Multiuser Precoding in Wireless Communication Systems

Parameter and Resource Optimization via Large System Analysis

Rusdha Muharar

Submitted in total fulfilment of the requirements of the degree of

Doctor of Philosophy

Department of Electrical and Electronic Engineering
THE UNIVERSITY OF MELBOURNE

October 2012

Abstract

MULTI-user Multi-Input Multi-Output (MU-MIMO) technologies have become an important feature of the physical layer in modern wireless communication systems, such as in wireless LAN 802.11 and 4G networks including LTE-Advanced and mobile-WiMax, that need throughput in the order of hundreds of megabits per second or more. MU-MIMO has become an important research area, particularly for downlink or broadcast transmissions. In that scenario, one transmitter or base station (BS) with multiple antennas sends different and independent messages to each user simultaneously. The resulting multi-user interference (MUI) can limit the maximum achievable sum rate. Designing a good precoder for the users' data symbols at the transmitter can reduce MUI.

In this thesis, we focus on finding the optimal design parameters that maximize the system performance, i.e., SINR or sum rate, in multiple-antenna broadcast channels (BC) employing a particular precoding strategy, called Regularization Channel Inversion. We investigate the system performance maximization under various scenarios. In single-cell BC, different forms of channel state information (CSI) at the transmitter such as perfect CSI in independent and identically distributed (i.i.d.) and spatially correlated as well as partial CSI with uncertainties are considered. We present the optimal strategy for the regularization parameter of the RCI precoder to adapt in those scenarios. For some cases we also study how we choose the cell-loading, defined as the ratio between the number of users and the number of transmit antennas, to maximize the system performance. Considering clustered or grouped users where each group has different path-losses, we explore the optimal power allocation across the groups that maximizes the sum rate per antenna. In two-cell BC, we investigate CSI feedback optimization for different levels of base station cooperation.

The analysis in this work is conducted in the large system limit where the number of single-antenna users and transmit antennas tend to infinity with their ratio being fixed. The mathematical tools are based on spectral theory of random matrices that explore the eigenvalues of large dimensional random matrices. Even though the analytical results are in the asymptotic regime, we show their validity for the finite size system design through numerical simulations.

Declaration

This is to certify that

(i) the thesis comprises only my original work towards the PhD,

(ii) due acknowledgement has been made in the text to all other material used,

(iii) the thesis is less than 100,000 words in length, exclusive of tables, maps, bibliographies and appendices.

Rusdha Muharar, October 2012

Acknowledgements

First and foremost, I would like to attribute my greatest gratitude to my advisor Prof. Jamie Evans for taking me as his student and supervising this thesis in the last four years. I thank him for introducing me into the area of MIMO wireless communications and random matrix theory that become the subjects of the present work. I am grateful to him for his patient guidance, insightful comments and advice during our meetings, and encouragement when I was down and facing difficult times. Without those, this thesis would not have been completed. I also thank him for providing the funding for my conference travels. He has the combination of sound technical knowledge and excellent inter-personal skills that I wish I had.

I am also indebted to Dr. Randa Zakhour. It is a pleasure to collaborate with her during her fellowship in the Department of Electrical and Electronic Engineering (EEE), The University of Melbourne. I recognize it was a productive period leading to the results presented in Chapter 6 and 7. She still helps me even though she is not in Melbourne anymore. Thanks, Randa.

My thanks also go to Dr. Tansu Alpcan who became my the thesis supervisor in the last year. I also thank Prof. Subhrakanti Dey and Dr. Peter Dower for becoming the committee members of this thesis. I also benefited from Subhra's lectures in Information Theory and Signal Processing. I also appreciate the discussions with Dr. Hazer Inaltekin and Dr. Vasanthan Raghavan during their stay in EEE that enriched my knowledge.

I acknowledge the Australian government for providing me the scholarship during the course of my study via the Australian Development Scholarship (ADS) program. I thank the AusAID Manager of the University of Melbourne, Lucia Wong, for administering the scholarship. I am also indebted to my employer, Indonesian Ministry of National

Education and Culture and also Syiah Kuala University, Indonesia, for allowing me to pursue the PhD study and granting me study leave.

I also would like to thank,

- Jackie Brisonette and Nasrin Hashemi for their help on the administrative tasks.
- all the fellow students, in particular Nasreen Badruddin, Nader Ghasemi, Paul Tune, Kenny Wu Wei Wu, Jack Chih Hong Wang, Tharaka Samarasinghe, Athipat Limmanee, Ehsan Nekaoi, Jack Hsu, 'what are you doing' Katrina Yuan Yuan He and Xiao Xi Gao for their friendship, humour, and discussions during my candidature.
- Mohd. Edil and Mohd. Asyraf for their friendship and help.
- All my Indonesian friends that I can not name one by one, for their warm friendship and for being helpful and supportive.

Last but not the least, I express my endless gratitude to my wife, Noraliyatun Jannah, for her unconditional love, patience, sacrifices and support. To my kids, Azzam and Akifah, you are the source of motivation for me to complete my study. This work also would not have been possible without the support, love and endless prayers from my parents: Umak Rosmiwati Adha and ayah Syahril (r.i.p.) as well as my parents in law, Abi Muhammad Amin and Ummi Wardah.

To my wife Nora, my lovely kids Azzam and Akifah, and my mother Rosmiwati, for everything.

Contents

List of Figures

Notations

Unless stated otherwise, boldface upper-case and lower-case letters denote matrices and vectors respectively.

$[\mathbf{X}]_{ij}$	the (i,j)th entry of the matrix $\mathbf{X}$
$\mathbf{X}^H$	the complex conjugate (Hermitian) transpose of matrix $\mathbf{X}$
$\mathbf{X}^T$	the transpose of matrix $\mathbf{X}$
$\mathrm{Tr}\,(\mathbf{X})$	trace of matrix $\mathbf{X}$
$\det(\mathbf{X})$	determinant of matrix $\mathbf{X}$
$\mathbf{X}^{-1}$	the inverse of matrix $\mathbf{X}$
$\|\mathbf{x}\|$	Euclidean norm of vector $\mathbf{x}$
x_j	the jth entry of vector $\mathbf{x}$
$\mathrm{diag}(\mathbf{x})$	$N \times N$ diagonal matrix with entries of $\mathbf{x}$
$\mathbf{I}_N$	Identity matrix of dimension $N \times N$
$\mathbb{R}, \mathbb{C}$	the real and complex numbers
$\mathbb{C}^{m\times n}$	set of $m \times n$ matrices with complex-valued entries
$\mathbb{C}^m$	set of column vectors of size m with complex-valued entries
$\Re[z]$	real part of $z \in \mathbb{C}$
$\Im[z]$	imaginary part of z
z^*	complex conjugate of z
$\|z\|$	absolute value of z
$1_{\{X\}}$	indicator function, where $1_{\{X\}} = 1$ iff X is true and $1_{\{X\}} = 0$ otherwise
$\max.$	maximize
$\mathbb{E}[X]$	expectation of X
$\sim$	distributed according to
$\mathcal{CN}(\mathbf{m}, \mathbf{C})$	complex Gaussian distribution with mean $\mathbf{m}$ and covariance $\mathbf{C}$
$\xrightarrow{a.s.}$	almost sure convergence
$\xrightarrow{i.p.}$	convergence in probability
$\xrightarrow{L_2}$	mean-square convergence

Chapter 1

Introduction

NOWADAYS, we start witnessing the rise of many mobile or wireless applications that demand high data rate exchanges such as high volume wireless data transfers, high-definition mobile television (TV), online gaming services, and video conferencing and streaming via mobile devices. To support this high data rate demands, *International Telecommunication Union-Radio communications sector* (ITU-R) released the International Mobile Telecommunications Advanced (IMT-Advanced) standard in 2008. This standard is also known as 4G, the successor of the current 3G standard, and sets the peak data rate requirements for 4G services: 100 megabits per second for low mobility communications and 1 gigabits per second for high mobility communications. Two technologies i.e., Mobile WiMAX (Worldwide Interoperability for Microwave Access) and LTE (Long Term Evolution) are often said to offer 4G services, even though their peak data rates are less than 4G specifications. To be compliant with IMT-Advanced requirements, Mobile WiMAX release 2 and LTE-Advanced (LTE-A) were then released and standardized in 2011.

To be able to produce high data rates, one of the core technologies for the physical layer that both 4G candidates rely on is the MIMO (Multiple-Input Multiple-Output) transmission techniques. MIMO in the communication systems refers to the use of multiple antennas at both the transmitter and the receiver. The pioneering works on MIMO systems can be dated back in mid 90s where Telatar [89,90] and Foschini and Gans [22,23] independently investigated the capacity of point to point or single-user MIMO systems. Since then, there has been a tremendous amount of work in this area.

In general, one of the main challenges in designing wireless communications systems,

besides interference, is multi-path fading. It is due to the scattering of the transmitted signals by the objects or obstacles that are present along the propagation channel between the transmitter and the receiver. Hence, multiple copies of the transmitted signals may arrive at the receiver at different times, directions (angles) and frequencies. This phenomenon will cause random fluctuations of the received signals. One may see it as an impairment. However, if the transmitter and/or the receiver, somehow have/has the knowledge about the fading states (propagation channels), then it can be exploited to increase the capacity or reliability of a wireless communication system [6]. In single-user MIMO systems, for an example, it can be used by the transmitter to determine the sending 'directions' and 'magnitudes' of parallel data streams over the same frequency band and consequently, the total data rate will increase. In information-theoretic point of view, it has been shown in [22, 23, 89, 90] that the capacity of single-user MIMO systems grows linearly with the minimum number of antennas at the transmitter and the receiver providing that the fading channel states, commonly called channel state information (CSI), are known perfectly by both ends and are statistically independent. This demonstrates the importance of CSI in the communication system designs.

In practical communication systems where multiple users are present such as in digital subscriber lines (DSL), wireless LAN 802.11 and wireless cellular networks, multi-user MIMO (MU-MIMO) has become an important research area recently, particularly for downlink transmissions, or also called broadcast channels (BC). In this scenario, a transmitter or base station (BS) with multiple antennas sends different and independent messages or data symbols to each user simultaneously. With a perfect knowledge of the CSI at the transmitter and the receiver, the sum-capacity of K users, where each user equipped with single antenna, grows linearly with the minimum number of transmit antennas (N) and users (see [26] and references therein for detail discussions regarding information theoretic aspects of MIMO systems). The capacity region of the BC has been characterized recently in [103]. To achieve the downlink sum- capacity, we have to deal with the multiuser interference (MUI) that presents because the communication between the transmitter and the users occurs at the same time and uses the same frequency band. Hence, transmitting messages to a particular user will cause interference to other users

in the system. The level of the interference can be influenced by some factors such as the number of users in the system, the magnitude (or power) and direction of the transmitted data that are transferred between the transmitter and the users. The transmitter can manage the MUI by adjusting those factors according to the CSI that is available at the transmitter (CSIT) . Specifically, based on the CSIT, the transmitter pre-processes or precodes the messages with a so called *precoder* or *beamformer* before transmissions in order to reduce the MUI. Furthermore, the CSIT can also be used to determine an optimal user scheduling and transmit power allocation for each user that maximize the downlink sum rate. However, obtaining a perfect CSIT in practical implementations is difficult. Any uncertainty that presents on the CSI may significantly impact the overall capacity of the system. The worst case is when the transmitter does not have the CSIT. In that case, if the users have the same number of receive antennas and the same fading statistic/distribution, then there is no advantage of performing multiuser communications and the single-user MIMO is optimal [6, Chapter 2], [9, 34].

The use of multiple antennas in MIMO systems also adds spatial dimension besides time, frequency and code to separate users in a wireless cellular network. This will lead to a more aggressive reuse of time and frequency resources in order to increase the network capacity [6, 25]. However, an aggressive frequency or time reuse will introduce more inter cell interference (ICI) especially for the users in the cell-edge. In the classical approach, where there is no cooperation between the base stations (BSs), the network reduces to an interference channel setup and its capacity still remains an open problem for over thirty years. In the current and emerging approach, the base stations are allowed to cooperate by sharing their control signals, transmission data and CSI via backhaul links. The level of cooperation is determined by the sharing amount amongst the BSs. It has been shown in several works [25, and references therein] that the base station cooperation, particularly with perfect CSIT available at the BSs, provides a significant increase in the system capacity compared to the systems adopting the classical approach.

Summarizing our discussions above, CSIT plays important roles in increasing the capacity of a communication system, from a single-user MIMO to a more complex multicell multiuser MIMO network. A precoder at the transmitter exploits the CSIT to control

the transmission strategies to achieve the maximum throughput. Thus, it is crucial to the determine the optimal design parameters of the precoder.

The channel gain matrix in multiuser MIMO communications represents propagation channels between the multiple antennas transmitter and all users. It has random entries because as previously mentioned the channel gains are random quantities. Thus, the study of the properties of random matrices is important in analyzing the performance of wireless communication systems. Some of the results of random matrix theory has been applied in [21, 89, 90] to derive the capacity of single-user MIMO systems. However, the performance analysis for finite-size systems, i.e., finite number of antennas and users, can be difficult or even intractable for a more complex and realistic system model. Performing the analysis in the asymptotic regime can provide an accurate approximation of the system performance with a reduced complexity [14, 95]. It can also be very useful and be applied to analyze the emerging research area of *Massive MIMO* or *very large MIMO* that employs a large number of small antennas at the base station to improve the rate and reliability of the MIMO communication systems [75]. It requires some advances on large dimensional random matrix theory. [91, 93, 97] are some works among others that initiated the application of large dimensional random matrix theory to analyze the performance of wireless communication systems, particularly for Code Division Multiple Access (CDMA) systems. Recently, it has been employed to analyze and design multiuser and multicell MIMO systems [14, 95].

1.1 Focus of Thesis

In this thesis, we consider Multiple-Input Single-Output (MISO) communication systems, where the transmitters have multiple antennas and each user in the system only has a single antenna. The main focus will be in the downlink transmissions or broadcast channels.

As mentioned earlier, the maximum system performance can be achieved by exploiting CSIT at the transmitter side via a precoder. Designing an optimal precoder that maximizes the system capacity could lead to a complicated task. A capacity achieving

precoder, called *dirty paper coding* (DPC) needs a perfect CSIT and its implementation in practice is computationally expensive [19, 86]. Linear-type precoders offer a lower complexity but with penalty in the system performance. One of the most popular precoders considered in the wireless communication literature is the Zero-Forcing (ZF) or Channel Inversion (CI) precoder. With the availability of a perfect CSIT, ZF can eliminate or null the MUI. However, channel inversion may lead to a significant penalty on the system performance when the channel gain matrix is rank deficient. To overcome this, a regularization parameter can be introduced in the channel inversion and the resulting precoder is called Regularized Zero-Forcing (RZF) or Regularized Channel Inversion (RCI). We particularly concentrate on the applications of this precoder for various channel and system models.

The performance measure analyzed in this document is the SINR or correspondingly the sum rate per-antenna. The analysis relies heavily on some results of large dimensional random matrix theory. We particularly perform the analysis in the large system limit regime, where the number of antennas at the transmitter or BS and the number of users go to infinity with their ratio being fixed. In that regime, as being shown later, the performance measure becomes a deterministic quantity and this could reduce the complexities or simplify the analysis. Moreover, the performance measure obtained in the analysis can give insights on its behavior in the finite-size or even in the small-size systems. Based on those facts, we then investigate the optimal values of system parameters, such as, regularization parameter and cell-loading i.e., the ratio between the number of users and the numbers of antenna.

1.2 Summary of Thesis and Contributions

In Chapter 2, a brief introduction on random matrix theory is presented. It discusses some basic definitions and important results on large dimensional random matrix theory. They will be used frequently in the analysis throughout of this document.

In Chapter 3, we start our investigation on the performance of a single-cell MISO broadcast channel with i.i.d. entries of the channel gain matrix. The transmitter is as-

sumed to have a perfect knowledge of CSI and this information is used to construct the RCI precoder. It is also assumed that the users have the same path-loss gain. First, we derive the SINR in the large system limit, also called the limiting SINR. It is deterministic quantity and the same for all users. Then, we derive the optimal regularization parameter that maximizes the limiting SINR or equivalently the limiting sum rate per-antenna of the precoder. It surprisingly takes a very simple form which is a ratio between the cell-loading and the received signal-to-noise ratio (SNR). A similar result is previously obtained by Peel et al., in [69]. The optimal cell-loading maximizing the sum rate per-antenna is also discussed. Some numerical simulations that are presented show that the asymptotic analysis can accurately approximate the finite-size systems. Moreover, we also consider the performance analysis of the Moore-Penrose Channel Inversion and single user (SUB) precoders.

The case of multiuser transmit beamforming in spatially correlated channel is considered in Chapter 4. The scattering effects from the propagation environment and the insufficient separation between antenna elements may cause fading correlation. It is assumed that the users are separated enough so that the correlation is only present at the transmitter side. Moreover, to simplify the problem, it is also assumed that the users see the same transmit- correlation. We adopt a separable correlation model, also called *Kronecker* model. Similar to the previous chapter, the base station is assumed to know the CSI of each user perfectly. We derive the limiting SINR and show that it is indeed affected by the correlation. On the contrary, we prove that the optimal regularization parameter of the RCI precoder is independent of the correlation and is the same as to that in the i.i.d. case. This is an important and a surprising result. With the same settings, we also analyze the performance of MPCI and SUB precoders.

In Chapter 5, we relax the assumption that the transmitter knows the CSI perfectly. To simplify the analysis, we consider the Time Division Duplex (TDD) scheme with a perfect channel reciprocity behavior. In that scenario, the uplink and downlink channel are the same regardless of the direction. It is assumed that each user knows its channel exactly and all users send their channel information to the transmitter via the multiple access channel (MAC). The BS estimates the CSI by the minimum mean square error (MMSE)

estimator. This will introduce channel estimation errors or uncertainties to the channel states owned by the transmitter. First, we consider a simple case where the communication has two phases: the channel estimation or training and the downlink transmission. The analysis starts with the derivation of the asymptotic sum rate per antenna. Based on that, for a given total communication period, determined by the coherence time, we investigate the optimal periods for the channel training and the data transmission that maximize the sum rate per antenna. Then, we move to a more complex situation where we add uplink data transmission as the third phase. Each user is assumed to have the same and a fixed uplink power. Then, we investigate how to split the uplink power for the channel training and uplink data transmission optimally. Thus, a joint optimization of the channel training and data transmission period as well as uplink power splitting is conducted. We show that the optimal training period is one symbol per user regardless optimal splitting uplink power. The optimal period for downlink (or relatedly uplink) data transmission and optimal uplink power splitting can not be determined explicitly. However, numerical simulations suggest that there is a trade-off between these two quantities.

In the previous chapters, it is assumed that the transmitter allocates the power equally for each user. In Chapter 6, we take a step further by investigating an optimal power allocation for the users that maximizes the achievable sum rate per antenna. The users are assumed to have different path-losses. The analysis is started by deriving the limiting SINR for each user. Then we divide all users into a finite number of groups based on their distance-dependent path losses. In that scenario, we show that the optimal power allocation follows the well- known *water-filling* strategy. Jointly with the power allocation, the optimal regularization parameter is also derived. The water-filling scheme suggests that the groups whose path-losses less that a certain threshold are allocated zero power. Therefore, it is natural to ask whether to include the channel gains of those groups in the precoder will improve the sum rate or not. This leads us to investigate the multi-mode transmission where we determine the optimal number of groups communicating with the BS. We show that it is optimal for the BS to transmit to some groups having best channels. We also provide the necessary conditions for the optimal cell-loading allocation

when the BS is allowed to transmit to only subsets of the users in the groups.

The last technical chapter, i.e., Chapter 7, discusses the feedback optimization in a symmetric two-cell network for various levels of base stations cooperation, namely Network MIMO or Multi-Cell Processing (MCP) and Coordinated Beamforming (CBf). We consider both the analog and the digital/limited feedback schemes. In the analog feedback case, we investigate how each user optimally allocate his/her transmit power to feed back the unquantized and uncoded CSI of the direct channel and the interfering channel for different BS cooperation schemes. In the limited feedback case, instead of power, we study the optimal bit allocation or partitioning for the direct and interfering channel. For both cooperation schemes, we show that when the cross channel gain is below a certain threshold, it is better for both BSs to perform the Single- Cell Processing (SCP), that is, there is no cooperation between BSs. For the MCP case, it has been shown in [109] that the SINR improves as the cross channel gain increases. However, in the presence of the channel uncertainty, we show that this occurs only when the cross channel gain is above a certain threshold.

The overall conclusion for the materials studied in this thesis appears in Chapter 8.

1.3 Publications

Below is the list of publications related to the investigations conducted during the course of completing the PhD candidature.

Journal

1. R. Muharar, R. Zakhour and J. Evans, "Base Station Cooperation with Feedback Optimization: A Large System Analysis", (In preparation for submission).

Conference

1. R. Muharar and J. Evans, "Optimal training for Time-Division Duplexed Systems with Transmit Beamforming ", *Proc. Australian Communications Theory Workshop (AusCTW)*, Melbourne, Australia, 2011, (Best Student Paper Award).

2. R. Muharar and J. Evans, "Downlink Beamforming with Transmit-Side Channel Correlation: A Large System Analysis", *Proc. International Conference on Communications (ICC)*, Kyoto, Japan, June 2011.
3. R. Muharar and J. Evans, "Optimal Power Allocation for Multiuser Transmit Beamforming via Regularized Channel Inversion", *Proc. Asilomar Conference on Signals, Systems, and Computers*, Pacific Groove, California, November 2011.
4. R. Muharar, R. Zakhour and J. Evans, "Base Station Cooperation with Noisy Analog Channel Feedback: A Large System Analysis", *Proc. International Conference on Communications (ICC)*, Ottawa, Canada, June 2012.
5. R. Muharar, R. Zakhour and J. Evans, "Base Station Cooperation with Limited Feedback: A Large System Analysis", *Proc. International Symposium on Information Theory (ISIT)*, Cambridge, MA, USA, July 2012.

Chapter 2
Asymptotic Random Matrix Theory

This chapter presents a brief introduction to random matrix theory, particularly the properties of the eigenvalues of matrices as the dimensions of the matrices grow to infinity with a constant ratio. A simple example of the application of random matrix theory to wireless communications, i.e. evaluating the ergodic capacity of a single-user MIMO system, is presented. This leads to the definition of the empirical eigenvalue distribution or empirical spectral distribution (e.s.d.) of a (Hermitian) matrix. As the dimensions of the matrix grow to infinity, the e.s.d. converges to what is called the limiting spectral distribution *(l.s.d.), or just limiting distribution. Some well known limiting distributions such as Wigner's semi-circle law, circular law, and Marčenko-Pastur law are presented. Moreover, some limiting distributions that are particularly important for the asymptotic analysis in the following chapters are presented in terms of their Stieltjes transform. The links between certain matrix quadratic forms and and the trace of that matrix are also discussed.*

2.1 Introduction

A random matrix can be defined as a matrix with random variable entries [88] or a matrix-valued random variable (see e.g., [14]). Until recently, random matrix theory has found its applications into numerous fields, such as physics, statistics, information theory and wireless communications, to name a few. The materials contained in this chapter are largely based on [14, 95] that provide a nice introduction into random matrix theory and its applications in information theory and communications which are the area of interests of this work. Some definitions and theorems are also adopted from [4, 96].

In 1928, John Wishart laid the foundation of the study of random matrices with his work on the density function of the random matrix $\sum_{k=1}^{N} \mathbf{x}_k \mathbf{x}_k^H$, where the elements of

vector $\mathbf{x}_k$ are independent and drawn from standard normal distribution [14, 95, 104]. The matrices with this structure are now called the (central) Wishart matrices. Since then, many studies have explored various aspects of Wishart matrices. For example, now we can find the joint distribution of (un)ordered eigenvalues and the distribution of the extremes eigenvalues of Wishart matrices [14, 95]. In information theory and wireless communications, these distributions are very important and useful in order to evaluate the performance, e.g., mutual information, capacity and SINR of a communication system. An example described below is taken from the work of Telatar in [90].

Let us consider a single-user MIMO communication system where the transmitter (base station) and receiver (user) have N and M antennas, respectively. The propagation channel between the transmitter and receiver is represented by an $M \times N$ complex-valued random matrix $\mathbf{H}$. In [90] the entries of $\mathbf{H}$ are assume to be i.i.d. and drawn from Gaussian distribution with zero mean and unit variance. The transmitted data vector from N antennas is denoted by $\mathbf{x} \in \mathbb{C}^N$ and satisfies the total power constraint $\mathbb{E}[\mathbf{x}^H\mathbf{x}] \leq P$. Considering a linear relation between $\mathbf{x}$ and the received signal vector, $\mathbf{y} \in \mathbb{C}^M$, the latter can be expressed as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n},$$

where $\mathbf{n}$ is the receiver or thermal noise vector and assumed to have complex Gaussian distribution with zero mean and $\mathbb{E}[\mathbf{n}\mathbf{n}^H] = \mathbf{I}_M$. The ergodic capacity of this setup is given by [90, Theorem 1],

$$C = \mathbb{E}\left[\log\det\left(\mathbf{I}_M + \frac{\mathsf{SNR}}{N}\mathbf{H}\mathbf{H}^H\right)\right], \tag{2.1}$$

where SNR is the received SNR. Note that the capacity above is achieved when $\mathbf{x}$ is complex Gaussian with covariance matrix $\frac{\mathsf{SNR}}{N}\mathbf{I}_N$. It is easy to see that $\mathbf{H}\mathbf{H}^H$ is a Wishart matrix with rank $r = \min\{M, N\}$. Let $\lambda_1, \lambda_2, \ldots, \lambda_r$ be the eigenvalues of $\mathbf{H}\mathbf{H}^H$. So, we can rewrite (2.1) as

$$\mathbb{E}\left[\sum_{j=1}^{r}\log\left(1 + \frac{\mathsf{SNR}}{N}\lambda_j\right)\right]. \tag{2.2}$$

Note that the (non-zero) eigenvalues of $\mathbf{H}^H\mathbf{H}$ and $\mathbf{H}\mathbf{H}^H$ are the same.

Evaluating (2.2) needs the knowledge of the joint density of the (un)ordered eigenvalues of the Wishart distribution, which can be presented in terms of the Laguerre polynomials. A closed form expression for (2.2) is unknown but can be obtained numerically (see [90, Theorem 2]). When the entries of $\mathbf{H}$ are not i.i.d. Gaussian, the system ergodic capacity is unknown. A simple example is when the entries are independent and drawn from Gaussian distribution but have different variances. Moreover, if the system performance involves a more complicated function of eigenvalues of a random matrix, then its *performance analysis in finite dimensions becomes intractable and only allows numerical evaluations that does not offer much insights*. As we will see later, evaluating (2.2) in the large system limit, where the dimensions of $\mathbf{H}$ go to infinity with their ratio being fixed, results in a closed form expression. Our results in the subsequent chapters also show that *the performance analysis for various channel conditions and functions of channel gain* $\mathbf{H}$ *in the large system limit can lead to closed form and compact expressions*. Hence, insights on the system behaviour can be obtained and we can determine the optimal design parameters that maximize system performances. Furthermore, simulation results presented in this and sequel chapters demonstrate that *the asymptotic analysis can give accurate approximations of the system performance even for small system dimensions* (of $\mathbf{H}$). All those asymptotic results are derived based on theory of spectral distribution of large random matrices that will be presented in the next section.

2.2 Spectral Distribution of Large Random Matrices

Let us start with the following definition.

Definition 2.1 ([95, 96]). *Let* $\mathbf{X}$ *be an* $N \times N$ *Hermitian matrix. The spectrum or empirical distribution of the eigenvalues of* $\mathbf{X}$*, denoted by* $F_{\mathbf{X}}^{N}$*, is defined as follows*

$$F_{\mathbf{X}}^{N}(x) = \frac{1}{N}\sum_{i=1}^{N} 1_{\{\lambda_i(\mathbf{X}) \leq x\}}$$

where $\lambda_1(\mathbf{X}), \ldots, \lambda_N(\mathbf{X})$ *are the eigenvalues of* $\mathbf{X}$ *and* $1_{\{.\}}$ *is the indicator function.*

When the actual distribution function is not known, then the empirical distribution

function is the natural estimator for it [96]. Note that $F_{\mathbf{X}}^N$ is random since the eigenvalues of $\mathbf{X}$ changes for each its realization and for different N. The convergence of the sequence $\{F_{\mathbf{X}}^N\}$ to a limit is one of the major topics in random matrix theory [4]. This limit is defined as follows.

Definition 2.2. *Let $F_{\mathbf{X}}^N$ be the empirical spectral distribution of $\mathbf{X}$. A non-random distribution $F_{\mathbf{X}}$ is defined as the limiting or asymptotic spectral distribution of $\mathbf{X}$ when $F_{\mathbf{X}}^N$ converges to $F_{\mathbf{X}}$ as $N \to \infty$.*

The initial works on the limiting distribution of large random matrices came from the area of physics. In a series of his paper [95, and references on Wigner's paper therein], Wigner investigated the e.s.d of a symmetric matrix that has a particular structure and showed that it converges to a *semicircle law*. By using this result, he was able to explain the experimental results regarding atomic energy levels [95]. *Wigner matrices*, attributed to his name, are defined as any Hermitian matrices whose upper triangular entries are independent and zero mean with identical variance. Let $\mathbf{W}$ be an $N \times N$ Wigner matrix whose entries are zero mean and variance $\frac{1}{N}$. Then, as $N \to \infty$, the limiting distribution of $\mathbf{W}$ follows the semicircle law with density [95],

$$f_{\mathbf{W}}(x) = \begin{cases} \frac{1}{2\pi}\sqrt{(4-x^2)}, & \text{for } |x| \leq 2, \\ 0, & \text{otherwise.} \end{cases}$$

Figure 2.1 illustrates the semi-circle law and the empirical distribution of the eigenvalues of matrix $\mathbf{W}$ with $N = 1000$. The diagonal entries of $\mathbf{W}$ are i.i.d. and distributed according to $\mathcal{N}(0, \frac{1}{N})$. The upper triangular entries are i.i.d. complex Gaussian random variable with zero mean and variance $\frac{1}{N}$.

If the square matrix is no longer Hermitian, but its entries are i.i.d., then the limiting distribution of that matrix follows the *full-circle* law or *circular* law where asymptotic spectrum of the matrix is uniformly distributed over the unit circle on the complex plane. A graphical presentation of this law can be seen in Figure 2.2.

Although matrices that lead to the semi-circle and full-circle laws are important and frequently used in the area of physics, their use in the area of wireless communications

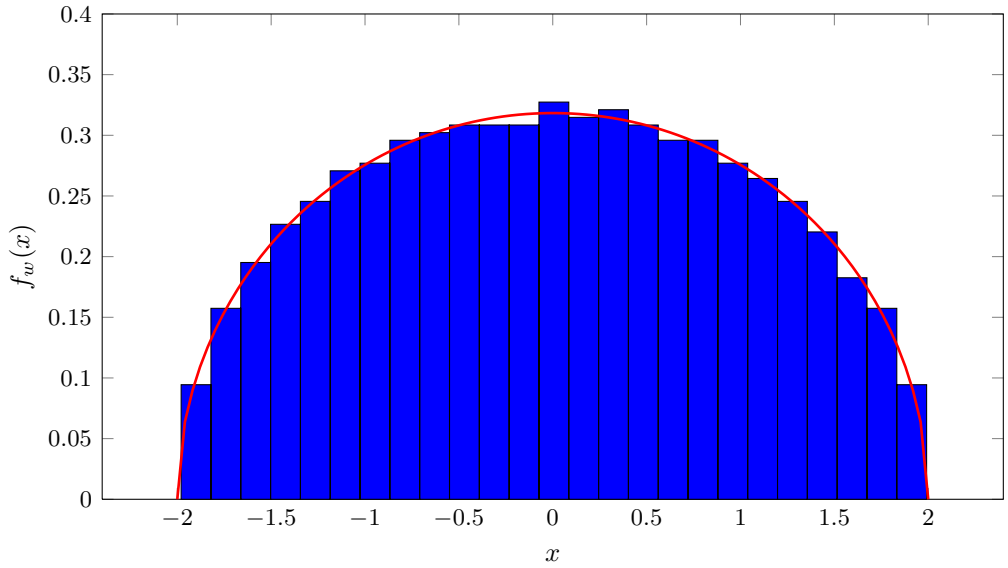


Figure 2.1: The histogram of the eigenvalues of $\mathbf{W}$ ($N = 1000$) vs. its semi-circle density.

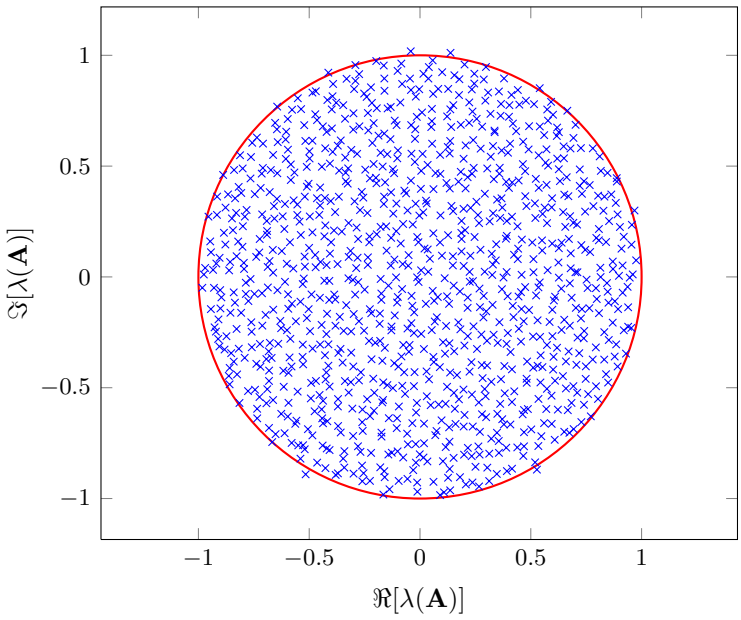


Figure 2.2: The full-circle law and the eigenvalues of matrix $\frac{1}{\sqrt{N}}\mathbf{A}$ with $N = 1000$. The entries of $\mathbf{A}$ are independent with zero mean and unit variance.

are rather limited. Generally, it can be said that the structure of sample covariance matrix $\mathbf{X}\mathbf{X}^H$ and its variations (generalizations), where $\mathbf{X}$ is a rectangular matrix with independent entries, are the objects of interest in this area. An obvious instance of application is our previous example where the asymptotic ergodic capacity per antenna can be evaluated based on the knowledge of the limiting distribution of $\mathbf{H}\mathbf{H}^H$. The work [53] by Marčenko and Pastur in 1967 was the first to consider the limiting spectrum distribution of sample covariance matrices. One of the main results is known as the Marčenko-Pastur law, as stated below [95]. See Figure 2.3 and 2.4 for some illustrations.

Theorem 2.1 (Marčenko-Pastur Law)**.** *Consider an $n \times N$ matrix $\mathbf{X}$ whose entries are i.i.d. complex or real random variable with zero mean and variance $\frac{1}{N}$. As $n, N \to \infty$ with $\frac{n}{N} \to \beta > 0$, the e.s.d. of $\mathbf{X}\mathbf{X}^H$ converges almost surely to a nonrandom limiting distribution F_β with density function*

$$f_\beta(x) = \left(1 - \frac{1}{\beta}\right)^+ \delta(x) + \frac{1}{2\pi\beta x}\sqrt{(x-a)^+(b-x)^+},$$

where $(y)^+ = \max\{0, y\}$, $a = (1-\sqrt{\beta})^2$, $b = (1+\sqrt{\beta})^2$ and $\delta(\cdot)$ is the Dirac delta function.

Equivalently, the e.s.d. of $\mathbf{X}^H\mathbf{X}$ converges almost surely to a nonrandom distribution $\breve{F}$ whose density [95]

$$\begin{aligned}\breve{f}_\beta(x) &= (1-\beta)\delta x + \beta f_\beta(x) \\ &= (1-\beta)^+ \delta(x) + \frac{1}{2\pi x}\sqrt{(x-a)^+(b-x)^+}. \end{aligned} \tag{2.3}$$

For $\beta = 1$, the singular values of $\frac{1}{\sqrt{N}}\mathbf{X}$ (or the eigenvalues of $\sqrt{\frac{1}{N}\mathbf{X}\mathbf{X}^H}$) are asymptotically distributed according to

$$f_q(x) = \frac{1}{\pi}\sqrt{4-x^2}, \quad 0 \le x \le 2,$$

which is also known as the *quarter circle* law.

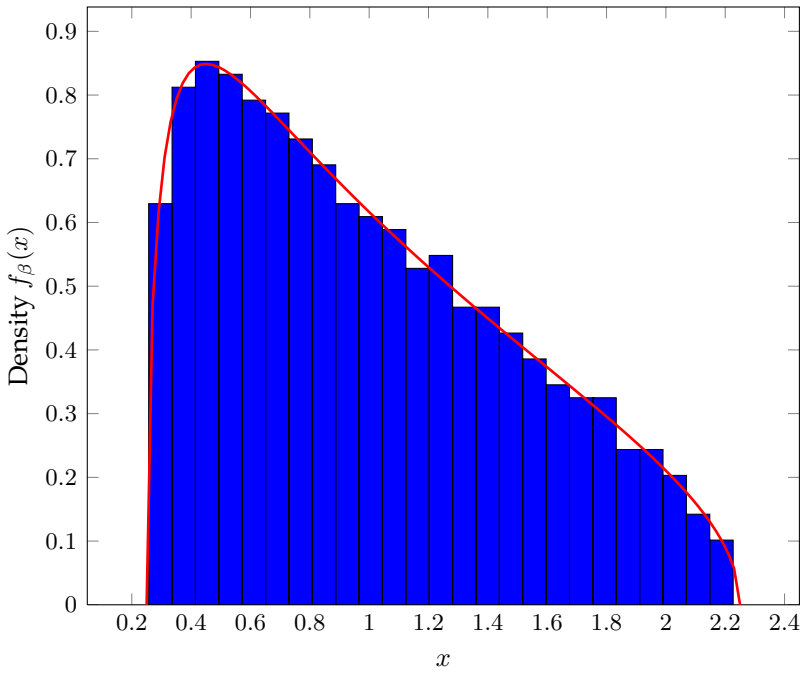


Figure 2.3: The histogram of the eigenvalues of $\mathbf{H}\mathbf{H}^H$ ($N = 2500$) vs. its Marčenko-Pastur density for $\beta = 0.25$.

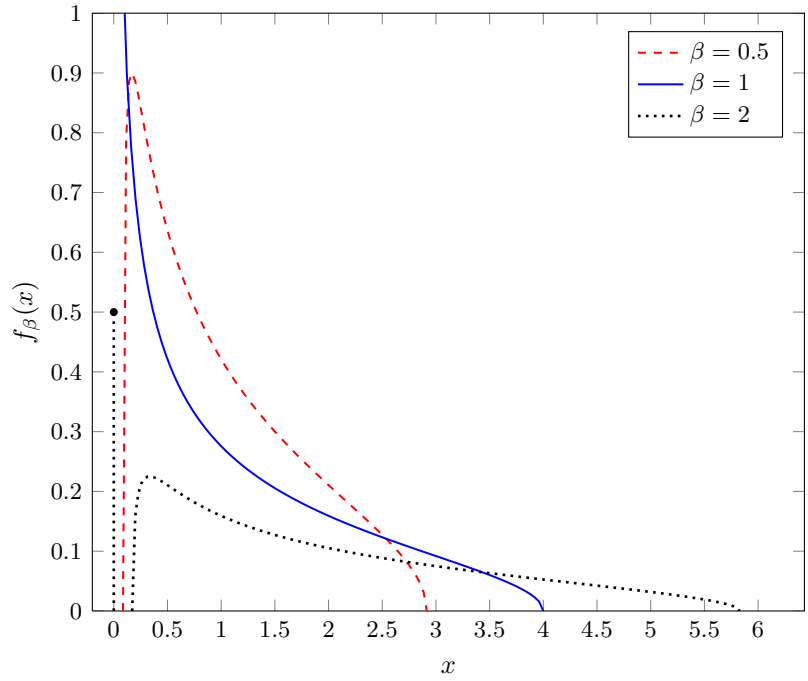


Figure 2.4: Marčenko-Pastur density for different values of β.

Now, let us consider the example in the previous section. First, it can be checked that

$$dF_{\mathbf{X}}^{N}(x) = \frac{1}{N}\sum_{j=1}^{N}\delta(x-\lambda_i(\mathbf{X})).$$

Thus, we can write the ergodic capacity per-transmit antennas as, i.e., C/N as

$$\frac{r}{N}\mathbb{E}\left[\int_0^\infty \log\left(1+\frac{r\mathsf{SNR}}{N}x\right)\, dF_{\frac{1}{r}\boldsymbol{\mathcal{W}}}^{r}(x)\right], \tag{2.4}$$

where $\boldsymbol{\mathcal{W}} = \mathbf{H}\mathbf{H}^H$ for $M < N$. Otherwise, $\boldsymbol{\mathcal{W}} = \mathbf{H}^H\mathbf{H}$. So, $\boldsymbol{\mathcal{W}}$ has size $r \times \ell$, where $\ell = \max\{M, N\}$. Let us denote $\gamma = \frac{r}{N}\mathsf{SNR}$ and $\beta = \ell/r \geq 1$. By applying Theorem 2.1, we can evaluate the integral in (2.4) as follows (see e.g., [14, 95]):

$$\begin{aligned}\int_0^\infty \log(1+\gamma x)\, dF_{\frac{1}{r}\boldsymbol{\mathcal{W}}}^{r}(x) &\xrightarrow{a.s.} \int_a^b \log(1+\gamma x)\, \breve{f}_\beta(x)\, dx \\ &= \beta\log\left(1+\gamma-\frac{1}{4}\mathcal{F}(\gamma,\beta)\right)+\log\left(1+\gamma\beta-\frac{1}{4}\mathcal{F}(\gamma,\beta)\right) \\ &\quad -\frac{\log e}{4\gamma}\mathcal{F}(\gamma,\beta),\end{aligned} \tag{2.5}$$

where $\sqrt{\mathcal{F}(x,y)} = \sqrt{x(1+\sqrt{y})^2+1}-\sqrt{x(1-\sqrt{y})^2+1}$. Let us denote the right-hand side of (2.5) as I^∞. Then, $C/N - \frac{r}{N}I^\infty \xrightarrow{a.s.} 0$. To test the validity of that approximation, let us consider the small size systems: $N=4$, $M=3$ ($\beta=4/3$) and $N=1$, $M=2$ ($\beta=2$). The ergodic capacity (2.1) has no closed-form expression but can be computed numerically according to [90, Theorem 2]. Figure 2.5 presents the ergodic capacity for each case and the corresponding asymptotic approximation for various values of SNR. We can see that the asymptotic results give a very accurate approximation. An almost perfect match is obtained at low SNRs. A very small gap can still be achieved for high SNR values. This simulation confirms one of the benefits applying large system approximations to analyze the system performance of communication systems.

In [53], Marčenko and Pastur considered a more general matrix structure which takes the form

$$\mathbf{Y} = \mathbf{A} + \mathbf{X}\mathbf{T}\mathbf{X}^H, \tag{2.6}$$

where $\mathbf{A}$ is a deterministic Hermitian matrix and $\mathbf{T}$ is a real diagonal matrix. When $\mathbf{A} = \mathbf{0}$

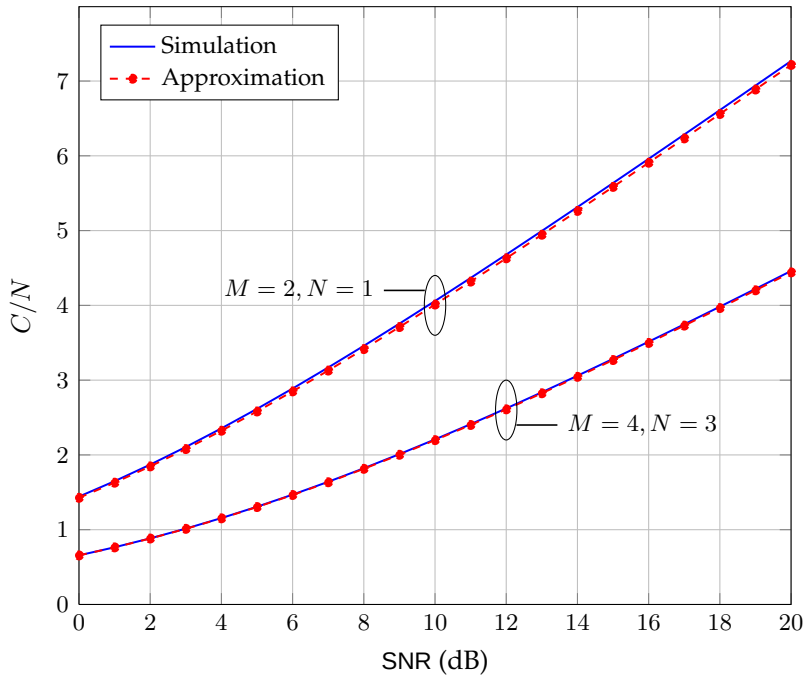


Figure 2.5: Ergodic capacity (2.1) per transmit antenna and its asymptotic approximation for various values of SNR.

and $\mathbf{T} = \mathbf{I}$, the limiting distribution of $\mathbf{Y}$ obeys the Marčenko-Pastur law. In general case, i.e., $\mathbf{A} \neq \mathbf{0}$ or $\mathbf{T} \neq \mathbf{I}$, the limiting distribution of $\mathbf{Y}$ is unknown explicitly. Instead, it is represented via the *Stieltjes transform* which is defined below.

Definition 2.3. *Let X be a real valued random variable with distribution F_X. The Stieltjes transform of F_X is defined as*

$$m(z) = \int_{-\infty}^{\infty} \frac{1}{\lambda - z}\, dF_X(\lambda) = \mathbb{E}_X\left[\frac{1}{X - z}\right],$$

for $z \in \mathbb{C}^+ = \{z \in \mathbb{C} : \Im[z] > 0\}$.

A more rigorous representation can be found in [14, Chapter 3]. We should note that m uniquely determines the distribution function F_X and the converse is also true. For a given m, the density and distribution function of X can be recovered by using the inverse transformation which are stated in the following [14, 95].

Definition 2.4. *The inverse Stieltjes transform of m_X, where X is a real valued random variable with distribution F_X, that yields the density f_X is given by*

$$f_X(x) = \lim_{\omega \to 0^+} \frac{1}{\pi} \Im \left[m(x + j\omega) \right].$$

Equivalently, the distribution function is given by

$$F_X(x) = \frac{1}{\pi} \lim_{\omega \to 0^+} \int_{-\infty}^{x} \Im \left[m(x + j\omega) \right] \, dx.$$

Here, x is a continuity points of F_X (also f_X).

Now, let $\mathbf{X} \in \mathbb{C}^{N \times N}$ be a Hermitian matrix. The Stieltjes transform of $F_{\mathbf{X}}$, denoted by $m_{\mathbf{X}}$, can be alternatively expressed as follows [4, 14]

$$\begin{aligned} m_{\mathbf{X}}(z) &= \int \frac{1}{\lambda - z} \, dF_{\mathbf{X}}(\lambda) = \frac{1}{N} \mathrm{Tr} (\mathbf{\Lambda} - z\mathbf{I}_N)^{-1} \\ &= \frac{1}{N} \mathrm{Tr} (\mathbf{X} - z\mathbf{I}_N)^{-1}, \end{aligned}$$

where $\mathbf{\Lambda}$ is a diagonal matrix that consist of the eigenvalues of $\mathbf{X}$. This implies that evaluating the (normalized) trace of $(\mathbf{X} - z\mathbf{I}_N)^{-1}$ is equivalent to evaluating the Stieltjes transform of $F_{\mathbf{X}}$ and vice-versa. This relation will be used frequently in obtaining the large system results throughout this document.

