is a point-to-point MIMO channel, rendering a simpler receiver structure. The resource allocation in multiuser MIMO systems aims to distribute the transmit power optimally such that a certain objective function, e.g. the sum capacity studied in this dissertation, is optimized. Although it has been shown that dirty paper coding is optimal for the sum capacity [88] of downlink multiuser MIMO systems, cost-effective coding schemes that approach the dirty paper coding sum capacity, however, are still unavailable. Therefore, block diagonalization, as a practically implementable technique for downlink multiuser MIMO systems, deserves a thorough study.

In this dissertation, I formulate an optimization problem for block diagonalization with both transmitter precoding and receiver post-processing to maximize the total system throughput [76]. While the optimal post-processing matrices at the receivers are difficult to obtain, I restrict myself to a set of selection matrices [76]. The selection matrices allow each user to select a subset of receive antennas to use. Although for a particular user, his/her throughput may be reduced by using fewer receive antennas, the system throughput can increase because additional spatial dimension is saved for other users. Further, since the precoding and post-processing matrices are designed at the basestation, the post-processing matrices should be conveyed to their own users, which increases system overhead. Due to the simple structure of the selection matrices, less system overhead is required for the selection matrices than the optimal post-processing matrices. Simulations show that with receive antenna selection, the total throughput of a downlink multiuser MIMO system with block diagonalization can be further improved.

I also analytically compare block diagonalization with dirty paper coding. It is shown that for a set of fixed channels, 1) if the user channels are orthogonal to each other, then block diagonalization achieves the same sum capacity as dirty paper coding; 2) if the user channels lie in the same subspace, then the gain of dirty paper coding over block diagonalization can be upper bounded by the minimum of the number of transmit and receive antennas. The ergodic sum capacity of block diagonalization in a Rayleigh fading channel is also studied. Simulations show that block diagonalization can achieve a significant part of the dirty paper coding sum capacity. An upper bound on the ergodic sum capacity gain of dirty paper coding over block diagonalization is proposed for easy estimation of the gap between the sum capacity of dirty paper coding and block diagonalization.

The third contribution of this dissertation consists of two low complexity user selection algorithms in downlink multiuser MIMO systems with block diagonalization [74] [75]. Due to the zero inter-user interference requirement imposed by block diagonalization, the maximum number of simultaneously supportable users is limited. For systems with a large number of users, optimal user set selection is necessary to increase the system throughput by exploiting the multiuser diversity. While the complete search method yields the optimal user set, its computational complexity is prohibitive if the number of users is large.

In this dissertation, I propose two suboptimal user selection algorithms, both of which aim to select a subset of users such that the total throughput is nearly maximized. The user selection criterion of the first algorithm is the total system throughput, and it greedily selects a user that provides the best total throughput along with those already scheduled users. While the computationally intensive matrix operations, such as Singular Value Decomposition, are frequently used in the first algorithm, the second algorithm selects users based on the channel energy, hence further reduces the computational complexity. It is shown that both algorithms have a linear computational complexity in the number of users and achieve around 95% of the total throughput of the complete search method. The first algorithm achieves slightly higher total throughput than the second algorithm since it directly uses the total throughput as the user selection criterion. The computational complexity, however, favors the second algorithm because Singular Value Decomposition is avoided as much as possible.

5.2 Future Research

In this section, I propose several future research topics for multicarrier and/or multi-antenna wireless systems, potentially for other researchers interested in this area.

- *Implementing Adaptive Resource Allocation with Proportional Data Rate Constraints in Multiuser OFDM systems*

  Several aspects on the proposed proportional data rate resource allocation algorithm need to be investigated before practical implementation. For example, the set of system parameters $\{\gamma_k\}_{k=1}^{K}$ should be determined based on users' target applications. A simple example is to let users choose their γ_k from a set pre-determined discrete values to represent their service priorities. The basestation, after receiving users' requests, can grant a subset of users for transmission based on the available resources. Other methods to determine the proportional data rate constraints need further study. Another implementation issue is on the solution to a system of non-linear equations, which is required for the optimal power distribution among users. In practical systems, the channel-to-noise ratio of different users can vary significantly, largely due to the different user locations and pathloss, which could make the system of non-linear equations ill-conditioned. Hence efficient and accurate implementation of the proposed algorithm is very important to obtain the optimal power allocation. Grouping users with similar channel-to-noise ratios and performing the proposed algorithm to each user group is a possible solution to make the system of non-linear equations less ill-conditioned, as the channel-to-noise ratio in each user group is about the same value. Another method [98] to lower the computational complexity is to allocate the subchannels such that the system of non-linear equations is reduced to the linear case as discussed in Section 2.4.2.

- *Adaptive Resource Allocation in Multiuser MIMO-OFDM Systems*

  The next generation of cellular systems is likely to be OFDM based with multiple antennas [65]. With OFDM, the wideband is divided into a number of parallel subchannels in the frequency domain. With multiple antennas, multiple users can be supported for simultaneous transmissions in each frequency subchannel. Resource allocation in multiuser MIMO-OFDM systems [109] is likely to be even more challenging because the limited resource shall be optimized in multiple dimensions. If block diagonalization is employed on each frequency subchannel, then the user selection algorithms proposed in Chapter 4 can be potentially used on a per-subchannel basis to select a good user set to optimize the total throughput. Power allocation shall be designed carefully as it involves

- Power allocation among different frequency subchannels;
- Power allocation among users in each frequency subchannel;
- Power allocation in the multi-dimensional spatial domain for each user.

Data rate fairness among users demands further study for adaptive resource allocation in multiuser MIMO-OFDM systems.

- *Semi-Adaptive Resource Allocation for Multiuser OFDM Systems*

  For multiuser OFDM systems discussed in this dissertation, it is assumed that adaptive resource allocation is performed as soon as the user channels are changed. The system overhead for conveying the channel state information from the users to the basestation and the resource allocation schemes from the basestation to the users has not been incorporated into the problem formulation. While this system overhead is negligible for slow varying channels, it may be large for systems with fast channel variations. One possible solution to reduce the system overhead is a semi-adaptive resource allocation, where the subchannel allocation among users is performed once and remains fixed throughout the whole transmission period, hence the subchannel allocation scheme only needs to be conveyed to users once. The subchannel allocation can be performed based on, e.g., the average channel condition of all users and/or the data rates, bit error rates, and service priorities required by different users. The subchannel allocation shall be carried out when one user's service is fulfilled or new service requests are admitted. Power allocation among the subchannels assigned to each user can still be adapted to the channel variations. Further, since the subchannel allocation is fixed, the resource allocation algorithm is much easier to realize in practical systems.