In what follows, some Stieltjes transforms that will be repeatedly used are listed. We start with the Stieltjes transform $m_{\mathbf{Y}}$ where $\mathbf{Y}$ has a structure given in (2.6).

Theorem 2.2. *Let $\mathbf{X} \in \mathbb{C}^{n \times N}$ be a matrix whose entries are i.i.d. with zero mean and variance $\frac{1}{n}$. Let $\mathbf{T} \in \mathbb{R}^{N \times N}$ be a diagonal matrix and its e.s.d. converges almost surely to a nonrandom distribution $F_{\mathbf{T}}$. Also, let $\mathbf{A} \in \mathbb{C}^{n \times n}$ be a deterministic Hermitian matrix and its e.s.d. converges almost surely to a nonrandom distribution whose Stieltjes transform is $m_{\mathbf{A}}$. Suppose that $\mathbf{X}$, $\mathbf{T}$ and $\mathbf{A}$ are independent. Then, as $n, N \to \infty$ with $\frac{n}{N} \to \beta$, the e.s.d. of*

$$\mathbf{Y} = \mathbf{A} + \mathbf{X}\mathbf{T}\mathbf{X}^H$$

converges almost surely to a nonrandom distribution whose Stieltjes transform is the unique so-

lution of

$$m_{\mathbf{Y}}(z) = m_{\mathbf{A}}\left(z - \frac{1}{\beta}\int \frac{t}{1+tm_{\mathbf{Y}}(z)}\, dF_{\mathbf{T}}(t)\right),$$

for $z \in \mathbb{C}^+$.

The case when $\mathbf{A} = \mathbf{0}$ and $\mathbf{T}$ is not a diagonal matrix is also of a particular interest in the subsequent chapters. Let us denote $\mathbf{Y}$ in that case as $\mathbf{Y}_1$.

Theorem 2.3. *Let* $\mathbf{X}$ *be defined as in Theorem 2.2 and* $\mathbf{T}$ *be an* $N \times N$ *Hermitian matrix and* $\mathbf{T} \geq \mathbf{0}$. *The e.s.d. of*

$$\mathbf{Y}_1 = \mathbf{X}\mathbf{T}\mathbf{X}^H$$

converges almost surely, as $n, N \to \infty$ *with* $\frac{n}{N} \to \beta$, *to a nonrandom distribution whose Stieltjes transform is the unique solution of*

$$m_{\mathbf{Y}_1}(z) = -\left(z - \frac{1}{\beta}\int \frac{t}{1+tm_{\mathbf{Y}_1}(z)}\, dF_{\mathbf{T}}(t)\right)^{-1},$$

for $z \in \mathbb{C}^+$.

Now, consider the following lemma.

Lemma 2.1. *Let* $\mathbf{B} \in \mathbb{C}^{n\times N}$. *Then the following holds*

$$m_{\mathbf{B}^H\mathbf{B}}(z) = \beta m_{\mathbf{B}\mathbf{B}^H}(z) + \frac{(\beta-1)}{z},$$

for $z \in \mathbb{C}^+$.

By using the lemma above, we can establish the following result which is also useful later.

Theorem 2.4. *The e.s.d. of* $\mathbf{Y}_2 = \mathbf{T}^{\frac{1}{2}}\mathbf{X}^H\mathbf{X}\mathbf{T}^{\frac{1}{2}}$ *converges almost surely, as* $n, N \to \infty$ *with* $\frac{n}{N} \to \beta$, *to a nonrandom distribution whose Stieltjes transform is the unique solution of*

$$m_{\mathbf{Y}_2}(z) = \int \frac{1}{t(1-\beta-\beta z m_{\mathbf{Y}_2}(z)) - z}\, dF_{\mathbf{T}}(t),$$

for $z \in \mathbb{C}^+$.

Proof. We can rewrite $m_{\mathbf{Y}_1}$ from Theorem 2.3 as follows

$$\begin{aligned} zm_{\mathbf{Y}_1}(z) &= -1 + \frac{1}{\beta}\int \frac{tm_{\mathbf{Y}_1}(z)}{1+tm_{\mathbf{Y}_1}(z)}\, dF_{\mathbf{T}}(t) \\ &= -1 + \frac{1}{\beta}\int \frac{-1+1+tm_{\mathbf{Y}_1}(z)}{1+tm_{\mathbf{Y}_1}(z)}\, dF_{\mathbf{T}}(t) \\ &= -1 + \frac{1}{\beta} - \frac{1}{\beta}\int \frac{1}{1+tm_{\mathbf{Y}_1}(z)}\, dF_{\mathbf{T}}(t). \end{aligned}$$

Hence,

$$\beta z m_{\mathbf{Y}_1}(z) + \beta - 1 = -\int \frac{1}{1+tm_{\mathbf{Y}_1}(z)}\, dF_{\mathbf{T}}(t).$$

From Lemma 2.1, it follows that

$$m_{\mathbf{Y}_2}(z) = \beta m_{\mathbf{Y}_1} + (\beta - 1)\frac{1}{z}. \tag{2.7}$$

Thus, we have

$$zm_{\mathbf{Y}_2}(z) = -\int \frac{1}{1+tm_{\mathbf{Y}_1}(z)}\, dF_{\mathbf{T}}(t).$$

Substituting $m_{\mathbf{Y}_1}(z)$ in the denominator of the integrand by (2.7) concludes the proof. □

In the Marčenko-Pastur law (Theorem 2.1), all the entries of $\mathbf{H}$ are assumed to have the same variance. The following result establishes the limiting distribution of the sample covariance matrix by allowing different variances for the independent entries of $\mathbf{X}$.

Theorem 2.5 ([30,95]). *Let $\mathbf{X}$ be a $\lfloor cN \rfloor \times \lfloor dN \rfloor$ random matrix with independent entries $[\mathbf{X}]_{ij}$ which are zero mean and variance $\mathbb{E}\left[|[\mathbf{X}]_{ij}|^2\right] = N^{-1}\mathsf{P}_{ij}$, such that P_{ij} are uniformly bounded from above. For each N, let*

$$v_N(x,y) : [0,c] \times [0,d] \to \mathbb{R}$$

be the variance profile function given by

$$v_N(x,y) = \mathsf{P}_{ij}, \quad \frac{i}{N} \le x \le \frac{i+1}{N}, \quad \frac{j}{N} \le y \le \frac{j+1}{N}.$$

Suppose that $v_N(x,y)$ converges uniformly to a limiting bounded function $v(x,y)$. Then, for each

$a, b \in [0, c], a < b$, and $z \in \mathbb{C}^+$

$$\frac{1}{N} \sum_{i=\lfloor aN \rfloor}^{\lfloor bN \rfloor} \left[\left(\mathbf{X}\mathbf{X}^H - z\mathbf{I} \right)^{-1} \right]_{ii} \xrightarrow{i.p.} \int_a^b u(x, z)\, dz,$$

where $u(x, z)$ satisfies

$$u(x, z) = \frac{1}{-z + \int_0^d \frac{v(x,y)dy}{1+\int_0^c u(w,z)v(w,y)dw}}$$

for every $x \in [0, c]$. The solution always exists and is unique in the class of functions $u(x, z) \geq 0$, analytic on $z \in \mathbb{C}^+$ and continuous on $x \in [0, c]$. Moreover, almost surely, the empirical eigenvalue distribution of $\mathbf{X}\mathbf{X}^H$ converges weakly to a limiting distribution whose Stieltjes transform is given by $\int_0^1 u(x, z)\, dx$

In the theorem above, x-axis and y-axis refer to the rows and columns of the matrix $\mathbf{X}$, respectively. This result will be used to prove some theorems presented in Chapter 7.

To this end, we already presented some key results of asymptotic random matrix theory that will be frequently or repeatedly used to derive main results discussed in this document. All those key results show that the limiting spectral distribution is *insensitive* to the type or shape of the density function of the random matrix entries [95]. Thus, the asymptotic limit of ergodic capacity (2.5) still remains the same even when the entries of $\mathbf{H}$ are not Gaussian or from other fading distributions.

In addition to those results, the following lemmas also play important roles or are the heart of the analyses in the following chapters. The lemmas mainly provide asymptotic results for particular quadratic matrix representations in the form $\mathbf{x}\mathbf{A}_N\mathbf{x}^H$. Here, $\mathbf{x} \in \mathbb{C}^{1\times N}$ is a random (row) vector with i.i.d. entries and $\mathbf{A}_N \in \mathbb{C}^{N\times N}$ has a uniformly bounded spectral norm. The first two lemmas are necessary as they show the relation between the quadratic form and the trace of $\mathbf{A}_N$. The first lemma is a well-known result.

Lemma 2.2 ([3, 14, 20]). *Let $\mathbf{x} \in \mathbb{C}^N$ be a random row vector whose entries are i.i.d. with zero mean, variance $1/N$ and a finite eight order moment. Let $\mathbf{A}_N \in \mathbb{C}^{N\times N}$ be a sequence of matrices with uniformly bounded spectral norm and independent of $\mathbf{x}$. Then, as $N \to \infty$,*

$$\mathbf{x}\mathbf{A}_N\mathbf{x}^H - \frac{1}{N} Tr(\mathbf{A}_N) \xrightarrow{a.s.} 0. \tag{2.8}$$

Lemma 2.3 ([20]). *Let* $\mathbf{x}$ *and* $\mathbf{A}_N$ *be as defined in Lemma 2.2. Let* $\mathbf{y}$ *be a random vector which is similar to and independent of* $\mathbf{x}$*. Then,*

$$\mathbf{x}\mathbf{A}_N\mathbf{y}^H \xrightarrow{a.s.} 0, \tag{2.9}$$

as $N \to \infty$*.*

An important quadratic expression is in the form $g_N = \mathbf{x}(\mathbf{B}\mathbf{B}^H + \rho\mathbf{I}_N)^{-1}\mathbf{x}^H$. This form may appear as a part of SINR or SIR expression of a communication system. For example, it appeared as a part of SIR in early papers, at the beginning of the 2000s, on the asymptotic system performance analysis of linear multiuser receivers in CDMA communications [20,43,91,92].

Lemma 2.4. *Let* $g_N = \mathbf{x}(\mathbf{B}\mathbf{B}^H + \rho\mathbf{I}_N)^{-1}\mathbf{x}^H$*, where* $\mathbf{B} \in \mathbb{C}^{N\times n}$ *and* $\rho > 0$*. Suppose that the elements of* $\mathbf{B}$ *are i.i.d. with zero mean and variance* $\frac{1}{N}$*,* $\mathbf{x}$ *and* $\mathbf{B}$ *are independent. Then, as* $n, N \to \infty$ *with* $n/N \to \beta$*,*

$$g_N - g \xrightarrow{a.s.} 0,$$

where g *is given by*

$$g = \left(\rho + \frac{\beta}{1+g}\right)^{-1} = \frac{1}{2}\left(\sqrt{\frac{(1-\beta)^2}{\rho^2} + \frac{2(1+\beta)}{\rho} + 1} + \frac{1-\beta}{\rho} - 1\right). \tag{2.10}$$

Proof. It is easy to see that g_N satisfies the conditions in Lemma 2.2. Hence,

$$g_N - \frac{1}{N}\mathrm{Tr}(\mathbf{B}\mathbf{B}^H + \rho\mathbf{I}_N)^{-1} \xrightarrow{a.s.} 0.$$

From the link between the trace of a matrix and the Stieltjes transform, as previously discussed, we can write

$$\frac{1}{N}\mathrm{Tr}(\mathbf{B}\mathbf{B}^H + \rho\mathbf{I}_N)^{-1} = \int \frac{1}{\lambda+\rho}\, dF_{\mathbf{B}\mathbf{B}^H}(\lambda).$$

Now assume that the entries of $\mathbf{B}$ are i.i.d. with zero mean and variance $\frac{1}{N}$. Moreover,

suppose that $F_{\mathbf{H}^H\mathbf{H}}$ converges almost surely to F. Then,

$$\int \frac{1}{\lambda+\rho}\, dF_{\mathbf{B}\mathbf{B}^H} \xrightarrow{a.s.} \int \frac{1}{\lambda+\rho}\, dF(\lambda),$$

and this establishes that

$$g_N \xrightarrow{a.s.} \int \frac{1}{\lambda+\rho}\, dF(\lambda) = g.$$

From Theorem 2.3 ($\mathbf{T} = \mathbf{I}_N$ in the current case), as $N, n \to \infty$ and $\frac{n}{N} \to \beta$, we have that $g = \lim_{z\to-\rho} m(z)$ is the unique solution of the quadratic equation

$$g = \left(\rho + \frac{\beta}{1+g}\right)^{-1}.$$

Solving for g, we have (2.10) □

Note that the limiting distribution F in the proof above follows the Marčenko-Pastur law. Thus, we can obtain (2.10) by directly evaluating the integral $\int(\lambda+\rho)^{-1}\, dF(\lambda)$. From (2.10), we can see that g is parametrized by β and ρ. Therefore, sometimes it is useful represent g by $g(\beta,\rho)$. For the rest of the thesis, we will use g and $g(\beta,\rho)$ interchangeably.

The next lemma can be thought as a generalization of Lemma 2.2.

Lemma 2.5 ([48, Lemma 5.1]). *Let $\mathbf{x}_m \in \mathbb{C}^N, m \le K$ be independent random vectors whose entries are i.i.d. with zero mean and variance $\frac{1}{N}$. Let $\mathbf{A}_{m,n}, n \le N^\eta, \eta > 0$ be a sequence of random matrices independent of $\mathbf{x}_n$ for each n and have a uniformly bounded spectral norm. Then,*

$$\max_{m,n} \left| \mathbf{x}_m \mathbf{A}_{m,n} \mathbf{x}_m^H - \frac{1}{N} Tr(\mathbf{A}_{m,n}) \right| \xrightarrow{a.s.} 0$$

as $N \to \infty$.

The lemma above can be used to prove the following lemma.

Lemma 2.6 (see also [68, Lemma 2.5]). *Let $m \le N$. Suppose that Lemma 2.5 holds. Then*

$$\left| \frac{1}{N} \sum_{m=1}^{N} \left(\mathbf{x}_m \mathbf{A}_{m,n} \mathbf{x}_m^H - \frac{1}{N} Tr(\mathbf{A}_{m,n}) \right) \right| \xrightarrow{a.s.} 0. \tag{2.11}$$

Proof. We can have the upper bound of the left-hand side (LHS) of (2.11) as follows

$$\left|\frac{1}{N}\sum_{m=1}^{N}\left(\mathbf{x}_m\mathbf{A}_{m,n}\mathbf{x}_m^H-\frac{1}{N}\mathrm{Tr}(\mathbf{A}_{m,n})\right)\right| \leq \frac{1}{N}\sum_{m=1}^{N}\left|\mathbf{x}_m\mathbf{A}_{m,n}\mathbf{x}_m^H-\frac{1}{N}\mathrm{Tr}(\mathbf{A}_{m,n})\right| \leq \max_{m\leq N,n}\left|\mathbf{x}_m\mathbf{A}_{m,n}\mathbf{x}_m^H-\frac{1}{N}\mathrm{Tr}(\mathbf{A}_{m,n})\right|.$$

Since

$$0\leq \max_{m\leq N,n}\left|\mathbf{x}_m\mathbf{A}_{m,n}\mathbf{x}_m^H-\frac{1}{N}\mathrm{Tr}(\mathbf{A}_{m,n})\right| \xrightarrow{a.s.} 0,$$

then (2.11) follows immediately. □

Suppose that $o_N = \frac{1}{N}\mathrm{Tr}(\mathbf{X}+\rho\mathbf{I}_N)^{-1}$, where $\mathbf{X}\in\mathbb{C}^{N\times N}$ is Hermitian. Let $p_N = \frac{1}{N}\mathrm{Tr}(\mathbf{X}+\mathbf{q}\mathbf{q}^H+\rho\mathbf{I}_N)^{-1}$, where $\mathbf{q}\in\mathbb{C}^N$. It is clear that p_N is o_N in which $\mathbf{X}$ is perturbed by a rank-1 matrix $\mathbf{q}\mathbf{q}^H$. The difference between of $o_N$and p_N will be the subject of interest of the following lemmas. The first lemma states that the difference between them is bounded by $(N\rho)^{-1}$ which implies that it gets smaller as N grows. The second lemma provides the generalization of the first lemma in which the difference is almost surely zero, as $N\to\infty$.

Lemma 2.7 (Rank-1 Perturbation Lemma, [83]). *Let $\rho>0$, $\mathbf{A}\in\mathbb{C}^{N\times N}$, $\mathbf{B}\in\mathbb{C}^{N\times N}$ be Hermitian, and $\mathbf{q}\in\mathbb{C}^N$. Then,*

$$\left|\frac{1}{N}Tr\left(\mathbf{A}(\mathbf{B}+\rho\mathbf{I}_N)^{-1}-\mathbf{A}(\mathbf{B}+\mathbf{q}\mathbf{q}^H+\rho\mathbf{I}_N)^{-1}\right)\right|\leq\frac{\|\mathbf{A}\|}{N\rho}.$$

Lemma 2.8 ([14]). *Let $\rho>0$, $\mathbf{A}\in\mathbb{C}^{N\times N}$, $\mathbf{B}\in\mathbb{C}^{N\times N}$ be Hermitian with eigenvalues $\lambda_1(\mathbf{B})\leq\cdots\leq\lambda_N(\mathbf{B})$, and $\mathbf{q}\in\mathbb{C}^N$. Suppose that there exists $\epsilon>0$ such that $\lambda_1(\mathbf{B})\geq\epsilon$ almost surely for all large N. Then,*

$$\frac{1}{N}Tr(\mathbf{A}(\mathbf{B}+\rho\mathbf{I}_N)^{-1}-\mathbf{A}(\mathbf{B}+\mathbf{q}\mathbf{q}^H+\rho\mathbf{I}_N)^{-1})\xrightarrow{a.s.}0. \tag{2.12}$$

To this end, we already presented some important mathematical tools for the asymptotic analysis in this document. The upcoming chapters discuss the application of those tools in analyzing the system performance of MISO broadcast channels for some channel conditions and system setups.

Chapter 3
Multiuser Precoding in i.i.d. Channel

In this chapter, the large system results for a single cell multiuser transmit via Regularized Channel Inversion (RCI) precoder with i.i.d. entries of the channel gain matrix are presented. In the analysis, the transmitter (BS) and the users are assumed to have a perfect CSI. The first result is the expression of the SINR in the large system limit, also called the limiting SINR. Then, the optimal regularization parameter of the RCI precoder that maximizes the limiting SINR, as well as the limiting sum rate per antenna, is derived. Its expression is remarkably simple. It is the ratio between the cell loading and the received SNR. Furthermore, we also derive the limiting SINR for the Moore-Penrose Channel Inversion (MPCI) and the Single User (SU) precoders. Those precoders are the special cases of the RCI precoder when the regularization parameter goes to zero and infinity, respectively. For each precoder, the characterization of the sum rate per-antenna with respect to the cell-loading is also presented.

3.1 Introduction

AS mentioned previously in the introductory chapter, one way to achieve the sum-capacity of MIMO broadcast channels is to employ a precoder at the transmitter side to reduce the MUI. Precoders exploit the CSI at the transmitter (CSIT) to adapt the transmission strategies or variables such as the 'direction' and 'magnitude' (or power) of the transmission for each user's data symbols. Perfect CSIT will be an ideal case and is often considered in the performance analysis of MIMO communication systems as a benchmark for the practical scenarios. Different forms of CSIT will lead to different ways of designing a precoder. Moreover, different performance criteria and channel conditions e.g., i.i.d. channel and spatially correlated channel will also influence the precoder design.

While the linear precoder can be optimal for single-user MIMO systems [6, Chapter

3],[100, and references therein], it is not the case for the multiuser MIMO settings. The sum capacity in multiuser MIMO downlink is achieved by using the Dirty Paper Coding (DPC) strategy [9,98,99,108]. In [9], the transmit signal is decomposed as a multiplication of a precoding matrix and an auxiliary input. A successive dirty paper encoding is applied to generate the auxiliary input. The precoding matrix is then optimized to achieve the optimal sum rate for a two-user MISO setup. The work by Viswanath and Tse [99] generalizes the previous result for an arbitrary number of single-antenna users. The sum capacity which is the maximum sum rate of the DPC-based transmission strategy is established by exploring the duality between MIMO-BC and MIMO-MAC. A similar technique is also used in an independent work [98] to derive the sum-capacity for the case where each user has multiple antennas. Yu and Cioffi in [108] use a DPC-based decision feedback equalizer at the transmitter to prove the same result.

DPC requires perfect CSIT and an entire knowledge of the transmitted signals. Therefore, the resulting interference at the receivers can be computed in advance. As a result, the interference can be pre-subtracted along with the transmitted messages so that the sum-capacity of using DPC is the same as if there were no interference. Regardless of the benefit in terms of the achieved sum- rate, its practical implementation is computationally intensive because it involves successive encoding at the transmitter and decoding at the receiver [86],[19]. Finding practical schemes for DPC implementations remains an active research area. The DPC could be categorized as a nonlinear precoder. Another example of the nonlinear precoding is the vector perturbation technique where the transmitted data or symbols are perturbed in order to prevent power enhancement caused by the channel inversion type precoders [32]. Although the vector perturbation gives better performance compared to linear precoders, finding the optimal perturbation vector still requires a high complexity search of the optimal point in lattice (an example of a well studied NP-hard problem)[74].

A lower complexity implementation is one of the advantages offered by linear precoders, but with the sum rate penalty compared to the DPC. An optimal linear precoder is hard to obtain. Some studies propose iterative algorithms for the precoder construction, but there is no proof for their convergences. One of the most well-known linear

precoders is the Zero-Forcing (ZF) or Channel Inversion (CI) precoder. It has a simple structure, i.e., the inverse of the users' channel gain matrix [87]. Employing this precoder will completely eliminate the MUI. Although it is a sub-optimal linear precoder, it has been shown in [107] that the combination of ZF and a particular user selection strategy, called semi-orthogonal user selection, can perform asymptotically (for a large number of users) close to the sum capacity of MISO BC. However, as shown in [69], the sum- capacity of this beamformer does not grow linearly with the number of users K, when $K = N$, where N is the number of transmit antennas. Moreover, if the channel gain matrix is ill-conditioned, the received SNR will drop significantly and this will affect the resulting SINR and sum rate.

To resolve that problem, a regularization parameter is introduced in the channel inversion and the corresponding sum-capacity scales linearly with $\min(N, K)$ but at a slower rate than that of DPC [69]. The resulting beamformer is called the Regularized Channel Inversion (RCI) or also known as Regularized Zero-Forcing (RZF). It does not fully cancel the MUI as ZF does but it offers a higher sum rate, particularly at low SNRs [69]. The regularization parameter controls the amount of interference introduced to each user. Therefore, this parameter should be chosen optimally to maximize some performance criteria such as the SINR. For the case $K = N$, the optimal parameter has been derived in [69] by assuming K is large. For general cases, it was derived in [62] by using the large system analysis (see also [101]). Finding it in the finite-size regime can be difficult.

In this chapter, we perform a large system analysis of a single-cell broadcast channels with the RCI precoder at the transmitter. The entries of the channel gain matrix are assumed to be i.i.d. with certain statistical moment conditions. Each user knows its channel perfectly and the transmitter is also assumed to have a perfect knowledge of the channel gain matrix. First, we derive the SINR in the large system limit which later is called the limiting SINR. Then, we obtain the expression for the optimal regularization parameter that maximizes the limiting SINR. As in [62], the analysis presented here holds for any cell loading, $\beta = K/N > 0$. Here, we strengthen the results presented in [62] by obtaining a nicer expression for the limiting SINR and a more concise derivation for the

optimal regularization parameter. We also provide a new characterization of the sum rate of the RCI over the cell-loading. We also derive the limiting SINR of the MPCI and SU from their finite-size SINR expressions while in [62,63], the limiting SINR expressions for those precoders are obtained from the limiting SINR of the RCI by setting the regularization parameter to zero and infinity, respectively. It turns out that both approaches produce the same results.

A closely related work is [101] that performs the large system analysis for the MISO downlink for various channel conditions. The current work and [101] are developed independently and use different approaches in deriving the large system results. Our approach relies on the behavior of the l.s.d. of random matrices in the model while the approach in [101] is mainly based on the deterministic equivalent of the involved random matrices (see [14] and references therein for detailed discussions on this subject).

The contributions of this chapter can be summarized as follows:

1. We perform the large system analysis for the SINR of the broadcast channel with the RCI precoder, for any $\beta > 0$. We show that the SINR converges (almost surely) to a deterministic (non-random) limiting SINR. The limiting SINR is the same for each user in the cell. It is the function of the cell loading, regularization parameter and the received SNR. We also demonstrate that the SINR of the finite-size system is well-approximated by the limiting SINR even for reasonably small system dimensions.
2. For a given received SNR and cell-loading, we derive the optimal regularization parameter that maximizes the limiting SINR. Its expression is simple and the same as that obtained in [69] for the particular case $K = N$. It is the ratio between the cell loading and the received SNR. Our numerical results show that employing the asymptotically optimal regularization parameter into the finite size systems only incurs a small throughput loss.
3. We also derive the limiting SINR of the broadcast channel with the MPCI and SU. The results are exactly the same as those obtained by evaluating the limiting SINR of the RCI for regularization parameter going to zero and infinity for MPCI and SU respectively.

4. For each considered precoder, we study the characterization of the sum rate per-antenna over different values of cell-loading. For the RCI precoder, the sum rate per-antenna is evaluated at the optimal value of the regularization parameter. We show that for the received SNR larger than one (zero decibel), the sum rate is unimodal function with respect to the cell loading. Thus, the optimal cell loading that maximizes the sum rate per-antenna is unique. For the SNR less or equal to one, the sum rate per antenna is an increasing function of the cell-loading. Consequently, the maximum sum rate per-antenna could be achieved with cell loading greater than one i.e., the number of users is larger that the number of transmit antenna. A similar phenomenon is also observed for the optimal cell-loading of the MPCI precoder. For the SU precoder, the sum rate per-antenna is an increasing function of the cell loading and converges to a certain value as the cell loading tends to infinity.

The rest of this chapter is organized as follows. In the next section, the system model is presented and the SINR expressions of the users when the BS employs one of the following precoders: RCI, CI or ZF, and SU, are derived. The analysis of these SINRs in the large system limit is presented in Section 3.3. The optimal regularization parameter of the RCI precoder maximizing the limiting SINR and the optimal cell-loading maximizing the sum rate per antenna are also discussed. The characterizations of the limiting SINR and sum rate per antenna over the cell-loading for MPCI and SU precoders are also studied. Section 3.4 provides some numerical results that compare the performance of different precoders and also show the applicability of the large system results for the finite-size systems. Section 3.5 concludes the chapter.

3.2 System Model

Figure 3.1 illustrates a MISO broadcast channel with a linear precoder at the transmitter end. The base station has N antennas and each of K users is equipped with a single antenna. The propagation channel coefficient between the transmit antenna $n \in \{1, \ldots, N\}$ and user $k \in \{1, \ldots, K\}$ is denoted by $h_{k,n}$. The channel gain vector between the BS and user k is represented by the row vector $\mathbf{h}_k = [h_{k,1}\ h_{k,2}\ \ldots\ h_{k,N}] \in \mathbb{C}^{1\times N}$. It is assumed that the entries of $\mathbf{h}_k$ are i.i.d. and $\mathbf{h}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$. Even though here, we assume a specific distribution for $\mathbf{h}_k$, the large system holds for any distribution of $\mathbf{h}_k$ if the entries of $\frac{1}{\sqrt{N}}\mathbf{h}_k$ are i.i.d. with zero mean, variance $\frac{1}{N}$ and have finite eighth moment (see e.g., [14]).

In broadcast transmissions, the BS sends the data symbols $\mathbf{s} = [s_1\ \ldots\ s_k\ \ldots\ s_K]^T$ intended for each user k simultaneously across its N antennas. Before transmission the data symbol vector $\mathbf{s}$ is precoded by a beamforming or precoder matrix $\mathbf{P}$. So, the transmitted data vector $\mathbf{x} = [x_1\ \cdots\ x_N]^T$ is given by $\mathbf{x} = \mathbf{P}\mathbf{s}$ and has a fixed transmit power constraint, $\mathbb{E}[\|\mathbf{x}\|^2] = P_d$. The received signal at user k can be written as

$$y_k = \sum_{n=1}^{N} h_{k,n}x_n + w_k = \mathbf{h}_k\mathbf{x} + w_k,$$

where w_k is the complex Gaussian receiver noise with zero mean and variance σ^2. Expanding $\mathbf{x}$, the received signal becomes

$$y_k = \mathbf{h}_k\mathbf{P}\mathbf{s} + w_k. \tag{3.1}$$

In this document, we focus on the RCI precoder proposed in [69]. This precoder takes the form

$$\mathbf{P} = c\mathbf{H}^H\left(\mathbf{H}\mathbf{H}^H + \alpha\mathbf{I}_K\right)^{-1}, \tag{3.2}$$

or equivalently,

$$\mathbf{P} = c\left(\mathbf{H}^H\mathbf{H} + \alpha\mathbf{I}_N\right)^{-1}\mathbf{H}^H, \tag{3.3}$$

where $\mathbf{H} = [\mathbf{h}_1^T\ \ldots\ \mathbf{h}_k^T\ \ldots\ \mathbf{h}_K^T]^T$ is the $K \times N$ complex channel gain matrix. We should

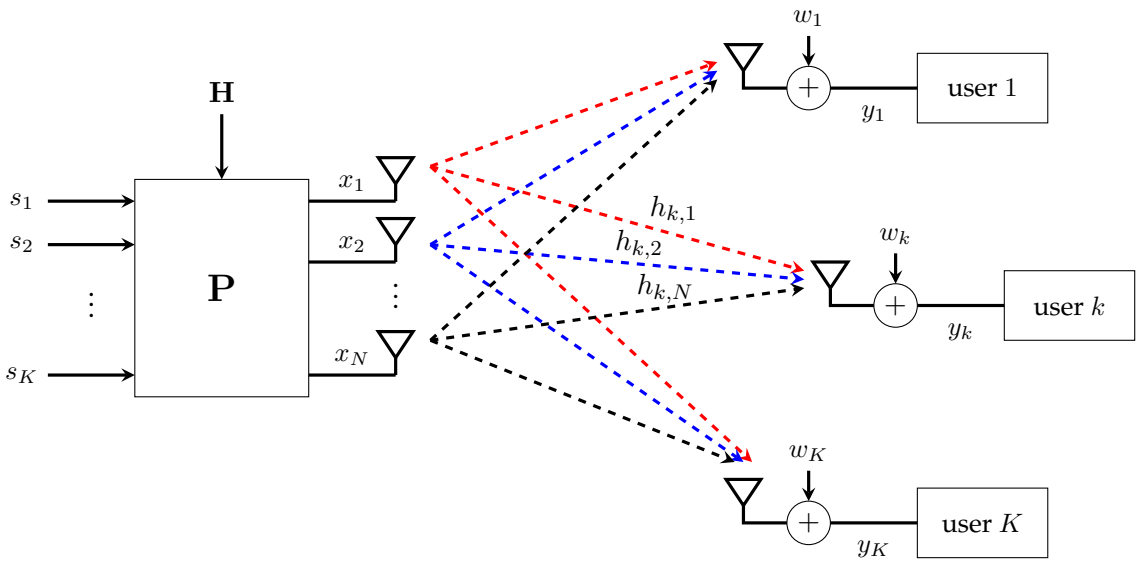


Figure 3.1: System model.

note that the channel gain vectors amongst the users are assumed to be independent. The normalizing constant c is chosen to satisfy the power constraint $\mathbb{E}[\|\mathbf{x}\|^2|\mathbf{H}] = P_d$. Assuming the data symbols $\{s_1, \ldots, s_K\}$ are independent and each symbol has unit power, c is given by

$$c^2 = \frac{P_d}{\mathrm{Tr}\left(\left(\mathbf{H}\mathbf{H}^H + \alpha\mathbf{I}_K\right)^{-2}\mathbf{H}\mathbf{H}^H\right)}, \tag{3.4}$$

or equivalently,

$$c^2 = \frac{P_d}{\mathrm{Tr}\left(\left(\mathbf{H}^H\mathbf{H} + \alpha\mathbf{I}_N\right)^{-2}\mathbf{H}^H\mathbf{H}\right)}. \tag{3.5}$$

Using the representation (3.3) for the precoder, the received signal (3.1) can be written as

$$\begin{aligned} y_k &= c\mathbf{h}_k\left(\mathbf{H}^H\mathbf{H} + \alpha\mathbf{I}_N\right)^{-1}\mathbf{H}^H\mathbf{s} + w_k \\ &= c\mathbf{h}_k\left(\mathbf{H}^H\mathbf{H} + \alpha\mathbf{I}_N\right)^{-1}\mathbf{h}_k^H s_k + c\sum_{j\neq k}^{K}\mathbf{h}_k\left(\mathbf{H}^H\mathbf{H} + \alpha\mathbf{I}_N\right)^{-1}\mathbf{h}_j^H s_j + w_k. \end{aligned} \tag{3.6}$$

The first term in the right-hand side is the desired signal for user k while the other terms are the interference and the receiver noise. Assuming single-user decoding at the receiver

and treating the interference as noise, the SINR of user k can be expressed as follows

$$\text{SINR}_k = \frac{c^2 \left| \mathbf{h}_k \left(\mathbf{H}^H \mathbf{H} + \alpha \mathbf{I}_N \right)^{-1} \mathbf{h}_k^H \right|^2}{\sigma^2 + c^2 \sum_{j \neq k} \left| \mathbf{h}_k \left(\mathbf{H}^H \mathbf{H} + \alpha \mathbf{I}_N \right)^{-1} \mathbf{h}_j^H \right|^2}. \tag{3.7}$$

It is in the standard form where the numerator and the second term in the denominator represent the signal and the interference energy respectively.

Now let us consider the MPCI and SU precoders that can be seen as the special cases of the RCI precoder. Letting $\alpha \to 0$, we have the MPCI precoder which is given by

$$\mathbf{P}_{\text{MPCI}} = \begin{cases} \lim\limits_{\alpha \to 0} c \mathbf{H}^H \left(\mathbf{H}\mathbf{H}^H + \alpha \mathbf{I}_K \right)^{-1} = c_1 \mathbf{H}^H \left(\mathbf{H}\mathbf{H}^H \right)^{-1}, & K/N \leq 1 \\ \lim\limits_{\alpha \to 0} c \left(\mathbf{H}^H \mathbf{H} + \alpha \mathbf{I}_N \right)^{-1} \mathbf{H}^H = c_2 \left(\mathbf{H}^H \mathbf{H} \right)^{-1} \mathbf{H}^H, & K/N > 1, \end{cases} \tag{3.8}$$

where $c_1^2 = \frac{P_d}{\text{Tr}\left((\mathbf{H}\mathbf{H}^H)^{-1}\right)}$ and $c_2^2 = \frac{P_d}{\text{Tr}\left((\mathbf{H}^H\mathbf{H})^{-1}\right)}$.

The first case, i.e., $\beta = \frac{K}{N} \leq 1$, is commonly considered in the wireless communication literatures. In practical situations, the number of users is larger than the number of antenna. So, user selections should be done accordingly. The received signal at the user k can be written as

$$y_k = c \mathbf{h}_k \mathbf{H}^H \left(\mathbf{H}\mathbf{H}^H \right)^{-1} \mathbf{s} + w_k.$$

Stacking $y_1, \ldots, y_K$ into a vector $\mathbf{y} = [y_1 \; y_2 \; \cdots \; y_K]^T$ leads to

$$\begin{aligned} \mathbf{y} &= c_1 \mathbf{H}\mathbf{H}^H \left(\mathbf{H}\mathbf{H}^H \right)^{-1} \mathbf{s} + \mathbf{w} \\ &= c_1 \mathbf{s} + \mathbf{w}, \end{aligned} \tag{3.9}$$

where $\mathbf{w} = [w_1 \; w_2 \; \cdots \; w_K]^T$ with $\mathbb{E}[\mathbf{w}\mathbf{w}^H] = \sigma^2 \mathbf{I}_K$. By using the precoder, it is obvious that the MUI is eliminated (zero-forced). The zero-forcing scheme can be seen to have the *altruistic* behaviour since the precoding vector for a particular user tries to eliminate the interference that it can cause to other users. From (3.9), the SINR of user k is simply given by

$$\text{SINR}_{\text{ZF},k} = \frac{c_1^2}{\sigma^2}. \tag{3.10}$$

Since the SINR is proportional to c_1^2, a rank deficiency on the matrix $\mathbf{H}\mathbf{H}^H$ will lead to a penalty to c^2 and also consequently to the SINR.

For the case $\beta > 1$, the received signal is given by

$$\begin{aligned} y_k &= c_2 \mathbf{h}_k \left(\mathbf{H}^H \mathbf{H}\right)^{-1} \mathbf{H}^H \mathbf{s} + w_k \\ &= c_2 \mathbf{h}_k \left(\mathbf{H}^H \mathbf{H}\right)^{-1} \mathbf{h}_k^H s_k + c_2 \sum_{j \neq k}^{K} \mathbf{h}_k \left(\mathbf{H}^H \mathbf{H}\right)^{-1} \mathbf{h}_j^H s_j + w_k \, . \end{aligned}$$

It shows us that the MUI still presents in the system. Hence, putting the term zero-forcing for the precoder for the current case could be misleading. The SINR of user k can be written as

$$\text{SINR}_{\text{NZF},k} = \frac{c_2^2 \left| \mathbf{h}_k \left(\mathbf{H}^H \mathbf{H}\right)^{-1} \mathbf{h}_k^H \right|^2}{\sigma^2 + c_2^2 \sum_{j \neq k} \left| \mathbf{h}_k \left(\mathbf{H}^H \mathbf{H}\right)^{-1} \mathbf{h}_j^H \right|^2}, \tag{3.11}$$

which is (3.7) with $\alpha \to 0$. Intuitively, the limiting SINR for (3.11) can also be obtained from the limiting SINR of (3.7).

The SU precoder is obtained from the RCI precoder by letting $\alpha \to \infty$ and we have

$$\begin{aligned} \mathbf{P}_{\text{SU}} &= \lim_{\alpha \to \infty} \mathbf{P} \\ &= \lim_{\alpha \to \infty} \sqrt{P_d} \frac{\mathbf{H}^H \left(\frac{1}{\alpha} \mathbf{H}\mathbf{H}^H + \mathbf{I}_K\right)^{-1}}{\sqrt{\text{Tr}\left(\left(\frac{1}{\alpha} \mathbf{H}\mathbf{H}^H + \mathbf{I}_K\right)^{-2} \mathbf{H}\mathbf{H}^H\right)}} \\ &= \lim_{\alpha \to \infty} \sqrt{P_d} \frac{\left(\frac{1}{\alpha} \mathbf{H}^H\mathbf{H} + \mathbf{I}_K\right)^{-1} \mathbf{H}^H}{\sqrt{\text{Tr}\left(\left(\frac{1}{\alpha} \mathbf{H}^H\mathbf{H} + \mathbf{I}_N\right)^{-2} \mathbf{H}^H\mathbf{H}\right)}} \\ &= \sqrt{\frac{P_d}{\text{Tr}\left(\mathbf{H}\mathbf{H}^H\right)}} \mathbf{H}^H. \end{aligned}$$

By using this precoder, the beamforming direction for the user k is given by $\sqrt{\frac{P_d}{\text{Tr}(\mathbf{H}\mathbf{H}^H)}} \mathbf{h}_k^H$. It will maximize the signal strength of the user k while ignoring the interference that it causes to other users. This can be considered as a *selfish* precoding scheme. The received

signal vector can be expressed as

$$\mathbf{y} = \sqrt{\frac{P_d}{\mathrm{Tr}\left(\mathbf{H}\mathbf{H}^H\right)}}\left(\mathbf{H}\mathbf{H}^H\mathbf{s}\right) + \mathbf{w},$$

and for user k, the received signal becomes

$$y_k = \sqrt{\frac{P_d}{\mathrm{Tr}\left(\mathbf{H}\mathbf{H}^H\right)}}\left(\mathbf{h}_k\mathbf{h}_k^H s_k + \sum_{j\neq k}\mathbf{h}_k\mathbf{h}_j^H s_j\right) + w_k.$$

Thus, the SINR is given by

$$\begin{aligned}\mathrm{SINR}_{\mathrm{SU},k} &= \frac{\frac{P_d}{\mathrm{Tr}(\mathbf{H}\mathbf{H}^H)}\left|\mathbf{h}_k\mathbf{h}_k^H\right|^2}{\sigma^2 + \frac{P_d}{\mathrm{Tr}(\mathbf{H}\mathbf{H}^H)}\sum_{j\neq k}\left|\mathbf{h}_k\mathbf{h}_j^H\right|^2} \\ &= \frac{\frac{P_d}{\mathrm{Tr}(\mathbf{H}\mathbf{H}^H)}\left|\mathbf{h}_k\mathbf{h}_k^H\right|^2}{\sigma^2 + \frac{P_d}{\mathrm{Tr}(\mathbf{H}\mathbf{H}^H)}\mathbf{h}_k\mathbf{H}_k^H\mathbf{H}_k\mathbf{h}_k}.\end{aligned} \tag{3.12}$$

Besides the SINR as the basic performance measure in this chapter, we also consider an information theoretic measure which is represented by achievable sum rate. It can be expressed as

$$R_{\mathrm{sum}} = \sum_{k=1}^{K}\log(1+\mathrm{SINR}_k), \tag{3.13}$$

where the SINR_k may represent (3.7), (3.10), (3.11) and (3.12). Clearly, the achievable rate for each user is a logarithmically monotonic function of the SINR. Also, there is one-to-one mapping between the achievable rate and the SINR for each user. The limiting SINR for each user under different precoders at the transmitter will be analyzed in the following section.

3.3 Large Sytem Analysis

The finite-size SINR of the broadcast channel with RCI precoder has been derived previously in [69] for the case $K = N$. It is the function of the eigenvalues of the random matrix $\mathbf{H}\mathbf{H}^H$ [69, cf. eq. (29)]. Thus, the SINR depends on the channel realization. Moreover, it is difficult to obtain the optimal regularization parameter that maximizes the SINR or sum rate in the finite-size regime.

Here, we derive the SINR in the large system limit where N and K tend to infinity with fixed $\beta = K/N$. The result is stated in the following theorem.

Theorem 3.1. *Let $\rho = \frac{\alpha}{N}$ be the normalized regularization parameter, $\gamma = \frac{P_d}{\sigma^2}$ be the received SNR and $g(\beta, \rho)$ be the function defined in (2.10). Then, desired signal converges almost surely to*

$$P_d \frac{g(\beta,\rho)}{(1+g(\beta,\rho))^2}\left(1+\frac{\rho}{\beta}(1+g(\beta,\rho))^2\right), \tag{3.14}$$

and the interference converges almost surely to

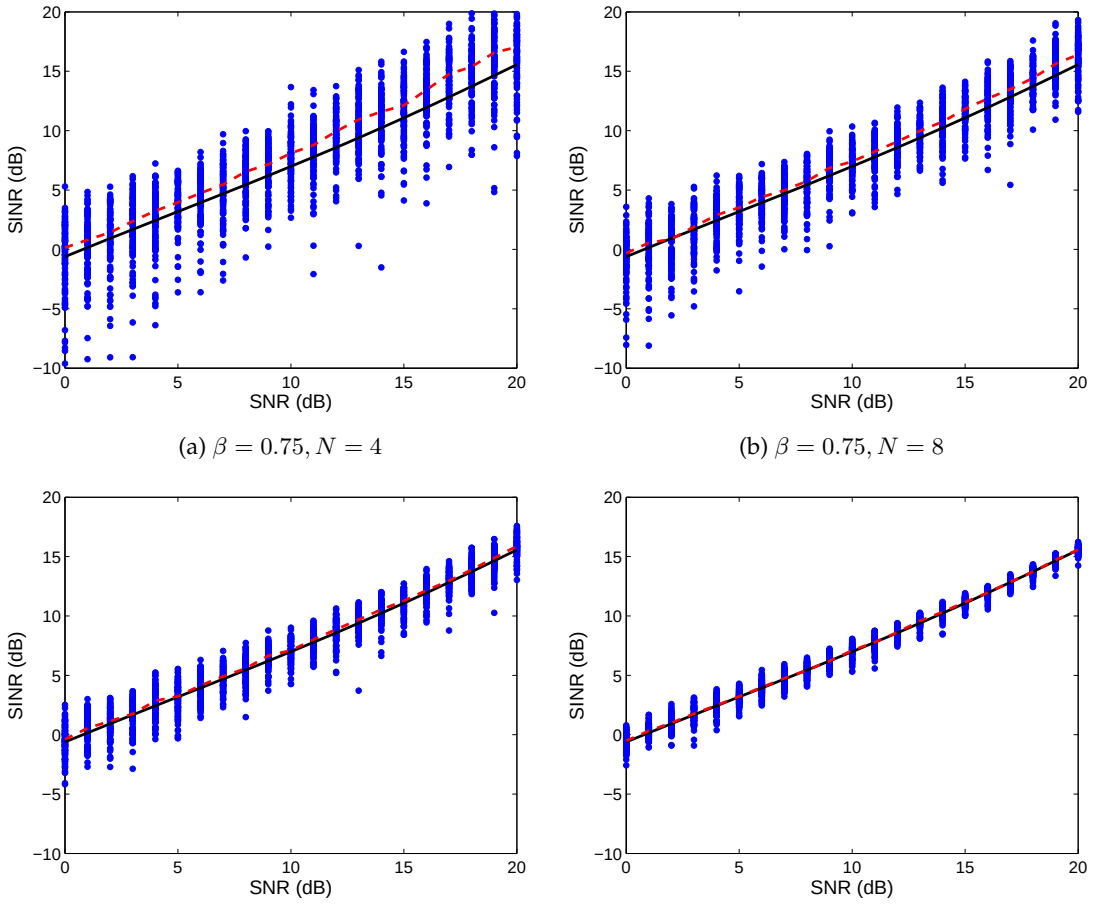
$$\frac{P_d}{(1+g(\beta,\rho))^2}. \tag{3.15}$$

Consequently, the $SINR_k$ converges almost surely to a deterministic limiting SINR, given by

$$SINR^{\infty} = g(\beta,\rho)\frac{\gamma+\frac{\gamma\rho}{\beta}\left(1+g(\beta,\rho)\right)^2}{\gamma+\left(1+g(\beta,\rho)\right)^2}. \tag{3.16}$$

Proof. Refer to Appendix 3.6.1 □

A quick glance at the limiting SINR expression shows us that it is the same for all users and depends only on system parameters: regularization parameter ρ, cell loading β and received SNR γ. Figure 3.2 depicts how the finite-size SNRs approach the limiting SINR for different values of SNR as the dimensions of the system get large. We set $\beta = 0.6$ and $\rho = \beta/\gamma$. For each SNR value, 100 channel realizations $\mathbf{H}$ are generated. For each channel realization the SINR is computed based on (3.7) and is marked by the dot. The dash line and solid line represent the average SINR over all 100 channel realizations and the limiting SINR, respectively. From the plots, we can see that even for a small-size



(a) $\beta = 0.75, N = 4$

(b) $\beta = 0.75, N = 8$

(b) $\beta = 0.75, N = 16$

(d) $\beta = 0.75, N = 64$

Figure 3.2: Comparison of the randomly generated SINR for user 1 (dot) with the average SINR (dash) and the asymptotic limit (3.16) (solid line) with $\rho = \beta/\gamma$.

system $N = 4$, the average and limiting SINR are quite close. However, the deviation or the spread of the SINR samples from the average and the asymptotic is still quite large. As N gets larger, the gap between the average and the limiting becomes smaller and so does the spread. For $N = 64$, the gap is negligible and the spread is about 2.5 dB.