- *Maximizing Ergodic Sum Capacity with Ergodic Proportional Rate Constraints in Multiuser OFDM Systems*

  The adaptive resource allocation in multiuser OFDM systems proposed in this dissertation is a static algorithm, i.e. for each channel realization, the algorithm should be carried out and the proportional rate constraints are strictly applied for each channel realization. Although the proposed algorithm guarantees proportional rates in any time scale, the ergodic sum capacity is not necessarily optimized. A future research is to optimize the ergodic sum capacity while maintaining users' ergodic rates proportional. Thus, multiuser diversity can be even further exploited to improve the ergodic sum capacity.

- *Impact of Imperfect Channel State Information for Adaptive Resource Allocation*

  Users' channel state information (CSI) is required at the basestation for adaptive resource allocation in both multiuser OFDM and multiuser MIMO systems. In this dissertation, it is assumed that channel station information is perfectly known at the basestation through a separate feedback channel. The CSI is usually estimated at the receivers and, hence, prone to estimation errors. Moreover, feedback delays may cause outdated CSI used by the adaptive resource allocation algorithm. The impact of imperfect CSI to the system performance with adaptive resource allocation needs further study. Channel prediction [71] [97] and limited feedback techniques [13] [50] can be combined with adaptive resource allocation to combat the effects of feedback delay and reduce the amount of feedback information.

- *Fixed-Point Implementation of Adaptive Resource Allocation Algorithms*

  Since adaptive resource allocation should be performed frequently to match the wireless channel variations, low complexity algorithms are desirable for practical implementations. However, even the low complexity algorithms require a certain amount of computational efforts. For example, the optimal power allocation for multiuser OFDM systems proposed in the dissertation requires the Newton-Raphson method to solve a set of nonlinear equations iteratively and Singular Value Decomposition is necessary for multiuser MIMO systems with block diagonalization to find the spatial eigenmodes of different users. Currently these algorithms are implemented with floating-point arithmetic. Future research shall map the proposed low complexity algorithms into fixed-point implementations and lower the memory footprint.

- *Sum Capacity of MIMO Broadcast Channels with Channel Frobenius Norm Constraints*

  It has been proven that the sum capacity of a MIMO broadcast channel can be achieved with dirty paper coding [88]. The duality results in [88] show that for a given set of channels $\{\mathbf{H}_k\}_{k=1}^{K}$, the sum capacity of a Gaussian broadcast MIMO channel can found by solving the following problem:

$$\max_{\{\mathbf{Q}_k\}} \log \left| \mathbf{I} + \sum_{k=1}^{K} \mathbf{H}_k^* \mathbf{Q}_k \mathbf{H}_k \right| \tag{5.1}$$
$$\text{subject to} \sum_{k=1}^{K} \mathrm{Tr}(\mathbf{Q}_k) \leq P$$

  where $\mathbf{H}_k$ is the MIMO channel matrix for user k; $\mathbf{Q}_k$ is the signal covariance matrix for user k in the dual multiple access channel; P is the total transmit power constraint; and K is the total number of users.

For future work, I propose to jointly optimize the user channel matrices $\{\mathbf{H}_k\}_{k=1}^K$ and their signal covariance $\{\mathbf{Q}_k\}_{k=1}^K$ with a set of channel Frobenius norm constraints as following

$$\max_{\{\mathbf{H}_k\}} \max_{\{\mathbf{Q}_k\}} \log \left| \mathbf{I} + \sum_{k=1}^{K} \mathbf{H}_k^* \mathbf{Q}_k \mathbf{H}_k \right| \tag{5.2}$$

$$\text{subject to} \quad \sum_{k=1}^{K} \mathrm{Tr}(\mathbf{Q}_k) \leq P$$

$$||\mathbf{H}_k||_F^2 \leq W_k \quad \text{for } k = 1, 2, \cdots, K.$$

The motivations for optimizing the channel jointly with the input covariance matrix under the transmit power and channel power constraints are: 1) for the class of power-constrained channels, an upper bound on the MIMO channel capacity can be found; and 2) the characteristics of the channels providing the maximum capacity can be obtained, which may be used to direct the adaptive antenna array configuration if possible. Initial results for some special cases can be found in [73]. The optimal solution in general is still for future study.

Appendix A

Convexity of the Objective Function in the Relaxed Optimization Problem in Section 2.3

First consider the following function

$$f\left(\rho_{k,n}, p_{k,n}\right) = \rho_{k,n} \log_2 \left(1 + \frac{p_{k,n} H_{k,n}}{\rho_{k,n}}\right) \tag{A.1}$$

where $H_{k,n} = \frac{h_{k,n}^2}{N_0 \frac{B}{N}}$.

The Jacobian of $f(\rho_{k,n}, p_{k,n})$ is calculated as

$$\nabla f\left(\rho_{k,n}, p_{k,n}\right) = \begin{bmatrix} \log_2\left(1 + \frac{p_{k,n} H_{k,n}}{\rho_{k,n}}\right) - \frac{1}{\ln 2} \frac{p_{k,n} H_{k,n}}{\rho_{k,n} + p_{k,n} H_{k,n}} \\ \frac{1}{\ln 2} \frac{\rho_{k,n} H_{k,n}}{\rho_{k,n} + p_{k,n} H_{k,n}} \end{bmatrix}. \tag{A.2}$$

The Hessian of $f(\rho_{k,n}, p_{k,n})$ is calculated as

$$\nabla^2 f\left(\rho_{k,n}, p_{k,n}\right) = \frac{1}{\ln 2} \frac{p_{k,n} H_{k,n}^2}{\left(\rho_{k,n} + p_{k,n} H_{k,n}\right)^2} \begin{bmatrix} -\frac{p_{k,n}}{\rho_{k,n}} & 1 \\ 1 & -\frac{\rho_{k,n}}{p_{k,n}} \end{bmatrix}. \tag{A.3}$$

Since $\rho_{k,n}, p_{k,n}, H_{k,n}$ are all positive, it is not difficult to see that the Hessian of $f(\rho_{k,n}, p_{k,n})$ is negative semi-definite and hence $f(\rho_{k,n}, p_{k,n})$ is concave. Thus the Hessian of $-f(\rho_{k,n}, p_{k,n})$ is positive semi-definite and $-f(\rho_{k,n}, p_{k,n})$ is convex. The objective function in (2.4) can be expressed as

$$\sum_{k=1}^{K}\sum_{n=1}^{N}\frac{1}{N}\left(-f\left(\rho_{k,n}, p_{k,n}\right)\right) \tag{A.4}$$

and thus is a summation of a set of convex functions, which is also convex.