Based on (3.16), the limiting sum rate per-antenna can then be defined as follows

$$R_{\text{sum}}^{\infty} = \beta \log(1 + \text{SINR}^{\infty}). \tag{3.17}$$

For the sake of simplicity in the naming, for the rest, we just call the limiting sum rate per-antenna as the limiting sum rate. We can see that there is a one-to-one monotonic

mapping between the limiting sum rate and the limiting SINR. Maximizing the limiting SINR via ρ and γ will equivalently maximize the limiting sum rate. Since β appears as the pre-log factor in (3.17), the link between the limiting sum rate and SINR is not monotonic w.r.t. β. Thus, characterizing the behaviour of the limiting sum rate and the limiting SINR should be considered separately.

First, let us discuss the role of the regularization parameter for the limiting SINR. From (3.15), it is clear that the interference energy is decreasing as $g(\beta, \rho)$ increases. However, as shown by (3.32), $g(\beta, \rho)$ is a decreasing function of ρ. So, increasing ρ will increase the level of interference. Now, how does ρ affect the (desired) signal strength?

Proposition 3.1. *In the large system limit, the desired signal strength is increasing in ρ.*

Proof. For brevity, we use g to denote $g(\beta, \rho)$. Let us represent (3.14) by $P_d S^{\infty}(\rho, g)$ (note that g is also a function of ρ) where $S^{\infty}(\rho, g)$ can be written as

$$S^{\infty}(\rho, g) = \frac{g}{(1+g)^2} + \frac{\rho}{\beta} g.$$

Then, the total derivative of $\partial S^{\infty}(\rho, g)$ w.r.t. ρ is given by

$$\frac{dS^{\infty}(\rho, g)}{d\rho} = \frac{\partial S^{\infty}(\rho, g)}{\partial \rho} + \frac{\partial S^{\infty}(\rho, g)}{\partial g} \frac{\partial g}{\partial \rho}.$$

The first term in the right-hand side is simply given by $\frac{g}{\beta}$. The expression for $\frac{\partial g}{\partial \rho}$ is given by (3.32). It is easy to show that

$$\frac{\partial S^{\infty}(\rho, g)}{\partial g} = \frac{1-g}{(1+g)^3} + \frac{\rho}{\beta}.$$

Thus,

$$\begin{aligned} \frac{dS^{\infty}(\rho, g)}{d\rho} &= \frac{g}{\beta} - \left(\frac{1-g}{(1+g)^3} + \frac{\rho}{\beta} \right) \frac{g(1+g)^2}{\beta + \rho(1+g)^2} \\ &= \frac{g - \frac{g(1-g)}{(1+g)}}{\beta + \rho(1+g)^2} \\ &= \frac{2g^2}{1+g} \frac{1}{\beta + \rho(1+g)^2} > 0, \end{aligned}$$

and this completes the proof. □

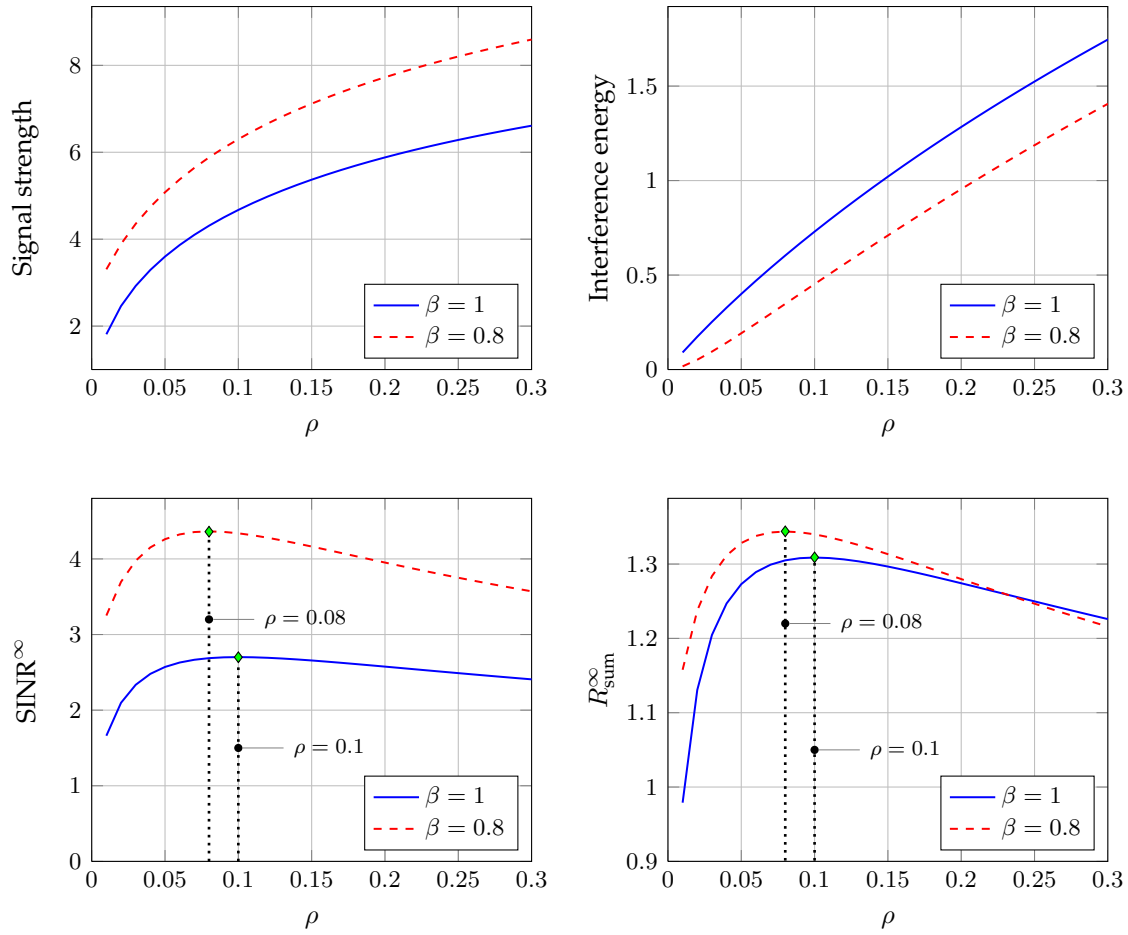


Figure 3.3: Signal strength, interference energy, SINR and sum rate per-antenna in the large system limit for different values of ρ. Parameters: $\gamma = 10$ dB.

Figure 3.3 illustrates the behavior of the desired signal and interference energy w.r.t. ρ. As predicted by the analysis, both are increasing with ρ. To get benefits for the SINR, we should, at the same time, reduce ρ to suppress the interference and increase ρ to strengthen the desired signal. Therefore, ρ provides a trade-off between reducing the interference level and increasing the signal energy and it should be chosen carefully in order to maximize the system performance. Figure 3.3 also depicts the optimal ρ that maximizes the limiting SINR and sum rate which is marked by the diamond. The maximizer for the limiting SINR and sum rate is the same. We can also see that both performance criteria are unimodal w.r.t. ρ. Thus, the maximizer is unique. The behavior of the

limiting SINR w.r.t. ρ is presented below.

Theorem 3.2. *The limiting SINR is a quasi-concave function of ρ. It is maximized by setting $\rho = \rho^*$ where ρ^* is unique and given by*

$$\rho^* = \frac{\beta}{\gamma}. \tag{3.18}$$

It follows that the maximum limiting SINR is

$$SINR^{*,\infty} = g(\beta, \rho^*). \tag{3.19}$$

Proof. Refer to Appendix 3.6.2. □

The expression of the optimal ρ that maximizes the limiting SINR turns out to be very simple. It is a ratio between the cell-loading and the received SNR. So, at high SNRs, ρ^* becomes very small and the RCI precoder behaves like the MPCI precoder. On the other hand, when the cell loading is very high, ρ^* also becomes large. We can expect that the precoder will perform like the SU precoder. We should also note that ρ^* obtained in our analysis coincides with that derived in [69] which is analyzed for the case $K = N$. Our results are also confirmed by [101] by using a different approach of large system analysis.

Now, let us consider how the cell loading affects the maximum limiting SINR in (3.19). $g(\beta, \rho^*)$ is the solution of

$$g(\beta, \rho^*) = \frac{\frac{1}{\beta}}{\frac{1}{\gamma} + \frac{1}{1+g(\beta,\rho^*)}}.$$

From the above, clearly, increasing the received SNR will improve the maximum limiting SINR and the limiting sum rate. It is easy to check that

$$\frac{\partial g(\beta, \rho^*)}{\partial \beta} = -\frac{1}{\beta^2}\left(\frac{1}{\gamma} + \frac{1}{(1+g(\beta,\rho^*))^2}\right)^{-1} < 0.$$

This shows us that the maximum limiting SINR is decreasing with the cell-loading. Considering the limiting sum rate, increasing β will increase the pre-log factor but decrease the log term. The following proposition states the characterization of the sum rate with respect to β.

Proposition 3.2. *For $\gamma > 1$, the limiting sum rate is a quasi-concave (unimodal) function of β. The unique stationary point, β^*, is given by the solution*

$$\beta^* = \gamma \frac{(1 + g(\beta^*, \beta^*/\gamma))}{(\gamma + (1 + g(\beta^*, \beta^*/\gamma))^2) \log(1 + g(\beta^*, \beta^*/\gamma))}. \tag{3.20}$$

For $\gamma \leq 1$, the sum rate per-antenna is an increasing function of β.

Proof. See Appendix 3.6.3. □

Since R_{sum} is unimodal w.r.t. β, β^* can be found effectively by using the bisection method [7].

So far, we already derived the limiting SINR for the broadcast channel with RCI precoder and characterized its behaviors with respect to involved design parameters such as the regularization parameter and the cell-loading. Now, let us consider the broadcast transmission with MPCI and SU. The following theorem presents the limiting SINR for the case of MPCI.

Theorem 3.3. *The SINR of the broadcast transmission stated in* (3.10) *and* (3.11)*, in the large system limit, converges almost surely respectively to*

$$SINR^{\infty}_{MPCI} = \begin{cases} \gamma\left(\frac{1}{\beta} - 1\right), & \beta < 1 \\ \frac{\gamma(\beta - 1)}{\gamma(\beta - 1)^2 + \beta^2}, & \beta > 1\,. \end{cases} \tag{3.21}$$

Proof. See Appendix 3.6.4. □

From the theorem, it is clear that the limiting SINR is increasing with the received SNR for different cases of β. For $\beta < 1$, it is also obvious that the limiting SINR decreases as β increases. For $\beta > 1$, it can be checked that the limiting SINR is a quasi-concave function of β where the stationary point that maximizes the limiting SINR is given by

$$\beta^* = 1 + \frac{1}{\sqrt{\gamma + 1}}.$$

Considering the limiting sum rate, it can be shown that it is a concave function of β for

$\beta < 1$. The optimizer satisfies,

$$\beta = \frac{\gamma}{\left(1+\gamma\left(\frac{1}{\beta}-1\right)\right)\log\left(1+\gamma\left(\frac{1}{\beta}-1\right)\right)},$$

and can be found by using line search methods.

For the downlink with SU precoder, the limiting SINR is stated below.

Theorem 3.4. *In the large system limit, the SINR of the downlink with SU precoder converges to*

$$SINR_{SU}^{\infty} = \frac{\gamma}{\beta(\gamma+1)}. \tag{3.22}$$

Proof. See Appendix 3.6.5 □

From the above, the limiting SINR is increasing with the SNR but decreasing over β. Considering the sum rate, it is a non-decreasing function of β since

$$\frac{\partial}{\partial\beta}\beta\log(1+\mathrm{SINR}_{\mathrm{SU}}^{\infty}) = \log(1+\mathrm{SINR}_{\mathrm{SU}}^{\infty}) - \frac{\mathrm{SINR}_{\mathrm{SU}}^{\infty}}{1+\mathrm{SINR}_{\mathrm{SU}}^{\infty}} \geq 0.$$

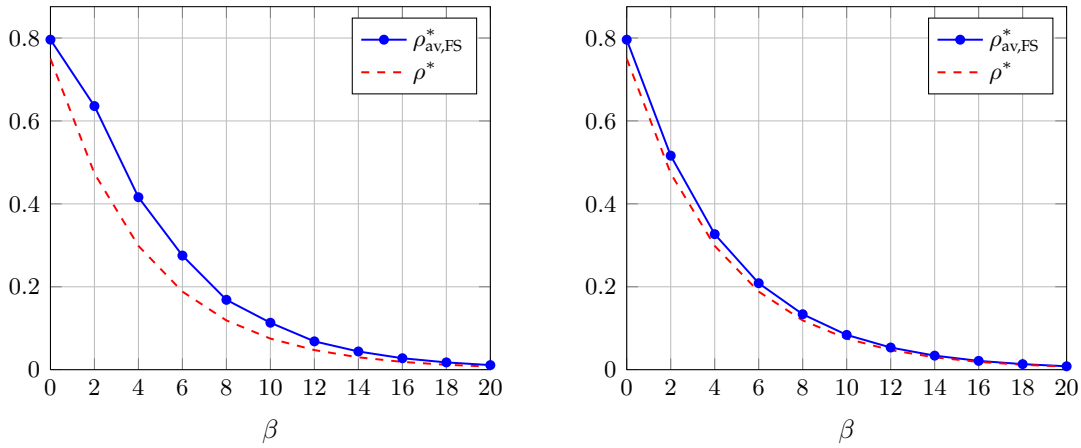
As $\beta \to \infty$, the limiting sum rate achieves the maximum value which is given by

$$\lim_{\beta\to\infty} \log(1+\mathrm{SINR}_{\mathrm{SU}}^{\infty})^{\beta} = \frac{\gamma}{\gamma+1}. \tag{3.23}$$

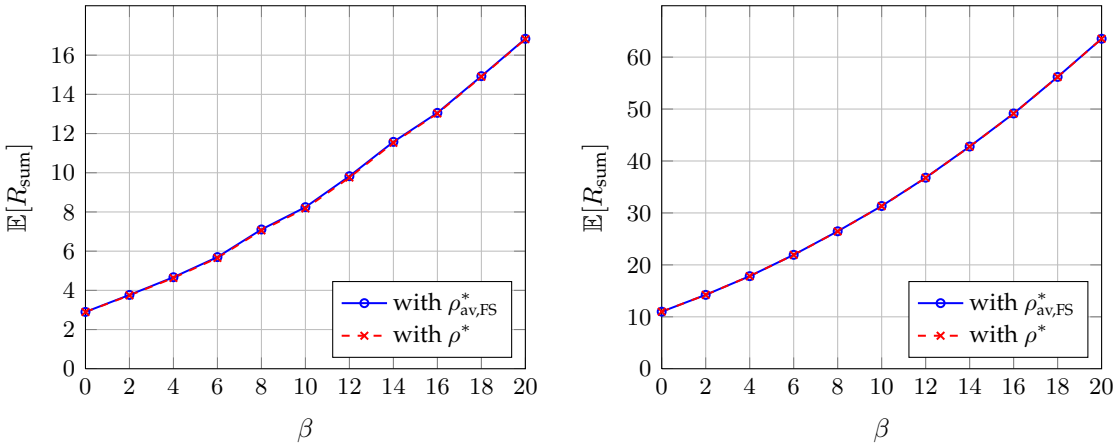
3.4 Performance Comparisons

In the previous we have seen that the limiting SINR accurately approximate the finite-size SINR even for small and moderate N, K. The optimal ρ that maximizes both the limiting SINR and sum rate is given by (3.18). In the finite size system, the SINR or the achievable rate for each user is different. Therefore, the optimal ρ that maximizes the SINR or the rate is also different for each user. In that case, one of the reasonable choices for the objective function is the achievable sum rate.

Figure 3.4(a) illustrate the comparison between the optimal ρ from the large system and finite-size system analysis for different SNR values and N but with a fixed β. In obtaining the optimal ρ for the finite systems, we generate N_s samples of the channel



(a) Optimal ρ vs. SNR with $\beta = 0.75$ and $N_s = 200$.

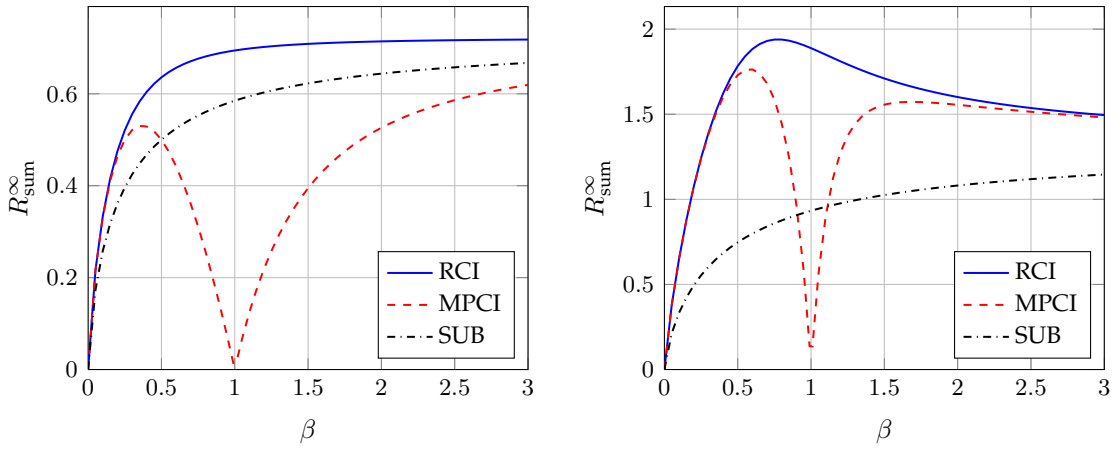


(b) Sum rate vs. SNR with $\beta = 0.75$ and $N_s = 200$.

Figure 3.4: The validity of the large system approximation for finite size systems.

gain matrix. For each channel realization, the optimal ρ that maximizes the sum rate is computed numerically by a grid search. Then the resulting optimal ρs are averaged over N_s, denoted by $\rho^*_{\text{av,FS}}$. We can observe that the optimal ρ from large system analysis follows the one from finite-size system. The gap between them is already small for a moderate system size $N = 16$.

Figure 3.4(b) demonstrates the applicability of the large system results into the design of finite-size systems. Computing the optimal ρ in the finite-size system by a grid search is time consuming. So, we would like to see what is the effect of using the optimal ρ from the large system limit on the finite-size sum rates. The solid and dash line respectively represent the average sum rates obtained by using the optimal ρ from the finite-size and



(a) $\gamma = 0$ dB (b) $\gamma = 10$ dB

Figure 3.5: The limiting downlink sum rate as a function of the cell-loading β for different employed precoders.

large system analysis, respectively. We can see that even for the small-size system $N = 4$, the difference is almost unnoticeable. For larger N, i.e. $N = 16$, the two lines coincide.

In the previous section, we already described analytically the behavior of the limiting SINR and sum rate of the downlink for each precoder. Figure 3.5 compares the limiting sum rate for different employed precoders. For the downlink with the RCI precoder, the sum rate is obtained by setting $\rho = \rho^*$. As predicted by the analysis, i.e. Proposition 3.2, the sum rate for the RCI-based downlink is increasing with β for SNR 0 dB ($\gamma = 1$) and a quasi-concave function of β for SNR 10 dB. We can also see that it beats the sum rate of the BC with other precoders. Figure 3.5(a) shows that the MPCI with $\beta > 1$ can gives higher sum rate compared to the commonly used MPCI precoder, i.e. the one with $\beta < 1$, even though it does not eliminate the MUI completely. It is also shown that setting β close to 1 will result in a poor performance. The downlink sum rates with the RCI and ZF become quite close, except near $\beta = 1$, for a moderate SNR 10 dB. We can expect that they are getting closer as SNR increases with the exception for $\beta \to 1$. We can also expect that for $\beta > 1$, the downlink sum rate with RCI will approach the limit (3.23), which is the maximum downlink sum rate with SU, as $\beta \to \infty$. This is because as $\beta \to \infty$, the optimal regularization parameter of the RCI also goes to ∞. Thus, the RCI will behave

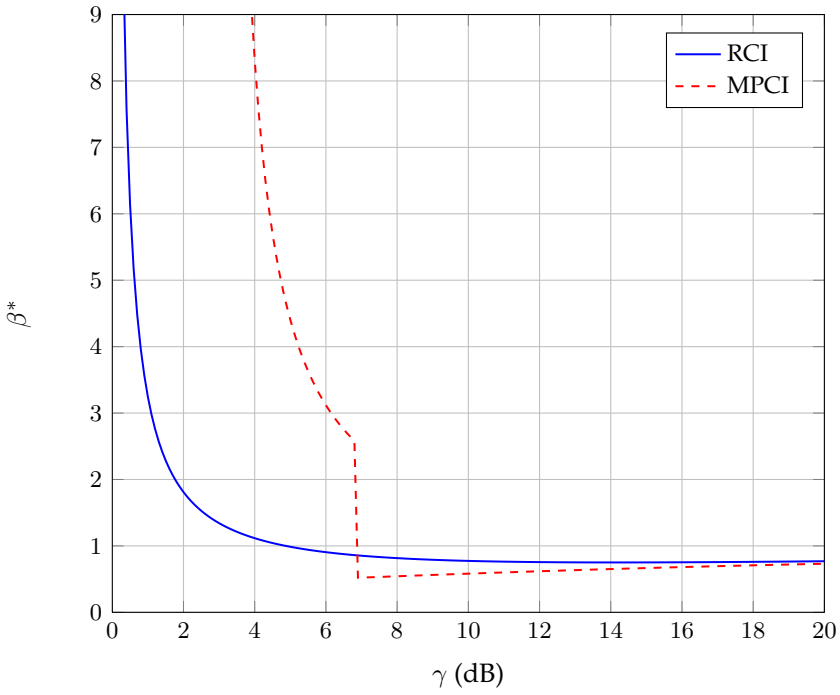


Figure 3.6: Optimal cell-loading β as a function of the SNR.

like the SU precoder. It can also be checked that the downlink sum rate with the MCPI also converges to (3.23) for the asymptotic value of β.

The optimal β as a function of the received SNR is depicted in Figure 3.6 (see also [63]). For the downlink with MPCI precoder, there is a threshold in the received SNR, about 5.85 dB, in which above this threshold, it is better to have less users that the number of antennas ($\beta < 1$) and consequently the ZF precoder is employed. Otherwise, the opposite is better. In other words, it is better to use ZF for moderate or high SNR values. For BCs with RCI precoder, the threshold takes at a lower SNR, about 4.87 dB.

3.5 Conclusion

This chapter presents the large system analysis of the MISO broadcast channels with RCI, MPCI and SU precoders. We have shown that the optimal regularization parameter of the RCI that maximizes the limiting SINR is the ratio between the cell-loading and the received SNR. We also presented the effect of the cell-loading on the limiting sum rate of RCI, MPCI and SU precoders. For the RCI precoder, we have demonstrated that the limiting sum rate is quasi-concave over the cell-loading. Thus, the optimal cell-loading can be found by line search algorithms. For the MPCI, our numerical simulations suggested the optimal cell loading can be larger than one when the SNR is below a certain threshold. For the SU precoder, the maximum limiting sum rate is achieved when the cell-loading tends to infinity.

3.6 Appendix

3.6.1 Proof of Theorem 3.1

We can write the SINR in (3.7) as

$$\text{SINR}_k = \frac{c^2 S_k}{\sigma^2 + c^2 I_k},$$

where $c^2 S_k$ and $c^2 I_k$ represent the signal strength and interference energy of user k respectively. The limiting SINR can be obtained by deriving the large system limit of each term in SINR_k. We start with S_k by first employing Lemma 2.2. However, to apply that lemma, we need to remove the dependency between $\mathbf{H}$ in $\left(\mathbf{H}^H\mathbf{H} + \alpha\mathbf{I}_N\right)^{-1}$ and $\mathbf{h}_k$. Now, let us write

$$\left(\mathbf{H}^H\mathbf{H} + \alpha\mathbf{I}_N\right)^{-1} = \left(\mathbf{H}_k^H\mathbf{H}_k + \alpha\mathbf{I}_N + \mathbf{h}_k^H\mathbf{h}_k\right)^{-1}, \tag{3.24}$$

where $\mathbf{H}_k$ is $\mathbf{H}$ with the k-th row removed. By applying the matrix inversion lemma (Sherman-Morrison Formula), (3.24) becomes

$$\left(\mathbf{H}_k^H\mathbf{H}_k + \alpha\mathbf{I}_N\right)^{-1} - \frac{\left(\mathbf{H}_k^H\mathbf{H}_k + \alpha\mathbf{I}_N\right)^{-1}\mathbf{h}_k^H\mathbf{h}_k\left(\mathbf{H}_k^H\mathbf{H}_k + \alpha\mathbf{I}_N\right)^{-1}}{1 + \mathbf{h}_k\left(\mathbf{H}_k^H\mathbf{H}_k + \alpha\mathbf{I}_N\right)^{-1}\mathbf{h}_k^H}.$$

It follows straightforwardly that

$$\left(\mathbf{H}^H\mathbf{H}+\alpha\mathbf{I}_N\right)^{-1}\mathbf{h}_k^H=\frac{\left(\mathbf{H}_k^H\mathbf{H}_k+\alpha\mathbf{I}_N\right)^{-1}\mathbf{h}_k^H}{1+\mathbf{h}_k\left(\mathbf{H}_k^H\mathbf{H}_k+\alpha\mathbf{I}_N\right)^{-1}\mathbf{h}_k^H}. \tag{3.25}$$

Thus, S_k can be rewritten as

$$S_k=\frac{\left|\mathbf{h}_k\left(\mathbf{H}_k^H\mathbf{H}_k+\alpha\mathbf{I}_N\right)^{-1}\mathbf{h}_k^H\right|^2}{\left(1+\mathbf{h}_k\left(\mathbf{H}_k^H\mathbf{H}_k+\alpha\mathbf{I}_N\right)^{-1}\mathbf{h}_k^H\right)^2}=\frac{|A_k|^2}{(1+A_k)^2},$$

where $A_k=\frac{1}{N}\mathbf{h}_k\left(\frac{1}{N}\mathbf{H}_k^H\mathbf{H}_k+\rho\mathbf{I}_N\right)^{-1}\mathbf{h}_k^H$ and $\rho=\frac{\alpha}{N}$. Note that the spectral norm of $\left(\frac{1}{N}\mathbf{H}_k^H\mathbf{H}_k+\rho\mathbf{I}_N\right)^{-1}$ is bounded above by ρ^{-1}.

Now, we can apply Lemma 2.2 to A_k and it yields

$$A_k-\frac{1}{N}\mathrm{Tr}\left(\left(\mathbf{H}_k^H\mathbf{H}_k+\alpha\mathbf{I}_N\right)^{-1}\right)\xrightarrow{a.s.}0.$$

Using Lemma 2.8, the above becomes

$$A_k-\frac{1}{N}\mathrm{Tr}\left(\left(\mathbf{H}^H\mathbf{H}+\alpha\mathbf{I}_N\right)^{-1}\right)\xrightarrow{a.s.}0.$$

Following the derivation in Lemma 2.4, we can show that

$$A_k-g(\beta,\rho)\xrightarrow{a.s.}0,$$

where $g(\beta,\rho)$ is the function defined in (2.10). Thus, it follows immediately

$$S_k-\frac{g^2}{(1+g)^2}\xrightarrow{a.s.}0, \tag{3.26}$$

where brevity, we use g to denote $g(\beta,\rho)$ and we do the same for the rest of this chapter. Note that the limiting value for S_k is the same for all users.

Now, let us move to the interference term. I_k can be expressed as

$$I_k=\sum_{j\neq k}\mathbf{h}_k\left(\mathbf{H}^H\mathbf{H}+\alpha\mathbf{I}_N\right)^{-1}\mathbf{h}_j^H\mathbf{h}_j\left(\mathbf{H}^H\mathbf{H}+\alpha\mathbf{I}_N\right)^{-1}\mathbf{h}_k^H$$

$$
\begin{aligned}
&= \mathbf{h}_k \left(\mathbf{H}^H\mathbf{H} + \alpha\mathbf{I}_N\right)^{-1} \mathbf{H}_k^H\mathbf{H}_k \left(\mathbf{H}^H\mathbf{H} + \alpha\mathbf{I}_N\right)^{-1} \mathbf{h}_k^H \\
&\overset{(a)}{=} \frac{\mathbf{h}_k \left(\mathbf{H}_k^H\mathbf{H}_k + \alpha\mathbf{I}_N\right)^{-1} \mathbf{H}_k^H\mathbf{H}_k \left(\mathbf{H}_k^H\mathbf{H}_k + \alpha\mathbf{I}_N\right)^{-1} \mathbf{h}_k^H}{\left(1 + \mathbf{h}_k \left(\mathbf{H}_k^H\mathbf{H}_k + \alpha\mathbf{I}_N\right)^{-1} \mathbf{h}_k^H\right)^2} \\
&= \frac{B_k}{(1+A_k)^2},
\end{aligned} \tag{3.27}
$$

where (a) is obtained by using (3.25) and its numerator is denoted by B_k. Let us write $B_k = \mathbf{h}_k\mathbf{B}_k\mathbf{h}_k^H$. We can show that

$$
\begin{aligned}
\mathbf{B}_k &= (\mathbf{H}_k^H\mathbf{H}_k + \rho\mathbf{I}_N)^{-1}\mathbf{H}_k^H\mathbf{H}_k(\mathbf{H}_k^H\mathbf{H}_k + \rho\mathbf{I}_N)^{-1} \\
&= (\mathbf{H}_k^H\mathbf{H}_k + \rho\mathbf{I}_N)^{-1}(\mathbf{H}_k^H\mathbf{H}_k + \rho\mathbf{I}_N - \rho\mathbf{I}_N)(\mathbf{H}_k^H\mathbf{H}_k + \rho\mathbf{I}_N)^{-1} \\
&= (\mathbf{H}_k^H\mathbf{H}_k + \rho\mathbf{I}_N)^{-1}\left[\mathbf{I}_N - \rho(\mathbf{H}_k^H\mathbf{H}_k + \rho\mathbf{I}_N)^{-1}\right] \\
&= (\mathbf{H}_k^H\mathbf{H}_k + \rho\mathbf{I}_N)^{-1} - \rho(\mathbf{H}_k^H\mathbf{H}_k + \rho\mathbf{I}_N)^{-2} \\
&= (\mathbf{H}_k^H\mathbf{H}_k + \rho\mathbf{I}_N)^{-1} + \rho\frac{\partial}{\partial\rho}(\mathbf{H}_k^H\mathbf{H}_k + \rho\mathbf{I}_N)^{-1}.
\end{aligned} \tag{3.28}
$$

Hence, $B_k = A_k + \rho\frac{\partial}{\partial\rho}A_k$ and we have (see also [48, Appendix A]),

$$
B_k - \left(g + \rho\frac{\partial g}{\partial\rho}\right) \xrightarrow{a.s.} 0. \tag{3.29}
$$

Consequently, it follows that

$$
I_k - \frac{g + \rho\frac{\partial g}{\partial\rho}}{(1+g)^2} \xrightarrow{a.s.} 0. \tag{3.30}
$$

To complete the proof, let us now consider the normalizing constant c^2. The denominator of c^2 in (3.5) can be expressed as

$$
\frac{1}{N}\text{Tr}\left(\left(\frac{1}{N}\mathbf{H}^H\mathbf{H} + \rho\mathbf{I}_N\right)^{-2}\frac{1}{N}\mathbf{H}^H\mathbf{H}\right).
$$

Following the same steps as in deriving the large system limit of B_k in (3.27), we obtain

$$
c^2 - \frac{P_d}{\left(g + \rho\frac{\partial g}{\partial\rho}\right)} \xrightarrow{a.s.} 0. \tag{3.31}
$$

Combining the large system results (3.26) and (3.31), the signal energy converges almost surely to a deterministic value given by

$$\frac{g^2}{(1+g)^2}\frac{P_d}{\left(g+\rho\frac{\partial g}{\partial \rho}\right)} = P_d\frac{g}{(1+g)^2}\left(1+\frac{\rho}{\beta}(1+g)^2\right),$$

where we can show that

$$\frac{\partial g}{\partial \rho} = -\frac{g(1+g)^2}{\beta+\rho(1+g)^2}. \tag{3.32}$$

Similarly, combining the results (3.27) and (3.31), the interference energy converges almost surely to a non-random quantity expressed by

$$\frac{P_d}{(1+g)^2}.$$

To this end, we can finally express the limiting SINR as in (3.16) with $\gamma = P_d/\sigma^2$ and this completes the proof.

3.6.2 Proof of Theorem 3.2

First, let us rewrite the limiting SINR as

$$\text{SINR}^{\infty} = \frac{\gamma}{\beta} g \Upsilon,$$

where

$$\Upsilon = \frac{\beta+\rho\left(1+g\right)^2}{\gamma+\left(1+g\right)^2}.$$

The first derivative of SINR^{∞} over ρ is given by

$$\frac{\partial \text{SINR}^{\infty}}{\partial \rho} = \frac{\gamma}{\beta}\left(\frac{\partial g}{\partial \rho}\Upsilon + g\frac{\partial \Upsilon}{\partial \rho}\right), \tag{3.33}$$

where

$$\frac{\partial \Upsilon}{\partial \rho} = \frac{[(1+g)^2+2\rho\frac{\partial g}{\partial \rho}(1+g)][\gamma+(1+g)^2]-[\beta+\rho(1+g)^2][2\frac{\partial g}{\partial \rho}(1+g)]}{[\gamma+(1+g)^2]^2}.$$

Performing further algebra manipulations on (3.33), we have the following steps

$$\begin{aligned}
\frac{\partial \text{SINR}^{\infty}}{\partial \rho} &= \frac{\gamma}{\beta}\left(\frac{\partial g}{\partial \rho}\Upsilon + g\frac{\partial \Upsilon}{\partial \rho}\right) \\
&= \frac{\gamma}{\beta} g\Upsilon\left[\frac{\frac{\partial g}{\partial \rho}}{g} + \frac{\frac{\partial \Upsilon}{\partial \rho}}{\Upsilon}\right] \\
&= \frac{\gamma}{\beta} g\Upsilon\left[\frac{\frac{\partial g}{\partial \rho}}{g} + \frac{[(1+g)^2 + 2\rho\frac{\partial g}{\partial \rho}(1+g)][\gamma + (1+g)^2] - [\beta + \rho(1+g)^2][2\frac{\partial g}{\partial \rho}(1+g)]}{[\gamma + (1+g)^2][\beta + \rho(1+g)^2]}\right] \\
&= \frac{\gamma}{\beta} g\Upsilon\left[\frac{\frac{\partial g}{\partial \rho}}{g} + \frac{(1+g)^2}{\beta + \rho(1+g)^2} + \frac{2\rho\frac{\partial g}{\partial \rho}(1+g)}{\beta + \rho(1+g)^2} + \frac{2\frac{\partial g}{\partial \rho}(1+g)}{\gamma + (1+g)^2}\right] \\
&= \frac{\gamma}{\beta} g\Upsilon\left[\frac{\frac{\partial g}{\partial \rho}}{g} - \frac{\frac{\partial g}{\partial \rho}}{g} + \frac{2\rho\frac{\partial g}{\partial \rho}(1+g)}{\beta + \rho(1+g)^2} + \frac{2\frac{\partial g}{\partial \rho}(1+g)}{\gamma + (1+g)^2}\right] \qquad (3.34) \\
&= \frac{2\gamma^2 g(1+g)^2}{\beta[\gamma + (1+g)^2]^2}\frac{\partial g}{\partial \rho}\left[\rho - \frac{\beta}{\gamma}\right] \\
&= \mathcal{K}\frac{\partial g}{\partial \rho}\left[\rho - \frac{\beta}{\gamma}\right], \qquad (3.35)
\end{aligned}$$

where (3.34) is obtained from (3.32), i.e.,

$$-\frac{\frac{\partial g}{\partial \rho}}{g} = \frac{(1+g)^2}{\beta + \rho(1+g)^2}\,. \qquad (3.36)$$

Since $\mathcal{K} > 0$ and $\frac{\partial g}{\partial \rho} < 0$, then the stationary point is given by

$$\rho^* = \frac{\beta}{\gamma}\,.$$

Moreover, since $\mathcal{K}\frac{\partial g}{\partial \rho} < 0$, then (3.35) is positive for $\rho < \rho^*$ and and negative for $\rho > \rho^*$. Thus, the the limiting SINR is increasing over ρ until reaching $\rho = \rho^*$ and decreasing after that. This concludes that the limiting SINR is a quasi-concave function of ρ ([7, pp. 99]) and ρ_{opt} is the global optimizer.

3.6.3 Proof of Proposition 3.2

The proof is divided into two parts. The first part is finding the stationary point which can be obtained by setting the first derivative of sum rate over β equal to zero. The second

part is to find the characterization of the sum rate w.r.t. β that use a similar approach to that in [109].

With $\rho = \beta/\gamma$, we can link β and g as follows

$$\beta = \frac{1}{g}\left(\frac{1}{\gamma} + \frac{1}{1+g}\right)^{-1}.$$

The first derivative of g w.r.t. β is given by

$$\beta\frac{\partial g}{\partial \beta} = -\frac{g\left(\frac{1}{1+g} + \frac{1}{\gamma}\right)}{\frac{1}{(1+g)^2} + \frac{1}{\gamma}}. \tag{3.37}$$

So, g is a decreasing function of β.

The derivative of the sum rate per antenna over β is given by

$$\frac{\partial R_{\text{sum}}}{\partial \beta} = \log(1+g) + \frac{\beta}{1+g}\frac{\partial g}{\partial \beta}.$$

Continuing further, we have

$$\begin{aligned}\frac{\partial R_{\text{sum}}}{\partial \beta} &= \log(1+g) - \frac{g\left(\frac{1}{1+g} + \frac{1}{\gamma}\right)}{\frac{1}{(1+g)} + \frac{1}{\gamma} + \frac{g}{\gamma}} \\ &\overset{(a)}{=} \log(1+g) - \frac{\frac{1}{\beta}}{\frac{1}{(1+g)} + \frac{1}{\gamma} + \frac{g}{\gamma}},\end{aligned}$$

where in (a), we use the fact that $\beta g = \left(\frac{1}{1+g} + \frac{1}{\gamma}\right)^{-1}$. Setting the derivative equal to zero will lead to a stationary point given by (3.20).

Since we know the behavior of g w.r.t β, hence the characterization of R_{sum} over g will indirectly describe the behavior R_{sum} w.r.t. β. We can rewrite R_{sum} as

$$R_{\text{sum}} = \frac{1}{g}\left(\frac{1}{\gamma} + \frac{1}{1+g}\right)^{-1}\log(1+g),$$

and correspondingly,

$$\frac{\partial R_{\text{sum}}}{\partial g} = -\frac{(1+g)^2+\gamma}{g^2(1+g+\gamma)^2}\log(1+g) + \frac{1}{g(1+g+\gamma)}.$$

The first derivative above has the same sign as

$$q = -\log(1+g) + \frac{g(1+g+\gamma)}{(1+g)^2+\gamma},$$

and $q = 0$ when $g = 0$, and $q \to -\infty$ as $g \to \infty$. It is easy to check that $q = -\frac{\partial R_{\text{sum}}}{\partial \beta}$. Taking the first derivative of q over g will lead to

$$\frac{\partial q}{\partial g} = \frac{-g^4 - 3g^3 - 3g^2 - g(1-\gamma^2)}{\left((1+g)^2+\gamma\right)^2}.$$

This shows that if $\gamma \leq 1$, $\frac{\partial q}{\partial g} < 0$ and correspondingly $\frac{\partial R_{\text{sum}}}{\partial g} < 0$. This implies that R_{sum} is decreasing over g but it is increasing over β. For $\gamma > 1$, $\frac{\partial q}{\partial g}$, hence also q, is positive for small g. So, R_{sum} increases up to $g(\beta^*, \beta^*/\gamma)$ and then decreases. This concludes that R_{sum} is a unimodal (quasi-concave) function over g or β for $\gamma > 1$.

3.6.4 Proof of Theorem 3.3

First, let us consider the case $\beta < 1$. From (3.10), the limiting SINR can be obtained by deriving the large system limit for c_1^2. Let $F_{\frac{1}{N}\mathbf{H}\mathbf{H}^H}$ be the empirical eigenvalue distribution of $\frac{1}{N}\mathbf{H}\mathbf{H}^H$. The denominator of c_1^2 can be expressed as

$$\frac{1}{N}\text{Tr}\left(\left(\frac{1}{N}\mathbf{H}\mathbf{H}^H\right)^{-1}\right) = \int \frac{1}{\lambda}\, F_{\frac{1}{N}\mathbf{H}\mathbf{H}^H}(\lambda).$$

Suppose that $F_{\frac{1}{N}\mathbf{H}\mathbf{H}^H}$ converges almost surely to the Marčenko Pastur limiting distribution F. Thus, the right hand side of the above equation converges to

$$\int \frac{1}{\lambda}\, F(\lambda) = \lim_{z \to 0} m_F(z),$$

where $m_F(z)$ is the Stieltjes transform of F. From Theorem 2.3, it follows that

$$m_F(z) = -\left(z - \frac{1}{\beta(1+m_F(z))}\right)^{-1}.$$

Thus, $\int \lambda^{-1}\, dF = \frac{\beta}{1-\beta}$. Hence, we can conclude

$$c^2 - P_d\left(\frac{1}{\beta} - 1\right)$$

and (3.21) for $\beta < 1$ follows immediately.

For the case $\beta > 1$, we can see that the SINR expression (3.11) is (3.7) with $\rho \to 0$. Therefore, the analyses in Appendix 3.6.1 hold directly. Then, the last equation of (3.21) can be obtained from (3.16) by letting $\rho \to 0$ with $g = \frac{1}{\beta-1}$ in this case.

3.6.5 Proof of Theorem 3.4

We can express (3.12) as follows

$$\text{SINR}_{\text{SU},k} = \frac{\frac{P_d}{\frac{1}{N}\text{Tr}\left(\frac{1}{N}\mathbf{H}^H\mathbf{H}\right)}\left|\frac{1}{N}\mathbf{h}_k\mathbf{h}_k^H\right|^2}{\sigma^2 + \frac{P_d}{\frac{1}{N}\text{Tr}\left(\frac{1}{N}\mathbf{H}^H\mathbf{H}\right)}\frac{1}{N}\mathbf{h}_k\frac{1}{N}\mathbf{H}_k^H\mathbf{H}_k\mathbf{h}_k}.$$

In proving the theorem, we only need to find the large system limit for $\frac{1}{N}\mathbf{h}_k\mathbf{h}_k^H$ and $\frac{1}{N}\text{Tr}\left(\frac{1}{N}\mathbf{H}^H\mathbf{H}\right)$.

Since $\mathbf{I}_N$ has a bounded spectral norm, then the following holds

$$\frac{1}{N}\mathbf{h}_k\mathbf{I}_N\mathbf{h}_K - \frac{1}{N}\text{Tr}(\mathbf{I}_N) \xrightarrow{a.s.} 0.$$

Moreover, $\frac{1}{N}\text{Tr}(\mathbf{I}_N) = 1$, thus $\frac{1}{N}\mathbf{h}_k\mathbf{h}_K - 1 \xrightarrow{a.s.} 0$.

For the trace term,

$$\begin{aligned}\frac{1}{N}\text{Tr}\left(\frac{1}{N}\mathbf{H}^H\mathbf{H}\right) &= \int \lambda\, dF_{\frac{1}{N}\mathbf{H}^H\mathbf{H}}(\lambda) \\ &\to \int \lambda\, dF(\lambda) \end{aligned} \tag{3.38}$$

which is the first moment of the Marčenko Pastur distribution F. We can evaluate (3.38) in several ways. First, we can express it in terms of the Stieltjes transform of F as

$$\int \lambda\, dF(\lambda) = \lim_{z\to 0}\frac{\partial}{\partial z}\frac{m_F(z^{-1})}{z},$$

where $m_F(z)$ is the solution of (see Theorem 2.3)

$$m_F(z) = -\frac{1}{z - \frac{\beta}{1+m_F(z)}}.$$

It is easy to show that

$$\begin{aligned}\frac{\partial}{\partial z}\frac{m_F(z^{-1})}{z} &= \beta\frac{\frac{\partial m_F(z^{-1})}{\partial z}}{(1+m_F(z^{-1}))^2} \\ &= \beta\frac{z^{-1}m_F(z^{-1})}{(1+m_F(z^{-1}))^2 - \beta z},\end{aligned}$$

where the last equation is obtained by using the fact that

$$\frac{\partial}{\partial z}m_F(z^{-1}) = \frac{z^{-1}m_F(z^{-1})}{1 - \frac{\beta z}{(1+m_F(z^{-1}))^2}}.$$

It can be checked that the following equations are true,

$$\lim_{z\to 0} m_F(z^{-1}) = 0, \quad \text{and} \quad \lim_{z\to 0} z^{-1}m_F(z^{-1}) = 0.$$

Thus,

$$\lim_{z\to 0}\frac{\partial}{\partial z}\frac{m_F(z^{-1})}{z} = \beta,$$

and consequently,

$$\frac{1}{N}\text{Tr}\left(\frac{1}{N}\mathbf{H}\mathbf{H}^H\right) - \beta \xrightarrow{a.s.} 0.$$

Alternatively, we can derive (3.38) by using the Marčenko-Partur density described in Theorem 2.1,

$$\begin{aligned}\int \lambda\, dF(\lambda) &= \int_a^b \lambda\,(1-\beta)^+\,\delta(\lambda)\,d\lambda + \int_a^b \frac{1}{2\pi}\sqrt{(\lambda-a)(b-\lambda)}\,d\lambda \\ &= \int_a^b \frac{1}{2\pi}\sqrt{(\lambda-a)(b-\lambda)}\,d\lambda, \\ &= \beta,\end{aligned}$$

where $a = (1-\sqrt{\beta})^2$, $b = (1+\sqrt{\beta})^2$. The last line is obtained by using [28, cf. 2.262(1)].

See also [95, cf. 2.102, p. 54] regarding the expression for the k-th moment of Marčenko-Pastur distribution.

Similarly, we can show

$$\frac{1}{N}\mathbf{h}_k\frac{1}{N}\mathbf{H}_k^H\mathbf{H}_k\mathbf{h}_k - \beta \xrightarrow{a.s.} 0.$$

Combining the results, (3.22) follows directly.

Chapter 4

Downlink Beamforming with Transmit-side Channel Correlation

This chapter focuses on the design of the RCI precoder for MISO broadcast channels with transmit-side channel correlation. We also investigate the effects of the correlation on the performance of the systems employing MPCI and SU precoders. The separable channel correlation model is adopted. The limiting SINR for each precoder is derived. It is a function of the eigenvalues of the correlation matrix. For the RCI, we derive the optimal regularization parameter of RCI that maximizes the (limiting) SINR. We show that it is not affected by the transmit correlation. This is a surprising result since the SINR itself is influenced by the correlation. We confirm this result through simulations where the channel has the exponential transmit-correlation model.

4.1 Introduction

THE system model in Chapter 3 assumes i.i.d. entries of the channel gain matrix. In practice, spatial fading correlation is present due to scattering in the propagation environment and spacing between antenna elements (see e.g., [6, 46, 67]. The spatial correlation between antenna elements (of the transmitters or receivers) will increase when the spacing between antennas or the number of scatterers around the antennas decreases [6, 46]. It has been shown in [12] that a separation of ten wavelengths between the adjacent antenna elements is enough for a decorrelation and to achieve a full-capacity in BLAST (Bell-labs LAyered SpaceTime) MIMO systems [22]. Reducing the separation to four wavelengths will cause the capacity to decrease about 20% [12]. Considering the scattering environment, there is a rich scattering around both transmitters and receivers in i.i.d. fading channel case. If a base station is located in an unobstructed environment

or with a relatively small number of local scatterers and the receivers/users are in a rich scattering environment, we may only observe the correlation at the transmitter side, also called the *transmit correlation*. Related to the distribution of the entries of the channel gain matrix, the impact and the capacity of spatially correlated point-to-point MIMO systems have been investigated in [10, 65, 81] for Rayleigh-fading channels and in [15, 55, 56] for Ricean channels.