Appendix B

Newton-Raphson Method for Nonlinear Equations

In the following, I outline the major steps to find the power allocation with Newton-Raphson method.

Denote the variables as

$$\mathbf{P} = [P_{1,tot} \;\; P_{2,tot} \;\; \ldots \;\; P_{K,tot}]^{+} \tag{B.1}$$

where $[\bullet]^{+}$ represents the operation of matrix transpose.

Also define a square system of equations $\mathbf{g}(\mathbf{P}) = \mathbf{0}$ where

$$g_1(\mathbf{P}) = \sum_{k=1}^{K} P_{k,tot} - P_{total} = 0 \tag{B.2}$$

and

$$\begin{aligned} g_k(\mathbf{P}) \;\; = \;\; & \frac{N_1}{N}\left(\log_2\left(1 + H_{1,1}\frac{P_{1,tot} - V_1}{N_1}\right) + \log_2 W_1\right) \\ & -\frac{\gamma_1}{\gamma_k}\cdot\frac{N_k}{N}\left(\log_2\left(1 + H_{k,1}\frac{P_{k,tot} - V_k}{N_k}\right) + \log_2 W_k\right) = 0 \end{aligned} \tag{B.3}$$

for $k = 2, ..., K$.

Denote $\triangle\mathbf{P}$ as the update direction. The major step in Newton-Raphson method is to solve the following equation to find $\triangle\mathbf{P}$

$$\mathcal{J}(\mathbf{P}) \triangle \mathbf{P} = -\mathbf{g}(\mathbf{P}) \tag{B.4}$$

and update $\mathbf{P}$ as

$$\mathbf{P} = \mathbf{P} + \triangle\mathbf{P} \tag{B.5}$$

where

$$\mathcal{J}(\mathbf{P}) = \begin{bmatrix} \frac{\partial g_1}{\partial P_{1,tot}} & \frac{\partial g_1}{\partial P_{2,tot}} & \cdots & \frac{\partial g_1}{\partial P_{K,tot}} \\ \frac{\partial g_2}{\partial P_{1,tot}} & \frac{\partial g_2}{\partial P_{2,tot}} & \cdots & \frac{\partial g_2}{\partial P_{K,tot}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_K}{\partial P_{1,tot}} & \frac{\partial g_K}{\partial P_{2,tot}} & \cdots & \frac{\partial g_K}{\partial P_{K,tot}} \end{bmatrix} \tag{B.6}$$

is the Jacobian matrix of $\mathbf{g}(\mathbf{P})$ evaluated at $\mathbf{P}$.

It is true that the computational complexity is still high since in general, a matrix inversion or LU decomposition has to be performed in order to get $\triangle\mathbf{P}$ each iteration. Fortunately, the Jacobain matrix of $\mathbf{g}(\mathbf{P})$ has a good structure which can be fully utilized to reduce the computational complexity.

$$\begin{aligned} \mathcal{J}(\mathbf{P}) &= \begin{bmatrix} \frac{\partial g_1}{\partial P_{1,tot}} & \frac{\partial g_1}{\partial P_{2,tot}} & \cdots & \frac{\partial g_1}{\partial P_{K,tot}} \\ \frac{\partial g_2}{\partial P_{1,tot}} & \frac{\partial g_2}{\partial P_{2,tot}} & \cdots & \frac{\partial g_2}{\partial P_{K,tot}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_K}{\partial P_{1,tot}} & \frac{\partial g_K}{\partial P_{2,tot}} & \cdots & \frac{\partial g_K}{\partial P_{K,tot}} \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 & \cdots & \cdots & 1 \\ \frac{H_{1,1}}{N\ln 2}\frac{1}{1+H_{1,1}\frac{P_{1,tot}-V_1}{N_1}} & -\frac{\gamma_1}{\gamma_2}\frac{H_{2,1}}{N\ln 2}\frac{1}{1+H_{2,1}\frac{P_{2,tot}-V_2}{N_2}} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{H_{1,1}}{N\ln 2}\frac{1}{1+H_{1,1}\frac{P_{1,tot}-V_1}{N_1}} & 0 & \cdots & 0 & -\frac{\gamma_1}{\gamma_K}\frac{H_{K,1}}{N\ln 2}\frac{1}{1+H_{K,1}\frac{P_{K,tot}-V_K}{N_K}} \end{bmatrix}. \end{aligned} \tag{B.7}$$

Every except the first row of the Jacobian matrix has only two non-zero entries. By substitution, $\triangle\mathbf{P}$ can be calculated with the complexity of $\mathcal{O}(K)$.

Appendix C

Derivation of the Maximum Deviation in Section 2.5.3

Notice that $\sum_{k=1}^{K} \widetilde{R}_{k,i} = 1$ and $\sum_{k=1}^{K} \widetilde{\gamma}_k = 1$. Hence there must exist some k, such that $\widetilde{R}_{k,i} - \widetilde{\gamma}_k < 0$. Without loss of generality, assume that $\widetilde{R}_{k,i} - \widetilde{\gamma}_k < 0$ for $k = 1, 2, ..., k^*$ and $\widetilde{R}_{k,i} - \widetilde{\gamma}_k \geq 0$ for $k = k^* + 1, ..., K$, then the objective function can be written as

$$\begin{aligned}
& \sum_{k=1}^{K} |\widetilde{R}_{k,i} - \widetilde{\gamma}_k| = \sum_{k=1}^{k^*} \left(\widetilde{\gamma}_k - \widetilde{R}_{k,i}\right) + \sum_{k=k^*+1}^{K} \left(\widetilde{R}_{k,i} - \widetilde{\gamma}_k\right) \\
= \quad & \sum_{k=1}^{k^*} \widetilde{\gamma}_k - \sum_{k=k^*+1}^{K} \widetilde{\gamma}_k + \sum_{k=k^*+1}^{K} \widetilde{R}_{k,i} - \sum_{k=1}^{k^*} \widetilde{R}_{k,i} \\
= \quad & \sum_{k=1}^{k^*} \widetilde{\gamma}_k + \sum_{k=k^*+1}^{K} \widetilde{\gamma}_k - 2 \sum_{k=k^*+1}^{K} \widetilde{\gamma}_k + \sum_{k=k^*+1}^{K} \widetilde{R}_{k,i} + \sum_{k=1}^{k^*} \widetilde{R}_{k,i} - 2 \sum_{k=1}^{k^*} \widetilde{R}_{k,i} \\
\leq \quad & 1 - 2 \min_k \widetilde{\gamma}_k + 1 \\
= \quad & 2 - 2 \min_k \widetilde{\gamma}_k \qquad \text{(C.1)}
\end{aligned}$$

Let $\text{argmin}_k \widetilde{\gamma}_k = k_{min}$, then one maximizer of the objective function in (2.24) is $\widetilde{R}_{k,i} = 1$ for $k = k_{min}$ and $\widetilde{R}_{k,i} = 0$ else.