There are several correlation models considered in the literature. The most well known for its analytical tractability is the separable correlation model, which is also known as the Kronecker model [13, 70, 82]. In this model, the transmit and the receive correlations are separable. The capacity scaling of point to point MIMO systems with N antennas at both transmitter and receiver ends under this model was investigated in [13]. The authors observed that as $N \to \infty$, the capacity still scales linearly with N for both with a perfect CSIT and no CSIT cases. However, the growth rate is affected by the correlation in both cases. It is about 10-20% smaller compared to the i.i.d. fading channel case. A similar problem is also considered in [57]. It studied the cases when the spatial correlation is present at either the transmitter or receiver side. By performing the large system analysis, the authors derived closed-form expressions for the capacity in each case. They showed that the correlation always degrades the capacity and this effect is more pronounced to the side with less antennas. When the correlation is present at both ends, the closed form approximations of the mutual information and outage capacity were derived in [52]. The approximations are obtained by first deriving the limiting eigenvalue distribution of the channel gain matrix. The authors proved that it follows the normal distribution.

Reference [94] proposed a more general channel correlation model in the framework of Unitary-Independent-Unitary (UIU) model. The unitary parts are represented by deterministic unitary matrices. The independent part consists of a matrix whose entries are independent, zero mean and can have arbitrary distributions and variances. The Kronecker model is a special case of the UIU framework where the unitary parts are the eigenvector matrices of the transmit and receive correlation matrices of the Kronecker model [94]. The unitary parts can also be considered to be equivalent to the Fourier

matrices in the virtual representation, another well-known scheme to model the spatial correlation, introduced in [79]. In [94], the impact of the correlation based on the UIU model on the capacity of single-user MIMO systems was studied. Depending on the capacity achieving input, the correlation can be detrimental or beneficial to the capacity. If the input is isotropic, i.e., its covariance is an identity matrix, then the correlation is detrimental for any SNR. On the other hand, if the input is non-isotropic, the correlation is beneficial only when the SNR is below a certain threshold. We should also note that besides the correlation, the keyhole or pinhole effects can cause the the channel gain matrices to be rank deficient and deteriorate the capacity of the multi-antenna systems [11, 24, 81].

The impact of the fading correlation on MIMO broadcast transmissions has been investigated in, for example, [47, 49]. The former presents the effects of the phase difference and the magnitude of the correlation coefficients on the sum rate of two-user MISO broadcast channels. The latter shows the beneficial impact of correlation on the capacity of multiuser MIMO with the maximum ratio combining (MRC) scheme. For a large number of users in the systems, the sum-rate scaling laws of the DPC and random beamforming schemes have been investigated in [1].

In this chapter, we study the impact of the transmit correlation on the performance of some well-known linear precoders, i.e., RCI, MPCI and SU, in MISO broadcast transmissions. For tractability in the analysis, we assume that the channel correlation is separable. Our contributions in this chapter can be summarized as follows. First, we derive the limiting SINR for each precoder. The eigenvalues of the correlation matrix determine how much the correlation affects the (limiting) SINR. For the RCI precoder, we derive the optimal regularization parameter that maximizes the limiting SINR. We demonstrate that it is not affected by the correlation. It is the ratio between the cell-loading and the downlink SNR, as expressed in (3.18). This is a surprising result because the correlation influences the corresponding limiting SINR. As far as our knowledge, we are the first to show and prove this result. In the end of the chapter, we provide some numerical simulations that illustrate the system performance when the transmit correlation follows the exponential correlation model.

At the completion of this work, we noticed a similar work in [101]. The analysis in [101] is based on the deterministic equivalent of large random matrices which is analogous to the Stieltjes transform-based approach which is adopted in our analysis. The effect of the transmit correlation on the optimal regularization parameter has not been originally discussed in [101] but has been included later. As previously mentioned, we are the first to show that the optimal regularization parameter is independent of the transmit correlation.

The rest of this chapter is organized as follows. The system model is described in the next section. Section 4.3 presents the limiting SINR expressions for the RCI, MPCI and SU precoders. The optimal regularization parameter of the RCI precoder maxmizing the limiting SINR is also presented. The section also discusses the behaviors of the limiting SINR w.r.t. the cell-loading. In Section 4.4, we verify our analysis by numerical simulations that are based on the exponential correlation model. The last section concludes the chapter.

4.2 System Model

We follow the same system model as in the i.i.d. channel case. As we consider the correlation amongst transmit antenna elements and assume a separable correlation model, the propagation channel matrix can be written as

$$\mathbf{H} = \widetilde{\mathbf{H}}\mathbf{R}_t^{1/2},$$

where the elements of $\widetilde{\mathbf{H}}$ are i.i.d. circularly-symmetric complex Gaussian (CSCG) random variables with zero mean and unit variance. The matrix $\mathbf{R}_t$ is the transmit correlation matrix and determines the correlation between the columns of $\mathbf{H}$ [46]. We assume that $\mathbf{R}_t$ is Hermitian and positive definite. Thus, the inverse of $\mathbf{R}_t$ exists and is also Hermitian. We also assume that each user experiences the same transmit correlation. We consider that $\mathbf{H}$ is known perfectly by the BS to construct the precoder. The corresponding channel vector for each user k is $\mathbf{h}_k = \widetilde{\mathbf{h}}_k\mathbf{R}_t^{1/2}$, where $\widetilde{\mathbf{h}}_k$ is the kth row of $\widetilde{\mathbf{H}}$. The user k is assumed to know this vector perfectly in order to decode the data sent by the BS.

The expression for each precoder follows the descriptions in Chapter 3. However, now each precoder is a function of the transmit correlation matrix. This implies that the presence of the correlation will change the beamforming direction for each user. Thus, we can strongly hypothesize that the correlation matrix will affect the performance of each precoder. For the RCI precoder, the resulting SINR for user k can be written as previously stated in (3.7),

$$\text{SINR}_k = \frac{c^2 \left| \mathbf{h}_k \left(\mathbf{H}^H \mathbf{H} + \alpha \mathbf{I}_N \right)^{-1} \mathbf{h}_k^H \right|^2}{\sigma^2 + c^2 \sum_{j \neq k} \left| \mathbf{h}_k \left(\mathbf{H}^H \mathbf{H} + \alpha \mathbf{I}_N \right)^{-1} \mathbf{h}_j^H \right|^2}, \tag{4.1}$$

where α as in the previous chapters denotes the regularization parameter and c^2 is defined as in (3.5). The numerator and denominator of (4.1) represents the signal strength and interference plus noise energy respectively. Similarly, the SINR for each MPCI and SU follow can be obtained by allowing $\alpha \to 0$ and $\alpha \to \infty$, respectively.

4.3 Large System Analysis

We can see immediately from (4.1) that SINR_k is a random quantity and depends on particular realizations of the propagation channel matrix. In this section, we derive the limiting SINR for each precoder.

First, let us consider the large system analysis for the RCI precoder.

Theorem 4.1. *Let $\rho = \frac{\alpha}{N}$ be the normalized regularization parameter and $\gamma = \frac{P_d}{\sigma^2}$ be the received SNR. In the large system limit, the SINR (4.1) converges almost surely to a deterministic quantity given by*

$$SINR^{\infty} = \xi \frac{\gamma E_{22} + \rho \frac{\gamma}{\beta} (1+\xi)^2 E_{12}}{\gamma E_{22} + (1+\xi)^2 E_{12}}, \tag{4.2}$$

where ξ is the positive solution of

$$\xi = \int \left(\frac{\rho}{t} + \frac{\beta}{1+\xi} \right)^{-1} d\Lambda(t) \tag{4.3}$$

$$= (1+\xi) E_{11}. \tag{4.4}$$

E_{ij} is given by

$$E_{ij} = \mathbb{E}\left[\frac{T^i}{(\rho(1+\xi)+\beta T)^j}\right] . \tag{4.5}$$

where the expectation is taken over the random variable T whose distribution function Λ is the limiting eigenvalue distribution of the correlation matrix $\mathbf{R}_t$.

Moreover, in the large system regime, the signal strength and the interference energy can be written as

$$-\frac{P_d\xi^2}{\beta\frac{\partial\xi}{\partial\rho}} = P_d\xi\left(\frac{\rho}{\beta}+\frac{E_{22}}{E_{12}(1+\xi)^2}\right) \tag{4.6}$$

and

$$-\frac{P_dE_{22}}{\frac{\partial\xi}{\partial\rho}(1-\beta E_{22})} = \frac{P_dE_{22}}{E_{12}(1+\xi)^2} , \tag{4.7}$$

respectively.

Proof. Refer to Appendix 4.6.1 □

Equation (4.2) shows that the limiting SINR is the same for all users. This is partly due to our assumption that all users experience the same transmit correlation and have the same distance dependent path-loss gain. (4.2) also has the same structure as the limiting SINR in i.i.d. channel case and in the presence of channel uncertainty (see (3.16) and (5.10) respectively). Comparing them, we can think of $\gamma\frac{E_{22}}{E_{12}}$ as the *effective SNR*. We can also check that $\xi = g(\beta,\rho)$ and $\frac{E_{22}}{E_{12}} = 1$ when $\mathbf{R}_t = \mathbf{I}_N$. Thus, in that case, (4.2) reduces to (3.16).

Now, let us consider the behavior of the signal and interference energy w.r.t. ρ.

Proposition 4.1. *Both the signal strength and interference energy are increasing in ρ.*

Proof. See Appendix 4.6.3. □

The proposition implies that ρ provides a trade-off between increasing signal strength and reducing the interference energy, as also observed in i.i.d. channel case. Thus, the optimal ρ should be determined to maximize the system performance and it is stated in the following.

Theorem 4.2. *The optimal regularization parameter, denoted by ρ^*, that maximizes the limiting SINR is given by*

$$\rho^* = \frac{\beta}{\gamma}. \tag{4.8}$$

The maximum limiting SINR is then given by

$$SINR^{\infty,*} = \xi^*, \tag{4.9}$$

where ξ^* is ξ evaluated at $\rho = \rho^*$.

Proof. Refer to Appendix 4.6.2. We should note that the proof is the adaptation of the proof for (3.18) to the current case. □

From the above, it is surprising and interesting that ρ^* is independent of the eigenvalue distribution of the correlation matrix. In other words, it is not affected by the correlation. On the contrary, the maximum limiting SINR (4.9) is influenced by the correlation.

Now, let us move to investigate the effect of the cell-loading on the system performance. Let

$$\breve{E}_{ij} = \int \frac{t}{\frac{1}{\gamma}(1+\xi^*)+t}\, d\Lambda(t). \tag{4.10}$$

Then, we can show that

$$\frac{\partial \xi^*}{\partial \beta} = -\frac{\xi^*}{\beta - \breve{E}_{22}} \tag{4.11}$$

$$= -\frac{(\xi^*)^2}{\frac{1}{\gamma}(1+\xi^*)\breve{E}_{12} + \breve{E}_{22}} \tag{4.12}$$

$$< 0.$$

So, $\text{SINR}^{\infty,*}$ is decreasing in β. Considering the limiting sum rate, defined in (3.17), the pre-log term is increasing in β. Therefore, for β, there is a trade-off between increasing the pre-log and decreasing the SINR.

Now let us consider the the large system analysis for MPCI and SU. The SINR for the MPCI takes the form as in (3.10) and (3.11) for $\beta < 1$ and $\beta > 1$, respectively. Its value in the large system limit is stated in the following theorem.

Theorem 4.3. *In the large system limit, the $SINR_k$ of MPCI converges to a non-random quantity given by*

$$SINR_{MPCI}^{\infty} = \begin{cases} \dfrac{\gamma}{\chi} & \beta < 1 \\ \dfrac{\gamma}{\gamma(\beta-1)+\psi\beta^2} & \beta > 1 \end{cases} \tag{4.13}$$

where χ is the positive solution of

$$\chi = \beta \left(\int \frac{t}{1+\chi \cdot t} \, d\Lambda(t) \right)^{-1}$$

and ψ is given by

$$\psi = \frac{1}{\beta - 1} \int \frac{d\Lambda(t)}{t}.$$

Proof. See Appendix 4.6.4 □

We can see clearly that the limiting SINR is increasing in γ. For $\beta < 1$, since $\partial\chi/\partial\beta = \left(\int \frac{t}{(1+\chi \cdot t)^2} \right)^{-1} > 0$, increasing β will cause the limiting SINR decreases. For $\beta > 1$, we can show that the limiting SINR is unimodal in β. It reaches the maximum point at

$$\beta = 1 + \frac{1}{\sqrt{\frac{\gamma}{\mathbb{E}[\mathsf{T}^{-1}]}} + 1}.$$

Note that in the uncorrelated channel case, $\mathbf{R}_t = \mathbf{I}_N$, $\mathbb{E}[\mathsf{T}^{-1}] = 1$. The characteristics of the limiting sum rates of the RCI and MPCI precoders w.r.t. β for a particular transmit correlation model are presented in the following section.

The SINR for user k of the SU precoder is given by (3.12). The following theorem states the limiting SINR for SU.

Theorem 4.4. *In the large system limit, the $SINR_k$ of SU converges to*

$$SINR_{SU}^{\infty} = \frac{\gamma(\mathbb{E}[\boldsymbol{T}])^2}{\beta\left(\gamma\mathbb{E}[\boldsymbol{T}^2] + \mathbb{E}[\boldsymbol{T}]\right)}. \tag{4.14}$$

Proof. See Appendix 4.6.5. □

From (4.14), one can observe that the limiting SINR of SU depends on the first and second moment of a random variable whose distribution function Λ. When $\mathbf{R}_t = \mathbf{I}_N$,

those moments are equal to one and (4.14) reduces to (3.22). It is obvious that the limiting SINR is decreasing in β. However, as in the i.i.d. channel case, we can show that the limiting sum rate is an increasing function of β since

$$\frac{\partial R_{\text{sum}}^{\infty}}{\partial \beta} = \log(1 + \text{SINR}_{\text{SU}}^{\infty}) - \frac{\text{SINR}_{\text{SU}}^{\infty}}{1 + \text{SINR}_{\text{SU}}^{\infty}} > 0,$$

for positive limiting SINR. The maximum limiting sum rate is then given by

$$\lim_{\beta \to \infty} \log \left(1 + \frac{\gamma(\mathbb{E}[\mathsf{T}])^2}{\beta \left(\gamma \mathbb{E}[\mathsf{T}^2] + \mathbb{E}[\mathsf{T}] \right)} \right)^{\beta} = \frac{\gamma(\mathbb{E}[\mathsf{T}])^2}{\left(\gamma \mathbb{E}[\mathsf{T}^2] + \mathbb{E}[\mathsf{T}] \right)}.$$

4.4 Examples and Numerical Simulations

In order to evaluate the limiting SINR, we obviously need the expression for the limiting eigenvalue distribution of the transmit correlation matrix. Here, we consider correlation matrices with a Toeplitz structure. Let $\mathbf{T}_N$ be an $N \times N$ Toeplitz matrix and t_{ij} be the element of $\mathbf{T}_N$ at row i and column j such that $t_{ij} = t_{i-j}$ [29]. We can illustrate the Toeplitz matrix as follows [29],

$$\mathbf{T}_N = \begin{bmatrix} t_0 & t_{-1} & t_{-2} & \cdots & t_{-(N-1)} \\ t_1 & t_0 & t_{-1} & \cdots & t_{-(N-2)} \\ t_2 & t_1 & t_0 & \cdots & t_{-(N-3)} \\ \vdots & \cdots & & \ddots & \\ t_{N-1} & & & \cdots & t_0 \end{bmatrix}.$$

Let $\{\lambda_k : k = 1, 2, \ldots, N\}$ be the eigenvalues of $\mathbf{T}_N$. Then, the following holds [29]

$$\lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} F(\lambda_k) \to \frac{1}{2\pi} \int_0^{2\pi} F(f(\omega))\, d\omega \tag{4.15}$$

for any function F that is continuous in the support of $f(\omega)$. Here, $f(\omega)$ denotes the spectral densities or Fouries series of $\{t_k\}$ and is given by

$$f(\omega) = \sum_{k=-\infty}^{\infty} t_k e^{ik\omega}, \qquad \omega \in [0, 2\pi]. \tag{4.16}$$

We should note that $\mathbf{R}_t$ needs to be Hermitian. Therefore, the Hermitian Toeplitz structure is of particular interest. From (4.15), we can express E_{ij} in (4.5) as follows

$$E_{ij} = \frac{1}{2\pi} \int_0^{2\pi} \frac{f^i(\omega)}{(\rho(1+\xi) + \beta f(\omega))^j} \, d\omega \, ,$$

and consequently, we can compute (4.2).

Numerical simulations in this section use a simple and well-known correlation model i.e., the exponential correlation model. The correlation matrix $\mathbf{R}_t$ has the ith row and jth column given by

$$r_{ij} = \nu^{|i-j|}, \tag{4.17}$$

where ν is the correlation coefficient. One can see that $\mathbf{R}_t$ is Hermitian Toeplitz. Using this model, $f(\omega)$ is given by

$$f(\omega) = \lim_{N\to\infty} \sum_{k=-N+1}^{N-1} \nu^{|k|} e^{ik\omega} = \frac{1-\nu^2}{1 - 2\nu\cos(\omega) + \nu^2}.$$

For the RCI, we can evaluate ξ as follows

$$\begin{aligned}
\xi &= \frac{(1+\xi)}{2\pi} \int_0^{2\pi} \frac{f(\omega)}{\rho(1+\xi) + \beta f(\omega)} \, d\omega \\
&= \frac{1}{2\pi} \int_0^{2\pi} \frac{1-\nu^2}{\rho(1+\nu^2) + \frac{\beta(1-\nu^2)}{1+\xi} - 2\rho\nu\cos(\omega)} \, d\omega \\
&= \frac{1-\nu^2}{\sqrt{\left(\rho(1+\nu)^2 + \frac{\beta(1-\nu^2)}{1+\xi}\right)\left(\rho(1-\nu)^2 + \frac{\beta(1-\nu^2)}{1+\xi}\right)}} \qquad (4.18) \\
&= \frac{1}{\sqrt{\left(\rho\frac{1+\nu}{1-\nu} + \frac{\beta}{1+\xi}\right)\left(\rho\frac{1-\nu}{1+\nu} + \frac{\beta}{1+\xi}\right)}}, \qquad (4.19)
\end{aligned}$$

where (4.18) is obtained by using [28, cf. eq. (3.661(4))]. One can see that (4.19) is a quartic equation w.r.t. ξ. By using Descartes' rule of signs, one can also check that it only has one positive solution. For E_{12} and E_{22}, we can express them as follows [28, cf. eq. (2.554(1))

and (2.554(3))]

$$E_{12} = \frac{1-\nu^2}{(1+\xi)^2}\frac{(1+\nu^2)a+2\nu b}{a^2-b^2}\left[\frac{1}{2\pi}\int_0^{2\pi}\frac{1}{a+b\cos(\omega)}\,d\omega\right] \tag{4.20}$$

$$E_{22} = \frac{(1-\nu^2)^2}{(1+\xi)^2}\frac{a}{a^2-b^2}\left[\frac{1}{2\pi}\int_0^{2\pi}\frac{1}{a+b\cos(\omega)}\,d\omega\right], \tag{4.21}$$

where $a = \rho(1+\nu^2) + \frac{\beta(1-\nu^2)}{1+\xi}$ and $b = -2\rho\nu$.
Substituting (4.20) and (4.21) into (4.2) gives

$$\text{SINR}^{\infty} = \xi\,\frac{\gamma\Theta_1 + \frac{\rho\gamma}{\beta}(1+\xi)^2\Theta_2}{\gamma\Theta_1 + (1+\xi)^2\Theta_2}, \tag{4.22}$$

where $\Theta_1 = \rho(1+\xi)(1+\nu^2) + \beta(1-\nu^2)$ and $\Theta_2 = [\rho(1+\xi)(1-\nu^2) + \beta(1+\nu^2)]$.

For the MPCI, we can show that the limiting SINR under the considered correlation model can be expressed as

$$\text{SINR}^{\infty}_{\text{MPCI}} = \begin{cases} \dfrac{\gamma(1-\beta^2)}{\beta^2(\mu+\sqrt{\mu^2-1+1/\beta^2})} & \beta < 1 \\ \dfrac{\gamma(\beta-1)}{\gamma(\beta-1)^2+\mu\beta^2} & \beta > 1, \end{cases} \tag{4.23}$$

where $\mu = \frac{1+\nu^2}{1-\nu^2}$. For the i.i.d. channel case $\nu = 0$ and hence $\mu = 1$. Similarly, for the SU, we can write the limiting SINR as

$$\text{SINR}^{\infty}_{\text{SU}} = \frac{\gamma}{\beta\,(\mu\gamma+1)}. \tag{4.24}$$

The results presented so far are illustrated in the figures below.

Figure 4.1 shows how the random SINR in (4.1) approaches the limiting SINR (4.2) as system dimensions (K, N) increase. For each value of γ, 500 realizations of $\widetilde{\mathbf{H}}$ are generated and the corresponding SINRs are computed. The correlation matrix has elements that follow (4.17) with $\nu = 0.4$. We also set $\rho = \beta/\gamma$. From the figure, we can see that the spread of the SINR for finite K and N is getting smaller as K and N increase. Considering the average SINRs, represented by the dash line, we can see that they are very close to the asymptotic limit even for small and moderate system sizes, e.g., $N = 8$ and $N = 16$, respectively. These observations are the same as attained in the previous chapter.

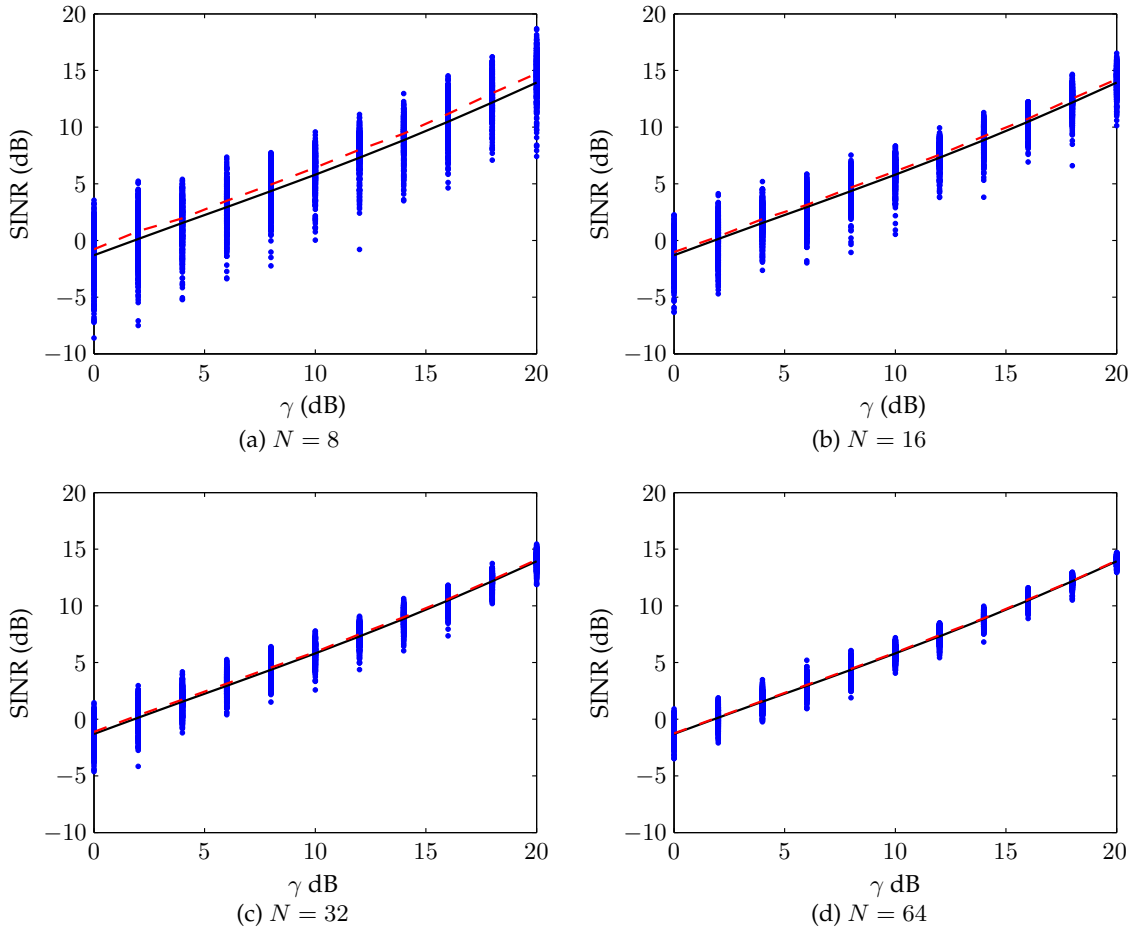


Figure 4.1: Comparison of the randomly generated SINR for user 1 (dot) with the average SINR (dash) and the asymptotic limit (4.2) (solid line) with $\beta = 0.75$, $\nu = 0.5$, $\rho = \beta/\gamma$.

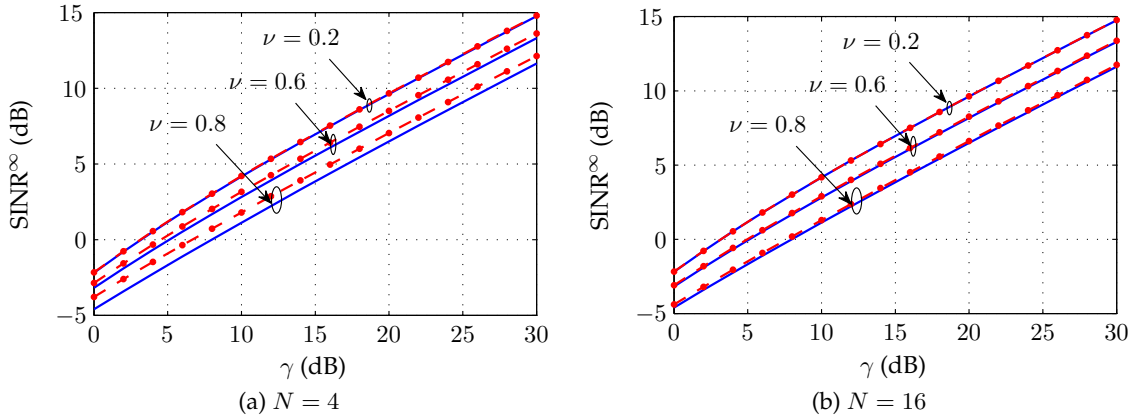
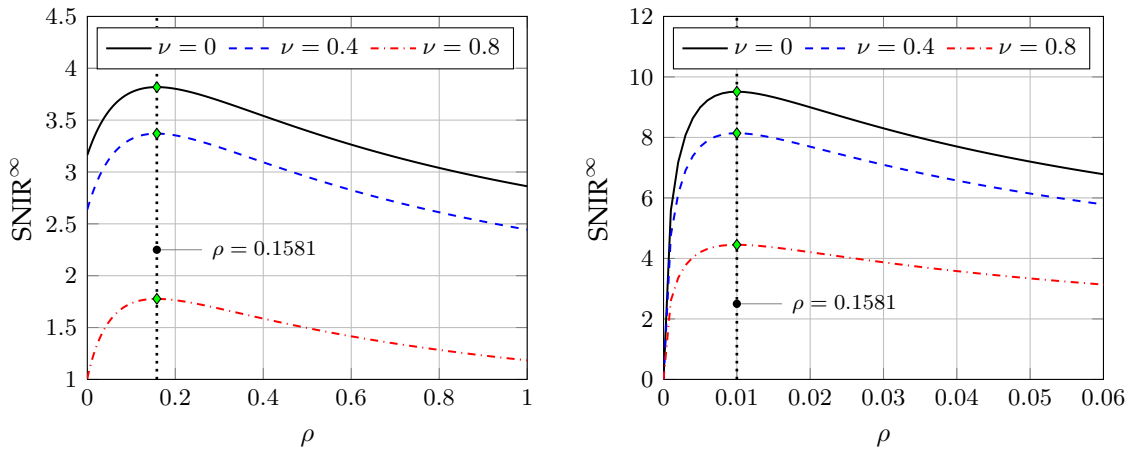


Figure 4.2: Comparison of the limiting SINR by using Λ (solid) and the approximation (4.25) (dash-dots) for different values of γ and ν with $\beta = 1$.



(a) $\beta = 0.5, \gamma = 5$ dB (b) $\beta = 1, \gamma = 20$ dB

Figure 4.3: Limiting SINR as a function ρ for different values of ν.

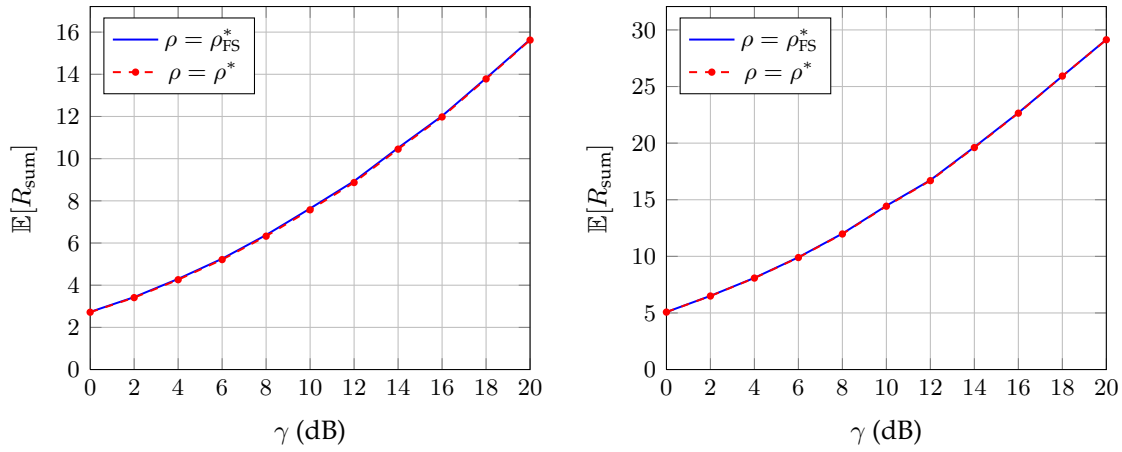
In practice, the l.s.d. of $\mathbf{R}_t$ is generally hard to obtain (as mentioned previously) or $\mathbf{R}_t$ does not have particular properties or structure such that E_{ij} can be evaluated. In that case, we can do an approximation for $\mathbb{E}[G(\mathsf{T})]$ as follows

$$\int G(t)\, d\Lambda(t) \approx \frac{1}{N} \sum_k G(\lambda_k), \tag{4.25}$$

where G is the function of the random variable T whose distribution Λ. Figure 4.2 compares the limiting SINR obtained between using the knowledge of Λ and the approximation (4.25). It demonstrates that for small system sizes, e.g., $N = 4$, the approximation is very good, particularly for small νs, such as for $\nu = 0.2$ and 0.4. For a strong channel correlation, e.g., $\nu = 0.8$, one can still notice a gap between the curves. For a moderate system size $N = 16$, we can observe that (4.25) gives an accurate approximation, even for a big ν.

The limiting SINR as a function of regularization parameter ρ is illustrated in Figure 4.3. We can observe that for different values of correlation coefficient ν, the limiting SINR achieves its maximum by choosing $\rho = \beta/\gamma$. This is consistent with the theoretical analysis: the correlation does not alter the optimal ρ.

Now, let us observe the applicability of the large system result for ρ to the finite size



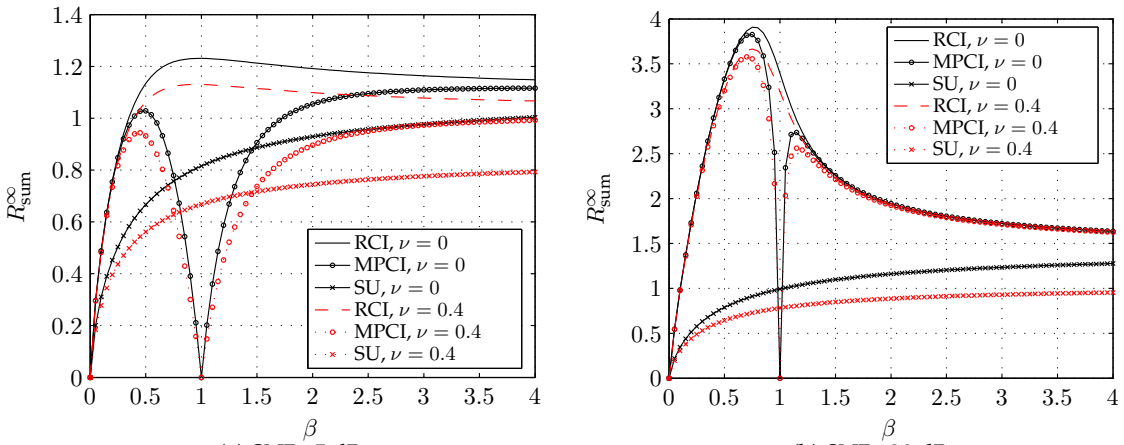
(a) $N = 4$ (b) $N = 8$

Figure 4.4: Comparison of the average sum rate between using $\rho = \rho^*_{\text{FS}}$ and $\rho = \rho^*$ with $\beta = 0.75$, $\nu = 0.5$.

system. We generate 500 channel realizations. For each channel realization, the optimal regularization parameter that maximizes the sum rate, denoted by ρ^*_{FS}, is evaluated numerically by a grid search. Similarly, the sum rates by using ρ^* are also computed. Figure 4.4 compares the average sum rates using ρ^* and ρ^*_{FS}. We can observe that even for a small $N = 4$, the gap between the curves is very small. For $N = 8$, the gap becomes unseen. This verifies the validity of using the large system analysis for finite size system designs.

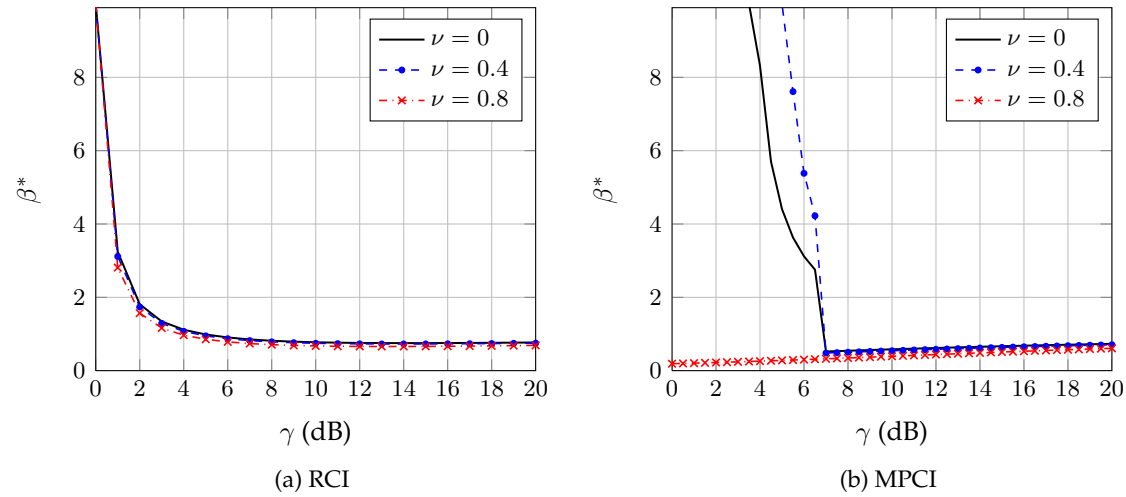
Figure 4.5 presents the limiting sum rates per antenna of the RCI, MPCI and SU precoders for different values of β and ν. The plots for the RCI precoder use $\rho = \rho^*$. For $\gamma = 5$ dB, we can see the effect of the transmit correlation on the limiting sum rate. This holds for all precoders. As we increase γ to 20 dB, the effect becomes unnoticeable for RCI and MPCI precoders except around the maximum limiting sum rates. For the SU precoder, we can still observe the effect even though it is smaller compared to the case of $\gamma = 5$ dB.

The plots of the optimal β that maximizes R^{∞}_{sum} for different values of γ and ν are shown in Figure 4.6. For the RCI precoder, a moderate correlation, e.g., $\nu = 0.4$, relatively does not affect the choice of β^* and a high correlation coefficient such as $\nu = 0.8$ reduces β^* slightly. A similar behavior is also observed for the MPCI when $\beta^* < 1$ (or after γ



(a) SNR=5 dB (b) SNR=20 dB

Figure 4.5: Limiting sum rate per-antenna as a function of cell loading.



(a) RCI (b) MPCI

Figure 4.6: Optimal β maximizing R_{sum}^{∞} for different values of γ and ν.

passes a certain threshold). Otherwise, one can see that the correlation affects the choice of β^* (see the case $\nu = 0$ and $\nu = 0.4$). Interestingly, when we have a strong correlation (e.g., $\nu = 0.8$), the optimal β is less than one even for low SNR values.

4.5 Conclusion

In this chapter, we derive the large system limit of the SINR of RCI, MPCI and SU precoders in the presence of correlation at the transmitter side. Not only does the limiting SINR depend on the cell-loading, SNR and regularization parameter (for the RCI precoder) as shown in Chapter 3, but it also depends on the function of the limiting eigenvalue distribution of the correlation matrix. We also prove a surprising result for the RCI precoder: the optimal regularization parameter maximizing the limiting SINR (or sum rate) is not affected by the transmit correlation. Thus, as in i.i.d. channel case, it is the ratio between the cell-loading and received SNR. By using the exponential correlation model as an example, we present some numerical simulations showing the impact of the correlation on the performance of the system in terms of the SINR and the sum rate per antenna.

4.6 Appendix

4.6.1 Proof of Theorem 4.1

Let $A_k = \mathbf{h}_k \left(\mathbf{H}_k^H \mathbf{H}_k + \alpha \mathbf{I}_N\right)^{-1} \mathbf{h}_k^H$. Then, the signal strength can be expressed as $c^2 S_k$ where $S_k = \frac{|A_k|^2}{(1+A_k)^2}$. Let $\widehat{\mathbf{H}} = \frac{1}{\sqrt{N}}\widetilde{\mathbf{H}}$ and $\widehat{\mathbf{h}}_k = \frac{1}{\sqrt{N}}\widetilde{\mathbf{h}}_k$. We can rewrite A_k as

$$A_k = \widehat{\mathbf{h}}_k \left(\widehat{\mathbf{H}}_k^H \widehat{\mathbf{H}}_k + \rho \mathbf{R}_t^{-1}\right)^{-1} \widehat{\mathbf{h}}_k^H.$$

Applying Lemma 2.2 to A_k yields

$$A_k - \frac{1}{N}\mathrm{Tr}\left(\widehat{\mathbf{H}}_k^H \widehat{\mathbf{H}}_k + \rho \mathbf{R}_t^{-1}\right)^{-1} \xrightarrow{a.s.} 0,$$

as $N \to \infty$. Note that the second term on the LHS is equal to

$$\frac{1}{N}\mathrm{Tr}\left(\mathbf{R}_t \left(\frac{1}{N}\mathbf{H}_k^H \mathbf{H}_k + \rho \mathbf{I}_N\right)^{-1}\right).$$

Thus, by using Lemma 2.8 (R1PL), we have

$$A_k - \frac{1}{N}\text{Tr}\left(\widehat{\mathbf{H}}^H\widehat{\mathbf{H}} + \rho\mathbf{R}_t^{-1}\right)^{-1} \xrightarrow{a.s.} 0.$$

The trace term can be evaluated as follows

$$\frac{1}{N}\text{Tr}\left(\widehat{\mathbf{H}}^H\widehat{\mathbf{H}} + \rho\mathbf{R}_t^{-1}\right)^{-1} = \int \frac{dF_{\mathbf{X}}^N(\lambda)}{\lambda}, \tag{4.26}$$

where we denote $\mathbf{X} = \widehat{\mathbf{H}}^H\widehat{\mathbf{H}} + \rho\mathbf{R}_t^{-1}$ and $F_{\mathbf{X}}^N$ is the e.s.d. of $\mathbf{X}$. Let F be the l.s.d. of $\mathbf{X}$ and suppose that $F_{\mathbf{X}}^N \xrightarrow{a.s.} F$. Then, (4.26) converges almost surely to

$$\lim_{z\to 0}\int \frac{dF(\lambda)}{\lambda - z},$$

which is the Stieltjes transform of F evaluated at 0. Let $F_{\mathbf{R}_t}^N$ be the e.s.d. of $\mathbf{R}_t$ and $F_{\mathbf{R}_t}^N \xrightarrow{a.s.} \Lambda$. Let $\xi = \lim_{z\to 0} m_F(z)$. Based on Theorem 2.2, it follows that ξ is the positive solution of

$$\xi = \int \left(\frac{\rho}{t} + \frac{\beta}{1+\xi}\right)^{-1} d\Lambda(t)\,. \tag{4.27}$$

Therefore, it follows that $A_k - \xi \xrightarrow{a.s.} 0$ and

$$S_k - \frac{\xi^2}{(1+\xi)^2} \xrightarrow{a.s.} 0.$$

The interference energy can be written as $c^2 I_k$, where I_k is defined and can be rewritten as follows

$$\begin{aligned}
I_k &= \sum_{j\neq k} |\widehat{\mathbf{h}}_k \left(\widehat{\mathbf{H}}^H\widehat{\mathbf{H}} + \rho\mathbf{R}_t^{-1}\right)^{-1} \widehat{\mathbf{h}}_j^H|^2 \\
&= \sum_{j\neq k} \widehat{\mathbf{h}}_k \left(\widehat{\mathbf{H}}^H\widehat{\mathbf{H}} + \rho\mathbf{R}_t^{-1}\right)^{-1} \widehat{\mathbf{h}}_j^H\widehat{\mathbf{h}}_j \left(\widehat{\mathbf{H}}^H\widehat{\mathbf{H}} + \rho\mathbf{R}_t^{-1}\right)^{-1} \widehat{\mathbf{h}}_k^H \\
&= \frac{1}{(1+A_k)^2}\sum_{j\neq k} \widehat{\mathbf{h}}_k \left(\widehat{\mathbf{H}}_k^H\widehat{\mathbf{H}}_k + \rho\mathbf{R}_t^{-1}\right)^{-1} \widehat{\mathbf{h}}_j^H\widehat{\mathbf{h}}_j \left(\widehat{\mathbf{H}}_k^H\widehat{\mathbf{H}}_k + \rho\mathbf{R}_t^{-1}\right)^{-1} \widehat{\mathbf{h}}_k^H \\
&= \frac{1}{(1+A_k)^2}\sum_{j\neq k} I_{k,j}.
\end{aligned}$$

Applying Lemma 2.5 for $I_{k,j}$ yields

$$\max_{j\leq K,k}\left|I_{k,j}-\frac{1}{N}\mathrm{Tr}\left(\left(\widehat{\mathbf{H}}_k^H\widehat{\mathbf{H}}_k+\rho\mathbf{R}_t^{-1}\right)^{-1}\widehat{\mathbf{h}}_j^H\widehat{\mathbf{h}}_j\left(\widehat{\mathbf{H}}_k^H\widehat{\mathbf{H}}_k+\rho\mathbf{R}_t^{-1}\right)^{-1}\right)\right|\xrightarrow{a.s.}0.$$

Let $\mathbf{D}_{k,j}$ be the matrix inside the trace. It is a rank-1 matrix. Thus, it can be rewritten as

$$\begin{aligned}\mathbf{D}_{k,j}&=\widehat{\mathbf{h}}_j\left(\widehat{\mathbf{H}}_k^H\widehat{\mathbf{H}}_k+\rho\mathbf{R}_t^{-1}\right)^{-2}\widehat{\mathbf{h}}_j^H\\&=\frac{\widehat{\mathbf{h}}_j\left(\widehat{\mathbf{H}}_{kj}^H\widehat{\mathbf{H}}_{kj}+\rho\mathbf{R}_t^{-1}\right)^{-2}\widehat{\mathbf{h}}_j^H}{\left(1+\widehat{\mathbf{h}}_j\left(\widehat{\mathbf{H}}_{kj}^H\widehat{\mathbf{H}}_{kj}+\rho\mathbf{R}_t^{-1}\right)^{-1}\widehat{\mathbf{h}}_j^H\right)^2},\end{aligned}$$

where $\widehat{\mathbf{H}}_{kj}$ is $\widehat{\mathbf{H}}_k$ with jth row removed. By using the same steps as in analyzing the asymptotic limit for A_k, one can show that for the term in the denominator

$$\max_{j\leq K,k}\left|\widehat{\mathbf{h}}_j\left(\widehat{\mathbf{H}}_{kj}^H\widehat{\mathbf{H}}_{kj}+\rho\mathbf{R}_t^{-1}\right)^{-1}\widehat{\mathbf{h}}_j^H-\xi\right|\xrightarrow{a.s.}0.$$

For the numerator of $\mathbf{D}_{k,j}$, after applying the rank-1 perturbation lemma twice, we have

$$\max_{j\leq K,k}\left|\widehat{\mathbf{h}}_j\left(\widehat{\mathbf{H}}_{kj}^H\widehat{\mathbf{H}}_{kj}+\rho\mathbf{R}_t^{-1}\right)^{-2}\widehat{\mathbf{h}}_j^H-\frac{1}{N}\mathrm{Tr}\left(\left(\widehat{\mathbf{H}}^H\widehat{\mathbf{H}}+\rho\mathbf{R}_t^{-1}\right)^{-2}\right)\right|\xrightarrow{a.s.}0.$$

One can check that

$$\begin{aligned}\frac{1}{N}\mathrm{Tr}\left(\left(\widehat{\mathbf{H}}^H\widehat{\mathbf{H}}+\rho\mathbf{R}_t^{-1}\right)^{-2}\right)&=\int\frac{1}{\lambda^2}\,dF_{\mathbf{X}}^N(\lambda)\\&\xrightarrow{a.s.}\lim_{z\to0}\frac{\partial}{\partial z}\int\frac{1}{\lambda-z}\,dF_(\lambda)\\&=\lim_{z\to0}\frac{\partial}{\partial z}m_F(z).\end{aligned}$$

Thus, we can write

$$\max_{j\leq K,k}\left|I_{k,j}-\frac{1}{N}\frac{\lim_{z\to0}\frac{\partial}{\partial z}m_F(z)}{(1+\xi)^2}\right|\xrightarrow{a.s.}0.$$

Collecting the large system results for the terms of I_k, we have

$$I_k-\beta\frac{\lim_{z\to0}\frac{\partial}{\partial z}m_F(z)}{(1+\xi)^4}\xrightarrow{a.s.}0. \tag{4.28}$$

Let $\vartheta = \lim_{z\to 0} \frac{\partial}{\partial z} m_F(z)$. It is easy to show that

$$\vartheta = \frac{(1+\xi)^2 E_{22}}{1-\beta E_{22}}, \tag{4.29}$$

where E_{ij} is already defined in (4.5).

Now, let us analyze the asymptotic limit for c^2. Its denominator (see (3.5)) can be written as

$$\begin{aligned}
&\frac{1}{N}\mathrm{Tr}\left(\left(\frac{1}{N}\mathbf{H}^H\mathbf{H}+\rho\mathbf{I}_N\right)^{-1}\left(\frac{1}{N}\mathbf{H}^H\mathbf{H}\right)\left(\frac{1}{N}\mathbf{H}^H\mathbf{H}+\rho\mathbf{I}_N\right)^{-1}\right)\\
&\quad=\frac{1}{N}\mathrm{Tr}\left(\widehat{\mathbf{H}}\left(\widehat{\mathbf{H}}^H\widehat{\mathbf{H}}+\rho\mathbf{R}_t^{-1}\right)^{-1}\mathbf{R}_t^{-1}\left(\widehat{\mathbf{H}}^H\widehat{\mathbf{H}}+\rho\mathbf{R}_t^{-1}\right)^{-1}\right)\\
&\quad=\frac{1}{N}\sum_{j=1}^{k}\widehat{\mathbf{h}}_j\mathbf{X}^{-1}\mathbf{R}_t^{-1}\mathbf{X}^{-1}\widehat{\mathbf{h}}_j^H\\
&\quad=\frac{1}{N}\sum_{j=1}^{k}\frac{\widehat{\mathbf{h}}_j\mathbf{X}_j^{-1}\mathbf{R}_t^{-1}\mathbf{X}_j^{-1}\widehat{\mathbf{h}}_j^H}{(1+A_j)^2},
\end{aligned}$$

where $\mathbf{X}_j = \left(\widehat{\mathbf{H}}_j^H\widehat{\mathbf{H}}_j+\rho\mathbf{R}_t^{-1}\right)^{-1}$. Using the previous result for A_k, it follows that

$$\max_{j\le K}|A_j-\xi| \xrightarrow{a.s.} 0.$$

By Lemma 2.5 and Lemma 2.8, $\max_{j\le K}\left|\widehat{\mathbf{h}}_j\mathbf{X}_j^{-1}\mathbf{R}_t^{-1}\mathbf{X}_j^{-1}\widehat{\mathbf{h}}_j^H-\frac{1}{N}\mathrm{Tr}\left(\mathbf{X}^{-2}\mathbf{R}_t^{-1}\right)\right| \xrightarrow{a.s.} 0$. It can be shown that

$$\frac{1}{N}\mathrm{Tr}\left(\mathbf{X}^{-2}\mathbf{R}_t^{-1}\right) = -\frac{\partial}{\partial\rho}\frac{1}{N}\mathrm{Tr}\left(\mathbf{X}^{-1}\right).$$

Since $\frac{1}{N}\mathrm{Tr}\left(\mathbf{X}^{-1}\right)\xrightarrow{a.s.}\xi$, then

$$\max_{j\le K}\left|\widehat{\mathbf{h}}_j\mathbf{X}_j^{-1}\mathbf{R}_t^{-1}\mathbf{X}_j^{-1}\widehat{\mathbf{h}}_j^H+\frac{\partial\xi}{\partial\rho}\right| \xrightarrow{a.s.} 0.$$

Thus, it follows that c^2 converges almost surely to

$$\frac{P_d(1+\xi)^2}{-\beta\frac{\partial\xi}{\partial\rho}}. \tag{4.30}$$

It is also easy to check that

$$\frac{\partial \xi}{\partial \rho} = -\frac{(1+\xi)^2 E_{12}}{1-\beta E_{22}} \,. \tag{4.31}$$

Combining the large system results above, (4.6) and (4.7) follow immediately. Moreover, the limiting SINR for user k can be written as

$$\text{SINR}^{\infty} = \frac{\gamma}{\beta} \frac{\xi^2 (1+\xi)^2}{\gamma\vartheta - (1+\xi)^2 \frac{\partial \xi}{\partial \rho}}, \tag{4.32}$$

where $\gamma = P_d/\sigma^2$ is the received SNR.

Substituting (4.29) and (4.31) into (4.32), we get

$$\begin{aligned} \text{SINR}^{\infty} &= \frac{\gamma}{\beta} \frac{\xi^2 (1-\beta E_{22})}{\gamma E_{22} + (1+\xi)^2 E_{12}} \\ &= \frac{\gamma}{\beta} \xi \frac{\xi - \beta\xi E_{22}}{\gamma E_{22} + (1+\xi)^2 E_{12}} \,. \end{aligned} \tag{4.33}$$

Now, let us consider ξ in (4.27). We can rewrite it as follows

$$\begin{aligned} \xi &= \int \frac{(1+\xi)t}{\rho(1+\xi)+\beta t} \, d\Lambda(t) \\ &= \int \frac{\rho(1+\xi)+\beta t}{\rho(1+\xi)+\beta t} \frac{(1+\xi)t}{\rho(1+\xi)+\beta t} \, d\Lambda(t) \\ &= \int \frac{\rho(1+\xi)^2 + \beta t^2 + \beta\xi t^2}{\left(\rho(1+\xi)+\beta t\right)^2} \, d\Lambda(t) \\ &= \rho(1+\xi)^2 E_{12} + \beta E_{22} + \beta\xi E_{22}. \end{aligned} \tag{4.34}$$

Thus, $\xi - \beta\xi E_{22} = \rho(1+\xi)^2 E_{12} + \beta E_{22}$ and (4.2) follows easily. This concludes the proof.

4.6.2 Proof of Theorem 4.2

We can rewrite the limiting SINR as, $\text{SINR}^{\infty} = \frac{\gamma}{\beta}\xi\Upsilon$, where

$$\Upsilon = \frac{\beta E_{22} + \rho(1+\xi)^2 E_{12}}{\gamma E_{22} + (1+\xi)^2 E_{12}}.$$

Let $x' = \partial x/\partial\rho$. Taking the derivative of SINR^{∞} over ρ yields

$$\frac{\partial\,\text{SINR}^{\infty}}{\partial\rho} = \frac{\gamma}{\beta}(\xi'\Upsilon + \xi\Upsilon'), \tag{4.35}$$

where Υ' is given by

$$\begin{aligned}\Upsilon' = &\frac{\left[(1+\xi)^2 + 2\rho(1+\xi)\xi'\right] E_{12} + \rho(1+\xi)^2 E'_{12} + \beta E'_{22}}{\gamma E_{22} + (1+\xi)^2 E_{12}} \\ &- \left[\beta E_{22} + \rho(1+\xi)^2 E_{12}\right] \times \\ &\frac{2(1+\xi)\xi' E_{12} + (1+\xi)^2 E'_{12} + \gamma E'_{22}}{\left(\gamma E_{22} + (1+\xi)^2 E_{12}\right)^2}.\end{aligned} \tag{4.36}$$

Putting (4.36) back into (4.35) results

$$\begin{aligned}\frac{\partial\,\text{SINR}^{\infty}}{\partial\rho} &= \frac{\gamma}{\beta}(\xi'\Upsilon + \xi\Upsilon') = \frac{\gamma}{\beta}\xi\Upsilon\left[\frac{\xi'}{\xi} + \frac{\Upsilon'}{\Upsilon}\right] \\ &= \frac{\gamma}{\beta}\xi\Upsilon\left[\frac{\xi'}{\xi} + \frac{(1+\xi)^2 E_{12}}{\beta E_{22} + \rho(1+\xi)^2 E_{12}}\right. \\ &\quad + \frac{2\rho(1+\xi)\xi' + \rho(1+\xi)^2 E'_{12} + \beta E'_{22}}{\beta E_{22} + \rho(1+\xi)^2 E_{12}} \\ &\quad \left. - \frac{2(1+\xi)\xi' E_{12} + (1+\xi)^2 E'_{12} + \gamma E'_{22}}{\gamma E_{22} + (1+\xi)^2 E_{12}}\right].\end{aligned} \tag{4.37}$$

By using (4.31) and (4.34), it is easy to show that

$$-\frac{\xi'}{\xi} = \frac{(1+\xi)^2 E_{12}}{\beta E_{22} + \rho(1+\xi)^2 E_{12}}, \tag{4.38}$$

which is the second term in the bracket of (4.37). Furthermore, by defining $\psi := 2(1+\xi)\xi' E_{12} + (1+\xi)^2 E'_{12}$, $\chi := (1+\xi)^2 E_{12}$ and $Z := (\beta E_{22} + \rho\chi)(\gamma E_{22} + \chi)$, we can rewrite (4.37) as

$$\begin{aligned}\frac{\partial \text{SINR}^{\infty}}{\partial \rho} &= \frac{\gamma}{\beta}\xi\Upsilon\left[\frac{\beta E'_{22}+\rho\psi}{\beta E_{22}+\rho\chi}-\frac{\gamma E'_{22}+\psi}{\gamma E_{22}+\chi}\right] \\ &= \frac{\gamma\xi\Upsilon}{\beta Z}\left[\gamma\rho E_{22}\psi+\beta E'_{22}\chi-\beta E_{22}\psi-\gamma\rho E'_{22}\chi\right] \\ &= \frac{\gamma^2\xi\Upsilon}{\beta Z}\left[E_{22}\psi\left(\rho-\frac{\beta}{\gamma}\right)+E'_{22}\chi\left(\frac{\beta}{\gamma}-\rho\right)\right] \\ &= \frac{\gamma^2\xi\Upsilon}{\beta Z}\left[\left(\rho-\frac{\beta}{\gamma}\right)\left(E_{22}\psi-E'_{22}\chi\right)\right]. \end{aligned} \tag{4.39}$$

It can be verified that

$$E_{22}\psi - E'_{22}\chi = 2(1+\xi)\xi' E_{22}E_{12} + (1+\xi)^2(E'_{12}E_{22} - E'_{22}E_{12}), \tag{4.40}$$

where $E'_{12} = -2(1+\xi+\rho\xi')E_{13}$ and $E'_{22} = -2(1+\xi+\rho\xi')E_{23}$. We can rewrite E_{12} as

$$\begin{aligned}E_{12} &= \mathbb{E}\left[\frac{\rho(1+\xi)+\beta\mathsf{T}}{\rho(1+\xi)+\beta\mathsf{T}}\frac{\mathsf{T}}{(\rho(1+\xi)+\beta\mathsf{T})^2}\right] \\ &= \rho(1+\xi)E_{13}+\beta E_{23}.\end{aligned}$$

Similarly, we can express

$$E_{22} = \rho(1+\xi)E_{23}+\beta E_{33}.$$

Putting these results together, we can establish

$$E'_{12}E_{22} - E'_{22}E_{12} = -2\beta(1+\xi+\rho\xi')(E_{13}E_{33}-E_{23}^2).$$

Since the eigen-values of $\mathbf{R}_t$ are positive, then by using the Cauchy-Schwarz inequality we have

$$\begin{aligned}E_{23}^2 &= \mathbb{E}\left[\frac{\mathsf{T}^{\frac{1}{2}}}{(\rho(1+\xi)+\beta\mathsf{T})^{\frac{3}{2}}}\frac{\mathsf{T}^{\frac{3}{2}}}{(\rho(1+\xi)+\beta\mathsf{T})^{\frac{3}{2}}}\right]^2 \\ &\leq E_{13}E_{33}.\end{aligned}$$

Hence, $E_{13}E_{33} - E_{23}^2 \geq 0$. Moreover, $E'_{12}E_{22} - E'_{22}E_{12} \leq 0$ since

$$\begin{aligned} 1 + \xi + \rho\xi' &= 1 + \xi - \rho\frac{(1+\xi)^2 E_{12}}{1 - \beta E_{22}} \\ &\overset{(a)}{=} \frac{(1+\xi)[\rho(1+\xi)^2 + \beta E_{22}] - \xi\rho(1+\xi)^2}{\xi(1-\beta E_{22})} \\ &\overset{(b)}{=} \frac{1}{1-\beta E_{22}} > 0, \end{aligned}$$

where (a) and (b) are obtained by incorporating (4.34) into the derivation. Finally, we have

$$E_{22}\psi - E'_{22}\chi < 0,$$

since $\xi' < 0$. So, we can rewrite (4.39) as

$$\frac{\partial \, \text{SINR}^{\infty}}{\partial \rho} = \mathcal{K}\left(\rho - \frac{\beta}{\gamma}\right). \tag{4.41}$$

Since $\mathcal{K} \neq 0$ for $\rho > 0$, the only stationary point, denoted by ρ°, is given by $\rho^{\circ} = \beta/\gamma$. Thus, $\rho^* = \rho^{\circ}$ is the optimal regularizing parameter that maximizes the limiting SNIR for any correlation matrices satisfying conditions explained in Section 4.2. We can also check that $\mathcal{K} < 0$, for all $\rho > 0$. So, from (4.41), SINR_{∞} is monotonically increasing for $\rho < \rho^*$ and decreasing for $\rho > \rho^*$. This concludes that the limiting SINR is a quasi-concave function of ρ. Thus, ρ^* is the global optimizer. By using ρ^*, we have $\frac{\gamma}{\beta}\Upsilon = 1$. As a result, ξ is the maximum limiting SINR. This completes the proof.

4.6.3 Proof of Proposition 4.3

Let us write the asymptotic signal strength (4.6) and the interference energy (4.7) as $P_d S^{\infty}$ and $P_d I^{\infty}$, respectively. To prove the proposition, we just need to show that $\frac{\partial S^{\infty}}{\partial \rho} > 0$ and $\frac{\partial I^{\infty}}{\partial \rho} > 0$. Evaluating the latter results

$$\frac{\partial I^{\infty}}{\partial \rho} = \frac{(1+\xi)^2(E'_{22}E_{12} - E_{22}E'_{12}) - 2(1+\xi)\xi' E_{12}E_{22}}{E_{12}^2(1+\xi)^4},$$

where $x' = \partial x/\partial\rho$. In Appendix 4.6.2, it is shown that $E'_{22}E_{12} - E_{22}E'_{12} \geq 0$. Moreover, from (4.31), $\xi' < 0$. Thus, $\frac{\partial I^{\infty}}{\partial \rho} > 0$ that implies I^{∞} is increasing in ρ.

For the signal strength, we can write $S^{\infty} = \frac{\rho}{\beta}\xi + \xi I^{\infty}$. Therefore,

$$\begin{aligned}\frac{\partial S^{\infty}}{\partial \rho} &= \frac{\rho}{\beta}\xi' + \frac{\xi}{\beta} + \xi' \frac{E_{22}}{E_{12}(1+\xi)^2} + \xi \frac{\partial I^{\infty}}{\partial \rho} \\ &\overset{(a)}{=} -\frac{\xi(1+\xi)^2 E_{12}}{\rho(1+\xi)^2 E_{12} + \beta E_{22}} \left(\frac{\rho}{\beta} + \frac{E_{22}}{E_{12}(1+\xi)^2} \right) + \frac{\xi}{\beta} + \xi \frac{\partial I^{\infty}}{\partial \rho} \\ &\overset{(b)}{=} -\frac{\xi}{\rho(1+\xi)^2 E_{12} + \beta E_{22}} \left(\frac{\rho(1+\xi)^2 E_{12}}{\beta} + E_{22} \right) + \frac{\xi}{\beta} + \xi \frac{\partial I^{\infty}}{\partial \rho} \\ &= \xi \frac{\partial I^{\infty}}{\partial \rho},\end{aligned}$$

where in (a), we use (4.38) for ξ'. The sum of the first and second terms of (b) is zero. Thus, the last equation follows. Since $\frac{\partial I^{\infty}}{\partial \rho} > 0$, then $\frac{\partial S^{\infty}}{\partial \rho} > 0$. This completes the proof.

4.6.4 Proof of Theorem 4.3

For $\beta < 1$, the SINR for user k can be expressed as

$$\begin{aligned}\text{SINR}_k &= \frac{\gamma}{\text{Tr}\left(\left(\mathbf{H}\mathbf{H}^H \right)^{-1} \right)} \\ &= \frac{\gamma}{\frac{1}{K}\text{Tr}\left(\left(\frac{1}{K}\widetilde{\mathbf{H}}\mathbf{R}\widetilde{\mathbf{H}}^H \right)^{-1} \right)}.\end{aligned}$$

Let $\mathbf{W} = \frac{1}{K}\widetilde{\mathbf{H}}\mathbf{R}\widetilde{\mathbf{H}}^H$. Then,

$$\frac{1}{K}\text{Tr}\left(\left(\frac{1}{K}\widetilde{\mathbf{H}}\mathbf{R}\widetilde{\mathbf{H}}^H \right)^{-1} \right) = \int \frac{1}{\lambda}\, dG^K_{\mathbf{W}}(\lambda), \tag{4.42}$$

where $G^K_{\mathbf{W}}$ is the e.s.d. of $\mathbf{W}$. Suppose that $G^K_{\mathbf{W}} \xrightarrow{a.s.} G$. In the large system limit, (4.42) converges almost surely to

$$\lim_{z \to 0} \int \frac{1}{\lambda - z}\, dG(\lambda) = \lim_{z \to 0} m_G(z). \tag{4.43}$$

Then, by Theorem 2.3, $m_G(z)$ is given by

$$m_G(z) = -\left(z - \frac{1}{\beta} \int \frac{t}{1 + t m_G(z)}\, d\Lambda(t) \right)^{-1}.$$

Let $\chi = \lim_{z\to 0} m_G(z)$. Then, it follows easily that χ is the positive solution of

$$\chi = \beta \left(\int \frac{t}{1+\chi \cdot t}\, d\Lambda(t) \right)^{-1},$$

and $\text{SINR}_k - \frac{\gamma}{\chi} \xrightarrow{a.s.} 0$.

For the case $\beta > 1$, the SINR for user k is given by (see (3.11)),

$$\frac{\gamma \mathcal{A}_k^2}{\gamma \mathcal{A}_k + \text{Tr}\left((\mathbf{H}^H \mathbf{H})^{-1} \right) (1+\mathcal{A}_k)^2},$$

where $\mathcal{A}_k = \mathbf{h}_k (\mathbf{H}_k^H \mathbf{H}_k)^{-1} \mathbf{h}_k^H$. It can be easily seen that $\mathcal{A}_k = \lim_{\alpha\to 0} A_k$, where A_k is defined as in Appendix 4.6.1. Hence, it follows that

$$\mathcal{A}_k - \frac{1}{\beta - 1} \xrightarrow{a.s.} 0.$$

Now, let $\mathbf{V} = \frac{1}{N} \mathbf{R}_t^{\frac{1}{2}} \widetilde{\mathbf{H}}^H \widetilde{\mathbf{H}} \mathbf{R}_t^{\frac{1}{2}}$. Suppose that that the e.s.d. of $\mathbf{V}$ converges to H almost surely. Thus, $\frac{1}{N} \text{Tr}\left(\left(\frac{1}{N} \mathbf{H}^H \mathbf{H} \right)^{-1} \right)$ converges almost surely to

$$\lim_{z\to 0} \int \frac{1}{\lambda - z}\, dH(\lambda) = \lim_{z\to 0} m_H(z).$$

Based on Theorem 2.4, $m_H(z)$ is given by

$$m_H(z) = \int \frac{1}{t(1-\beta-\beta z m_H(z)) - z}\, d\Lambda(t).$$

Let $\psi = \lim_{z\to 0} m_H(z)$. Thus,

$$\psi = \frac{1}{\beta - 1} \int \frac{d\Lambda(t)}{t}.$$

Combining the results, we obtain the second equation of (4.13) and and we complete the proof.

4.6.5 Proof of Theorem 4.4

The SINR for user k can be expressed as (cf. eq.(3.12))

$$\text{SINR}_k = \frac{\gamma(\mathbf{h}_k\mathbf{h}_k^H)^2}{\gamma\mathbf{h}_k\mathbf{H}_k^H\mathbf{H}_k\mathbf{h}_k^H + \text{Tr}\,(\mathbf{H}\mathbf{H}^H)}.$$

For the numerator, we can write $\frac{1}{N}\mathbf{h}_k\mathbf{h}_k^H = \widehat{\mathbf{h}}_k\mathbf{R}_t\widehat{\mathbf{h}}_k$. By Lemma 2.3, $\widehat{\mathbf{h}}_k\mathbf{R}_t\widehat{\mathbf{h}}_k^H - \frac{1}{N}\text{Tr}\,(\mathbf{R}_t) \xrightarrow{a.s.} 0$ where $\frac{1}{N}\text{Tr}\,(\mathbf{R}_t)$ represents the average of the eigenvalues of the correlation matrix. Hence, we have $N^{-1}\mathbf{h}_k\mathbf{h}_k^H - \int t\, d\Lambda(t) \xrightarrow{a.s.} 0$.

Now let us consider the terms in the denominator. We can express $\frac{1}{N^2}\mathbf{h}_k\mathbf{H}_k^H\mathbf{H}_k\mathbf{h}_k^H$ as $\widehat{\mathbf{h}}_k\mathbf{R}_t\widehat{\mathbf{H}}_k^H\widehat{\mathbf{H}}_k\mathbf{R}_t\widehat{\mathbf{h}}_k^H$. By using Lemma 2.3, $\widehat{\mathbf{h}}_k\mathbf{R}_t\widehat{\mathbf{H}}_k^H\widehat{\mathbf{H}}_k\mathbf{R}_t\widehat{\mathbf{h}}_k^H - \frac{1}{N}\text{Tr}\left(\mathbf{R}_t\widehat{\mathbf{H}}_k^H\widehat{\mathbf{H}}_k\mathbf{R}_t\right) \xrightarrow{a.s.} 0$. We can write

$$\frac{1}{N}\text{Tr}\left(\mathbf{R}_t\widehat{\mathbf{H}}_k^H\widehat{\mathbf{H}}_k\mathbf{R}_t\right) = \frac{1}{N}\sum_{j\neq k}\widehat{\mathbf{h}}_j\mathbf{R}_t^2\widehat{\mathbf{h}}_j^H.$$

By applying Lemma 2.5, we obtain

$$\max_{j\leq K}\left|\widehat{\mathbf{h}}_j\mathbf{R}_t^2\widehat{\mathbf{h}}_j^H - \frac{1}{N}\text{Tr}\left(\mathbf{R}_t^2\right)\right| \xrightarrow{a.s.} 0.$$

In the large system limit, $N^{-1}\text{Tr}\left(\mathbf{R}_t^2\right) \xrightarrow{a.s.} \int t^2\, d\Lambda(t)$. We can rewrite the last as $\mathbb{E}[\mathsf{T}^2]$. Thus, $\widehat{\mathbf{h}}_k\mathbf{R}_t\widehat{\mathbf{H}}_k^H\widehat{\mathbf{H}}_k\mathbf{R}_t\widehat{\mathbf{h}}_k^H - \beta\mathbb{E}[\mathsf{T}^2] \xrightarrow{a.s.} 0$.

Considering the second term in the denominator of the SINR expression, we have

$$\begin{aligned}\frac{1}{N^2}\text{Tr}\left(\mathbf{H}\mathbf{H}^H\right) &= \frac{1}{N}\text{Tr}\left(\widehat{\mathbf{H}}_k\mathbf{R}_t\widehat{\mathbf{H}}_k^H\right)\\ &= \frac{1}{N}\sum_{j\neq k}\widehat{\mathbf{h}}_j\mathbf{R}_t\widehat{\mathbf{h}}_j^H.\end{aligned}$$

By using Lemma 2.5, we have

$$\max_{j\leq K}\left|\widehat{\mathbf{h}}_j\mathbf{R}_t\widehat{\mathbf{h}}_j^H - \frac{1}{N}\text{Tr}\,(\mathbf{R}_t)\right| \xrightarrow{a.s.} 0.$$

Hence, it follows that $N^{-2}\text{Tr}\left(\mathbf{H}\mathbf{H}^H\right) - \beta\int t\, d\Lambda(t) \xrightarrow{a.s.} 0$. Arranging the large system results together, (4.14) follows immediately.

Chapter 5

Optimal Training for Time-Division Duplexed Systems with Multiuser Precoding

In the previous chapters, it is assumed that the base station has perfect CSIT which is an ideal assumption and is hard to obtain in practice. Here, we investigate the effect of channel uncertainty on the design of MISO broadcast channels. For CSI acquisitions, we consider the Time Division Duplex scheme where a perfect reciprocity between uplink and downlink channels is assumed. We study the weighted sum rate optimization of the uplink and downlink data transmissions by finding the optimal period for the uplink training and the data transmission for the uplink and the downlink. Furthermore, for a fixed uplink power, we also explore the optimal uplink power allocation for the training and uplink data transmission. Neglecting uplink data transmission and hence, there is no power adaptation in the training process, it is shown that the optimal training period maximizing the downlink sum rate is equal to or larger than one symbol or channel use per user. On the other hand, by incorporating the uplink data communication, regardless the sum rate weighting, it is one symbol per user. The optimal period for the downlink or the uplink data transmission and the optimal uplink power allocation can be obtained by using standard numerical optimization algorithms.

5.1 Introduction

CONSIDERING the MISO broadcast channel setup used in the previous chapters, the full multiplexing gain $\min(N, K)$ can be achieved if the system has both perfect CSIT and CSIR [26],[6, Ch. 2]. Perfect knowledge of the CSI is an ideal scenario and is hard to obtain in practical situations. The quality of the CSI at both communication ends definitely affects the system performance. Thus, the optimal CSI acquisition maximizing

the system performance is an important research topic and has been addressed in a large volume of work, see for example, [50,51, and references therein], [54], [31] and [8].

One of the most common strategies to acquire CSIR is the pilot-based training, where the transmitter sends pilot symbols and the receiver estimates the downlink channel based on the received signals. There is quite a long history and considerably rich literature discussing pilot-based training in single-user MIMO systems. Here, we just mention a seminal paper in this area, which is due to Hassibi et al. [31]. The authors investigate the impact of channel uncertainties or channel estimation errors at the receiver on the capacity of single-user MIMO in the fading channel. They show that an orthonormal training sequence is optimal in maximizing the capacity lower bound (i.e., the worst case noise capacity). By performing the power allocation in the downlink for training and data transmissions, they prove that the optimal number/period for the training symbols is equal to the number of transmit antennas. Without power allocation, it is larger than the number of transmit antennas. The extension of this work for the MISO broadcast channels with perfect CSIT, reaches the same conclusions [17].

In general, there are two signaling schemes considered in the literature for the CSIT acquisition: the Time-Division Duplex (TDD) and Frequency-Division Duplex (FDD) schemes. In TDD, it is assumed that the uplink and downlink communications share the same frequency but use different time slots and there is a reciprocity between these two communication links. Thus, the BS learns the users' downlink channels via the pilot-based training in the uplink. In FDD, the downlink channels are first estimated by the users based on the training symbols sent by the BS. Then, the users send these estimates to the BS via the feedback link (uplink). Considering the feedback signal, it can be categorized into two types: analog and digital. The latter is also called the limited, quantized, or finite-rate feedback [50].

The use of the analog feedback in multiuser systems becomes popular after the publication of [54] (see also [76]). In this feedback scheme, the users send unquantized and uncoded signals for the feedback transmissions. This allows a fast transfer of the CSI [54]. In the digital feedback, on the contrary, the users feed back the quantized and encoded signals representing their downlink channels. Various digital feedback schemes in single-

user and multiuser systems are nicely summarized in [50, 51, and references therein]. In the following, we will discuss some relevant references.

The focus of [54] is the analog feedback transmission in both TDD and FDD schemes. It is assumed that each user has perfect CSIR. In TDD, even though the users do not send their CSIR for the uplink training, they need it to decode their downlink data. The efficiency of the analog feedback transmission in the TDD and the FDD based on the mean-square error of the downlink channel estimates is compared. [40, 41] extend [54] by analyzing the impact of the channel uncertainty on the system performance. Besides deriving the optimal training period, the authors also propose a linear precoding design and user selection mechanisms that maximize the achievable sum rates.

A recent work by Caire et al. in [8] analyzes the impact of downlink training and channel state feedback on the achievable ergodic sum rate of multiuser MIMO systems with the ZF precoder. Both the analog feedback and the digital feedback via Random Vector Quantization (RVQ) are considered in the system modeling. The precoder is constructed based on the CSI feedback estimated by the BS. Often ignored, another round of training, called the dedicated training, is also performed so as the users can estimate the coupling between their downlink channel and the beamforming vector. Ergodic sum rate bounds for various scenarios such as: feedback transmission through additive white Gaussian noise (AWGN) and MAC channels as well as with delayed feedback are derived. The authors conclude that a proper and careful design of the channel state feedback can yield a significant downlink sum rate. The extension of this work studies the optimal training period and feedback duration that maximize the ergodic sum rates [44, 45].

In this chapter, we consider the MISO broadcast channel as described in the previous chapter. For the CSI acquisition, we assume the TDD scheme with a perfect reciprocity between the uplink and the downlink channels. Different from other works mentioned above, except [45] but in the context of the FDD, we include the uplink data transmission in the analysis. The BS estimates the users' downlink channels from the uplink training transmission. Then, the BS uses the channel estimates to construct the RCI precoder (for the downlink data transmission) and also to decode the uplink data transmission via the MMSE receiver. Thus, the channel estimation errors will affect both the effective

downlink and uplink throughput. Our goal is to find the optimal system design parameters that maximize the weighted sum rate of the uplink and downlink transmissions. As in the previous chapters, the analysis throughout this chapter is performed in the large system regime. Our contributions can be summarized as follows:

1. We derive the limiting SINR expressions for the uplink and downlink transmissions. We also derive the optimal regularization parameter of the RCI that maximizes the downlink SINR and show that it is the ratio between the cell-loading and the effective SNR. The latter is the function of the mean-square error of the downlink channel estimates. This demonstrates that *the RCI precoder adapts to the CSIT quality*.
2. Ignoring the uplink data transmission, we perform the optimization of the training period that maximizes the downlink sum rate. We show that it is a convex program. We can also determine the uplink training power such that the optimal training period is one symbol per user. A similar problem is also considered in [101], but the authors use different approach in obtaining the large system results.
3. Incorporating the uplink data transmission in the analysis, we derive the optimal period for the uplink training, uplink and downlink data transmission, and determine the optimal uplink power allocation for the training and the data transmission. As a result, we show that the optimal training period is always one symbol per user. The optimal parameter for others can be obtained by using standard numerical optimization algorithms. Part of the analysis in [45] also considers the uplink data transmission, but focuses only in obtaining the optimal training period. Moreover, that analysis is intended for the FDD scheme.

The rest of the chapter is organized as follows. The next section give the system descriptions including the models for the uplink training, downlink and uplink data transmissions. In Section 5.3, the large system results for the SINR of the downlink and uplink data transmission are derived. They will be used to determine the optimal system parameters maximizing the weighted sum rate of the downlink and uplink transmissions for various scenarios, as discussed in Section 5.4. Some numerical results are also presented. The conclusions can be found in Section 5.5.

5.2 System Model

Similar to the previous chapters, we consider a MISO broadcast channel with N transmit antennas at the base station and K users, where each user is equipped with a single antenna. To mitigate the multi-user interference, the BS precodes or beamforms the data symbols before transmission. In order to design the precoder, the BS needs to know the channel of each user. Here, we focus on the TDD with a perfect reciprocity for the uplink and downlink transmissions. Thus, for the CSI acquisition, the BS learns the CSI from pilots or training symbols sent by each user in the uplink transmission. If the training symbols are sent orthogonally by each user (orthogonal MAC), then the training period is at least K symbols time (see [31, 54]). Intuitively, increasing the number of training symbols will provide a better channel estimate at BS, as confirmed later in the following section. For the uplink data transmission,the channel estimates are used by the BS to decode the data sent by the user coherently. For the downlink transmission, they are employed by the BS to construct the precoding vectors for each user. To decode the data transmitted by the BS during this transmission, each user needs to know the coupling between its downlink channel and the corresponding precoding vector which can be obtained by performing a dedicated training [8]. However, it only takes one symbol and can be considered very small in the system with large number of users or BS's antennas. Therefore, for the rest of this chapter, the dedicated training is omitted and we assume that each user has a perfect knowledge of the coupling coefficient.

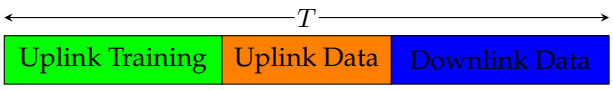


Figure 5.1: Transmission phases.

In this work, we adopt the block-fading channel model where the channel is constant for one block that has length T channel uses (symbols) and changes independently between the blocks. Thus, from our discussions above, each block consists of three transmission types: an uplink training, an uplink data transmission and a downlink transmission, as illustrated in Figure 5.1. The data models for each type or phase are described in details in the following subsections.

5.2.1 Uplink Channel Training

Let T_τ be the training period or the number of the training symbols and $\mathbf{s}_t$ be the vector of size K that represents the training symbols from all users at integer time t. The collective training symbols during T_τ are denoted by $\mathbf{S}_\tau = [\mathbf{s}_1, \cdots, \mathbf{s}_{T_\tau}]$. Here, we assume that each user has the same transmit power for the training, denoted by P_τ. The received signal at the BS can be expressed as

$$\mathbf{Y}_{\text{BS}} = \mathbf{F}\mathbf{S}_\tau + \mathbf{N},$$

where $\mathbf{Y}_{\text{BS}} = [\mathbf{y}_{1,\text{BS}}, \ldots, \mathbf{y}_{T_\tau,\text{BS}}]$ is the collective receive signal of size $N \times T_\tau$ and $\mathbf{N} = [\mathbf{n}_1, \cdots, \mathbf{n}_{T_\tau}]$ is the matrix of receiver (BS) noise. The elements of $\mathbf{n}_t$ are i.i.d. with distribution $\mathcal{CN}(0, \sigma_u^2)$. The matrix $\mathbf{F}$ is the uplink channel gain matrix where each element $f_{n,k}$ represents the uplink channel between user k and the n-th antenna of the BS. The elements of $\mathbf{F}$ are modeled as i.i.d. complex Gaussian with zero mean and unit variance.

For analytical tractability, we consider an orthogonal training scheme proposed previously in [54]. To be able to estimate $\mathbf{F}$, we need $T_\tau \geq K$ symbols. The training sequences can be written as [54]

$$\mathbf{S}_\tau = \sqrt{T_\tau P_\tau}\boldsymbol{\Theta},$$

where $\boldsymbol{\Theta}$ is a $K \times T_\tau$ unitary matrix ($\boldsymbol{\Theta}\boldsymbol{\Theta}^H = \mathbf{I}_K$). Following [54], the minimum mean-square error (MMSE) estimate of the uplink channel is given by

$$\hat{\mathbf{F}} = \frac{\sqrt{T_\tau P_\tau}}{\sigma_u^2 + T_\tau P_\tau}\mathbf{Y}_{\text{BS}}\boldsymbol{\Theta}^H.$$

The relation between the actual channel $\mathbf{F}$ and its estimates $\hat{\mathbf{F}}$ can be modeled as follows

$$\mathbf{F} = \hat{\mathbf{F}} + \tilde{\mathbf{F}}, \tag{5.1}$$

where $\tilde{\mathbf{F}}$ is the channel estimation error and independent from the channel estimate $\hat{\mathbf{F}}$. The entries of $\tilde{\mathbf{F}}$ are i.i.d. zero mean circularly symmetric complex Gaussian (ZMCSCG) with variance

$$\sigma_\tau^2 = \frac{1}{1 + \gamma_\tau T_\tau}, \tag{5.2}$$

where $\gamma_\tau = P_\tau/\sigma_u^2$ is the SNR per user in the uplink. The entries of $\hat{\mathbf{F}}$ are also i.i.d. ZMCSCG with variance $1-\sigma_\tau^2$. We can reduce σ_τ^2 by increasing the training power or/and the training period. Note that $\hat{\mathbf{F}}$ will be used to construct the precoder for the downlink transmission and to decode the users' data in the uplink transmission.

5.2.2 Downlink Data Transmission

The model for the downlink data transmission is similar to the one in Section 3.2. As we assume a perfect reciprocity between the uplink and the downlink channels, the relation between the uplink and downlink channel matrices is given by

$$\mathbf{F} = \mathbf{H}^T.$$

Thus, we can write

$$\mathbf{H} = \widehat{\mathbf{H}} + \widetilde{\mathbf{H}}, \quad \text{and} \quad \mathbf{h}_k = \widehat{\mathbf{h}}_k + \widetilde{\mathbf{h}}_k, \tag{5.3}$$

where $\mathbf{h}_k$ is the kth row of $\mathbf{H}$. The entries of $\widehat{\mathbf{H}}$ and $\widetilde{\mathbf{H}}$ have the probability distributions that follow those of $\widehat{\mathbf{F}}^T$ and $\widetilde{\mathbf{F}}^T$, respectively. The BS uses $\widehat{\mathbf{H}}$ to construct the RCI precoder which is given by

$$\mathbf{P} = c(\widehat{\mathbf{H}}^H\widehat{\mathbf{H}} + \alpha\mathbf{I}_N)^{-1}\widehat{\mathbf{H}}^H, \tag{5.4}$$

where α is the regularization parameter and c is the normalizing constant to meet the transmit power constraint $\mathbb{E}[\|\mathbf{x}\|_2^2] = P_d$. Hence, c^2 can be expressed as

$$c^2 = \frac{P_d}{\mathrm{Tr}\left((\widehat{\mathbf{H}}^H\widehat{\mathbf{H}} + \alpha\mathbf{I}_N)^{-2}\widehat{\mathbf{H}}^H\widehat{\mathbf{H}}\right)}. \tag{5.5}$$

The received signal model is the same as in (3.6) but $\mathbf{h}_k$ and $\mathbf{H}$ follows the channel model (5.3). Similar to (3.7), we can express the downlink SINR as follows

$$\mathrm{SINR}_{k,\mathrm{dl}} = \frac{c^2\left|\mathbf{h}_k\left(\widehat{\mathbf{H}}^H\widehat{\mathbf{H}} + \alpha\mathbf{I}_N\right)^{-1}\widehat{\mathbf{h}}_k^H\right|^2}{\sigma^2 + c^2\sum_{j\neq k}\left|\mathbf{h}_k\left(\widehat{\mathbf{H}}^H\widehat{\mathbf{H}} + \alpha\mathbf{I}_N\right)^{-1}\widehat{\mathbf{h}}_j^H\right|^2}, \tag{5.6}$$

where σ^2 is the users' receiver noise.

5.2.3 Uplink Data Transmission

For the uplink transmission, it is assumed that each user synchronously transmits to the BS. The received signal at the base station can be expressed as

$$\mathbf{y}_{\text{BS}} = \sum_{k=1}^{K} \mathbf{f}_k r_k + \mathbf{n}_k = \mathbf{F}\mathbf{r} + \mathbf{n}, \tag{5.7}$$

where $\mathbf{r} = [r_1, \ldots, r_K]^T$ and r_k is the uplink data symbol of user k. We assume that the data symbols $\{r_k\}$ are independent. $\mathbf{f}_k$ denotes the uplink channel between BS and user k and is the kth column of $\mathbf{F}$ (or kth row of $\mathbf{H}$).

Using (5.1), we can express (5.7) as

$$\mathbf{y}_{\text{BS}} = (\widehat{\mathbf{F}} + \widetilde{\mathbf{F}})\mathbf{r} + \mathbf{n} = \widehat{\mathbf{F}}\mathbf{r} + \underbrace{\widetilde{\mathbf{F}}\mathbf{r} + \mathbf{n}}_{\mathbf{z}}, \tag{5.8}$$

where $\mathbf{z}$ is the sum of the receiver noise and the residual channel estimation error. We can consider the last as the additional noise since the BS only knows $\widehat{\mathbf{F}}$ to decode $\mathbf{r}$. The covariance matrix of $\mathbf{z}$ can be derived as follows,

$$\begin{aligned}
\mathbf{K}_z &= \mathbb{E}[\mathbf{z}\mathbf{z}^H] = \mathbb{E}[(\widetilde{\mathbf{F}}\mathbf{r} + \mathbf{n})(\widetilde{\mathbf{F}}\mathbf{r} + \mathbf{n})^H] \\
&= P_u \left(\sum_{j=1}^{K} \mathbb{E}[\tilde{\mathbf{f}}_j \tilde{\mathbf{f}}_j^H] \right) + \mathbb{E}[\mathbf{n}\mathbf{n}^H] \\
&= \left(P_u K \sigma_\tau^2 + \sigma_u^2\right) \mathbf{I}_N.
\end{aligned}$$

In the second line, we use the fact that $\mathbb{E}[\mathbf{r}\mathbf{r}^H] = P_u \mathbf{I}_K$, and in the last line, from (5.2), it follows that $\mathbb{E}[\tilde{\mathbf{f}}_j \tilde{\mathbf{f}}_j^H] = \sigma_\tau^2 \mathbf{I}_N$.

In order to decode the data symbols $\{r_k\}$, we employ the MMSE estimation based on the received signal vector, $\mathbf{y}_{\text{BS}}$ and the channel estimate, $\widehat{\mathbf{F}}$, which is known at the BS. Let $\gamma_u = P_u/\sigma_u^2$ be the uplink SNR. The MMSE receiver for user k is given by [20]

$$\begin{aligned}
\mathbf{m}_k &= \left(\widehat{\mathbf{F}}\widehat{\mathbf{F}}^H + (K\sigma_\tau^2 + \gamma_u^{-1})\mathbf{I}_N \right)^{-1} \hat{\mathbf{f}}_k \\
&= \text{constant} \times \left(\widehat{\mathbf{F}}_k \widehat{\mathbf{F}}_k^H + (K\sigma_\tau^2 + \gamma_u^{-1})\mathbf{I}_N \right)^{-1} \hat{\mathbf{f}}_k,
\end{aligned}$$

where $\widehat{\mathbf{F}}_k$ is $\widehat{\mathbf{F}}$ with the kth column removed. Note that the MMSE receiver also has the knowledge of σ_τ^2. The uplink SINR then can be expressed as (see for an example [20])

$$\text{SINR}_{k,\text{up}} = \hat{\mathbf{f}}_k^H \left(\widehat{\mathbf{F}}_k \widehat{\mathbf{F}}_k^H + (K\sigma_\tau^2 + \gamma_u^{-1})\mathbf{I}_N \right)^{-1} \hat{\mathbf{f}}_k. \tag{5.9}$$

From the above, it is clear that the uplink SINR is the function of the (uplink) channel estimates, channel estimation variance and the uplink SNR. Moreover, since the channel estimates are random then (5.9) and also (5.6) are also random quantities. In the next section, we can see that the randomness disappears in the large system regime.

5.3 Large System Analysis

In this section, we will derive the limiting SINR expressions for the uplink and downlink data transmissions. We show how the channel uncertainty, represented by the channel estimation error variance, affects those limiting SINRs. We also derive the optimal regularization parameter of the RCI precoder that maximizes the limiting SINR. We show that it adapts to the changes of σ_τ^2.

5.3.1 Downlink Transmission

Theorem 5.1. *Let $\rho = \frac{\alpha}{N(1-\sigma_\tau^2)}$ be the normalized or effective regularization parameter and $\gamma_d = P_d/\sigma^2$ be the downlink SNR. Let $g(x, y)$ be the function as defined in (2.10). In the large system limit, the downlink SINR, (5.6), converges almost surely to a deterministic quantity given by*

$$SINR_{dl}^{\infty} = \gamma_e g(\beta, \rho) \frac{1 + \frac{\rho}{\beta}\left(1 + g(\beta, \rho)\right)^2}{\gamma_e + \left(1 + g(\beta, \rho)\right)^2}, \tag{5.10}$$

where

$$\gamma_e = \frac{\gamma_d(1 - \sigma_\tau^2)}{\gamma_d \sigma_\tau^2 + 1} \tag{5.11}$$

is defined as the effective (downlink) SNR.

Proof. See Apppendix 5.6.1. □

It is obvious from (5.10) that the limiting SINR is deterministic and is the same for all users. The randomness introduced by the channel estimates disappears. It is the function of the effective SNR, the regularization parameter, and the cell-loading. Specifically, it is increasing in γ_e. Thus, increasing γ_d and/or decreasing σ_τ^2 will improve the effective SNR and consequently the limiting SINR. It can be checked that ρ and β affect both the signal strength and the interference energy. As discussed in the previous chapters, ρ controls the amount of interference introduced to the users and hence, should be determined optimally as stated in the following.

Corollary 5.1. *The optimal choice of ρ, denoted by ρ_d, that maximizes $SINR_{dl}^{\infty}$ is*

$$\rho_d = \frac{\beta}{\gamma_e}. \tag{5.12}$$

Consequently, the maximum limiting SINR can be expressed as

$$SINR_{dl}^{\infty,*} = g(\beta, \rho_d). \tag{5.13}$$

Proof. The limiting SINR (5.10) has the same structure as (3.16). Thus, the proof follows the steps in Appendix 3.6.1. □

The corollary above can be perceived as the extension of Theorem 3.2 for the channel model (5.3) that accommodates the channel estimation error. As mention earlier, reducing σ_τ^2 will increase γ_e and thus, will decrease ρ_d. This will improve $\text{SINR}_{\text{dl}}^{\infty,*}$ since $g(\beta, \rho_d)$ is decreasing in ρ_d.

5.3.2 Uplink Transmission

Theorem 5.2. *Let*

$$\rho_u = \frac{\beta}{1-\sigma_\tau^2}\left(\sigma_\tau^2 + \overline{\gamma}_u^{-1}\right). \tag{5.14}$$

Then, the uplink SINR (5.9) converges almost surely to a deterministic quantity given by

$$SINR_{up}^{\infty} = g(\beta, \rho_u). \tag{5.15}$$

Proof. We can write (5.6) as

$$\frac{1}{N(1-\sigma_\tau^2)}\hat{\mathbf{f}}_k^H\left(\frac{1}{N(1-\sigma_\tau^2)}\widehat{\mathbf{F}}_k\widehat{\mathbf{F}}_k^H+\frac{1}{N(1-\sigma_\tau^2)}(K\sigma_\tau^2+\gamma_u^{-1})\mathbf{I}_N\right)^{-1}\hat{\mathbf{f}}_k.$$

The entries of $\sqrt{N^{-1}(1-\sigma_\tau^2)^{-1}}\hat{\mathbf{f}}_k$ and $\sqrt{N^{-1}(1-\sigma_\tau^2)^{-1}}\widehat{\mathbf{F}}_k$ are now zero mean with variance $1/N$. Defining ρ_u as in (5.14) and applying Lemma 2.4, we have

$$\text{SINR}_{k,\text{up}}-\text{SINR}_{\text{up}}^{\infty}\xrightarrow{a.s.}0,$$

and this completes the proof. □

Similar to the downlink transmission case, it is obvious from the expression of ρ_u that reducing the channel estimation error variance will decrease ρ_u. As a result, the uplink SINR will increase.