Appendix D

Proof of Lemma 2 in Section 3.4

Proof: Let the SVD of $\mathbf{H}_i$ be

$$\mathbf{H}_i = \mathbf{U}_i \mathbf{\Lambda}_i \mathbf{V}_i^* \tag{D.1}$$

where $\mathbf{U}_i$ is of size $N_r \times N_r$ and $\mathbf{U}_i \mathbf{U}_i^* = \mathbf{I}$; $\mathbf{\Lambda}_i = \text{diag}\{\lambda_{i,1}, \lambda_{i,2}, \cdots, \lambda_{i,N_r}\}$ is a diagonal matrix of size $N_r \times N_r$; and $\mathbf{V}_i$ is of size $N_t \times N_r$ and $\mathbf{V}_i^* \mathbf{V}_i = \mathbf{I}$. Furthermore, $\mathbf{H}_i^* \mathbf{H}_i \mathbf{V}_i = \mathbf{V}_i \mathbf{\Lambda}_i^2$. For $i \neq j$, $\mathbf{V}_j^* \mathbf{V}_i = (\mathbf{\Lambda}_j^2)^{-1} \mathbf{V}_j^* \mathbf{H}_j^* \mathbf{H}_j \mathbf{H}_i^* \mathbf{H}_i \mathbf{V}_i (\mathbf{\Lambda}_i^2)^{-1} = 0$ because $\mathbf{H}_j \mathbf{H}_i^* = 0$.

Let $\mathbf{H} = [\mathbf{H}_1^* \ \mathbf{H}_2^* \ \cdots \ \mathbf{H}_K^*]^*$, then the SVD of $\mathbf{H}$ can be expressed as $\mathbf{H} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^*$, where $\mathbf{U} = \text{bdiag}\{\mathbf{U}_1, \mathbf{U}_2, \cdots, \mathbf{U}_K\}$ is an unitary block diagonal matrix of size $KN_r \times KN_r$; $\mathbf{\Lambda} = \text{bdiag}\{\mathbf{\Lambda}_1, \mathbf{\Lambda}_2, \cdots, \mathbf{\Lambda}_K\}$ is a diagonal matrix of size $KN_r \times KN_r$; and $\mathbf{V} = [\mathbf{V}_1 \ \mathbf{V}_2 \ \cdots \ \mathbf{V}_K]$ is of size $N_t \times KN_r$ and $\mathbf{V}^*\mathbf{V} = \mathbf{I}$ because $\mathbf{V}_j^* \mathbf{V}_i = 0$ for $i \neq j$ and $\mathbf{V}_j^* \mathbf{V}_i = \mathbf{I}$ for $i = j$.

The capacity of the point-to-point MIMO channel $\mathbf{H}$ can be regarded as an upper bound on the sum capacity of the broadcast channel because user cooperation is allowed with $\mathbf{H}$. Hence

$$\begin{aligned} C_{DPC}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2) &\leq C_{coop}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2) && \text{(D.2)} \\ &= \sum_{i=1}^{K} \sum_{n=1}^{N_r} \log\left(1 + \frac{P_{i,n}}{\sigma^2} \lambda_{i,n}^2\right) && \text{(D.3)} \end{aligned}$$

where $P_{i,n}$ is the power allocated to user i's nth eigenmode and $P_{i,n}$ is obtained by the water-filling algorithm with total power constraint $\sum_{i=1}^{K}\sum_{n=1}^{N_r} P_{i,n} = P$.

On the other hand, since $\mathbf{V}_j^*\mathbf{V}_i = 0$ for $i \neq j$, we have $\mathbf{H}_j^*\mathbf{V}_i = 0$ for $i \neq j$. Thus we can set $\mathbf{T}_j = \mathbf{V}_j$ to satisfy the null constraint in (3.2). Notice the effective channel $\overline{\mathbf{H}}_j = \mathbf{H}_j\mathbf{V}_j$ has the same singular values as $\mathbf{H}_j$. Hence

$$\begin{aligned} C_{BD}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2) &\geq \max_{\{\mathbf{Q}_j:\, \mathbf{Q}_j \geq 0,\, \sum_{j\in\mathcal{K}} \mathrm{Tr}(\mathbf{Q}_j) \leq P\}} \sum_{j\in\mathcal{K}} \log\left|\mathbf{I} + \frac{1}{\sigma^2}\overline{\mathbf{H}}_j\mathbf{Q}_j\overline{\mathbf{H}}_j^*\right| && \text{(D.4)} \\ &= \sum_{i=1}^{K}\sum_{n=1}^{N_r} \log\left(1 + \frac{P_{i,n}}{\sigma^2}\lambda_{i,n}^2\right). && \text{(D.5)} \end{aligned}$$

With (D.3), (D.5), and the fact that $C_{DPC} \geq C_{BD}$, we have $C_{DPC} = C_{BD}$ as the conditions in Lemma 3 are satisfied. □

Appendix E

Proof of Lemma 3 in Section 3.4

Proof: Let $\mathbf{E} = [\mathbf{e}_1^* \ \mathbf{e}_2^* \ \cdots \mathbf{e}_{N_r}^*]^*$ be a basis in $\mathcal{W}$, which is the row vector space spanned by $\{\mathbf{H}_i\}_{k=1}^K$, where $\mathbf{e}_i$ is of size $1 \times N_t$. Hence $\mathbf{E}\mathbf{E}^* = \mathbf{I}$. Let the SVD of $\mathbf{H}_i$ be $\mathbf{H}_i = \mathbf{U}_i \mathbf{\Lambda}_i \mathbf{V}_i^*$. There exists a unitary matrix $\mathbf{F}_i$ of size $N_r \times N_r$ such that $\mathbf{V}_i^* = \mathbf{F}_i \mathbf{E}$. Then $\mathbf{H}_i = \mathbf{U}_i \mathbf{\Lambda}_i \mathbf{F}_i \mathbf{E}$. Denote $\mathbf{H}_i^{(\mathcal{W})} = \mathbf{U}_i \mathbf{\Lambda}_i \mathbf{F}_i$, it is easy to see that $\mathbf{H}_i^{(\mathcal{W})}$ has the same singular values of $\mathbf{H}_i$. Hence