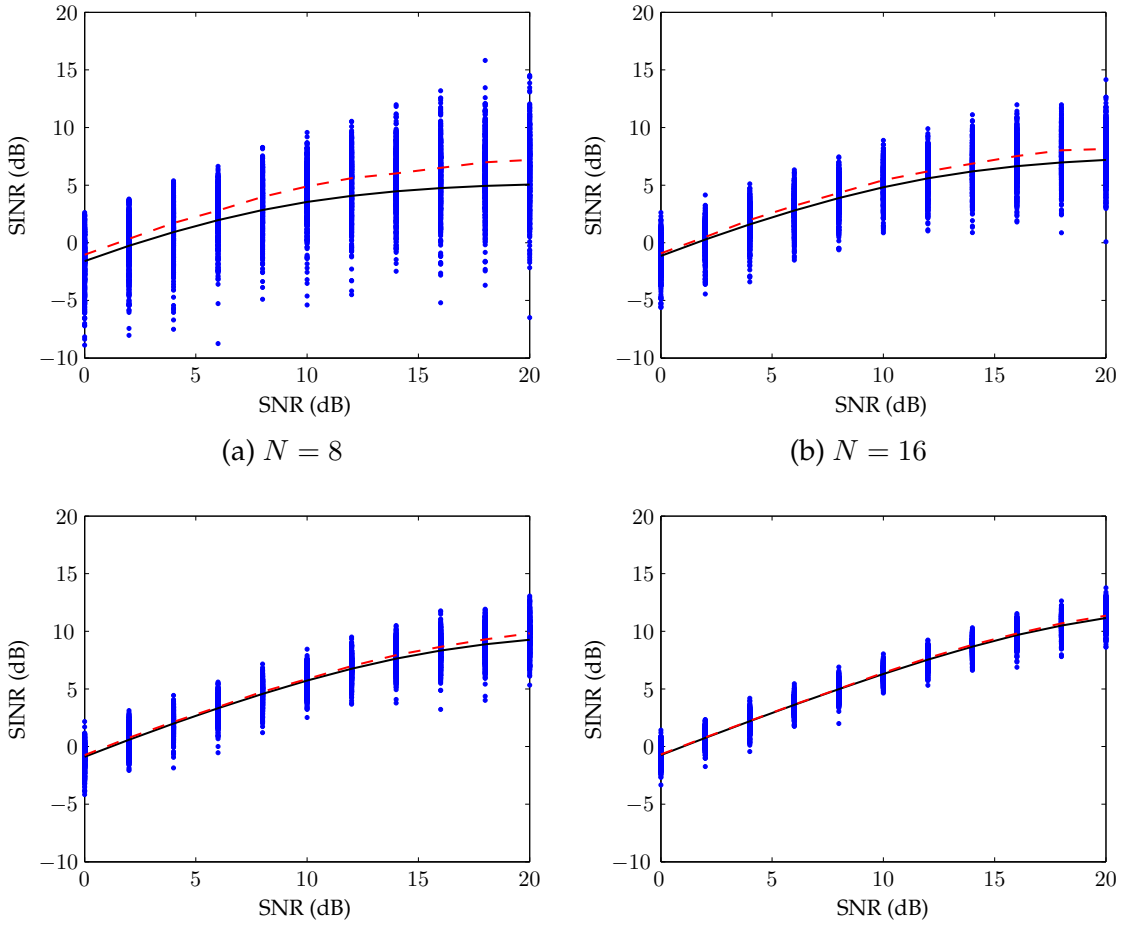
By comparing (5.15) and (5.13), we can see that they have a common structure. Moreover, by rewriting ρ_d as

$$\frac{\beta}{1-\sigma_\tau^2}\left(\sigma_\tau^2+\gamma_d^{-1}\right),$$

we can see that it has the same form as ρ_u where γ_d can be considered to be equivalent to $\overline{\gamma}_u$. Thus, $\frac{\overline{\gamma}_u(1-\sigma_\tau^2)}{\overline{\gamma}_u\sigma_\tau^2+1}$ can be perceived as the effective uplink SNR. For the rest of the chapter, we just call $\overline{\gamma}_u$ as the uplink SNR.

5.3.3 Numerical Simulations

Theorem 5.1 (and also Corollary 5.1) already provided the expression for the limiting downlink SINR. Figure 5.2 shows how the random *downlink* SINRs of (5.6) converge to the limiting SINR. The random SINRs are obtained by using the same procedure as in obtaining Figure 3.2. Considering user 1, for different system sizes N and for each γ_d, 500 channel realizations are generated and the corresponding SINRs are computed. The dash line in the figure represents the average of the random SINRs. The difference between this average SINR and its asymptotic limit is about 2.2 dB (for $\gamma_d = 20$ dB) when $N = 8$ and becomes around 0.25 dB when $N = 64$. One can see that as the system size increases



(a) $N = 8$ (b) $N = 16$

(c) $N = 32$ (d) $N = 64$

Figure 5.2: Comparison of the randomly generated downlink SINR for user 1 (dot) with the average SINR (dash) and the asymptotic limit (5.10) (solid line) with $\rho = \rho_d$, $T_\tau = K$, $\gamma_\tau = 0$ dB and $\beta = 0.75$.

the spread of the random SINRs becomes smaller. It is already about 2 dB around the average when $N = 64$.

Figure 5.3 illustrates the applicability of the large system results for the finite-size system design. Similar to the previous simulations, we generate 500 channel realizations. For each channel realization, we compute the optimal regularization parameter, denoted by $\rho^*_{d,\text{FS}}$, that maximizes the achievable downlink sum rate, $R_{\text{sum}} = \sum_{k=1}^{K} \log_2(1 + \text{SINR}_{k,\text{dl}})$. Then, its average sum rate is compared the one that apply the large system limit result, i.e., ρ_d. As one can observe from the figure, even for a reasonable small system $N = 8$, the gap between the curves is very small.

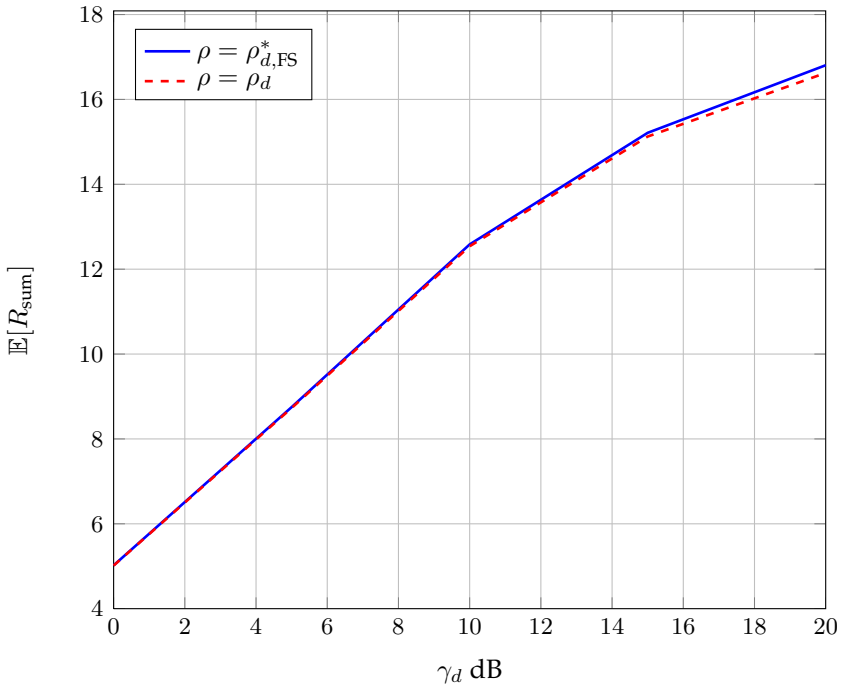
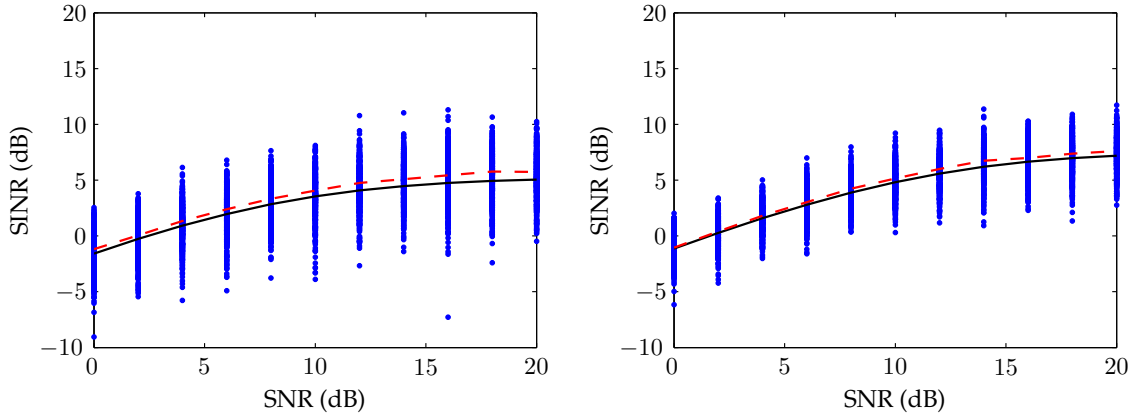


Figure 5.3: Comparison of the average sum rate of the downlink by using $\rho = \rho_d$ and $\rho = \rho^*_{d,\mathrm{FS}}$ with $N, \beta = 0.75, \sigma^2_\tau = 0.1$.

Similar to Figure 5.2, Figure 5.4 demonstrates the convergence of the random *uplink* SINRs to its asymptotic limit. Our observations from this figure follow to those for the downlink case. Our numerical simulations so far show the validity of using the large system results in Theorem 5.1 and 5.2 for the finite-size system. Thus, these asymptotic results will be used in the following section to derive the optimal parameters that maximize system performance.



(a) $N = 8$ (b) $N = 16$

(c) $N = 32$ (d) $N = 64$

Figure 5.4: Comparison of the randomly generated uplink SINR for user 1 (dot) with the average SINR (dash) and the asymptotic limit (5.15) (solid line) with $\rho = \rho_u$, $T_\tau = K$, $\gamma_\tau = 0$ dB and $\beta = 0.75$.

5.4 Asymptotically Optimal Parameters in TDD Training-based RCI Beamforming Scheme

As mentioned previously in the introduction section, most of the works in the area considered in this chapter focus on maximizing the downlink throughput. Similar to [45], we include the uplink data transmission in our analysis. So, in this section, we concern in maximizing the weighted (limiting) sum rate of the downlink and uplink by optimally choosing the following parameters: the training period, the uplink and downlink data transmission periods, the power for the training and the uplink data transmission.

As illustrated in Figure 5.1, for one channel block of length T, we have an uplink training, a downlink transmission and an uplink data transmission. The last two have transmission duration of $\overline{T}_d$ and T_u, respectively. Let $\overline{T} = \frac{T}{K}$ be the normalized block length w.r.t. the number of users. Then, we can write

$$\overline{T}_\tau + \overline{T}_u + \overline{T}_d = \overline{T},$$

where $\overline{T}_\bullet = \frac{T_\bullet}{K}$. We should note that $\overline{T} \leq W_c T_c$ where W_c and T_c are the coherence bandwidth and coherence time respectively [45]. Now, let $\overline{T}_{\text{up}} = \overline{T}_\tau + \overline{T}_u$ be the total period for the uplink transmission. Then, for a fixed uplink power or SNR $\overline{\gamma}_{\text{up}}$, we can define the following relationship (see [31])

$$\overline{\gamma}_\tau \overline{T}_\tau + \overline{\gamma}_u \overline{T}_u = \overline{\gamma}_{\text{up}} \overline{T}_{\text{up}},$$

where $\overline{\gamma}_\tau = K\gamma_\tau$. Moreover, let $\nu \in (0, 1)$ be the fraction of the uplink SNR which is allocated for the uplink data transmission. Then, we can write [31]

$$\overline{\gamma}_u \overline{T}_u = \nu \overline{\gamma}_{\text{up}} \overline{T}_{\text{up}} \tag{5.16}$$

$$\overline{\gamma}_\tau \overline{T}_\tau = (1-\nu) \overline{\gamma}_{\text{up}} \overline{T}_{\text{up}}. \tag{5.17}$$

For the downlink transmission of T_d channel uses, the following sum rate is achiev-

able

$$R_d = \frac{T_d}{T}\sum_{k=1}^{K}\log(1+\text{SINR}_{k,\text{dl}}).$$

From Theorem 5.1, $\text{SINR}_{k,\text{dl}} - \text{SINR}_{\text{dl}}^{\infty} \xrightarrow{a.s.} 0$ and the maximum limiting SINR is given by (5.13). Thus, we can define the limiting sum rate per antenna as follows

$$R_d^{\infty} = \beta\frac{\overline{T}_d}{\overline{T}}\log\left(1+\text{SINR}_{\text{dl}}^{\infty,*}\right). \tag{5.18}$$

Similarly, we can also define the limiting sum rate for the uplink transmission as follows

$$R_u^{\infty} = \beta\frac{\overline{T}_u}{\overline{T}}\log\left(1+\text{SINR}_{\text{up}}^{\infty}\right). \tag{5.19}$$

Let $w \in [0,1]$ and $R_w^{\infty} = wR_d^{\infty} + (1-w)R_u^{\infty}$ be the weighted limiting sum rate. In the remainder of this chapter, we consider to solve the following optimization problem

$$\begin{aligned}
\mathbf{P1}: \quad \max_{\overline{T}_\tau,\overline{T}_d,\overline{\gamma}_\tau} \quad & R_w && \text{(5.20a)}\\
\text{s.t.} \quad & \overline{T}_\tau + \overline{T}_u + \overline{T}_d = \overline{T} && \text{(5.20b)}\\
& \overline{\gamma}_\tau\overline{T}_\tau + \overline{\gamma}_u\overline{T}_u = \overline{\gamma}_{\text{up}}\overline{T}_{\text{up}}. && \text{(5.20c)}
\end{aligned}$$

A somewhat similar optimization problem is also considered in [45] but in the context of FDD systems. In what follows, we will discuss several scenarios of the optimization problem above.

5.4.1 Two Phases of Transmission: No Uplink Data Transmission

This is the simplest case of **P1**. The optimization problem becomes

$$\begin{aligned}
\mathbf{P2}: \quad \max_{\overline{T}_\tau} \quad & R_d^{\infty} && \text{(5.21)}\\
\text{s.t.} \quad & \overline{T}_\tau + \overline{T}_d = \overline{T}.
\end{aligned}$$

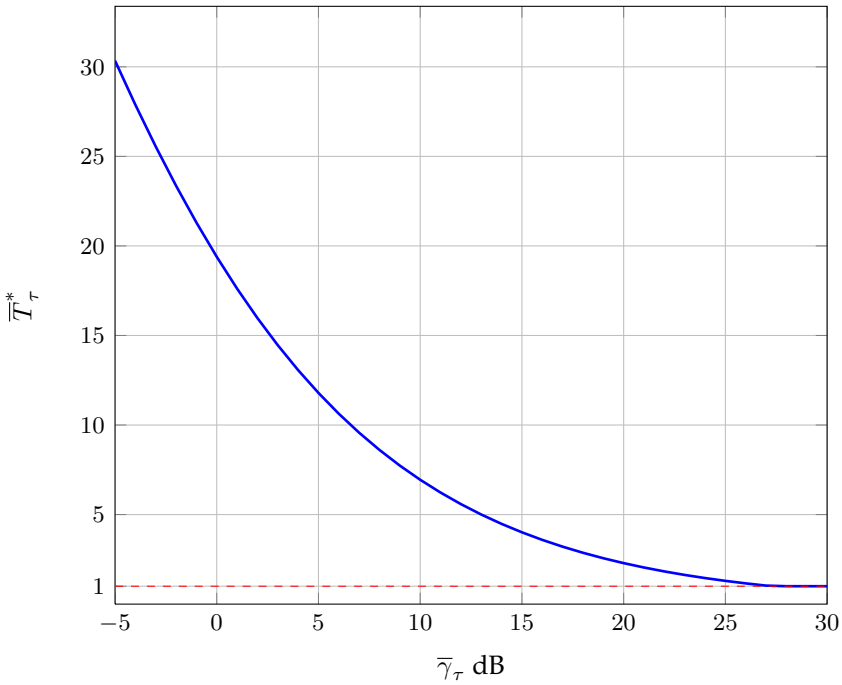


Figure 5.5: Optimal training symbols per user vs. uplink training power ($\gamma_d = 20$ dB, $\overline{T} = 200$).

Here, we focus on finding the optimal training period that maximizes the limiting downlink sum rate. A similar problem is considered in [40] and also [45] but the authors employ the ZF precoder at the BS. Independently, [101] considers the same settings as ours but the large system results are obtained by using a different approach.

The objective function of **P1** is a strictly concave function as proved in Appendix 5.6.2. Moreover, the constraint of **P1** is linear. Thus, **P1** is a convex program. The unique stationary point can be obtained by setting $\frac{\partial R_d^\infty}{\partial \overline{T}_\tau}$, in (5.26), equal to zero. However, it can not be expressed in a closed form and can be found by using standard convex optimization or line search algorithms. Since **P1** is convex, we can also determine the uplink training power such that the optimal training symbols per user, denoted by $\overline{T}^*_\tau$, is one symbol per user by finding $\bar{\gamma}_\tau$ that satisfies $\frac{\partial R_d^\infty}{\partial \overline{T}_\tau}\Big|_{\overline{T}^*_\tau=1} = 0$. Figure 5.5 illustrates $\overline{T}^*_\tau$ as a function of $\bar{\gamma}_\tau$. It shows that the optimal training symbol is decreasing in $\bar{\gamma}_\tau$. In our example, where we set $\gamma_d = 20$ dB, $\overline{T} = 200$, we need a training SNR which is larger than 28 dB to get one pilot or training symbol per user.

5.4.2 Fixed Period of the Downlink Data Transmission

The goal of this subsection is to investigate how the training period and the training power affect the downlink transmission. Since $\overline{T}_d$ is fixed then $\overline{T}_{\text{up}}$ is also fixed. The weighted limiting sum rate maximization the training can be formulated as follows

$$\begin{aligned} \mathbf{P3}: \quad \max_{\overline{\gamma}_\tau, \overline{T}_\tau} \quad & R_w \\ \text{s.t.} \quad & \overline{T}_\tau + \overline{T}_u = \overline{T}_{\text{up}} \\ & \overline{\gamma}_\tau \overline{T}_\tau + \overline{\gamma}_u \overline{T}_u = \overline{\gamma}_{\text{up}} \overline{T}_{\text{up}}. \end{aligned} \tag{5.22}$$

Our optimizing variables in **P3** are $\overline{\gamma}_\tau$ and $\overline{T}_\tau$. Both will affect the downlink (limiting) SINR via σ_τ^2. Since $\overline{T}_d$ is already fixed, then pre-log term of R_d^∞ does not change with $\overline{T}_\tau$. Hence, R_d^∞ is increasing in both $\overline{\gamma}_\tau$ and $\overline{T}_\tau$. For the uplink data transmission, increasing $\overline{T}_\tau$ will increase the uplink SINR but on the other hand decrease the pre-log factor of R_u^∞. Similarly, increasing $\overline{\gamma}_\tau$ will improve the CSI at the BS but with the cost of decreasing the power for the uplink data transmission. Hence, both $\overline{\gamma}_\tau$ and $\overline{T}_\tau$ provide a trade-off between improving the CSI at the BS and decreasing the performance of the uplink data transmission. Solving **P3** will give the optimal choice of $\overline{\gamma}_\tau$ and $\overline{T}_\tau$ and its solution is summarized in the following.

Theorem 5.3. *R_w in **P3** reaches its maximum by choosing $\overline{T}_\tau = 1$. For any given feasible $\overline{T}_\tau$, R_w is also a strictly concave function of $\overline{\gamma}_\tau$. Thus, the optimal $\overline{\gamma}_\tau$, denoted by $\overline{\gamma}_\tau^*$, is unique and is given by the solution of $\frac{\partial R_w}{\partial \overline{\gamma}_\tau}\Big|_{\overline{T}_\tau = 1} = 0$.*

Proof. We can solve **P3** in two steps: first, with a fixed $\overline{\gamma}_\tau$, we derive the optimal $\overline{T}_\tau$ and by substituting it back into R_w we then solve for the optimal $\overline{\gamma}_\tau$. Let us consider R_d^∞ in (5.18). From (5.17), we can write

$$\rho_d = \frac{\beta}{\gamma_d} \left(\frac{\gamma_d + 1}{(1-\nu)\overline{\gamma}_{\text{up}} \overline{T}_{\text{up}}} + 1 \right). \tag{5.23}$$

Since $\overline{T}_d$ and $\overline{T}_{\text{up}}$ are fixed then R_d^∞ is not affected by $\overline{T}_\tau$ ($\forall \nu$). Then, let us consider R_d^∞

in (5.19) with $\overline{T}_u = \overline{T}_{\text{up}} - \overline{T}_\tau$. ρ_u in the uplink SINR can be expressed as

$$\rho_u = \frac{\beta}{(1-\nu)\overline{\gamma}_{\text{up}}\overline{T}_{\text{up}}}\left(1 + \frac{\overline{T}_u}{\nu\overline{\gamma}_{\text{up}}\overline{T}_{\text{up}}}\right) + \frac{\beta\overline{T}_u}{\nu\overline{\gamma}_{\text{up}}\overline{T}_{\text{up}}}. \tag{5.24}$$

Thus, for a fixed ν, as $\overline{T}_\tau$ increases, the (limiting) uplink SINR increases logarithmically. On the other hand, the pre-log factor of R_u^∞ decreases linearly. This suggests that R_u^∞ is decreasing in $\overline{T}_\tau$. Analytically, it is stated in following proposition.

Lemma 5.1. *The uplink sum rate is a monotonically decreasing function over* $\overline{T}_\tau$

Proof. See Appendix 5.6.3 □

Based on the lemma above, the optimal training period is one symbol per user. Now, we only have ν as the design variable. By using the same steps as those in Appendix 5.6.2, it can be easily shown that $g(\beta, \rho_d)$ is strictly concave over ν. It is also proved in Appendix 5.6.4 that $g(\beta, \rho_u)$ is concave in ν. Since both the $\log$ operation and linear combination do not change the concavity, therefore the weighted sum rate is strictly concave over ν. From (5.17), we can see that $\overline{\gamma}_\tau$ is affine in ν. Thus, R_w is also strictly concave over $\overline{\gamma}_\tau$ [7]. The optimal $\overline{\gamma}_\tau$ can then be found efficiently by using line search or other existing algorithms for solving convex/concave optimization problems. □

Comparing the optimal $\overline{T}_\tau$ obtained in the theorem above and the one from Subsection 5.4.1, we can conclude that involving power allocation between the uplink training and data transmission in maximizing the weighted sum rate will lead to a minimum number of training period. A similar conclusion is also stated in [31] in the context of *single-user* MIMO systems. In [44], it can be achieved if a certain condition is satisfied.

5.4.3 Variable Downlink and Uplink Transmission Periods

In this subsection, we consider the original problem **P1**. It is the same as **P3** except we have $\overline{T}_d$ as another design variable. Now, ρ_d in (5.23) is a function of both ν and $\overline{T}_d$ while ρ_u is a function of ν, $\overline{T}_d$ and also $\overline{T}_\tau$. By using the same arguments as in Theorem 5.3, we can state the following.

Theorem 5.4. *The maximum R_w in* **P1** *can be achieved by choosing* $\overline{T}_\tau = 1$.

The above applies for all $\nu, \overline{T}_d > 0$. Our previous discussions regarding the optimal $\overline{T}_\tau$ still hold.

Now, we only have to determine the other two optimal design variables. Unfortunately, one can show that the weighted sum rate is not jointly concave over ν and $\overline{T}_d$. However, since this problem involves only two design variables we can solve it by using a brute-force method. Since the design variables are continuous, we should first discretize the design variables and then do an exhaustive search. This leads to an approximate solution to the original problem. We can expect this solution to be closer to the true optimum when we use a finer discretization. A more directed approach is by using the alternating optimization technique although a global optimum solution is not guaranteed. In this method, we find an optimum point over one variable while keeping other variables fixed. In particular, for our problem, the algorithm can be described as follows [64]: start with an arbitrary initial point value $\overline{T}_d^{(0)}$; for $\ell \geq 1$, perform the following steps iteratively

$$\nu^{(\ell)} = \arg\max_{\nu \in \mathcal{V}} R_w(\overline{T}_d^{(\ell-1)}, \nu)$$
$$\overline{T}_d^{(\ell)} = \arg\max_{\overline{T}_d \in \mathcal{T}_d} R_w(\overline{T}_d, \nu^{(\ell)})$$

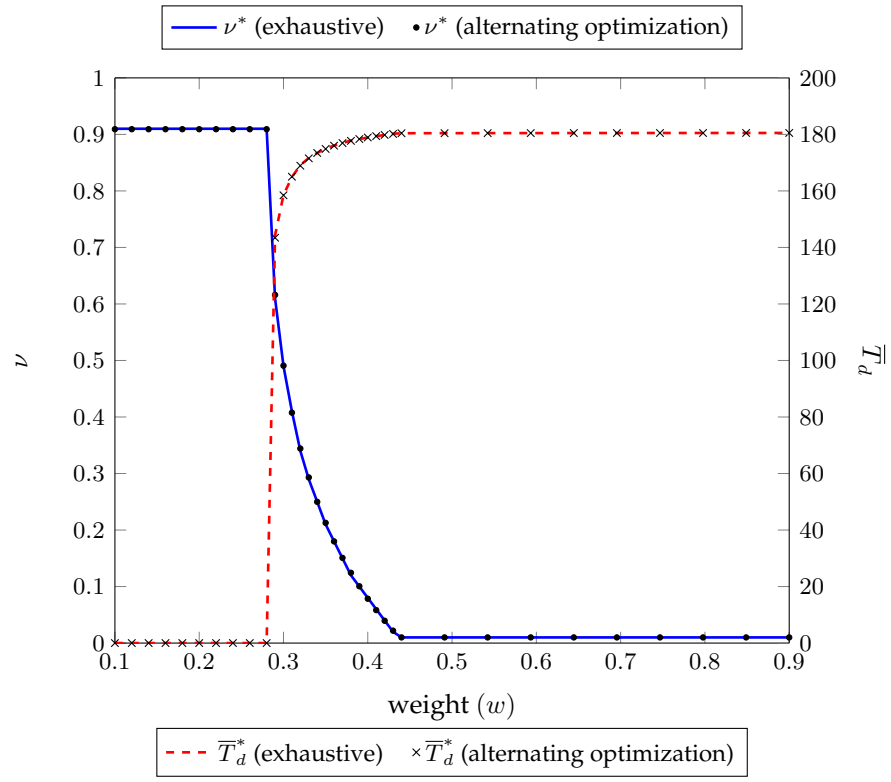
with $\mathcal{V} = \{\nu : \nu_{\text{min}} \leq \nu \leq \nu_{\text{max}}\}$ and $\mathcal{T}_{\text{dl}} = \{\overline{T}_d : \overline{T}_{d,\text{min}} \leq \overline{T}_d \leq \overline{T}_{d,\text{max}}\}$, until ν and $\overline{T}_d$ converge under a specified termination condition. We should note that we need the upper bound and lower bound for ν in the above due to ν exists in the interval $(0, 1)$. This will avoid the training or uplink data transmission to have zero power. Similarly, we also have the bounds for T_d to prevent zero period for the uplink and downlink data transmissions.

Now let us consider the steps in the alternating optimization. Since $R_w(\overline{T}_d, \nu)$ is concave over ν, we can obtain $\nu^{(\ell)}$ by employing existing algorithms for solving the convex programming. Finding $\overline{T}_d^{(\ell)}$ is a little bit more complicated since it is a difference of convex (dc) problem for some values of $\nu^{(\ell)}$. However, since it is a one dimensional optimization problem we can use existing algorithms of line search with constraints

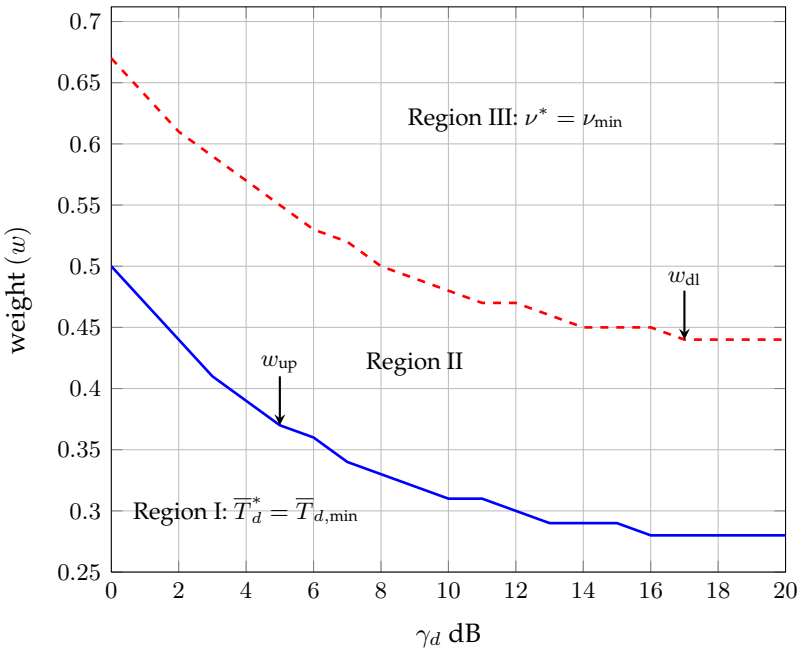
such as the projected gradient algorithm (albeit its slow convergence), or the projected Broyden-Fletcher-Goldfarb-Shanno (BFGS). The later is one of the members of quasi-Newton methods. These algorithms can only be guaranteed to converge to a local optimum point. Moreover, as we can see later in the simulation results (Figure 5.6(a)), the resulting optimum points using the BFGS algorithm are relatively the same as those obtained from the exhaustive search.

The optimal design variables as function of weight w can be seen in Figure 5.6(a). In general, as we increase the weight w, the optimal ν is non-increasing while the optimal $\overline{T}_d$ is non-decreasing. We can divide w in Figure 5.6(a) into three regions: $w \leq w_{\text{up}}$, $w_{\text{up}} < w < w_{\text{dl}}$ and $w \geq w_{\text{dl}}$. We refer to these regions as region I, II and III respectively. We denote w_{up} as the upper limit of w where $\overline{T}_d^* = \overline{T}_{d,\text{min}}$ or equivalently, $\nu^* = \nu_{\text{max}}$, and w_{dl} as the lower limit of w where $\nu^* = \nu_{\text{min}}$ or $\overline{T}_d^* = \overline{T}_{d,\text{max}}$. In region I, the uplink (downlink) period is maximum (minimum) and ν^* is chosen to maximize the sum rate. We can say that in this region, the uplink sum rate strongly dominates the downlink sum rate. If we allow $\overline{T}_{\text{dl,min}} = 0$ then we only have the uplink transmission. On the contrary, in the region III, the downlink sum rate strongly dominates the uplink sum rate. It is also apparent from the figure that the optimal $\overline{T}_d$ is increasing while optimal ν is decreasing in region II. In this region, we see a trade-off between $\overline{T}_d$ and optimal ν which consequently affects the downlink and uplink sum rates.

It is also interesting to see how the regions change as we vary the downlink or uplink SNR. As shown in Figure 5.6(b), with a fixed uplink SNR, as the downlink SNR increases, region I is getting smaller while region III becomes larger. This result is intuitive because increasing downlink SNR implies that we put more weight (less weight) indirectly to the downlink (uplink) sum rate. For the middle region, it grows slightly.



(a) Optimal ν and $\overline{T}_d$ as function w at $\overline{\gamma}_d = 20$ dB.



(b) Region of ν^* and $\overline{T}_d^*$ based on the values of w and $\overline{\gamma}_d$.

Figure 5.6: Characterization of ν^* and $\overline{T}_d^*$ for $\overline{\gamma}_u = 0$ dB, $\overline{T} = 200$, $\nu_{\min} = 0.01$ and $\overline{T}_{d,\min} = 0.01$.

5.5 Conclusion

In this chapter, we perform the weighted sum rate maximization of the uplink and downlink transmissions in a TDD training-based MISO broadcast channel consisting of K single-antenna users and N antennas at the base station. The design parameters are the training period, the uplink and downlink data transmission periods, the training power and the uplink data transmission power. Considering only the downlink transmission, the optimal training period per user, $\overline{T}_\tau^*$, maximizing the downlink sum rate is larger than or equal to one. Increasing the training power will reduce $\overline{T}_\tau^*$ and after a certain threshold of the training power $\overline{T}_\tau^*$ becomes one symbol per user. Incorporating the uplink transmission and conducting the power allocation between the training and the uplink data transmission, we demonstrate that $\overline{T}_\tau^* = 1$. This holds for any values of other design parameters. For a fixed downlink transmission period, the weighted sum rate is a concave function of ν where ν controls the uplink power splitting between the training and data transmission. For a fixed downlink transmission period, the optimal downlink period and ν can be obtained, for an example, by using the alternating optimization scheme. We also show, by numerical simulations, the trade-off between the optimal power allocated to uplink data transmission and the optimal period of the downlink data transmission.

5.6 Appendix

5.6.1 Proof of Theorem 5.1

We start the proof by rewriting (5.6) based on the channel model (5.3) as follows

$$\text{SINR}_k = \frac{c^2[\hat{A}_k + \tilde{A}_k]^2}{c^2\left[B_k + 2(1+\hat{A}_k)\text{Re}\{Q_k\} + (1+\hat{A}_k)^2 R_k\right] + (1+\hat{A}_k)^2\sigma^2}, \tag{5.25}$$

where the terms $\hat{A}_k$, $\hat{B}_k$, Π_k, Q_k and R_k are defined as follows

$$\begin{aligned}
\hat{A}_k &:= \widehat{\mathbf{h}}_k(\widehat{\mathbf{H}}_k^H\widehat{\mathbf{H}}_k + \alpha\mathbf{I}_N)^{-1}\widehat{\mathbf{h}}_k^H \\
B_k &:= \widehat{\mathbf{h}}_k(\widehat{\mathbf{H}}_k^H\widehat{\mathbf{H}}_k + \alpha\mathbf{I}_N)^{-1}\widehat{\mathbf{H}}_k^H\widehat{\mathbf{H}}_k(\widehat{\mathbf{H}}_k^H\widehat{\mathbf{H}}_k + \alpha\mathbf{I}_N)^{-1}\widehat{\mathbf{h}}_k^H
\end{aligned}$$

$$\begin{aligned}
\tilde{A}_k &:= \widetilde{\mathbf{h}}_k(\widehat{\mathbf{H}}_k^H\widehat{\mathbf{H}}_k + \alpha\mathbf{I}_N)^{-1}\widehat{\mathbf{h}}_k^H \\
Q_k &:= \widehat{\mathbf{h}}_k\mathbf{O}_k\widehat{\mathbf{H}}_k^H\widehat{\mathbf{H}}_k\left\{\mathbf{O}_k - \frac{\mathbf{O}_k\widehat{\mathbf{h}}_k^H\widehat{\mathbf{h}}_k\mathbf{O}_k}{1+\hat{A}_k}\right\}\widetilde{\mathbf{h}}_k^H \\
R_k &:= \widetilde{\mathbf{h}}_k\left(\mathbf{O}_k - \frac{\mathbf{O}_k\widehat{\mathbf{h}}_k^H\widehat{\mathbf{h}}_k\mathbf{O}_k}{1+\hat{A}_k}\right)\widehat{\mathbf{H}}_k^H\widehat{\mathbf{H}}_k\left(\mathbf{O}_k - \frac{\mathbf{O}_k\widehat{\mathbf{h}}_k^H\widehat{\mathbf{h}}_k\mathbf{O}_k}{1+\hat{A}_k}\right)\widetilde{\mathbf{h}}_k^H
\end{aligned}$$

with $\mathbf{O}_k = (\widehat{\mathbf{H}}_k^H\widehat{\mathbf{H}}_k + \tilde{\alpha}\mathbf{I}_N)^{-1}$. $\widehat{\mathbf{H}}_k$ is $\widehat{\mathbf{H}}$ with kth row removed. In the following, the large system limit for each term in (5.25) is derived.

1. $\hat{A}_k$: It can be rewritten as

$$\hat{A}_k = \frac{1}{N(1-\sigma_\tau^2)}\hat{\mathbf{h}}_k\left(\frac{1}{N(1-\sigma_\tau^2)}\hat{\mathbf{H}}_k^H\hat{\mathbf{H}}_k + \rho\mathbf{I}_N\right)^{-1}\hat{\mathbf{h}}_k^H,$$

   where $\rho = \frac{\alpha}{N(1-\sigma_\tau^2)}$. Here, the entries $\sqrt{1/N(1-\sigma_\tau^2)^{-1}}\widehat{\mathbf{h}}$ and $\sqrt{1/N(1-\sigma_\tau^2)^{-1}}\widehat{\mathbf{H}}$ are zero mean with variance $1/N$. Applying Lemma 2.4 leads to

$$\hat{A}_k - g(\beta,\rho) \xrightarrow{a.s.} 0,$$

   where g is defined in (2.10).

2. $\tilde{A}_k$: Similar to the above, we can rewrite this term as

$$\tilde{A}_k = \frac{1}{N(1-\sigma_\tau^2)}\widetilde{\mathbf{h}}_k\left(\frac{1}{N(1-\sigma_\tau^2)}\widehat{\mathbf{H}}_k^H\widehat{\mathbf{H}}_k + \rho\mathbf{I}_N\right)^{-1}\widehat{\mathbf{h}}_k^H.$$

   Since the entries $\sqrt{\frac{1}{N}\sigma_\tau^{-2}}\widetilde{\mathbf{h}}_k$ and $\sqrt{\frac{1}{N}(1-\sigma_\tau^2)^{-1}}\widehat{\mathbf{h}}_k$ are independent and zero mean with variance $1/N$, then by Lemma 2.3 we can show $\tilde{A}_k \xrightarrow{a.s.} 0$.

3. B_k: It can be rewritten as

$$\begin{aligned}
\hat{B}_k = &\frac{1}{N(1-\sigma_\tau^2)}\widehat{\mathbf{h}}_k\left(\frac{1}{N(1-\sigma_\tau^2)}\widehat{\mathbf{H}}_k^H\widehat{\mathbf{H}}_k + \rho\mathbf{I}_N\right)^{-1}\left(\frac{1}{N(1-\sigma_\tau^2)}\widehat{\mathbf{H}}_k^H\widehat{\mathbf{H}}_k\right) \\
&\times\left(\frac{1}{N(1-\sigma_\tau^2)}\widehat{\mathbf{H}}_k^H\widehat{\mathbf{H}}_k + \rho\mathbf{I}_N\right)^{-1}\widehat{\mathbf{h}}_k^H.
\end{aligned}$$

   It is similar to the expression of B_k in (3.27). Thus, we can show

$$B_k - \left(g(\beta,\rho) + \rho\frac{\partial g(\beta,\rho)}{\partial\rho}\right) \xrightarrow{a.s.} 0.$$

4. Q_k: We can rewrite this term as

$$\begin{aligned} Q_k &= \widehat{\mathbf{h}}_k \mathbf{O}_k \widehat{\mathbf{H}}_k^H \widehat{\mathbf{H}}_k \mathbf{O}_k \widetilde{\mathbf{h}}^H - \frac{B_k \tilde{A}_k^*}{1+\hat{A}_k} \\ &= Q_k^{(1)} - \frac{B_k \tilde{A}_k^*}{1+\hat{A}_k}. \end{aligned}$$

By using Lemma 2.3, it follows that

$$\sqrt{\frac{1}{N(1-\sigma_\tau^2)}} \widehat{\mathbf{h}}_k \mathbf{O}_k \widehat{\mathbf{H}}_k^H \widehat{\mathbf{H}}_k \overline{\mathbf{O}}_k \left(\sqrt{\frac{1}{N\sigma_\tau^2}} \widetilde{\mathbf{h}}^H \right) \xrightarrow{a.s.} 0,$$

where $\overline{\mathbf{O}}_k = N(1-\sigma_\tau^2)\mathbf{O}_k$. Then, it follows easily that $Q_k^{(1)} \xrightarrow{a.s.} 0$. Previously, we have shown $\tilde{A}_k \xrightarrow{a.s.} 0$. Thus, $Q_k \xrightarrow{a.s.} 0$.

5. R_k: It can be written as

$$\begin{aligned} R_k &= \widetilde{\mathbf{h}} \mathbf{O}_k \widehat{\mathbf{H}}_k^H \widehat{\mathbf{H}}_k \mathbf{O}_k \widetilde{\mathbf{h}} - \frac{2\Re\left[Q_k^{(1)} \tilde{A}_k^*\right]}{1+\hat{A}_k} + \frac{B_k |\tilde{A}_k|^2}{(1+\hat{A}_k)^2} \\ &= \frac{\sigma_\tau^2}{1-\sigma_\tau^2} \left(\frac{1-\sigma_\tau^2}{\sigma_\tau^2} \widetilde{\mathbf{h}} \mathbf{O}_k \widehat{\mathbf{H}}_k^H \widehat{\mathbf{H}}_k \overline{\mathbf{O}}_k \widetilde{\mathbf{h}} \right) - 2\frac{\Re\left[Q_k^{(1)} \tilde{A}_k^*\right]}{1+\hat{A}_k} + \frac{B_k |\tilde{A}_k|^2}{(1+\hat{A}_k)^2} \\ &= \frac{\sigma_\tau^2}{1-\sigma_\tau^2} R_k^{(1)} - 2\frac{\Re\left[Q_k^{(1)} \tilde{A}_k^*\right]}{1+\hat{A}_k} + \frac{B_k |\tilde{A}_k|^2}{(1+\hat{A}_k)^2}. \end{aligned}$$

By using the previous large system results, the second term and the third term of R_k converge almost surely to 0. It can be checked that $R_k^{(1)}$ has the same structure as B_k. Then, it follows that, $R_k^{(1)} - \left(g(\beta,\rho) + \rho \frac{\partial g(\beta,\rho)}{\partial \rho} \right) \xrightarrow{a.s.} 0$. Thus,

$$R_k - \frac{\sigma_\tau^2}{1-\sigma_\tau^2} \left(g(\beta,\rho) + \rho \frac{\partial g(\beta,\rho)}{\partial \rho} \right) \xrightarrow{a.s.} 0.$$

6. c^2: We can write this term as

$$c^2 = \frac{P_d(1-\sigma_\tau^2)}{\frac{1}{N}\mathrm{Tr}\left(\frac{1}{N(1-\sigma_\tau^2)} \widehat{\mathbf{H}}^H \widehat{\mathbf{H}} \left(\frac{1}{N(1-\sigma_\tau^2)} \widehat{\mathbf{H}}^H \widehat{\mathbf{H}} + \rho \mathbf{I}_N \right)^{-2} \right)}.$$

Following the same step as in deriving the large system limit for B_k in (3.27), the

denominator of the equation above converges almost surely to $g(\beta,\rho)+\rho\frac{\partial g(\beta,\rho)}{\partial\rho}$. Thus,

$$c^2-\frac{P_d(1-\sigma_\tau^2)}{g(\beta,\rho)+\rho\frac{\partial g(\beta,\rho)}{\partial\rho}}\xrightarrow{a.s.}0.$$

Combining the large system results above, (5.10) follows immediately and this concludes the proof.

5.6.2 The Concavity of R_d^∞ over $\overline{T}_\tau$

We will prove the concavity of the sum rate per antenna over $\overline{T}_\tau$ in two-phase training by showing that the second derivative of R_d^∞ (over $\overline{T}_\tau$) is nonpositive for all values of $\overline{T}_\tau$ ($1\le\overline{T}_\tau\le T$). For notational simplicity, we denote $g=g(\cdot,\cdot)$ as just g, $x'=\partial x/\partial\overline{T}_\tau$, $x''=\partial^2x/\partial\overline{T}_\tau^2$ and $\rho=\rho_d$.

The first derivative of R_d^∞ over $\overline{T}_\tau$ is given by

$$\frac{\partial R_d^\infty}{\partial\overline{T}_\tau}=-\frac{\beta}{\overline{T}}\log(1+g)+\beta\left(1-\frac{\overline{T}_\tau}{\overline{T}}\right)\frac{g'}{1+g},\tag{5.26}$$

where

$$g'=-g\frac{\rho'(1+g)^2}{\beta+\rho(1+g)^2}.$$

We can check that $g'>0$ since

$$\rho'=-\frac{\beta}{\overline{\gamma}_\tau\overline{T}_\tau^2}\left(1+\frac{1}{\gamma_d}\right)<0.$$

The stationary point can be obtained by setting the derivative equal to zero. The second derivative of R_d^∞ is given by

$$\frac{\partial^2R_d^\infty}{\partial\overline{T}_\tau^2}=-\frac{2\beta}{\overline{T}}\frac{g'}{1+g}+\beta\left(1-\frac{\overline{T}_\tau}{\overline{T}}\right)\left[\frac{g''(1+g)-(g')^2}{(1+g)^2}\right].\tag{5.27}$$

The first term in (5.27) is always positive. Hence, to prove the concavity of R_d^∞ we only need to show that $g''\le0$. After some algebraic manipulations, g'' can be expressed

as

$$\begin{aligned}
\frac{g''}{g} &= -\frac{2\beta\rho'(1+g)g'}{[\beta+\rho(1+g)^2]^2} - \frac{\rho''(1+g)^2}{\beta+\rho(1+g)^2} + 2\frac{(g')^2}{g^2} \\
&\overset{a}{=} 2\beta\left(\frac{-\rho'(1+g)}{\beta+\rho(1+g)^2}\right)\left(\frac{g'}{\beta+\rho(1+g)^2}\right) - 2\frac{g'}{g\overline{\overline{T}}_\tau} + 2\frac{(g')^2}{g^2} \\
&= 2\left(\frac{g'}{g(1+g)}\right)\left(\frac{g'}{1+\frac{1}{\beta}\rho(1+g)^2}\right) - 2\frac{g'}{g\overline{\overline{T}}_\tau} + 2\frac{(g')^2}{g^2} \\
&\leq 2\frac{g'}{g}\left(\frac{g'}{g}\left(1+\frac{1}{1+z}\right) - \frac{1}{\overline{\overline{T}}_\tau}\right) \\
&\overset{b}{\leq} 2\frac{g'}{g}\left(\frac{1}{\overline{\overline{T}}_\tau}\left(1-\frac{1}{1+z}\right)\left(1+\frac{1}{1+z}\right) - \frac{1}{\overline{\overline{T}}_\tau}\right) \\
&= 2\frac{g'}{g}\left(\frac{1}{\overline{\overline{T}}_\tau}\left(1-\frac{1}{(1+z)^2}\right) - \frac{1}{\overline{\overline{T}}_\tau}\right) \\
&\overset{c}{<} 0
\end{aligned}$$

where

$$z = \frac{1}{\beta}\rho(1+g)^2 = \frac{1}{\gamma_d}\left(\frac{\gamma_d+1}{\overline{\gamma}_\tau\overline{\overline{T}}_\tau}+1\right)(1+g)^2$$

and (a) is obtained by using the relation

$$\rho'' = -\frac{2}{\overline{\overline{T}}_\tau}\rho'$$

and (b) from

$$\begin{aligned}
\frac{g'}{g} &= -\frac{\rho'(1+g)^2}{\beta+\rho(1+g)^2} \\
&= \frac{1}{\overline{\overline{T}}_\tau}\frac{\frac{1}{\gamma_d}\left(\frac{\gamma_d+1}{\overline{\gamma}_\tau\overline{\overline{T}}_\tau^2}\right)(1+g)^2}{1+\frac{1}{\gamma_d}\left(\frac{\gamma_d+1}{\overline{\gamma}_\tau\overline{\overline{T}}_\tau}+1\right)(1+g)^2} \\
&\leq \frac{1}{\overline{\overline{T}}_\tau}\frac{\frac{1}{\gamma_d}\left(\frac{\gamma_d+1}{\overline{\gamma}_\tau\overline{\overline{T}}_\tau}+1\right)(1+g)^2}{1+\frac{1}{\gamma_d}\left(\frac{\gamma_d+1}{\overline{\gamma}_\tau\overline{\overline{T}}_\tau}+1\right)(1+g)^2} \\
&= \frac{1}{\overline{\overline{T}}_\tau}\frac{z}{1+z} = \frac{1}{\overline{\overline{T}}_\tau}\left(1-\frac{1}{1+z}\right)
\end{aligned}$$

and (c) from the fact that

$$1 - \frac{1}{(1+z)^2} < 1.$$

Since $g'' < 0$, therefore R_d^{∞} is strictly concave over $\overline{T}_\tau$.