$$\begin{aligned}
C_{DPC}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2) &= \max_{\{\mathbf{S}_j:\, \mathbf{S}_j \geq 0, \sum_{j=1}^K \mathrm{Tr}(\mathbf{S}_j) \leq P\}} \log \left| \mathbf{I} + \frac{1}{\sigma^2} \sum_{j=1}^K \mathbf{H}_j^* \mathbf{S}_j \mathbf{H}_j \right| \\
&= \max_{\{\mathbf{S}_j:\, \mathbf{S}_j \geq 0, \sum_{j=1}^K \mathrm{Tr}(\mathbf{S}_j) \leq P\}} \log \left| \mathbf{I} + \frac{1}{\sigma^2} \sum_{j=1}^K \mathbf{E}^* (\mathbf{H}_i^{(\mathcal{W})})^* \mathbf{S}_j \mathbf{H}_i^{(\mathcal{W})} \mathbf{E} \right| \\
&= \max_{\{\mathbf{S}_j:\, \mathbf{S}_j \geq 0, \sum_{j=1}^K \mathrm{Tr}(\mathbf{S}_j) \leq P\}} \log \left| \mathbf{I} + \frac{1}{\sigma^2} \sum_{j=1}^K (\mathbf{H}_i^{(\mathcal{W})})^* \mathbf{S}_j \mathbf{H}_i^{(\mathcal{W})} \right| \\
&= C_{DPC}(\mathbf{H}_{1,\cdots,K}^{(\mathcal{W})}, P, \sigma^2). \qquad \text{(E.1)}
\end{aligned}$$

Since the size of $\mathbf{H}_i^{(\mathcal{W})}$ (for $i = 1, 2, \cdots, K$) is $N_r \times N_r$, analogous to Theorem 1 in [39], we can obtain

$$\begin{aligned} C_{DPC}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2) &= C_{DPC}(\mathbf{H}^{(\mathcal{W})}_{1,\cdots,K}, P, \sigma^2) \\ &\leq N_r \log\left(1 + \frac{P}{\sigma^2}\lambda^2_{max}\right) \end{aligned} \tag{E.2}$$

where $\lambda_{max} = \max\limits_{1 \leq i \leq K, 1 \leq n \leq N_r} \lambda_{i,n}$ where $\lambda_{i,n}$ is the ith user's nth singular value.

On the other hand, if $\text{span}(\mathbf{H}_1) = \text{span}(\mathbf{H}_2) = \cdots = \text{span}(\mathbf{H}_K)$ and only one user is supported with BD, we have

$$C_{BD}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2) = C_{TDMA}(\mathbf{H}_{1,\cdots,K}, P, \sigma^2) \tag{E.3}$$

$$\geq \log\left(1 + \frac{P}{\sigma^2}\lambda^2_{max}\right) \tag{E.4}$$

Then we can immediately obtain

$$G(\mathbf{H}_{1,\cdots,K}, P, \sigma^2) \leq \min\{N_r, K\} \tag{E.5}$$

by Theorem 3 in [39]. □

Bibliography

[1] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, New York: Dover, 1972.

[2] M. Airy, A. Forenza, R. W. Heath Jr., and S. Shakkottai, "Practical Costa precoding for the multiple antanna broadcast channel," in *Proc. IEEE Global Communications Conf.*, vol. 6, pp. 3942–3946, Dec. 2004.

[3] N. Al-Dhahir and j. M. Cioffi, "Optimum finite-length equalization for multi-carrier transceivers," *IEEE Transactions on Communications*, vol. 44, no. 1, pp. 56–64, Jan. 1996.

[4] G. Arslan, B. L. Evans, and S. Kiaei, "Equalization for discrete multitone transceivers to maximize bit rate," *IEEE Transactions on Signal Processing*, vol. 49, no. 12, pp. 3123–3135, Dec. 2001.

[5] R. Baldick, "Optimization of engineering systems course notes," the Univeristy of Texas at Austin, http://www.ece.utexas.edu/˜baldick/.

[6] D. Bartolome and A. L. Pirez-Neira, "Performance analysis of scheduling and admission control for multiuser downlink SDMA," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 333–336, May 2004.

[7] E. Biglieri, G. Caire, and G. Taricco, "Limiting performance of block-fading channels with multiple antennas," *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1273–1289, May 2001.

[8] J. A. C. Bingham, "Multicarrier modulation for data transmisson: an idea whose time has come," *IEEE Communications Magazine*, vol. 28, no. 5, pp. 5–14, May 1990.

[9] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1691–1706, Jul. 2003.

[10] R. Cendrillon, M. Moonen, J. Verliden, T. Bostoen, and W. Yu, "Optimal multiuser spectrum management for digital subscriber lines," in *Proc. IEEE International Conference on Communications*, vol. 1, pp. 1–5, Jun. 2004.

[11] R. Chen, J. G. Andrews, and R. W. Heath Jr., "Transmit selection diversity for multiuser spatial multiplexing systems," in *Proc. IEEE Global Communications Conference*, vol. 4, pp. 2625–2629, Dec. 2004.

[12] M. Chiani, M. Z. Win, and A. Zanella, "On the capacity of spatially correlated MIMO Rayleigh-fading channels," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2363–2371, Oct. 2003.

[13] J. Choi and R. W. Heath Jr., "Interpolation based transmit beamforming for MIMO-OFDM with limited feedback," *IEEE Transactions on Signal Processing*, vol. 53, no. 11, pp. 4125–4135, Nov. 2005.

[14] L. U. Choi and R. D. Murch, "A transmit preprocessing technique for multiuser MIMO systems using a decomposition approach," *IEEE Transactions on Wireless Communications*, vol. 3, no. 1, pp. 20–24, Jan. 2004.

[15] J. M. Cioffi, G. P. Dudevoir, M. V. Eyuboglu, and G. D. Forney Jr., "MMSE decision-feedback equalizers and coding. i. equalization results," *IEEE Transactions on Communications*, vol. 43, no. 10, pp. 2582–2594, Oct. 1995.

[16] J. M. Cioffi, G. P. Dudevoir, M. V. Eyuboglu, and G. D. Forney Jr., "MMSE decision-feedback equalizers and coding. ii. coding results," *IEEE Transactions on Communications*, vol. 43, no. 10, pp. 2595–2604, Oct. 1995.

[17] M. Costa, "Writing on dirty paper," *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

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

[19] A. Edelman, *Eigenvalue and Condition Numbers of Random Matrices*, Ph.D. thesis, MIT, May 1989.

[20] R. Fantacci, F. Chiti, D. Marabissi, G. Mennuti, S. Morosi, and D. Tarchi, "Perspectives for present and future CDMA-based communications systems," *IEEE Communications Magazine*, vol. 43, no. 2, pp. 95–100, Feb. 2005.

[21] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multiple antennas," *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–59, Autumn 1996.

[22] G. J. Foschini and M. Gans, "On the limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, Mar. 1998.

[23] M. Gharavi-Alkhansari and A. B. Gershman, "Fast antenna subset selection in MIMO systems," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 339–347, Feb. 2004.

[24] A. Ghosh, D. R. Wolter, J. G. Andrews, and R. Chen, "Broadband wireless access with WiMax/802.16: current performance benchmarks and future potential," *IEEE Communications Magazine*, vol. 43, no. 2, pp. 129–136, Feb. 2005.

[25] G. D. Golden, G. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky, "Detection algorithm and initial laboratory results using the V-BLAST space-time communication architecture," *Electronics Letters*, vol. 35, no. 1, pp. 14–15, Jan. 1999.

[26] A. Goldsmith, *Wireless Communications*, Cambridge University Press, 2005.

[27] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE Journal on Selected Area in Communications*, vol. 21, no. 5, pp. 684–702, Jun. 2003.

[28] G. H. Golub and C. F. Van Loan, *Matrix Computations*, The John Hopkisns Univ. Press, 3rd edition, 1996.

[29] A. Gorokhov, D. A. Gore, and A. J. Paulraj, "Receive antenna selection for MIMO spatial multiplexing: theory and algorithms," *IEEE Transactions on Signal Processing*, vol. 51, no. 11, pp. 2796–2807, Nov. 2003.

[30] K. A. Hamied and G. L. Stuber, "Performance of trellis-coded modulation for equalized multipath fading ISI channels," *IEEE Transactions on Vehicular Technology*, vol. 44, no. 1, pp. 50–58, Feb. 1995.

[31] Z. Han, Z. Ji, and K. J. R. Liu, "Fair multiuser channel allocation for OFDMA networks using Nash bargaining solutions and coalitions," *IEEE Transactions on Communications*, vol. 53, no. 8, pp. 1366–1376, Aug. 2005.

[32] B. Hassibi and T. L. Marzetta, "Multiple-antennas and isotropically-random

unitary inputs: The received signal density in closed-form," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1473–1484, Jun. 2002.

[33] B. Hochwald and S. Vishwanath, "Space-time multiple access: Linear growth in the sum rate," in *Proc. of the 40'th Annual Allerton Conference*, 2002.

[34] B. M. Hochwald, T. L. Marzetta, and V. Tarokh, "Multiple-antenna channel-hardening and its implications for rate feedback and scheduling," *IEEE Transactions on Information Theory*, vol. 50, no. 9, pp. 1893–1909, Sep. 2004.

[35] W. C. Jakes, Ed., *Microwave Mobile Communications*, New York: Wiley, 1974.

[36] J. Jang and K. B. Lee, "Transmit power adaptation for multiuser OFDM systems," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 2, pp. 171–178, Feb. 2003.

[37] S. K. Jayaweera and H. V. Poor, "Capacity of multiple-antenna systems with both receiver and transmitter channel state information," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2697–2709, Oct. 2003.

[38] N. Jindal, "MIMO broadcast channels with finite rate feedback," in *Proc. IEEE Global Telecommunications Conference*, Nov. 2005.

[39] N. Jindal and A. Goldsmith, "Dirty-paper coding versus TDMA for MIMO broadcast channels," *IEEE Transactions on Information Theory*, vol. 51, no. 5, pp. 1783–1794, May 2005.

[40] N. Jindal, W. Rhee, S. Vishwanath, S. A. Jafar, and A. Goldsmith, "Sum power iterative water-filling for multi-antenna Gaussian broadcast channels," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1570–1580, Apr. 2005.

[41] R. W. Heath Jr. and D. J. Love, "Multi-mode antenna selection for spatial multiplexing with linear receivers," *IEEE Trans. on Signal Processing*, vol. 53, no. 8, part 2, pp. 3042–3056, Aug. 2005.

[42] R. W. Heath Jr. and A. J. Paulraj, "Switching between diversity and multiplexing in MIMO systems," *IEEE Transactions on Communications*, vol. 53, no. 6, pp. 962–968, Jun. 2005.

[43] K. Kar, S. Sarkar, and L. Tassiulas, "Achieving proportional fairness using local information in Aloha networks," *IEEE Transactions on Automatic Control*, vol. 49, no. 10, pp. 1858–1863, Oct. 2004.

[44] F. P. Kelly, "Charging and rate control for elastic traffic," *European Transactions on Telecommunications*, vol. 8, pp. 33–37, 1997.

[45] I. Kim, H. L. Lee, B. Kim, and Y. H. Lee, "On the use of linear programming for dynamic subchannel and bit allocation in multiuser OFDM," in *Proc. IEEE Global Communications Conf.*, vol. 6, pp. 3648–3652, Nov. 2001.

[46] D. Kivanc, G. Li, and H. Liu;, "Computationally efficient bandwidth allocation and power control for OFDMA," *IEEE Transactions on Wireless Communications*, vol. 2, no. 6, pp. 1150–1158, Nov. 2003.

[47] E. Lawrey, "Multiuser OFDM," in *Proc. International Symposium on Signal Processing and its Applications*, Brisbane, Australia, Aug. 1999.

[48] G. Li and H. Liu, "On the optimality of the OFDMA network," *IEEE Communications Letters*, vol. 9, no. 5, pp. 438–440, May 2005.

[49] K. Liu, V. Raghavan, and A. M. Sayeed, "Capacity scaling and spectral efficiency in wide-band correlated MIMO channels," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2504–2526, Oct. 2003.

[50] D. J. Love and R. W. Heath Jr., "Limited feedback unitary precoding for spatial multiplexing systems," *IEEE Trans. on Information Theory*, vol. 51, no. 8, pp. 2967–2976, Aug. 2005.

[51] R. Lui and W. Yu;, "Low-complexity near-optimal spectrum balancing for digital subscriber lines," in *Proc. IEEE International Conference on Communications*, vol. 3, pp. 1947–1951, May 2005.

[52] V. H. MacDonald, "The cellular concept," *The Bell Systems Technical Journal*, vol. 58, no. 1, pp. 15–43, Jan. 1979.

[53] C. Martin and B. Ottersten, "Asymptotic eigenvalue distributions and capacity for MIMO channels under correlated fading," *IEEE Transactions on Wireless Communications*, vol. 3, no. 4, pp. 1350–1359, Jul. 2004.