5.6.3 Proof of Lemma 5.1

In the following, we denote $x' = \partial x / \partial \overline{T}_\tau$. To prove the lemma, we just need to show $\partial R_u^{\infty} / \partial \overline{T}_\tau < 0, \forall \overline{T}_\tau$. The derivative is given by

$$\frac{\partial R_u^{\infty}}{\partial \overline{T}_\tau} = -\frac{\beta}{\overline{T}} \log(1 + g(\beta, \rho_u)) + \beta \frac{\overline{T}_{\text{up}} - \overline{T}_\tau}{\overline{T}} \frac{g'(\beta, \rho_u)}{1 + g(\beta, \rho_u)}.$$

Since $g(\beta, \rho_u) > 0$ and

$$\rho_u' = -\frac{\beta}{\nu \overline{\gamma}_{\text{up}} \overline{T}_{\text{up}}} \left(1 + \frac{1}{(1-\nu)\overline{\gamma}_{\text{up}} \overline{T}_{\text{up}}} \right) < 0,$$

then it follows that

$$\frac{g'(\beta, \rho_u)}{g(\beta, \rho_u)} = -\frac{\rho_u' (1 + g(\beta, \rho_u))^2}{\beta + \rho_u (1 + g(\beta, \rho_u))^2} > 0.$$

Continuing on, we have

$$\begin{aligned} \frac{\partial R_u^{\infty}}{\partial \overline{T}_\tau} &= \frac{\beta}{\overline{T}} \left(-\log(1 + g(\beta, \rho_u)) + (\overline{T}_{\text{up}} - \overline{T}_\tau) \frac{g'(\beta, \rho_u)}{g(\beta, \rho_u)} \frac{g(\beta, \rho_u)}{1 + g(\beta, \rho_u)} \right) \\ &\overset{(a)}{<} \frac{\beta}{\overline{T}} \left(-\log(1 + g(\beta, \rho_u)) + \frac{g(\beta, \rho_u)}{1 + g(\beta, \rho_u)} \right) \\ &\overset{(b)}{<} 0. \end{aligned}$$

The line (a) is obtained by using the following result

$$\begin{aligned} (\overline{T}_{\text{up}} - \overline{T}_\tau) \frac{g'(\beta, \rho_u)}{g(\beta, \rho_u)} &< \frac{\overline{T}_{\text{up}} \beta}{\nu \overline{\gamma}_{\text{up}} \overline{T}_{\text{up}}} \left(1 + \frac{1}{(1-\nu)\overline{\gamma}_{\text{up}} \overline{T}_{\text{up}}} \right) \frac{(1 + g(\beta, \rho_u))^2}{\beta + \rho_u (1 + g(\beta, \rho_u))^2} \\ &< 1, \end{aligned}$$

where the last line is obtained by using the fact that

$$\frac{\beta}{\nu\overline{\gamma}_{\text{up}}}\left(1+\frac{1}{(1-\nu)\overline{\gamma}_{\text{up}}\overline{T}_{\text{up}}}\right)<\rho_u.$$

The line (b) is easily obtained by using the well-known inequality

$$\log(1+x)\geq\frac{x}{1+x},$$

and this line concludes the proof.

5.6.4 The Concavity of $g(\beta,\rho_u)$ over ν

Here, we denote $x'=\partial x/\partial\rho_u$, $x''=\partial^2 x/\partial\rho_u^2$ and $g=g(\beta,\rho_u)$. In what follows, we show that $\partial^2 g/\partial\rho_u^2\leq 0$. The first and second derivatives of ρ_{up} over ν are respectively given by

$$\begin{aligned}
\rho_u' &= \frac{\beta}{(1-\nu)^2\overline{\gamma}_{\text{up}}\overline{T}_{\text{up}}}+\frac{\beta\overline{T}_u}{\nu\overline{\gamma}_{\text{up}}\overline{T}_{\text{up}}}\left(\frac{1}{(1-\nu)^2\overline{\gamma}_{\text{up}}\overline{T}_{\text{up}}}-\frac{1}{\nu}-\frac{1}{\nu(1-\nu)\overline{\gamma}_{\text{up}}\overline{T}_{\text{up}}}\right)\\
\rho_u'' &= \frac{2\beta}{(1-\nu)^3\overline{\gamma}_{\text{up}}\overline{T}_{\text{up}}}+\frac{\beta\overline{T}_u}{\nu^2\overline{\gamma}_{\text{up}}\overline{T}_{\text{up}}}\left(\frac{3\nu-1}{(1-\nu)^3\overline{\gamma}_{\text{up}}\overline{T}_{\text{up}}}+\frac{2}{\nu}-\frac{3\nu-2}{\nu(1-\nu)^2\overline{\gamma}_{\text{up}}\overline{T}_{\text{up}}}\right).
\end{aligned}$$

Furthermore, it is easy to show that ρ_u is a convex function over ν, since

$$\begin{aligned}
\rho_u'' &= \frac{2\beta}{(1-\nu)^3\overline{\gamma}_{\text{up}}\overline{T}_{\text{up}}}+\frac{\beta\overline{T}_u}{\nu^2\overline{\gamma}_{\text{up}}\overline{T}_{\text{up}}}\left(\frac{6\nu^2-6\nu+2}{\nu(1-\nu)^3\overline{\gamma}_{\text{up}}\overline{T}_{\text{up}}}+\frac{2}{\nu}\right)\\
&\geq 0.
\end{aligned}$$

The relation between ρ_u' and ρ_u can be expressed as

$$\begin{aligned}
\rho_u' &= \frac{1}{1-\nu}\rho_u-\frac{\beta\overline{T}_u}{\nu^2(1-\nu)\overline{\gamma}_{\text{up}}\overline{T}_{\text{up}}}\left(1+\frac{1}{\overline{\gamma}_{\text{up}}\overline{T}_{\text{up}}}\right)\\
&= \frac{1}{1-\nu}\rho_u-d(\nu) \qquad (5.28)
\end{aligned}$$

and between ρ_u'' and ρ_u' as

$$\begin{aligned}\rho_u'' &= \frac{2}{1-\nu}\rho_u' + \frac{2\beta\overline{T}_u}{\nu^3(1-\nu)\overline{\gamma}_{\text{up}}\overline{T}_{\text{up}}}\left(1+\frac{1}{\overline{\gamma}_{\text{up}}\overline{T}_{\text{up}}}\right)\\ &= \frac{2}{1-\nu}\rho_u' + c(\nu).\end{aligned} \tag{5.29}$$

The second derivative of g over ν is given by

$$\frac{g''}{g} = -\frac{2\beta\rho_u'(1+g)g'}{[\beta+\rho(1+g)^2]^2} - \frac{\rho_u''(1+g)^2}{\beta+\rho(1+g)^2} + 2\frac{(g')^2}{g^2}.$$

Continuing on, we have

$$\begin{aligned}\frac{g''}{g} &= 2\left(\frac{(g')^2}{g(1+g)}\right)\left(\frac{1}{1+\frac{1}{\beta}\rho(1+g)^2}\right) - \frac{\rho_u''(1+g)^2}{\beta+\rho(1+g)^2} + 2\frac{(g')^2}{g^2}\\ &\leq 2\frac{(g')^2}{g^2}\left(1+\frac{1}{1+\frac{1}{\beta}\rho(1+g)^2}\right) - \frac{\frac{1}{\beta}\rho_u''(1+g)^2}{1+\frac{1}{\beta}\rho(1+g)^2}\\ &= \frac{2(g')^2(2+\frac{1}{\beta}\rho(1+g)^2) - \frac{1}{\beta}\rho_u''g^2(1+g)^2}{g^2\left(1+\frac{1}{\beta}\rho(1+g)^2\right)}.\end{aligned} \tag{5.30}$$

The term $(g')^2$ is given by

$$(g')^2 = \frac{(\rho_u')^2g^2(1+g)^4}{(\beta+\rho(1+g)^2)^2},$$

and substituting it to (5.30), we have (after some algebraic manipulations)

$$\frac{g''}{g} \leq \frac{2g^2(1+g)^4(2(\rho_u')^2-\rho_u\rho_u'') + \frac{1}{\beta}\rho_u g^2(1+g)^6(2(\rho_u')^2-\rho_u\rho_u'') - \beta\rho_u''g^2(1+g)^2}{g^2\left(1+\frac{1}{\beta}\rho(1+g)^2\right)(\beta+\rho(1+g)^2)^2}.$$

Now, we will prove that $2(\rho_u')^2 - \rho_u\rho_u'' \leq 0$. By rewriting

$$\rho_u = (1-\nu)(\rho_u' + d(\nu))$$

we have

$$\rho_u\rho_u'' = 2(\rho_u')^2 + (1-\nu)c(\nu)(\rho_u' + d(\nu)) + 2d(\nu)\rho_u'.$$

Therefore,

$$2(\rho_u')^2 - \rho_u\rho_u'' = -(1-\nu)c(\nu)(\rho_u' + d(\nu)) - 2d(\nu)\rho_u'.$$

By substituting $\rho_u' = \frac{1}{1-\nu}\rho_u - d(\nu)$ to the above equation, we get

$$\begin{aligned}
2(\rho_u')^2 - \rho_u\rho_u'' &= -\rho_u c(\nu) - 2\frac{\rho_u d(\nu)}{1-\nu} + 2d^2(\nu) \\
&= -\rho_u c(\nu) - 2d(\nu)\left(\frac{\rho_u}{1-\nu} - d(\nu)\right) \\
&= -\rho_u c(\nu) - 2d(\nu)\left(\frac{\overline{T}_u}{\nu^2(1-\nu)\overline{\gamma}_{\text{up}}\overline{T}_{\text{up}}}\left(\nu + \frac{\nu}{(1-\nu)\overline{\gamma}_{\text{up}}\overline{T}_{\text{up}}} - 1 - \frac{1}{\overline{\gamma}_{\text{up}}\overline{T}_{\text{up}}}\right)\right. \\
&\quad \left. + \frac{x(\nu)}{1-\nu}\right) \\
&= -\left(\frac{\beta\overline{T}_u}{\overline{\gamma}_{\text{up}}\overline{T}_{\text{up}}}\right)^2 \frac{2}{\nu^4(1-\nu)^2}\left(1-\nu+\frac{1}{\overline{\gamma}_{\text{up}}\overline{T}_{\text{up}}}\right)\left(1+\frac{1}{\overline{\gamma}_{\text{up}}\overline{T}_{\text{up}}}\right) \\
&\quad - \left(\frac{\beta\overline{T}_u}{\overline{\gamma}_{\text{up}}\overline{T}_{\text{up}}}\right)^2 \frac{2}{\nu^4(1-\nu)^2}\left(\nu + \frac{\nu}{(1-\nu)\overline{\gamma}_{\text{up}}\overline{T}_{\text{up}}} - 1 - \frac{1}{\overline{\gamma}_{\text{up}}\overline{T}_{\text{up}}}\right) \\
&\quad \times \left(1+\frac{1}{\overline{\gamma}_{\text{up}}\overline{T}_{\text{up}}}\right) - 2x(\nu)\left(c(\nu) + \frac{d(\nu)}{1-\nu}\right) \\
&= -\left(\frac{\beta\overline{T}_u}{\overline{\gamma}_{\text{up}}\overline{T}_{\text{up}}}\right)^2 \frac{2}{\nu^4(1-\nu)^2}\left(\nu + \frac{\nu}{(1-\nu)\overline{\gamma}_{\text{up}}\overline{T}_{\text{up}}}\right)\left(1+\frac{1}{\overline{\gamma}_{\text{up}}\overline{T}_{\text{up}}}\right) \\
&\quad - 2x(\nu)\left(c(\nu) + \frac{d(\nu)}{1-\nu}\right) \\
&\leq 0,
\end{aligned}$$

where

$$x(\nu) = \frac{\beta}{(1-\nu)\overline{\gamma}_{\text{up}}\overline{T}_{\text{up}}} \geq 0.$$

From (5.28) and (5.29) it is obvious that $c(\nu)$, $d(\nu)$ and ρ_u'' are nonnegative. Thus, it is straightforward to see that $g'' \leq 0$ and therefore $g(\beta, \rho_u)$ is a concave function over ν. This completes the proof.

Chapter 6

Optimal Power Allocation for Multiuser Precoding via RCI

In this chapter, we consider an optimal power allocation problem that maximizes the sum rate of a single-cell MISO broadcast channel with the RCI precoder at the base station. Unlike the channel inversion or zero forcing beamforming, the optimal power allocation with RCI precoder is a nonlinear non-convex optimization problem with many local optima. Here, we investigate this problem in the large system limit. We assume that user k has data symbols transmitted with power p_k, and slow-varying path-loss a_k. Using results from previous chapters, we obtain the expression for the limiting SINR. Then, we divide all K users into L groups where L is finite, and all users in each group are assumed to be co-located or to have approximately the same distance from the BS. In other words, all users in one group have the same path-loss which is distance dependent. Based on this system model, we investigate optimal power allocation schemes and optimal regularization parameter of the RCI that maximize the limiting sum rate per antenna under an average power constraint. We also study a multi-mode transmission where the number of groups that the BS communicates to may change as the system parameters such as groups' cell loadings also change.

6.1 Introduction

IN previous chapters we have assumed that all users are equidistant from the BS and therefore have the same distance-dependent path-loss gain. Moreover, equal power allocation across all users (data symbols) is also assumed. In this chapter, we allow power allocation across the data symbols where the users may have different path-losses. Unlike the channel inversion or zero forcing beamforming, the optimal power allocation for the RCI precoder is a non-convex optimization problem with many local optima [35,105,106], even in the case of all users having the same path-losses. In [105,106], the authors inves-

tigated the sum rate maximization of MIMO broadcast channels with RCI under a total power constraint. They showed that the problem is a global d.c. optimization problem and proposed the local gradient method to solve the problem. Their numerical results suggests that employing an RCI precoder with power allocation give a better sum rate compared to ZF. Reference [35] extends the previous works, but in the MISO broadcast channels setting, by putting additional quality of service (QoS) constraints where each user's data rate should be above a specified minimum rate. The authors re-casted the optimization problem as a series of geometric programming (GP) problems, called iterative GP (IGP), and solved it iteratively.

Besides power allocation, selecting the users for transmission can improve the system performance. A user selection scheme, called semi-orthogonal user selection (SUS), is proposed in [107]. The main idea behind using SUS is to increase the effective channel gain which is equivalent to c_1^2 in (3.8). It has been shown in [107] that a combination of water-filling based power allocation scheme and SUS in MISO BC systems with ZF precoder can achieve the sum rate obtained by employing DPC when the number of users is large. A similar conclusion is also presented in [18, 102] but using greedy search algorithms for the user selection. The performance analysis of that algorithm for the case of finite (at most two) scheduled users was carried out in [66]. Besides for ZF, the authors in [84] also proposed greedy user selection based on the closed form approximation of the expected sum rate for the RCI precoder. Dai et al. in [16] studied MISO BC systems with ZF precoder under finite-rate or quantized feedback. The proposed power allocation scheme is binary or on/off. They showed that the feedback rate and received SNR affect the optimal number of active ('on') users. Moreover, their scheme can be applied in heterogeneous environments where the users may have different path-losses. A similar problem is also considered in [112, 113] but with different settings. Besides considering finite-rate feedback, the authors take into account the feedback delay by using the Gauss-Markov model. The authors also assume a homogeneous environment and an equal power allocation across the users. A sum rate approximation expression as a function of the number of users is derived. Then, the number of users can be adjusted adaptively based on the feedback delay and channel quantization error (or feedback rate). This

strategy is similar to the multi-mode transmission, considered later.

The first part of the work in this chapter considers the joint optimization of the power allocation and regularization parameter of the RCI precoder in MISO BCs with heterogeneous users. The analysis is performed in the large system limit. First, we derive the users' limiting SINRs and show that the limiting SINR for each user differs in the power allocated to that user and user's path-losses. Then we divide all users into a finite number of groups, where all users in each group have approximately the same distance from the base station and therefore share the same distance dependent path-loss. Based on this model description, for a fixed regularization parameter, we show that the optimal power allocation problem under the average power constraint that maximizes the limiting sum rate per antenna is *convex* and the *power allocation follows the familiar water-filling strategy* [27, 107]. Similar results are also obtained in [101] but with different approaches in the large system analysis. We derive the optimal regularization parameter by substituting back the power allocation scheme to the limiting sum rate expression. Even though a closed form expression is not obtained, this substitution yields a one dimensional optimization problem which can be solved by existing line search algorithms.

In the second part, we consider a multi-mode transmission where for a given total number of groups (L) and cell loading for each group we determine the optimal number of groups for transmission and also which groups the BS should communicate with. We arrange or sort the groups based on their path-losses in a descending order. We investigate two cases. In the first case, for each group, the BS can only decide between transmitting to all the users in the group or to none of them. In the second case, the BS is allowed to communicate with any subset of the users in a group. For the first case with uniform cell-loading over the groups, it is optimal for the BS to transmit to the first $m \leq L$ groups in the mode m transmission. The optimal mode can be determined by comparing the maximum limiting sum rate of each mode. For the second case, we provide a necessary condition for the optimal cell-loading allocation for each group. Assuming that $M \leq L$ groups are allocated positive power, the cell loadings of the first $M-1$ groups should be set at their maximum value and the cell loading for the M-th group can be in between zero and its maximum value. We also propose an algorithm to solve the opti-

mization problem. Considering the cell loading allocation, the algorithm offers a lower complexity in comparison to a brute force search method. In both cases, the optimal power allocation and regularization parameter are also considered.

The rest of this chapter is structured as follows. The next section describes the finite size system model and presents the corresponding SINR expression. A brief derivation of the limiting SINR is also presented. Section 6.3 discusses a joint optimization of the allocated power across the users and the optimal regularization of the RCI. It is followed by the investigation on the multi-mode transmissions in Section 6.4. Section 6.5 concludes the chapter.

6.2 System Model

6.2.1 Finite-size System Model

The descriptions and notations for the received signal model largely follow those in Chapter 3. Now, by allowing power allocation across the users, we model the data symbol as $\mathbf{s} = \mathbf{\Lambda}^{1/2}\bar{\mathbf{s}}$, where $\bar{\mathbf{s}}$ is the normalized (power) data symbol , i.e., $\mathbb{E}[\bar{\mathbf{s}}\bar{\mathbf{s}}^H] = \mathbf{I}_K$. Let $\mathbf{\Lambda} = \text{diag}(p_1, p_2, \cdots, p_K)$ where p_k denotes the power allocated to user k. The transmitted data vector can written as $\mathbf{x} = \mathbf{P}\mathbf{\Lambda}^{1/2}\bar{\mathbf{s}}$ and has a power constraint $\mathbb{E}[\|\mathbf{x}\|_2^2] = P_d$. The received signal for user k is given by

$$y_k = a_k \mathbf{h}_k \mathbf{x} + w_k,$$

where a_k^2 denotes the slow-varying path-loss between the base station and the receiver of user k. The received signal vector then can be written as

$$\mathbf{y} = \mathbf{A}\mathbf{H}\mathbf{x} + \mathbf{w} = \mathbf{A}\mathbf{H}\mathbf{P}\mathbf{s} + \mathbf{w},$$

where $\mathbf{A} = \text{diag}(a_1, a_2, \cdots, a_K)$. The RCI precoder takes the form in (3.3). The normalizing constant c is chosen to meet the transmit power constraint $\mathbb{E}[\|\mathbf{x}\|_2^2] = \mathbb{E}[\text{Tr}\left(\mathbf{x}\mathbf{x}^H\right)] =$

P_d. Hence, c^2 can be expressed as

$$c^2 = \frac{P_d}{\text{Tr}\left(\mathbf{\Lambda}\mathbf{H}(\mathbf{H}^H\mathbf{H} + \alpha\mathbf{I}_N)^{-2}\mathbf{H}^H\right)}. \tag{6.1}$$

Note that (6.1) is (3.5) with the additional $\mathbf{\Lambda}$ in the trace. The received signal vector $\mathbf{y}$ is then given by

$$\mathbf{y} = c\mathbf{A}\mathbf{H}(\mathbf{H}^H\mathbf{H} + \alpha\mathbf{I}_N)^{-1}\mathbf{H}^H\mathbf{\Lambda}^{1/2}\bar{\mathbf{s}} + \mathbf{w}.$$

The corresponding received signal for user k can be easily shown to be

$$\begin{aligned} y_k = {} & ca_k\sqrt{p_k}\mathbf{h}_k(\mathbf{H}^H\mathbf{H} + \alpha\mathbf{I}_N)^{-1}\mathbf{h}_k^H\bar{s}_k \\ & + \sum_{j\neq k}^{K} ca_k\sqrt{p_j}\mathbf{h}_k(\mathbf{H}^H\mathbf{H} + \alpha\mathbf{I}_N)^{-1}\mathbf{h}_j^H\bar{s}_j. \end{aligned}$$

Based on the expression above, the SINR attained by user k can be expressed as follows

$$\text{SINR}_k = \frac{c^2a_k^2p_k|\mathbf{h}_k(\mathbf{H}^H\mathbf{H} + \alpha\mathbf{I}_N)^{-1}\mathbf{h}_k^H|^2}{\sum_{j\neq k}^{K} c^2a_k^2p_j|\mathbf{h}_k(\mathbf{H}^H\mathbf{H} + \alpha\mathbf{I}_N)^{-1}\mathbf{h}_j^H|^2 + \sigma^2}. \tag{6.2}$$

It is clear that the SINR_k is a random quantity since the propagation channels fluctuate randomly. In the large system limit, as we see in the next section, this randomness disappears.

6.2.2 Large System Regime SINR

In this section, we will derive the limiting SINR for (6.2). The large system limit derivation for each term of (6.2) borrows the techniques and results from previous chapters, particularly Chapter 3. In what follows, we will show that SINR_k converges almost surely to a deterministic quantity denoted by SINR_k^{∞}.

First, as in Appendix 3.6.1, we can write $S_k = |\mathbf{h}_k(\mathbf{H}^H\mathbf{H} + \alpha\mathbf{I}_N)^{-1}\mathbf{h}_k^H|^2 = \frac{A_k^2}{(1+A_k)^2}$, where $A_k = \frac{1}{N}\mathbf{h}_k\left(\frac{1}{N}\mathbf{H}_k^H\mathbf{H}_k + \rho\mathbf{I}_N\right)^{-1}\mathbf{h}_k^H$ with $\rho = \alpha/N$. It has been shown in that appendix that A_k converges almost surely to $g(\beta,\rho)$. We can write $\sum_{j\neq k}^{K} p_j|\mathbf{h}_k(\mathbf{H}^H\mathbf{H} +$

$\alpha\mathbf{I}_N)^{-1}\mathbf{h}_j^H|^2$ in the denominator as

$$\mathbf{h}_k(\mathbf{H}^H\mathbf{H}+\alpha\mathbf{I}_N)^{-1}\mathbf{H}_k^H\mathbf{\Lambda}_k\mathbf{H}_k(\mathbf{H}^H\mathbf{H}+\alpha\mathbf{I}_N)^{-1}\,,$$

where $\mathbf{\Lambda}_k$ is $\mathbf{\Lambda}$ with k-th column and row removed. By following the steps in Appendix 3.6.1, we can show that it converges almost surely to

$$\frac{\mathcal{P}}{(1+g(\beta,\rho))^2}\left(g(\beta,\rho)+\rho\frac{\partial}{\partial\rho}g(\beta,\rho)\right)\,,$$

where we define $\mathcal{P}=\lim_{K\to\infty}1/K\sum_{k=1}^{K}p_k$. $\mathcal{P}$ can be interpreted as the empirical mean of the users' power or just average power. Here, we assume that the limit $\mathcal{P}$ exists and is bounded. For c^2, we can also show that it converges almost surely to

$$\frac{P_d/\mathcal{P}}{g(\beta,\rho)+\rho\frac{\partial}{\partial\rho}g(\beta,\rho)}.$$

Therefore, the interference energy converges almost surely to $P_d a_k^2(1+g(\beta,\rho))^{-2}$.

Combining the large system limit results above, the SINR of user k then converges almost surely to

$$\begin{aligned}\text{SINR}_k^{\infty}&=\bar{p}_k g(\beta,\rho)\frac{\gamma_k+\frac{\gamma_k\rho}{\beta}(1+g(\beta,\rho))^2}{\gamma_k+(1+g(\beta,\rho))^2} &(6.3)\\ &=\bar{p}_k f_k(\beta,\rho), &(6.4)\end{aligned}$$

where $\gamma_k=\frac{P_d a_k^2}{\sigma^2}$ is defined as the effective SNR and $\bar{p}_k=\frac{p_k}{\mathcal{P}}$ is the normalized power w.r.t. $\mathcal{P}$. Different from the limiting SINR in previous chapters, here we can see that the limiting SINR is different for each user and depends on a_k and p_k. From (6.4), one can see that f_k is independent of $\bar{p}_k$. This property will ease the analysis in finding the optimal power allocation maximizing (limiting) sum rate discussed in the next section. It is also easy to see that assuming an equal power allocation across the user, e.g., $\bar{p_k}=1,\forall k$, f_k represents the limiting SINR. As shown later, f_k is vital in determining the optimal power allocation.

6.3 Optimal Power Allocation and Regularization Parameter

Let us consider the following scenario. We divide all K users into a finite number L of groups. All users in each group are assumed to have the same path-loss. Here, we assume that $a_1 \geq a_2 \geq \cdots \geq a_L$. The number of users in group j is denoted by K_j, with $\sum_{j=1}^{L} K_j = K$. Since the path-loss and other parameters β, ρ as well as SNR are the same for all users in a group, then based on (6.3), we can reasonably assume that the power allocated to each user in that group is also the same. This assumption holds for the rest of this chapter.

Based on the scenario above, we can define the limiting achievable sum rate per antenna as follows

$$R_{\text{sum}}^{\infty} = \sum_{j=1}^{L} \beta_j \log\left(1 + \text{SINR}_j^{\infty}\right), \tag{6.5}$$

where $\beta_j = \frac{K_j}{N}$ denotes the cell-loading of group j. Our goal in this section is to find the optimal power allocation that maximizes R_{sum}^{∞}. Moreover, it is also interesting to explore how the regularization parameter of the RCI precoder adapts to different path-losses and also users' power. Denoting $\bar{\mathbf{p}} = [\bar{p}_1, \bar{p}_2, \cdots, \bar{p}_L]^T$, a joint optimization problem can be formulated as follows,

$$\mathbf{P1}: \quad \max_{\bar{\mathbf{p}} \succeq 0, \rho \geq 0} \quad R_{\text{sum}}^{\infty}$$
$$\text{s.t.} \quad \sum_{j=1}^{L} \beta_j \bar{p}_j \leq \beta \,,$$

where $\succeq$ denotes the element-wise inequality for vectors. Note that the first constraint in $\mathbf{P1}$ is the large system average power constraint. The second constraint ensures that the normalized powers are non-negative.

Before addressing the solution of $\mathbf{P1}$, we characterize the objective function as a function of $\bar{p}_j$. Let $R_{\text{sum},j}^{\infty} = \beta_j \log\left(1 + \text{SINR}_j^{\infty}\right)$ denote the sum rate for group j. It can be checked that it is an increasing function in p_j. Moreover, we can show that the following lemma holds.

Lemma 6.1. *The sum rate per antenna R_{sum}^{∞} is concave in $\bar{\mathbf{p}}$.*

Proof. The second derivative of the limiting SINR w.r.t. $\bar{p}_j$ is

$$\frac{\partial^2 \text{SINR}_j^{\infty}}{\partial \bar{p}_j^2} = -\frac{f_j^2(\beta,\rho)}{(1+\bar{p}_j f_j(\beta,\rho))^2} < 0.$$

This implies that SINR_j^{∞} is concave in $\bar{p}_j$. Since the $\log$ operation does not change the concavity, therefore $R_{\text{sum},j}^{\infty}$ is also concave in $\bar{p}_j$. Moreover, R_{sum}^{∞} is a linear combination of $R_{\text{sum},j}^{\infty}$ and this operation preserves the concavity. □

From the lemma above, we can see that for a fixed ρ, **P1** is a convex program because $-R_{\text{sum}}^{\infty}$ is convex in $\bar{\mathbf{p}}$ and the constraints are linear. For a fixed $\bar{\mathbf{p}}$, it has been shown in Chapter 3 that SINR_k^{∞} is not concave in ρ but quasi-concave. Since $\log$ is a non-decreasing function then $R_{\text{sum},j}^{\infty}$ is also quasi-concave (not concave) in ρ. Since a linear combination operation does not preserve the quasi-concavity, the sum rate needs not be quasi-concave.

Now, let us consider the Lagrangian for **P1**, as stated below

$$\mathcal{L} = \sum_{j=1}^{L} \beta_j (1+\bar{p}_j f_j(\beta,\rho)) - \lambda \sum_{j=1}^{L} \beta_j (\bar{p}_j - 1) + \mu_j \bar{p}_j + \kappa\rho, \tag{6.6}$$

where λ and μ_j are the Lagrange multipliers for the average power and non-negative power constraints, and κ is the Lagrange multiplier for the constraint $\rho \geq 0$. The associated Karush-Kuhn-Tucker (KKT) necessary conditions are given by

$$\frac{\partial \mathcal{L}}{\partial \rho} = \sum_{j=1}^{L} \frac{\beta_j \bar{p}_j}{1+\bar{p}_j f_j(\beta,\rho)} \frac{\partial f_j(\beta,\rho)}{\partial \rho} + \kappa = 0 \tag{6.7}$$

$$\frac{\partial \mathcal{L}}{\partial \bar{p}_j} = \beta_j \left(\frac{f_j}{1+\bar{p}_j f_j} - \lambda \right) + \mu_j = 0, \tag{6.8}$$

and

$$\lambda \sum_{j=1}^{L} \beta_j (\bar{p}_j - 1) = 0, \ \mu_j \bar{p}_j = 0, \tag{6.9}$$

$$\kappa\rho = 0, \tag{6.10}$$

for all $j = 1, \ldots, L$.

Recall that for a given ρ, **P1** reduces to a convex program. Thus, the KKT conditions (6.7) and (6.9) lead to the optimal power allocation strategy maximizing the limiting sum rate, as presented in the following theorem.

Theorem 6.1. *For a fixed ρ, the optimal power allocation for the optimization problem* **P1** *follows the water-filling (WF) scheme and is given by*

$$\bar{p}_j = \left[\frac{1}{\lambda} - \frac{1}{f_j(\beta,\rho)}\right]_+ \tag{6.11}$$

where $[x]_+ = \max(0, x)$. The constant (Lagrange multiplier) λ is the solution of

$$\sum_{j=1}^{L} \beta_j \bar{p}_j = \beta,$$

for which the average power constraint is satisfied with equality.

In the WF scheme above, $1/\lambda$ can be perceived as the water level. It determines how power is poured to each user and is based on the value of $f_j(\beta,\rho)$. Recall that the limiting SINR is given by $\bar{p}_j f_j(\beta,\rho)$. It can be checked that $f_j(\beta,\rho)$ is increasing in a_j. Thus, more power will be allocated for the users with better channels which can be represented by the path losses $\{a_j\}$. Note that in this case, fairness amongst users could be an issue since some users might have zero rate.

Since we assume $a_1 \geq a_2 \geq \cdots \geq a_L$, then $\bar{p}_1 \geq \bar{p}_2 \geq \cdots \geq \bar{p}_L$. Now let us assume that the first m groups have non-zero power. To determine λ, we just need to solve $\sum_{j=1}^{m} \beta_j \bar{p}_j = \beta$. Using $\bar{p}_j$ in (6.11), it is easy to show that

$$\lambda = \frac{\sum_{j=1}^{m} \beta_j}{\beta + \sum_{j=1}^{m} \frac{\beta_j}{f_j(\beta,\rho)}}. \tag{6.12}$$

The power allocated to group j is then given by

$$\bar{p}_j = \frac{\beta + \sum_{j=1}^{m} \frac{\beta_j}{f_j(\beta,\rho)}}{\sum_{j=1}^{m} \beta_j} - \frac{1}{f_j(\beta,\rho)}. \tag{6.13}$$

To determine m, we just need to find m such that $\bar{p}_m > 0$ and $\bar{p}_{m+1} \leq 0$.

Related to the KKT (stationary) conditions (6.21), we have (for a fixed $\bar{\mathbf{p}}$)

$$\frac{\partial R_{\text{sum}}^{\infty}}{\partial \rho} = \sum_{j=1}^{L} \frac{\beta_j \bar{p}_j}{1+\bar{p}_j f_j(\beta,\rho)} \frac{\partial f_j(\beta,\rho)}{\partial \rho},$$

where

$$\frac{\partial f_j(\beta,\rho)}{\partial \rho} = \frac{\gamma_j^2}{[\gamma_j+(1+g)^2]^2} 2g(1+g)\left(\frac{\rho}{\beta}-\frac{1}{\gamma_j}\right)\frac{\partial g}{\partial \rho} \tag{6.14}$$

$$= f_j^2(\beta,\rho)\frac{2\left(\frac{1}{g}+1\right)}{[1+\frac{\rho}{\beta}(1+g)^2]^2}\left(\frac{\rho}{\beta}-\frac{1}{\gamma_j}\right)\frac{\partial g}{\partial \rho}, \tag{6.15}$$

and g represents $g(\beta,\rho)$. Thus,

$$\frac{\partial R_{\text{sum}}^{\infty}}{\partial \rho} = \frac{2\left(\frac{1}{g}+1\right)}{[1+\frac{\rho}{\beta}(1+g)^2]^2}\sum_{j=1}^{L}\frac{\beta_j \bar{p}_j f_j^2(\beta,\rho)}{1+\bar{p}_j f_j(\beta,\rho)}\left(\frac{\rho}{\beta}-\frac{1}{\gamma_j}\right)\frac{\partial g}{\partial \rho}.$$

Recall that $\frac{\partial g}{\partial \rho} < 0$ (see (3.32)). Let $q_j = \frac{\rho}{\beta} - \frac{1}{\gamma_j}$. It is also obvious that q_j is decreasing in j. Thus, for $q_L > 0$, $\frac{\partial R_{\text{sum}}^{\infty}}{\partial \rho}$ is negative. This implies that $\frac{\partial R_{\text{sum}}^{\infty}}{\partial \rho}$ can not be zero for $\rho > \frac{\beta}{\gamma_L}$. For $q_1 < 0$, $\frac{\partial R_{\text{sum}}^{\infty}}{\partial \rho}$ is positive and consequently, can not be zero for $\rho < \frac{\beta}{\gamma_1}$. Therefore, the optimal ρ must be in the interval of

$$\frac{\beta}{\gamma_1} \le \rho \le \frac{\beta}{\gamma_L}. \tag{6.16}$$

When we only have one group then the expression for ρ^* is the same as the one obtained in Chapter 3. We can also remove the boundary point $\rho = 0 < \frac{\beta}{\gamma_1}$ since as previously discussed, $\frac{\partial R_{\text{sum}}^{\infty}}{\partial \rho} > 0$ at that point. Thus, from (6.21) with $\kappa = 0$ or by evaluating $\frac{\partial R_{\text{sum}}^{\infty}}{\partial \rho} = 0$, the optimal ρ must satisfy

$$\sum_{j=1}^{L}\frac{\beta_j \bar{p}_j f_j^2(\beta,\rho)}{1+\bar{p}_j f_j(\beta,\rho)}\left(\frac{\rho}{\beta}-\frac{1}{\gamma_j}\right) = 0. \tag{6.17}$$

By substituting (6.11) into (6.17), it is straightforward to see that (6.17) becomes a one-dimensional zero/root-finding problem Thus, the optimal ρ can be found by using existing line search algorithms for the interval given in (6.16).

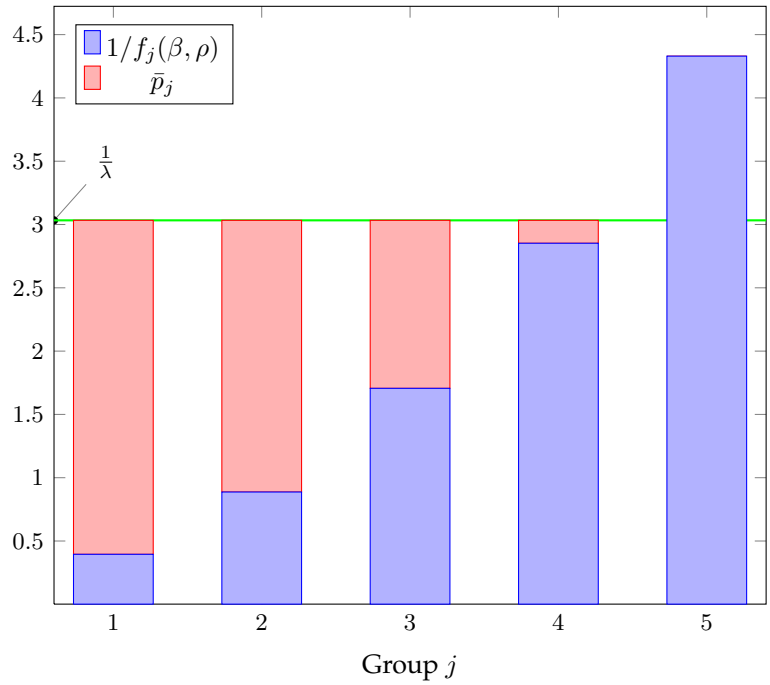


Figure 6.1: Power allocation scheme for $L = 5$, $\beta = 1$, $\beta_j = 1/5$, $\forall j = 1, \ldots, 5$, $P_d/\sigma^2 = 10$ dB and $a_j^2 = 1/j^2$.

Figure 6.1 illustrates the power allocation scheme based on $\mathbf{P1}$. In generating the plot, we set $L = 5$, $\beta = 1$, $\beta_j = 1/5$, $\forall j = 1, \ldots, 5$, $P_d/\sigma^2 = 10$ dB. The path-loss gain are set according to $a_j^2 = 1/j^2$. Our numerical simulation resulted $\frac{1}{\lambda} \approx 3.033$ and $\rho^* \approx 0.2265$. It can be checked that the latter is between $1/\gamma_1$ and $1/\gamma_4$, as predicted by the analysis. From the plot, we can see that the last group is allocated zero power since $1/f_5(\beta, \rho)$ exceeds the water level $1/\lambda$. This indicates that the users' channel in the group is not so good that allocating positive power to this group will not increase the sum rate. In Figure 6.2, we study the validity of using the large system results for the finite size system. We set the system parameters: $L = 2$, $\beta = 1$, $N = 8$, $\beta_j = 1/2$, $\rho = \rho^*$ and $a_j^2 = 1/j^2$. We generate 500 channel realizations and for each realization we compute the optimal power allocation, denoted by $\bar{\mathbf{p}}_{\text{FS}}^*$, by a grid search. In the plot, we compare the average sum rate, denoted by $\mathbb{E}[R_{\text{sum}}]$, between using the power allocation $\bar{\mathbf{p}}$ in (6.11) and $\bar{\mathbf{p}}_{\text{FS}}^*$. The gap between the curves in the figure is very small and can be said negligible.

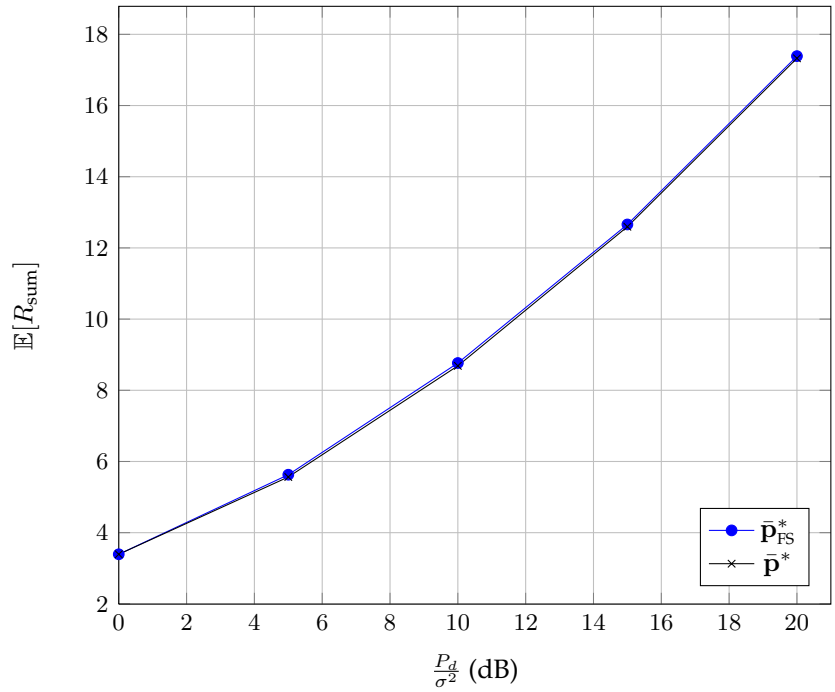


Figure 6.2: Comparison of the average sum rate between using $\bar{\mathbf{p}} = \bar{\mathbf{p}}^*_{\text{FS}}$ and $\bar{\mathbf{p}} = \bar{\mathbf{p}}^*$ for $L = 2$, $\beta = 1$, $N = 8$, $\beta_j = 1/2$, $\rho = \rho^*$ and $a_j^2 = 1/j^2$.

6.4 Multimode Broadcast Channels

In the previous sections, we consider an optimal power allocation that maximizes the (limiting) sum rate where the base station (BS) communicates to all L groups simultaneously. In that setting, we include the channels from all groups of users in the precoding design even though we allocate zero power to some of the groups. This leads us to ask: how about if we allow the BS to communicate to only some of the groups such as the groups that have users with positive power? Could this scenario give a higher sum rate? For an example, let us consider the case of $L = 3$. Let R be the radius of the cell. We assume the path-losses has the form $a_j^2 = 1/r_j^2$, where $r_j = jR/L$ for $j = 1, \ldots, L$ is the distance between the users in group j to the BS. We set the cell loading for each group to be uniform i.e., $\beta_j = \beta/L$. Figure 6.3 shows the limiting sum rate obtained when the BS communicates to only the first $m \leq L$ groups, denoted by $R_{\text{sum}}^{(m),\infty}$. This means that we only include the channel of the users from these m groups in the system model and in designing the precoder. We call this scheme as mode-m transmission.The figure demon-

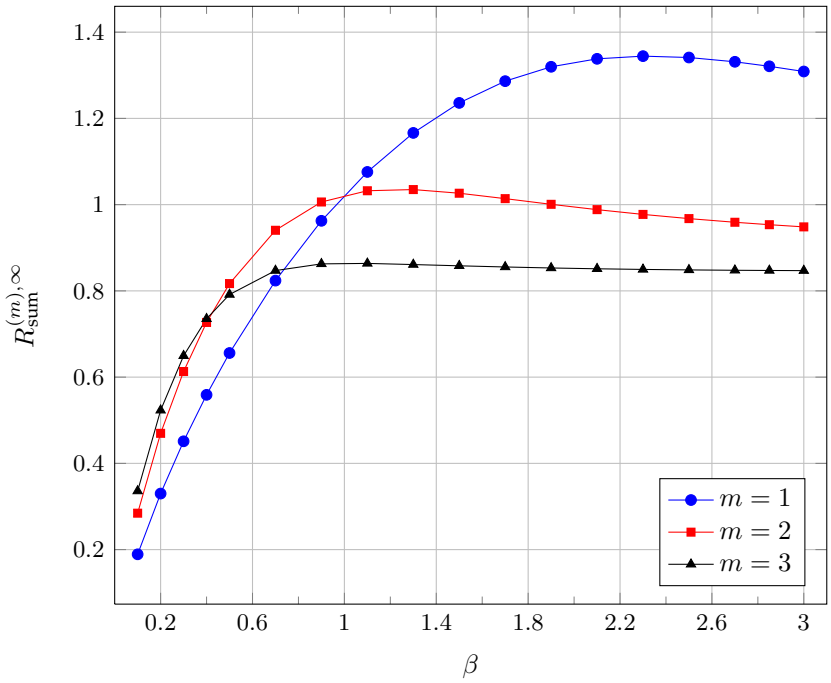


Figure 6.3: Multimode Transmission for $L = 3$ and $\beta_j = \beta/L$.

strates that for some values of cell-loading β, the maximum sum rate is achieved when $m < L$ groups. The optimal m is also changing with β. So we call this as the multi-mode transmission.

In the multimode transmission, it is clear that there are $\binom{L}{m}$ combinations of the groups that can be chosen by the base station to communicate to. The question is which mode and group combination will give the highest sum rate? To answer that question, we can formulate the following optimization problem,

$$\begin{aligned} \mathbf{P2}: \quad & \max_{\bar{\mathbf{p}} \succeq 0, \boldsymbol{\beta}, \rho, \beta \geq 0} \quad R_{\text{sum}}^{\infty} \\ & \text{s.t.} \quad \sum_{j=1}^{L} \beta_j \bar{p}_j \leq \beta, \\ & \quad \sum_{j=1}^{L} \beta_j = \beta, \\ & \quad \beta_j \in \{0, \beta_{j,\max}\}, \ \forall j = 1, 2, \ldots, L \ , \end{aligned}$$

where $\boldsymbol{\beta} = [\beta_1 \ \beta_2 \ \ldots \ \beta_L]^T$. So, **P2** is **P1** with additional design variables, $\boldsymbol{\beta}$ and β and

additional constraints related to them. In **P2**, β_j is only allowed to have value either 0 or $\beta_{j,\max}$. So β_j will determine whether the BS transmits to users in group j or not. The latter occurs when $\beta_j = 0$. In that case, the channel gain matrix of the users in group j is not included in the precoder design.

First, let us investigate the optimal strategies for **P2** when $\beta_{j,\max}$ is the same for all groups, i.e., $\beta_{j,\max} = \breve{\beta}$. Let us consider the mode-m transmission. In that case, we have m groups with $\beta_j = \breve{\beta}$ and the remaining groups have $\beta_j = 0$. Let $\mathcal{G} \subset \{1, 2, \ldots, L\}$, $|\mathcal{G}| = m$ be the set of the group indexes that the BS communicates to ($\beta_j > 0, j \in \mathcal{G}$). Then, the maximum limiting sum rate achieved for a given $\mathcal{G}$ can be obtained by solving

$$\begin{aligned} \max_{\bar{\mathbf{p}},\rho} \quad & R_{\text{sum}}^{(m),\infty}(\mathcal{G}) = \sum_{j\in\mathcal{G}} \beta_j \log(1 + \bar{p_j} f_j(\beta, \rho)) \\ \text{s.t.} \quad & \frac{1}{m}\sum_{j\in\mathcal{G}} \bar{p}_j \leq 1 \\ & \bar{p}_j \geq 0, \quad j \in \mathcal{G}. \end{aligned} \tag{6.18}$$

We should note that in the average power constraint we use the fact that the total cell-loading β is $\sum_{j\in\mathcal{G}} \beta_j = m\breve{\beta}$. We can also see that (6.18) is equivalent to **P1**. Thus, its solutions can be obtained by using the same strategies as in solving **P1**. The maximum limiting sum rate for mode-m transmission can be attained by evaluating (6.18) for every possible choice of group combinations $\mathcal{G}$, i.e.,

$$\breve{R}_{\text{sum}}^{(m),\infty} = \max_{\mathcal{G}\subset\{1,\ldots,L\},|\mathcal{G}|=m} R_{\text{sum}}^{(m),\infty}(\mathcal{G}). \tag{6.19}$$

By using the formulation (6.19), we can rewrite **P2** as

$$\mathbf{P2}: \max_{m\leq L} \breve{R}_{\text{sum}}^{(m),\infty}. \tag{6.20}$$

As mentioned earlier, for (6.19) there are $\binom{L}{m}$ possible choices or candidates for the optimal $\mathcal{G}$. For the problem (6.20), the number of candidates becomes $\sum_{i=1}^{L}\binom{L}{i}$. It grows when L is large. Hence, reducing the complexity of (6.20) is of our interest. In doing that, let us solve (6.19) and its solution is summarized by the following lemma.