[54] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Transactions on Information Theory*, vol. 45, no. 1, pp. 139–157, Jan. 1999.

[55] L. Massoulie and J. Roberts, "Bandwidth sharing: objectives and algorithms," *IEEE/ACM Transactions on Networking*, vol. 10, no. 3, pp. 320–328, Jun. 2002.

[56] C. Mohanram and S. Bhashyam, "A sub-optimal joint subcarrier and power allocation algorithm for multiuser OFDM," *IEEE Communications Letters*, vol. 9, no. 8, pp. 685–687, Aug. 2005.

[57] A. F. Molisch, M. Z. Win, Y.-S. Choi, and J. H. Winters, "Capacity of MIMO systems with antenna selection," *IEEE Transactions on Wireless Communications*, vol. 4, no. 4, pp. 1759–1772, Jul. 2005.

[58] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*, John Wiley & Sons, Inc., 1982.

[59] Z. Pan, K. K. Wong, and T. S. Ng, "Generalized multiuser orthogonal space-division multiplexing," *IEEE Transactions on Wireless Communications*, vol. 3, no. 6, pp. 1969–1973, Nov. 2004.

[60] A. S. Y. Poon, D. N. C. Tse, and R. W. Brodersen, "An adaptive multi-antenna transceiver for slowly flat fading channels," *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1820–1827, Nov. 2003.

[61] T. S. Rappaport, *Wireless Communications: Principles and Practice*, Prentice Hall PTR, 2002.

[62] T. S. Rappaport, A. Annamalai, R. M. Buehrer, and W. H. Tranter, "Wireless communications: Past events and a future perspective," *IEEE Communications Magazine*, vol. 40, no. 5, pp. 148–161, May 2002.

[63] V. Reddy, A. Paulraj, and T. Kailath, "Performance analysis of the optimum beamformer in the presence of correlated sources and its behavior under spatial smoothing," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, no. 7, pp. 927–936, Jul. 1987.

[64] W. Rhee and J. M. Cioffi, "Increasing in capacity of multiuser OFDM system using dynamic subchannel allocation," in *Proc. IEEE Int. Vehicular Tech. Conf.*, Tokyo, Japan, May 2000.

[65] H. Sampath, S. Talwar, J. Tellado, V. Erceg, and A. Paulraj, "A fourth-generation MIMO-OFDM broadband wireless system: Design, performance, and field trial results," *IEEE Communications Magazine*, vol. 40, no. 9, pp. 143–149, Sep. 2002.

[66] S. Shakkottai and R. Srikant, "Mean FDE models for internet congestion control under a many-flows regime," *IEEE Transactions on Information Theory*, vol. 50, no. 6, pp. 1050–1072, Jun. 2004.

[67] M. Sharif and B. Hassibi, "Scaling laws of sum rate using time-sharing, DPC, and beamforming for MIMO broadcast channels," in *Proc. International Symposium on Information Theory*, p. 175, Jun. 2004.

[68] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channel with partial side information," *IEEE Transactions on Information Theory*, vol. 51, no. 2, pp. 506–522, Feb. 2005.

[69] K. Sheikh, D. Gesbert, D. Gore, and A. Paulraj, "Smart antennas for broadband wireless access networks," *IEEE Communications Magazine*, vol. 37, no. 11, pp. 100–105, Nov. 1999.

[70] Z. Shen, J. G. Andrews, and B. L. Evans, "Optimal power allocation in multiuser OFDM systems," in *Proc. IEEE Global Communications Conference*, vol. 1, pp. 337–341, Dec. 2003.

[71] Z. Shen, J. G. Andrews, and B. L. Evans, "Short range wireless channel prediction using local information," in *Proc. IEEE Asilomar Conf. on Signals, Systems, and Computers*, vol. 1, pp. 1147–1151, Nov 2003.

[72] Z. Shen, J. G. Andrews, and B. L. Evans, "Adaptive resource allocation in multiuser OFDM systems with proportional rate constraints," *IEEE Transactions on Wireless Communications*, vol. 4, no. 6, pp. 2726–2737, Nov. 2005.

[73] Z. Shen, J. G. Andrews, and B. L. Evans, "Upper bounds on MIMO channel capacity with channel Frobenius norm constraints," in *Proc. IEEE Global Communications Conference*, Dec. 2005.

[74] Z. Shen, R. Chen, J. G. Andrews, R. W. Heath Jr., and B. L. Evans, "Low complexity user selection algorithms for multiuser MIMO systems with block diagonalization," *IEEE Transactions on Signal Processing*, Nov. 2005, accepted for publication.

[75] Z. Shen, R. Chen, J. G. Andrews, R. W. Heath Jr., and B. L. Evans, "Low complexity user selection algorithms for multiuser MIMO systems with block diagonalization," in *Proc. IEEE Asilomar Conf. on Signals, Systems, and Computers*, Nov. 2005.

[76] Z. Shen, R. Chen, J. G. Andrews, R. W. Heath Jr., and B. L. Evans, "Sum capacity of multiuser MIMO broadcast channels with block diagonalization," *IEEE Transactions on Wireless Communcations*, Oct. 2005, submitted.

[77] Z. Shen, R. Chen, J. G. Andrews, R. W. Heath Jr., and B. L. Evans, "Sum capacity of multiuser MIMO broadcast channels with block diagonalization," in *IEEE International Symposium on Information Theory*, Jun. 2006, submitted.

[78] Z. Shen, R. W. Heath Jr., J. G. Andrews, and B. L. Evans, "Comparison of space-time water-filling and spatial water-filling for MIMO fading channels," in *Proc. IEEE Global Communications Conference*, vol. 1, pp. 431–435, Dec. 2004.

[79] P. Smith and M. Shafi, "An approximate capacity distribution for MIMO systems," *IEEE Transactions on Communications*, vol. 52, no. 6, pp. 887–890, Jun. 2004.

[80] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 461–471, Feb. 2004.

[81] D. Stiliadis and A. Varma, "Efficient fair queueing algorithms for packet-switched networks," *IEEE/ACM Transactions on Networking*, vol. 6, no. 2, pp. 175–185, Apr. 1998.

[82] M. Stojnic, H. Vikalo, and B. Hassibi, "Rate maximization in multi-antenna

broadcast channels with linear preprocessing," in *Proc. IEEE Global Telecommunications Conference*, vol. 6, pp. 3957–3961, Dec. 2004.

[83] K. Tachikawa, "A perspective on the evolution of mobile communications," *IEEE Communications Magazine*, vol. 41, no. 10, pp. 66–73, Oct. 2003.

[84] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, Nov./Dec. 1999.

[85] M. Tuchler, R. Koetter, and A. C. Singer, "Turbo equalization: principles and new results," *IEEE Transactions on Communications*, vol. 50, no. 5, pp. 754–767, May 2002.

[86] G. De Veciana, T.-J. Lee, and T. Konstantopoulos, "Stability and performance analysis of networks supporting elastic services," *IEEE/ACM Transactions on Networking*, vol. 9, no. 1, pp. 2–14, Feb. 2001.

[87] V. V. Veeravalli, Y. Liang, and A. M. Sayeed, "Correlated MIMO wireless channels: capacity, optimal signaling, and asymptotics," *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 2058–2072, Jun. 2005.

[88] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. on Information Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.

[89] P. Viswanath and D. N. C. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality," *IEEE Transactions on Information Theory*, vol. 49, no. 8, pp. 1912–1921, Aug. 2003.

[90] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1277–1294, Jun. 2002.

[91] H. Viswanathan and S. Venkatesan, "Asymptotics of sum rate for dirty paper coding and beamforming in multiple-antenna broadcast channels," in *Proc. of the 41'st Annual Allerton Conference*, 2003.

[92] Z. Wang and G. B. Giannakis, "Outage mutual information of space-time MIMO channels," *IEEE Transactions on Information Theory*, vol. 50, no. 4, pp. 657–662, Apr. 2004.

[93] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "The capacity region of the Gaussian MIMO broadcast channel," in *Proc. IEEE International Symposium on Information Theory*, p. 174, Jun. 2004.

[94] J. Winters, "Optimum combining in digital mobile radio with cochannel interference," *IEEE Journal on Selected Areas in Communications*, vol. 2, no. 4, pp. 528–539, Jul. 1984.

[95] J. Winters, "On the capacity of radio communication systems with diversity in a Rayleigh fading environment," *IEEE Journal on Selected Areas in Communications*, vol. 5, no. 5, pp. 871–878, Jun. 1987.

[96] C. Y. Wong, R. S. Cheng, K. B. Letaief, and R. D. Murch, "Multicarrier OFDM with adaptive subcarrier, bit, and power allocation," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 10, pp. 1747–1758, Oct. 1999.

[97] I. C. Wong, A. Forenza, R. W. Heath Jr., and B. L. Evans, "Long range channel prediction for adaptive OFDM systems," in *Proc. IEEE Asilomar Conf. on Signals, Systems, and Computers*, pp. 732–736, Nov. 2004.

[98] I. C. Wong, Z. Shen, J. G. Andrews, and B. L. Evans, "A low complexity algorithm for proportional resource allocation in OFDMA systems," in *Proc. IEEE Int. Workshop on Signal Processing Systems*, pp. 1–6, Oct. 2004.

[99] K. K. Wong, R. D. Murch, and K. B. Letaief, "A joint-channel diagonalization for multiuser MIMO antenna systems," *IEEE Transactions on Wireless Communications*, vol. 2, no. 4, pp. 773–786, Jul. 2003.

[100] H. Yin and H. Liu, "Performance of space-division multiple-access (SDMA) with scheduling," *IEEE Transactions on Wireless Communications*, vol. 1, no. 4, pp. 611–618, Oct. 2002.

[101] T. Yoo and A. J. Goldsmith, "Optimality of zero-forcing beamforming with multiuser diversity," in *Proc. IEEE International Conference on Communications*, vol. 1, pp. 542–546, May 2005.

[102] K. Yu and B. Ottersten, "Models for MIMO propagation channels, a review," *Wiley Journal on Wireless Communications and Mobile Computing*, vol. 2, no. 7, pp. 653–666, Nov. 2002.

[103] W. Yu and J. M. Cioffi, "Sum capacity of Gaussian vector broadcast channels," *IEEE Transactions on Information Theory*, vol. 50, no. 9, pp. 1875–1892, Sep. 2004.

[104] W. Yu, G. Ginis, and J. Cioffi, "Distributed multiuser power control for digital subscriber lines," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 5, pp. 1105–1115, Jun. 2002.

[105] W. Yu, R. Lui, and R. Cendrillon, "Dual optimization methods for multiuser orthogonal frequency division multiplex systems," in *Proc. IEEE Global Telecommunications Conference*, vol. 1, pp. 225–229, Dec. 2004.

[106] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi, "Iterative water-filling for Gaussian vector multiple-access channels," *IEEE Transactions on Information Theory*, vol. 50, no. 1, pp. 145–152, Jan. 2004.

[107] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1250–1276, Jun. 2002.

[108] Y. J. Zhang and K. B. Letaief, "Multiuser adaptive subcarrier-and-bit allocation with adaptive cell selection for OFDM systems," *IEEE Transactions on Wireless Communications*, vol. 3, no. 4, pp. 1566–1575, Sep. 2004.

[109] Y. J. Zhang and K. B. Letaief, "An efficient resource-allocation scheme for spatial multiuser access in MIMO/OFDM systems," *IEEE Transactions on Communications*, vol. 53, no. 1, pp. 107–116, Jan. 2005.

[110] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels," *IEEE Transactions on Information Theory,*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[111] Y. Zhou and H. Sethu, "On achieving fairness in the joint allocation of processing and bandwidth resources: principles and algorithms," *IEEE/ACM Transactions on Networking*, vol. 13, no. 5, pp. 1054–1067, Oct. 2005.

Vita

Zukang Shen received his B.S.E.E. from Tsinghua University, Beijing, China in 2001 and M.S.E.E. from the The University of Texas at Austin. From January to August of 2004, Zukang was an intern at Freescale Semiconductor in Austin, Texas. For summer 2005, he worked at Texas Instruments in Dallas, Texas. Zukang was awarded the David Bruton, Jr. Graduate Fellowship for the 2004-2005 academic year by The Office of Graduate Studies at The University of Texas at Austin. He also received UT Austin Texas Telecommunications Engineering Consortium fellowships for the 2001-2002 and 2003-2004 academic years. Zukang has been a member of the Institute of Electrical and Electronics Engineers (IEEE) since 2001.

Permanent Address: Wenyunli 45-408
Jiaxing, Zhejiang, 314000, P. R. China

This dissertation was typeset with LaTeX 2_ε[1] by the author.

[1] LaTeX 2_ε is an extension of LaTeX. LaTeX is a collection of macros for TeX. TeX is a trademark of the American Mathematical Society. The macros used in formatting this dissertation were written by Dinesh Das, Department of Computer Sciences, The University of Texas at Austin, and extended by Bert Kay and James A. Bednar.