Lemma 6.2. *$\breve{R}_{sum}^{(m),\infty}$ is achieved by choosing $\mathcal{G} = \mathcal{G}^*$ where $\mathcal{G}^* = \{1, 2, \ldots, m\}$.*

Proof. Let $\mathcal{G}^* = \{1, 2, \ldots, m\}$. Also, let $\mathcal{S} \subset \{1, \ldots, L\}$ with $|\mathcal{S}| = m$ such that $\mathcal{G}^* \neq \mathcal{S}$. Moreover, the elements of $\mathcal{S}$ is arranged in an increasing order. Let $\mathbf{a}_{\mathcal{G}^*}$ and $\mathbf{a}_{\mathcal{S}}$ be the path-loss gain vector for group combinations $\mathcal{G}^*$ and $\mathcal{S}$, respectively. It is clear that $\mathbf{a}_{\mathcal{G}^*} \succeq \mathbf{a}_{\mathcal{S}}$. Thus, for a fixed power and regularization parameter, it follows that $R_{\text{sum}}^{(m),\infty}(\mathcal{G}^*) \geq R_{\text{sum}}^{(m),\infty}(\mathcal{S})$. Now, suppose that $\bar{\mathbf{p}}_{\mathcal{S}}^*$ and $\rho_{\mathcal{S}}^*$ are the optimal power allocation and ρ under $\mathcal{S}$. Let us denote the corresponding limiting sum rate as $R_{\text{sum}}^{(m),\infty}(\mathcal{S}, \rho_{\mathcal{S}}^*, \bar{\mathbf{p}}_{\mathcal{S}}^*)$. Under $\mathcal{G}^*$, let us choose $\bar{\mathbf{p}}_{\mathcal{G}} = \bar{\mathbf{p}}_{\mathcal{S}}^*$ and $\rho_{\mathcal{G}^*} = \rho_{\mathcal{S}}^*$ for the power allocation and ρ, respectively. Even though those choices are not optimal in maximizing $R_{\text{sum}}^{(m),\infty}(\mathcal{G}^*)$, they satisfy the constraint in (6.18). Since both $\mathcal{G}^*$ and $\mathcal{S}$ have the same allocations for power and ρ, then it follows that $R_{\text{sum}}^{(m),\infty}(\mathcal{G}^*, \rho_{\mathcal{S}}^*, \bar{\mathbf{p}}_{\mathcal{S}}^*) \geq R_{\text{sum}}^{(m),\infty}(\mathcal{S}, \rho_{\mathcal{S}}^*, \bar{\mathbf{p}}_{\mathcal{S}}^*)$. This concludes the proof. □

It is clear from the lemma above that we greatly reduce the complexity of **P2**. Now, we only need to compare L limiting sum rates, $\breve{R}_{\text{sum}}^{(m),\infty}$. It is also easy to see that Lemma 6.2 also holds when $\beta_{1,\max} \geq \beta_{2,\max} \geq \cdots \geq \beta_{L,max}$. For arbitrary structures of $\{\beta_{j,\max}\}$, we hypothesize that some results in solving **P3**, as discussed later, can be applied. This would be a subject for a future investigation.

Rather than restricting β_j s.t. $\beta_j \in \{0, \beta_{j,\max}\}$, we can relax it so that $0 \leq \beta_j \leq \beta_{j,\max}$. This allows the BS to transmit not to all the users in the groups but some of them. This scenario is more general compared to **P2** and can be written as follows,

$$\begin{aligned} \mathbf{P3}: \quad \underset{\bar{\mathbf{p}} \succeq 0, \boldsymbol{\beta}, \rho, \beta \geq 0}{\text{max.}} \quad & R_{\text{sum}}^{\infty} \\ \text{s.t.} \quad & \sum_{j=1}^{L} \beta_j \bar{p}_j \leq \beta, \\ & \sum_{j=1}^{L} \beta_j = \beta, \\ & 0 \leq \beta_j \leq \beta_{j,\max}\,. \end{aligned}$$

In the following, we investigate the solution for **P3**. We start by writing the La-

grangian of **P3** as follows

$$\mathcal{L} = \sum_{j=1}^{L} \beta_j(1+\bar{p}_j f_j(\beta,\rho)) - \lambda \sum_{j=1}^{L} \beta_j(\bar{p}_j - 1) + \mu_j \bar{p}_j + \nu_j \beta_j - \eta_j(\beta_j - \beta_{j,\max}) + \kappa\rho + \mu\left(\sum_{j=1}^{L} \beta_j - \beta\right) + \eta\beta,$$

where $\lambda, \{\mu_j\}, \{\nu_j\}, \{\eta_j\}, \kappa, \mu$ and η are the Lagrange multipliers for the constraints of **P3**. The KKT necessary conditions are then given by

$$\frac{\partial \mathcal{L}}{\partial \rho} = \sum_{j=1}^{m} \frac{\beta_j p_j}{1+p_j f_j(\beta,\rho)} \frac{\partial f_j(\beta,\rho)}{\partial \rho} + \kappa = 0 \tag{6.21}$$

$$\frac{\partial \mathcal{L}}{\partial \bar{p}_j} = \beta_j \left(\frac{f_j}{1+\bar{p}_j f_j} - \lambda\right) + \mu_j = 0 \tag{6.22}$$

$$\frac{\partial \mathcal{L}}{\partial \beta_j} = \log(1+\bar{p}_j f_j) - \lambda(\bar{p}_j - 1) + \nu_j - \eta_j + \mu = 0 \tag{6.23}$$

$$\frac{\partial \mathcal{L}}{\partial \beta} = \sum_{j=1}^{m} \frac{\beta_j p_j}{1+p_j f_j(\beta,\rho)} \frac{\partial f_j(\beta,\rho)}{\partial \beta} - \mu + \nu = 0 \tag{6.24}$$

and

$$\lambda \sum_{j=1}^{L} \beta_j(\bar{p}_j - 1) = 0,\ \mu_j \bar{p}_j = 0,\ \nu_j \beta_j = 0, \tag{6.25}$$

$$\eta_j(\beta_j - \beta_{j,\max}) = 0,\ \mu\left(\sum_{j=1}^{L} \beta_j - \beta\right) = 0,\ \kappa\rho = 0,\ \eta\beta = 0 \tag{6.26}$$

$$[\lambda\ \kappa\ \mu\ \eta]^T \succeq 0, \mu_j \geq 0 \nu_j \geq 0, \eta_j \geq 0, \tag{6.27}$$

for all $j = 1, \ldots, L$.

Let us consider the stationary condition (6.21). In solving **P1**, we have shown that $f_j(\beta,\rho)$ in increasing in ρ up to $\rho = \beta/\gamma_j$ and then decreasing. Thus, the optimal ρ can not be zero ($\kappa = 0$) and at the optimum,

$$\sum_{j=1}^{m} \frac{\beta_j \bar{p}_j}{1+\bar{p}_j f(a_j)} \frac{\partial f(a_j)}{\partial \rho} = 0. \tag{6.28}$$

Looking at (6.22), one can see that when $\bar{p}_j > 0$ ($\mu_j = 0$), it satisfies

$$\bar{p}_j = \left[\frac{1}{\lambda} - \frac{1}{f_j(\beta,\rho)}\right]_+$$

which is the same as obtained in the solution for **P1**. Since $a_1 \geq \ \ldots \ \geq a_L$, then $\bar{p}_1 \geq \ldots \geq \bar{p}_L$. At the optimum, the following holds

$$\sum_{j=1}^{L} \beta_j \left(\left[\frac{1}{\lambda} - \frac{1}{f_j(\beta,\rho)}\right]_+ - 1 \right) = 0$$

and it can be used to determine λ.

Exploring the stationary condition (6.23) will lead us to the following result.

Lemma 6.3. *The optimal $\{\beta_j\}$ allocation is such that*

(i) *the first M groups, for some $M \leq L$, will be allocated non-zero power*

(ii) *$\beta_1, \beta_2, \ldots, \beta_{M-1}$ are all at the maximum possible values*

(iii) $0 \leq \beta_M \leq \beta_{M,\max}$

(iv) *the remaining groups are allocated zero power*

Proof. In the first part, we will prove (i) - (iii). We show those by considering any two groups l and j such that $l < j$, such that the current allocation has $\beta_j > 0$ and $\bar{p}_j > 0$ and proving that we can improve performance by having β_l at its maximum possible value. Let us assume an assignment $(\beta_l, \bar{p}_l)$ and $(\beta_j, \bar{p}_j)$ such that $\beta_l \leq \beta_{l,\max}$ and $\beta_j \leq \beta_{j,\max}$. In that case, the combined cell-loading is $\beta_l + \beta_j$. Now, let x_l be the new cell-loading allocation for group l and y_l be the corresponding assigned power. In the following we will show that the optimal x_l maximizing the sum rate of of users in group j and l is $\beta_{l,\max}$ by solving the following optimization problem

$$\begin{aligned}
\mathbf{P4}: \quad \max_{x_l, y_l, y_j} \quad & x_l \log(1 + y_l f_l(\beta,\rho)) + (\beta_l + \beta_j - x_l)\log(1 + y_j f_j(\beta,\rho)) \\
\text{s.t.} \quad & \max(0, \beta_l + \beta_j - \beta_{j,\max}) \leq x_l \leq \min(\beta_l + \beta_j, \beta_{l,\max}) \\
& y_l x_l + y_j(\beta_l + \beta_j - x_l) \leq \beta_l \bar{p}_l + \beta_j \bar{p}_j \\
& y_l \geq 0, y_j \geq 0.
\end{aligned}$$

The Lagrangian is given by

$$\begin{aligned}\mathcal{L} = & x_l \log(1 + y_l f_l(\beta, \rho)) + (\beta_l + \beta_j - x_l) \log(1 + y_j f_j(\beta, \rho)) \\ & + \mu_{x_l} \left(x_l - \max(0, \beta_l + \beta_j - \beta_{j,\max})\right) + \nu_{x_l} \left(\min(\beta_l + \beta_j, \beta_{l,\max}) - x_l\right) \\ & + \lambda \left(\beta_l \bar{p}_l + \beta_j \bar{p}_j - y_l x_l - y_j(\beta_l + \beta_j - x_l)\right) + \mu_{y_l} y_l + \mu_{y_j} y_j,\end{aligned}$$

where $\mu_{x_l}, \nu_{x_l}, \mu_{y_l}, \mu_{y_j}, \lambda$ are the Lagrange multipliers associated to the constraints on x_l, y_l, y_j and the second constraint, respectively. The stationary conditions are then given by

$$\frac{\partial \mathcal{L}}{\partial x_l} = \log(1 + y_l f_l(\beta, \rho)) - \log(1 + y_j f_j(\beta, \rho)) + \mu_{x_l} - \nu_{x_l} - \lambda(y_l - y_j) = 0 \tag{6.29}$$

$$\frac{\partial \mathcal{L}}{\partial y_l} = \frac{x_l f_l(\beta, \rho)}{1 + y_l f_l(\beta, \rho)} + \mu_{y_l} - \lambda x_l = 0 \tag{6.30}$$

$$\frac{\partial \mathcal{L}}{\partial y_j} = (\beta_l + \beta_j - x_l) \frac{f_j(\beta, \rho)}{1 + y_j f_j(\beta, \rho)} + \mu_{y_j} - \lambda(\beta_l + \beta_j - x_l) = 0. \tag{6.31}$$

From the last two stationary conditions, it follows that

$$y_l = \left[\frac{1}{\lambda} - \frac{1}{f_l(\beta, \rho)}\right]_+, \tag{6.32}$$

$$y_j = \left[\frac{1}{\lambda} - \frac{1}{f_j(\beta, \rho)}\right]_+. \tag{6.33}$$

One can check that $y_l = 0$ will never be the optimal solution of $\mathbf{P4}$. For $y_l > 0$, two cases arise depending on whether y_j is strictly positive or not.

- Case $y_j = 0$. To satisfy the KKT conditions, the second constraint is met with equality, for $\lambda > 0$. Thus, we have $y_l = \frac{\beta_l \bar{p}_l + \beta_j \bar{p}_j}{x_l}$. From (6.32), we can express

$$\frac{1}{\lambda} = \frac{\beta_l \bar{p}_l + \beta_j \bar{p}_j}{x_l} - \frac{1}{f_l(\beta, \rho)}.$$

When $y_j = 0$, it also holds $1/\lambda - 1/f_j(\beta, \rho) \leq 0$. Consequently, from the equation above, we can write

$$\frac{\beta_l \bar{p}_l + \beta_j \bar{p}_j}{\frac{1}{f_j(\beta,\rho)} - \frac{1}{f_l(\beta,\rho)}} \leq x_l \,.$$

From (6.29), we can obtain

$$\log\left(1+\left(\frac{\beta_l\bar{p}_l+\beta_j\bar{p}_j}{x_l}\right)f_l(\beta,\rho)\right)-\frac{1}{1+\frac{x_l}{\beta_l\bar{p}_l+\beta_j\bar{p}_j}\frac{1}{f_l(\beta,\rho)}}=\nu_{x_l}-\mu_{x_l}. \tag{6.34}$$

The LHS of (6.34) is a function of the form $f(x)=\log(1+x)-\frac{x}{1+x}$, which can be easily shown to be strictly increasing in x. Moreover, at $x=0$, $f(x)=0$. So, the LHS of (6.34) is positive. Thus, ignoring the constraint on x_l, the objective function is strictly increasing for $\frac{\beta_l\bar{p}_l+\beta_j\bar{p}_j}{\frac{1}{f_j(\beta,\rho)}-\frac{1}{f_l(\beta,\rho)}}\leq x_l$.

- Case $y_j>0$. For $\gamma>0$, the average power constraint is met with equality and we have

$$\begin{aligned}y_j&=\frac{\beta_l\bar{p}_l+\beta_j\bar{p}_j-(y_l-y_j)x_l}{\beta_l+\beta_j}\\&=\frac{\beta_l\bar{p}_l+\beta_j\bar{p}_j-\left(\frac{1}{f_j(\beta,\rho)}-\frac{1}{f_l(\beta,\rho)}\right)x_l}{\beta_l+\beta_j}.\end{aligned}$$

Then, we can express

$$\frac{1}{\lambda}=\frac{1}{f_j(\beta,\rho)}+\frac{\beta_l\bar{p}_l+\beta_j\bar{p}_j-\left(\frac{1}{f_j(\beta,\rho)}-\frac{1}{f_l(\beta,\rho)}\right)x_l}{\beta_l+\beta_j}.$$

Since for $y_j>0$, $\frac{1}{\lambda}>\frac{1}{f_j(\beta,\rho)}$, then we obtain

$$\frac{\beta_l\bar{p}_l+\beta_j\bar{p}_j}{\frac{1}{f_j(\beta,\rho)}-\frac{1}{f_l(\beta,\rho)}}>x_l\ . \tag{6.35}$$

Using the expression for $1/\lambda$, we can rewrite (6.29) as

$$\log\left(\frac{f_l(\beta,\rho)}{f_j(\beta,\rho)}\right)-\frac{\frac{1}{f_j(\beta,\rho)}-\frac{1}{f_l(\beta,\rho)}}{\frac{1}{f_j(\beta,\rho)}+\frac{\beta_l\bar{p}_l+\beta_j\bar{p}_j-\left(\frac{1}{f_j(\beta,\rho)}-\frac{1}{f_l(\beta,\rho)}\right)x_l}{\beta_l+\beta_j}}=\nu_{x_l}-\mu_{x_l}. \tag{6.36}$$

It is clear that the LHS of (6.36) is decreasing in x_l. Moreover, for $x_l\to\infty$, its value is $\log\left(\frac{f_l(\beta,\rho)}{f_j(\beta,\rho)}\right)>0$. Therefore, without the constraints on x_l, the objective function is also strictly increasing in x_l when the condition (6.35) holds.

Combining the two cases, the optimal x_l is equal to its maximum allowable value. By using this fact repeatedly, starting from group 1, we establish (i)-(iii).

Now, it remains to show that if no power is allocated to a group, it must be that the corresponding $\beta_j = 0$ (see (iv)). Let us consider the stationary conditions for β_j and β which are given by (6.23) and (6.24), respectively. We can rewrite them as

$$\log(1+\bar{p}_j f_j) - \lambda(\bar{p}_j - 1) + \nu_j + \mu = \eta_j \tag{6.37}$$

and

$$\sum_{j=1}^{L} \frac{\beta_j \bar{p}_j}{1+\bar{p}_j f_j(\beta,\rho)} \frac{\partial f_j(\beta,\rho)}{\partial \beta} = \mu, \tag{6.38}$$

respectively. In ontaining (6.38), we use the fact that β must be positive, i.e., $\eta = 0$. The first derivative of $f_j(\beta,\rho)$ over β in (6.38) can be shown to take the form

$$\frac{\partial f_j(\beta,\rho)}{\partial \beta} = -\frac{f_j(\beta,\rho)}{\beta}\left[1 + \frac{g}{1+\frac{\rho}{\beta}(1+g)^2} + \frac{2g(1+g)^2(\frac{\rho}{\beta}\gamma_j - 1)}{[\gamma_j + (1+g)^2][1+\frac{\rho}{\beta}(1+g)^2]^2}\right], \tag{6.39}$$

where for brevity we denote $g = g(\beta,\rho)$. The derivative of $f_j(\beta,\rho)$ w.r.t. ρ in (6.21) can be written as follows

$$\frac{\partial f_j(\beta,\rho)}{\partial \rho} = -\frac{f_j(\beta,\rho)}{\beta}\frac{2g(1+g)^3(\frac{\rho}{\beta}\gamma_j - 1)}{[\gamma_j + (1+g)^2][1+\frac{\rho}{\beta}(1+g)^2]^2}.$$

So we can rewrite (6.39) in terms of $\frac{\partial f_j(\beta,\rho)}{\partial \rho}$ as

$$\frac{\partial f_j(\beta,\rho)}{\partial \beta} = -\frac{f_j(\beta,\rho)}{\beta}\left[1 + \frac{g}{1+\frac{\rho}{\beta}(1+g)^2}\right] + \frac{1}{1+g}\frac{\partial f_j(\beta,\rho)}{\partial \rho}. \tag{6.40}$$

Recall that $1 + p_j f_j(\beta,\rho) = f_j(\beta,\rho)/\lambda$. Substituting (6.40) into (6.38) yields

$$\begin{aligned}\mu &= -\frac{\lambda}{\beta}\left[1 + \frac{g}{1+\frac{\rho}{\beta}(1+g)^2}\right]\sum_{j=1}^{L}\beta_j\bar{p}_j + \frac{1}{1+g}\sum_{j=1}^{L}\frac{\beta_j\bar{p}_j}{1+\bar{p}_j f_j(\beta,\rho)}\frac{\partial f_j(\beta,\rho)}{\partial \rho} \\ &\overset{(a)}{=} -\lambda\left[1 + \frac{g}{1+\frac{\rho}{\beta}(1+g)^2}\right],\end{aligned} \tag{6.41}$$

where in (a) we use the fact that $\sum_{j=1}^{L} \beta_j \bar{p}_j = \beta$ and the second term of the r.h.s. is zero due to (6.21). Moreover, (a) give the expression for μ at the optimal operating points. Plugging (a) into (6.37) with $p_j = 0$, we obtain

$$-\lambda \frac{g}{1 + \frac{\rho}{\beta}(1+g)^2} + \nu_j = \eta_j.$$

As a result, ν_j must be strictly positive. This implies that $\beta_j = 0$ and the proof is completed. □

To this end, we already know the necessary conditions for the optimal solutions of **P3**. We should note that in the lemma above, we do not know the optimal value of M maximizing the limiting sum rate since there are several values of M that satisfy to the lemma. Let $R_{\text{sum}}^{(i),\infty}$ be the achieved limiting sum rate with $M = i$. Let $\mathcal{M} = \{1, 2, \ldots, \breve{M}\}$ with $\breve{M} \leq L$ be the set of possible values of M. Then, the optimal M is given by

$$M^* = \arg\max_{i \in \mathcal{M}} R_{\text{sum}}^{(i),\infty}. \tag{6.42}$$

We should note that in evaluating $R_{\text{sum}}^{(i),\infty}$, we use $\{\beta_j\}$ allocation scheme in Lemma 6.3, $\beta = \sum_{j=1}^{i} \beta_j$ and also the stationary conditions in (6.21) and (6.22) to determine the optimal ρ and power allocation respectively. The value for β_M must satisfy (6.23) with $\nu_M = 0$ and $\eta_M = 0$, i.e.,

$$\log(1 + \bar{p}_M f_M) - \lambda(\bar{p}_M - 1) + \mu = 0, \tag{6.43}$$

where μ is given by (6.41). Thus, solving (6.42) correspondingly solves **P3**. Steps in solving it are described in Algorithm 6.4.1. We have L iteration when in a particular iteration, say iteration j, the first j groups are considered. Assuming those group to have their cell loading at the maximum value, the corresponding optimal power allocation (i.e., solving **P1**) is computed. Then, the value of $M \leq j$ for that iteration can be determined by using the fact that $\bar{p}_{M+1} = 0$. We should note that different js may give the same M and hence, we need only to consider one of them. After having M, we can set $\beta_{M+1} = \cdots = \beta_j = 0$. To determine the optimal value for β_M, we need to compute η_M. If $\eta_M > 0$, $\beta_M = \beta_{M,\max}$

Algorithm 6.4.1 Algorithm for Solving **P3**

1: $\mathcal{M} = \{\}$ ▷ Contain possible values for M
2: **for** $j = 1$ **to** L **do**
3: $\beta_i = \beta_{i,\max}, \forall\, i = 1, \ldots, j$ ▷ Assume that $\beta_j^* = \beta_{j,\max}$
4: $\lambda, \rho^*, [\bar{p}_1^* \ \ldots \bar{p}_j^*]^T \leftarrow$ Solving **P1** with $\beta = \sum_i^j \beta_i$
5: Determine M s.t. $\bar{p}_{M+1} = \cdots = \bar{p}_j = 0$ ▷ $M \geq 1$
6: **if** $M \in \mathcal{M}$ **then**
7: **continue** ▷ Skip the remaining steps and go to the next iteration
8: **end if**
9: $\mathcal{M} \leftarrow M$
10: $\beta_{M+1} = \cdots = \beta_j = 0$
11: Compute μ according to (6.41)
12: $\eta_M = \log(1 + \bar{p}_M f_M) - \lambda(\bar{p}_M - 1) + \mu$
13: **if** $\eta_M < 0$ **then**
14: $\beta_M \in [0, \beta_{M,\max}] \leftarrow$ Solving **P1** and (6.43) with $\beta = \sum_i^{M-1} \beta_{i,\max} + \beta_M$
15: **end if**
16: Compute $R_{\text{sum}}^{(M),\infty}$ with the updated β and $\{\beta_j\}$
17: **end for**
18: $M^* \leftarrow$ Solving (6.42)

(we already set this in the first step). Otherwise, $0 \leq \beta_M \leq \beta_{M,\max}$. In the latter case, we need to solve **P1** and (6.43) simultaneously. Then, we can update the value for $\{\beta_j\}_{j=1}^M$ and β and also compute the corresponding limiting sum rate. In the final steps, we compare the limiting sum rates for different M and the maximum is the solution of **P3**.

Figure 6.4 illustrates the implementation results of algorithm 6.4.1 for the case: $L = 5$, $a_j^2 = 1/j^2,\ j = 1, \ldots, L$, $\boldsymbol{\beta}_{\max} = [0.1\ 0.7\ 0.1\ 0.05\ 0.05]^T$ where the j-th element corresponds to $\beta_{j,\max}$ and $P_d/\sigma^2 = 10$ dB. From the (upper-left) plot, we can see that we only have three possible values for M, i.e., $\mathcal{M} = \{1, 2, 3\}$. For $M = 1$, we have a positive η_M while for $M = 2$ and $M = 3$, η_M is negative. We should note that for $M = 3$, η_2 is slightly above zero (0.0028). Executing step 14 in the algorithm 6.4.1 yields $\beta_2 = 0.6393$ and $\beta_3 = 0$ for $M = 2$ and $M = 3$, respectively. Even though $M = 2$ and $M = 3$ have the same two groups with positive cell-loading, they have different total cell-loadings, i.e., 0.7393 and 0.8, respectively and consequently have different sum rates. The last plot in the bottom-right shows that the maximum limiting sum rate is achieve when $M = 2$. To validate the result from Algorithm 6.4.1, we perform a grid search where β takes values between 0 and 1 with 0.001 increment. For each value of β, the corresponding limiting

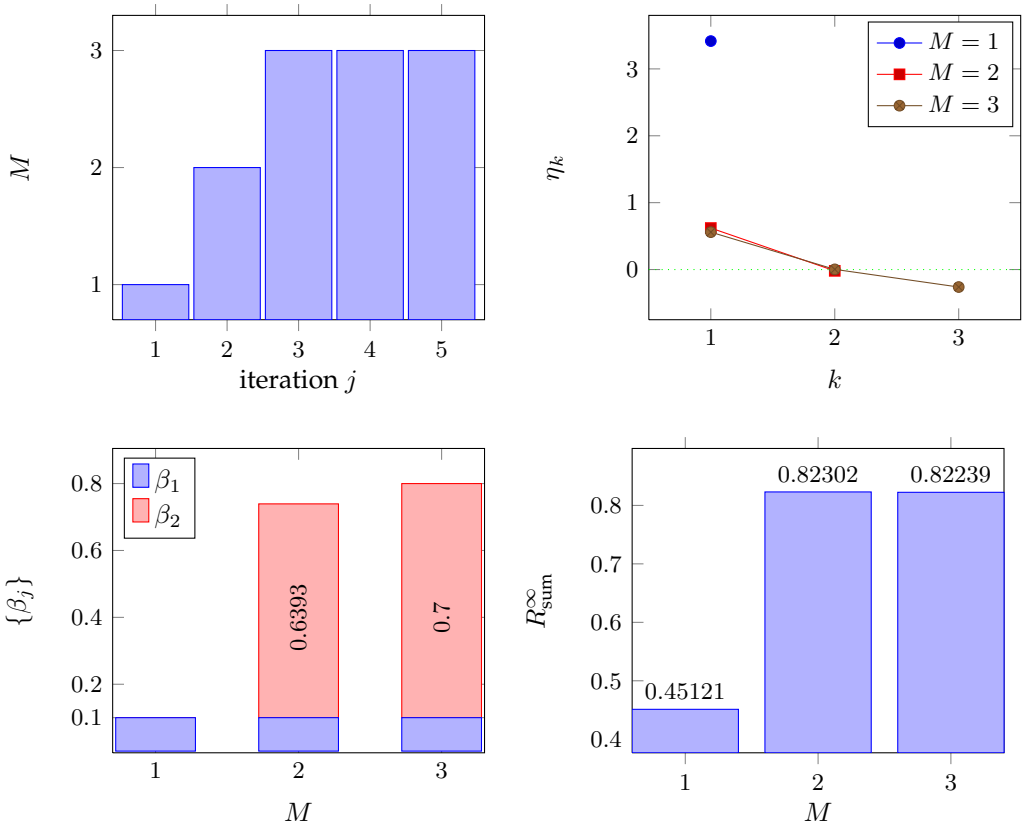


Figure 6.4: Algorithm 6.4.1 implementation for $L = 5$, $\boldsymbol{\beta}_{\max} = [0.1\ 0.7\ 0.1\ 0.05\ 0.05]^T$, $a_j^2 = 1/j^2$ and $P_d/\sigma^2 = 10$ dB.

sum rate is compute. The results are plotted in Figure 6.5. The plot shows that the maximum limiting sum rates and the optimal β obtained from the grid search and Algorithm 6.4.1 are identical. This confirm our theoretical analysis and the proposed algorithm. We should note that even though the line around the optimal β looks flat, the limiting sum rates in that region, by inspecting their numerical values, are actually increasing until reaching the optimal β and then decreasing.

We can also observe from the implementation of Algorithm 6.4.1 in Figure 6.4 that when at the first time (a certain iteration) we find $\eta_M < 0$ (and set $\beta_M \in [0, \beta_{M,\max}]$) we can stop the iterations. This is because continuing the iteration (or adding more groups/increasing M) will not increase the limiting sum rate. We can see this by realizing that the limiting sum rate obtained by increasing β_M by, say β_δ, is greater or equal to that obtained by adding one more group with group loading β_δ. Moreover, increasing

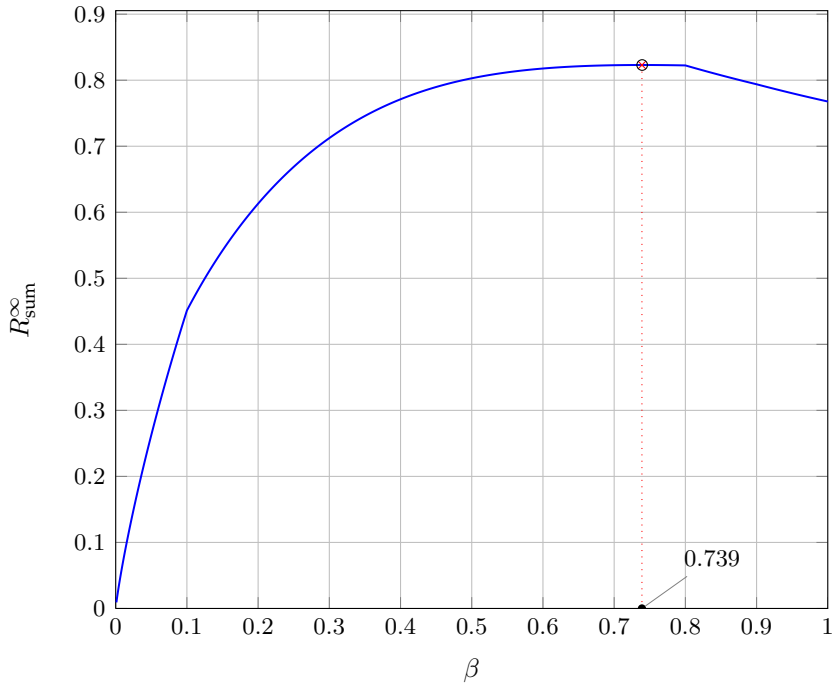


Figure 6.5: The maximum limiting sum rate obtained from the grid search ($\times$) and Algorithm 6.4.1 ($\circ$).

β_M still gives a negative η_M which does not satisfy the KKT necessary condition ($\eta_M \geq 0$). So, we can modify Algorithm 6.4.1 by adding a '**break**' instruction after line 14. That will stop the iteration and jump directly to line 18. This will reduce the number of iterations and computations.

6.5 Conclusion

In this chapter, we have investigated problems related to determining the optimal power allocation, regularization parameter and cell-loading of a finite group or cluster of users so as to maximize the sum rate in MISO broadcast channels. Even though the analysis was performed in the large system limit, our numerical simulations shows its validity for finite-size system designs. Considering the power allocation problem only, we show that it is optimal for the BS to allocate the power to the users by using a water-filling scheme. Applying that scheme, we can not obtain a closed form expression for the optimal regularization parameter but it can be found by using line search methods. For some cases

considered in this chapter we show that it is optimal for the BS to communicate to some groups having best channels (highest path-loss gains). We also provide the KKT necessary conditions for the optimal cell-loading allocation when the BS is allowed to transmit to only subsets of the users in the groups.

Chapter 7

Base Station Cooperation with Feedback Optimization

In this chapter, we study feedback optimization problems that maximize the users' signal to interference plus noise ratio (SINR) in a two-cell MIMO broadcast channel. Assuming the users learn their direct and interfering channels perfectly, they can feed back this information to the base stations (BSs) over the uplink channels. The BSs then use the channel information to design their transmission scheme. Two types of feedback are considered: analog and digital. In the analog feedback case, the users send their unquantized and uncoded CSI over the uplink channels. In this context, given a user's fixed transmit power, we investigate how he/she should optimally allocate it to feed back the direct and interfering (or cross) CSI for two types of base station cooperation schemes, namely, Multi-Cell Processing (MCP) and Coordinated Beamforming (CBf). In the digital feedback case, the direct and cross link channel vectors of each user are quantized separately, each using RVQ, with different size codebooks. The users then send the index of the quantization vector in the corresponding codebook to the BSs. Similar to the feedback optimization problem in the analog feedback, we investigate the optimal bit partitioning for the direct and interfering link for both types of cooperation.

We focus on RCI precoding structures and perform our analysis in the large system limit. We show that for both types of cooperation, for some values of interfering channel gain, usually at low values, no cooperation between the base stations is preferred: This is because, for these values of cross channel gain, the channel estimates for the cross link are not accurate enough for their knowledge to contribute to improving the SINR and there is no benefit in doing base station cooperation under that condition. We also show that for the MCP scheme, unlike in the perfect CSI case, the SINR improves only when the interfering channel gain is above a certain threshold.

7.1 Introduction

7.1.1 Background

RECENTLY, researchers have started to put more attention to investigate how to maximize data rates in *multicell* MIMO networks, particularly in the downlink [25, and references therein]. The main challenge that limits the spectral efficiency in the downlink of multicell networks, besides intra-cell interference, is the inter-cell interference (ICI). The conventional approach to mitigate this interference is to use spatial reuse of resources such as frequency and time [25]. The move towards aggressive frequency or time reuse will cause the networks to be interference limited especially for the users at the cell- edge. The current view is to mitigate ICI through base station (BS) cooperations. Within this scheme, the BSs share the control signal, channel state information (CSI) and data symbols for all users via a central processing unit or wired backhaul links [5].

It has been established in [23,38,42,80,85,110], to name a few, that MIMO cooperation schemes provide a significant increase in spectral efficiency compared to conventional cellular networks. BS cooperation can be implemented at different levels [25]. In the *MultiCell Processing* setup, also known as *Network MIMO* or *Coordinated Multi-Point* (CoMP) transmission, the BSs fully cooperate and share both the channel state information (CSI) and transmission data. This full cooperation requires high capacity backhaul links which is sometimes not viable in practical settings. To alleviate this requirement, only CSI (including direct and interfering channels) is shared amongst base stations in the *interference coordination* scheme [25]. Several works have addressed coordinated beamforming and power control schemes to improve the spectral efficiency in interference- limited downlink multicell networks. Detailed discussions regarding these topics can be found in [25] and references therein.

In both base station cooperation schemes, the CSI at the base stations play an important role in maximizing the system performance. The base stations use this information to adapt their transmission strategies to the channel conditions. The benefit of having CSI at the transmitter (CSIT) with respect to the capacity in single and multicell multiantenna systems is nicely summarized in [6, 26]. However, these advantages are also

accompanied by the overhead cost for the CSI acquisition via channel training and feedback in frequency division duplex (FDD) systems. It needs to scale proportionally to the number of transmit and receive antennas and the number of users in the system in order to maintain a constant gap of the sum rate with respect to the full CSI case [50]. Moreover, in practical systems, the backhaul-link capacity for CSI and user data exchanges and feedback-link bandwidth are limited [5]. Considering the CSI signaling overhead from channel training and CSI feedback, references [71, 72] (see also [111]) suggested that the conventional single-cell processing (SCP) without coordination may outperform the cooperative systems, even the MCP scheme. Here, to reduce the complexity in the analysis, we ignore the (important) constraints of limited backhaul-link and CSI training overhead. We assume a perfect CSI training so that all the users know their CSI perfectly. We focus on studying how to allocate feedback resources, that depends on the feedback schemes, to send the CSI for the direct channel and interfering (cross) channel to BSs so that the users' SINR are maximized. Two feedback schemes are considered in our study: the analog feedback scheme, introduced in [54] and the limited (quantized) feedback via random vector quantization (RVQ), introduced in [77]. In the analog feedback scheme, each user sends its unquantized and uncoded channel state information through the uplink channel. Hence, we ask the question, for a given uplink power constraint, what fraction of this uplink power is allocated optimally to transmit the direct and interfering channel information? For the digital feedback scheme, the number of feedback bits determines the quality of the CSI. Hence, we can ask, how many bits are optimally needed to feedback the direct and cross CSI?

7.1.2 Contributions

The main goal of the work in this chapter is to optimize and investigate the effect of feedback for MCP and CBf cooperation schemes under analog and quantized feedback (via RVQ). We consider a *symmetric* two-cell Multi-Input Single-Output (MISO) network where the base stations have multiple antennas and each user has a single antenna. We assume that the users in each cell know their own channel perfectly: they feed back this information through the uplink channel and the base stations form the users' channel

estimates. The BSs use these estimates to construct a regularized channel inversion (RCI) type beamformer, also called regularized zero-forcing (RZF), to precode the data symbols of the users. The precoders follow the structures proposed in [109]. Unlike [5, 111], we assume several users are simultaneously active in each cell so that users experience both intra- and inter-cell interference. To mitigate ICI through base station cooperation, we consider both full cooperation (MCP) and interference coordination via CBf.

Our contributions can be summarized as follows. First, under both feedback models and both cooperation schemes, we derive the SINR expression in the large system limit, also called the *limiting SINR*, where the number of antennas at base stations and the number of users in each cell go to infinity with their ratio kept fixed: As our numerical results will show, *this is indicative of the average performance for even finite numbers of antennas*. Then, we formulate a joint optimization problem that performs the feedback optimization for both feedback models and both cooperation schemes and finds the optimal regularization parameter of the corresponding RZF structure precoder. The regularization parameter is an important design parameter for the precoder because it controls the amount of interference introduced to the users. Optimizing this parameter, as discussed later, will allow the precoder to adapt to the changes of the CSIT quality and consequently produces a '*robust beamformer*'.

We analyze the behavior of the maximum limiting SINR as a function of the cross channel gains and the available feedback resources, and identify, for both the analog and quantized feedback models, regions where SCP processing is optimal. We also show that whereas in the perfect CSI case, MCP performance always improves with epsilon, this only occurs after a certain threshold is crossed in both analog and limited feedback cases.

Parts of this work appeared in [60, 61], but without the proofs.

7.1.3 Related Works

In the last decade, there has been a large volume of research discussing feedback schemes in multi-antenna systems. A summary of digital feedback (also known as limited or finite-rate feedback) schemes in multi-antenna (also single-antenna) and multi-user systems in the single-cell setup can be found in [50]. Since the optimal codebook for the

limited feedback is not known yet [8, 36, 77], the use of RVQ, which is based on a random codebook, as the feedback scheme becomes popular. Furthermore, the RVQ-based system performance analysis is also more tractable. In multiple-antenna and multi-user systems, works on the analog feedback commonly refer to [54] (sometimes [76]).

The paper by Jindal [36] sparked the use of RVQ in analyzing broadcast channels. Considering a MISO broadcast channel with a zero-forcing (ZF) precoder and assuming that each user knows its own channel, the main result in the paper is that the feedback rate should be increased linearly with the signal-to-noise ratio (SNR) to maintain the full multiplexing gain. Caire et al. in [8] investigate achievable ergodic sum rates of BC with ZF precoder under several practical scenarios. The CSI acquisition involves four steps; downlink training, CSI feedback, beamformer selection and dedicated training where each user will try to estimate the coupling between its channel and the beamforming vectors. They derive and compare the lower bound and upper bound of the achievable ergodic sum- rate of the analog feedback as in [54] and RVQ-based digital feedback under different considerations, e.g., feedback transmission over AWGN and MAC channel, feedback delay and feedback errors for the digital feedback scheme. A subsequent work by Kobayashi et al. in [45] studies training and feedback optimizations for the same system setup as in [8] except without dedicated training. The optimal period for the training and feedback that minimized achievable rate gap (with and without perfect CSI) are derived under different scenarios as in [8]. The authors also show that the digital feedback can give a significant advantage over the analog feedback. In the same spirit as [36] , reference [101] discusses the feedback scaling (as SNR increases) in order to maintain a constant rate gap for a broadcast channel with regularized zero-forcing or RCI precoder. The analysis has been done in the large system limit since the analysis the finite-size turns out to be difficult [36]. Moreover, besides analyzing for the case $K = N$, as in [36], the authors also investigate the case $K < N$.

While channel state feedback in the single-cell system has received a considerable amount of attention so far, fewer works have addressed this problem in multicell settings. The effect of channel uncertainty, specifically the channel estimation error, in the multicell setup is studied in [33, 39]. In [39], the authors conclude that when channel es-

timates at one base station contain interference from the users in other cells, also called as pilot contamination phenomenon, the inter-cell interference increases. Thus, this phenomenon could severely impacts the performance of the systems. Huh et al. in [33] investigate optimal user scheduling strategies to reduce the feedback and also the effects of channel estimation error on the ergodic sum rate of the clustered Network MIMO systems. They consider the ZF precoder at the base stations and derive the optimal power allocation that maximizes the (instantaneous) weighted sum rate. In deriving the results, it is assumed that the BSs received perfectly (error-free) the CSI fed back by the users. The overhead caused by the channel training is also investigated and they observe that there is a trade-off between the number of cooperating antennas and the cost of estimating the channel. Based on the trade-off, the optimal cooperation cluster size can be determined. By incorporating the channel training cost, no-coordination amongst the base stations could be preferable. The same conclusion is also obtained in [71,72].

For the interference coordination scheme, [5] investigated the RVQ-based limited feedback in an infinite Wyner cellular model using generalized eigenvector beamforming at the base stations. The work adopts the intra-cell TDMA mechanism where a single user is active in each cell per time slot. Each user in each cell is also assumed to know its downlink channel perfectly. Based on that system model, an optimal bit partitioning strategy for direct and interfering channels that minimizes the sum rate gap is proposed. Explicitly, it is a function of the received SNR from the direct and cross links. It is observed that as the received SNR from the cross link increases, more bits are allocated to quantize the cross channel. A better quality of the cross channel estimate will help to reduce the inter-cell interference. The authors also show that the proposed bit partitioning scheme reduce the average sum rate loss. Also in the interference coordination setting, [111] takes into account both CSI training and feedback in analyzing the system what they called inter- cell interference cancellation (ICIC). In ICIC, the precoding of a user is the projection of its channel in the null-space of the others users' channels in other cells so that the transmission form this will not cause interference to the users in other cells. The work also assumes the intra-cell TDMA and presents the training optimization and feedback optimization for both analog and digital feedback (RVQ). Based on that system

setup, the most interesting result is that training optimization is more important than the feedback optimization for the analog feedback while the opposite holds for the digital feedback.

For different levels of cooperation, i.e., MCP, CBf and SCP, [109] investigates an optimization problem to minimize the total downlink transmit power while satisfying a specified SINR target. The authors derived the optimal transmit power, beamforming vectors, cell loading and achieved SINR for those different cooperation schemes in a symmetric two-cell network. The resulted optimal beamforming vectors have a structure related to RCI.

The current work is closely related to [109] in the sense we use the same cooperative schemes and precoder structure. We extend the work by analyzing the optimal feedback strategies for analog and digital feedback under MCP and CBf schemes. The results in this work are obtained by performing the analysis in the large system limit where the dimensions of the system i.e., the numbers of users and transmit antennas tend to infinity with their ratio being fixed. The large system analysis mainly exploits the eigenvalue distribution of large random matrices. For examples, it has been used to derive the asymptotic performance of linear multiuser receivers in CDMA communications in early 2000 (see [95]), single-cell broadcast channels with RCI for various channel conditions [58, 59, 62, 101], base station cooperations in downlink multicell networks (see e.g., [33, 109]). The asymptotic performance measure becomes a deterministic quantity and can have close-form/compact expressions. Hence, it can be used to derive the optimal parameters for the system design. Moreover, it can provide a good approximation of, hence insights on, the system performance in the finite-size or even small systems.

Similar to [5] and [111], we perform the feedback optimization in interference-coordination scheme (CBf). As in [111], we also investigate the feedback optimization for the analog and digital feedback schemes. However, different from those works, we do not assume the intra-cell TDMA in each cell, and hence each user experiences both intra-cell and inter-cell interference. We also consider a different type of precoder i.e., the RCI. Moreover, we also analyze the feedback optimization for different level of cooperations between the base stations, including the MCP setup, and try to capture how we allocate

resources available at the user side as the the interfering channel gain varies.

The rest of the chapter is structured as follows. The system model is described in Section II. It starts with the channel model, and the expressions of the transmit signal, precoder and the corresponding SINR for each MCP and CBf. In the end of the section, the feedback schemes and true channel model in terms of the channel estimate at the BSs and the channel uncertainty for the analog and digital feedback are presented. The main results for the noisy analog feedback and digital feedback and for different types of coordination are discussed in Section III and IV, respectively. In each section, we begin by discussing the large system result of the SINR for the MCP and CBf and then followed by deriving the corresponding optimal feedback allocation; optimal (uplink) power for the analog feedback and optimal pit partitioning for the digital feedback. The optimal regularization parameter for the RCI precoder is also derived for both types of feedback and cooperation. The end of each section provides numerical results that depict how the optimal feedback allocation and the SINR of each user behave as the interfering channel gain varies. In Section V, we provide some numerical simulations that compare the performance of the system under the analog feedback and digital feedback. The conclusion are drawn in the Section VI and some of the proofs go to the appendices.

7.2 System Model

We consider a symmetric two-cell broadcast channel, as shown in Figure 7.1, where each cell has K single antenna users and a base station equipped with N antennas. The channel between user k in cell j and the BS in cell i is denoted by row vector $\mathbf{h}_{k,j,i}$ where $\mathbf{h}_{k,j,j} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$ and $\mathbf{h}_{k,j,\bar{j}} \sim \mathcal{CN}(\mathbf{0}, \epsilon\mathbf{I}_N)$, for $j = 1, 2$ and $\bar{j} = \mod(j, 2) + 1$. We refer to the $\mathbf{h}_{k,j,j}$ as direct channels and $\mathbf{h}_{k,j,\bar{j}}$ as cross or "interfering" channels. We find it useful to group these into a single channel vector $\mathbf{h}_{k,j} = [\mathbf{h}_{k,j,1}\ \mathbf{h}_{k,j,2}]$.

We consider an FDD system and assume that the users to have perfect knowledge of their downlink channels, $\mathbf{h}_{k,j,j}$ and $\mathbf{h}_{k,j,\bar{j}}$. Each user feeds back the channel information to the direct BS and neighboring BS through the corresponding uplink channels. The BSs estimate or recover these channel states and use them to construct the precoder.

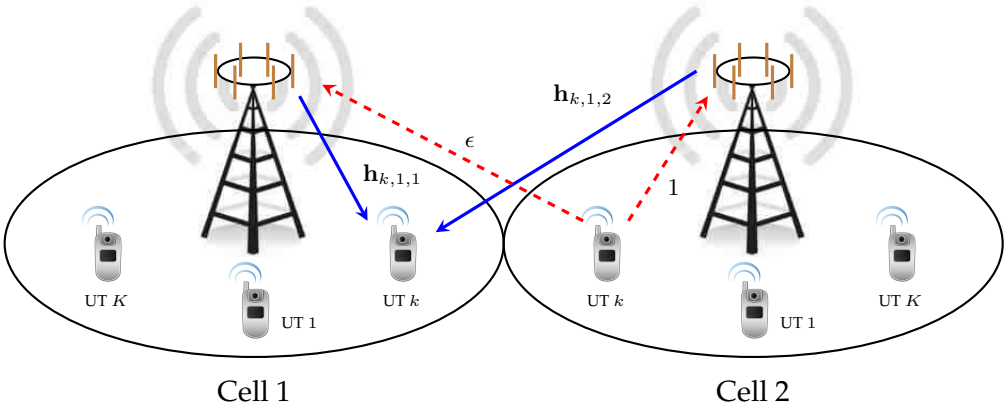


Figure 7.1: System model for a symmetric two-cell broadcast channel.

The received signal of user k in cell j can be written as

$$y_{k,j} = \mathbf{h}_{k,j,1}\mathbf{x}_1 + \mathbf{h}_{k,j,2}\mathbf{x}_2 + n_{k,j},$$

where $\mathbf{x}_i \in \mathbb{C}^{N\times 1}$, $i = 1, 2$ is the transmitted data from BS i, and $n_{k,j} \sim \mathcal{CN}(0, \sigma_d^2)$ is the noise at the user's receiver. The transmitted data $\mathbf{x}_i$ depends on the level of cooperation assumed, and will be described in more details in Sections 7.2.1 and 7.2.2: we restrict ourselves to linear precoding schemes, more specifically RCI precoder. We assume each BS's transmission is subject to a power constraint $\mathbb{E}\left[\|\mathbf{x}_i\|^2\right] = P_i$. In the MCP case, we relax this constraint to a sum power constraint so that $\mathbb{E}\left[\|\mathbf{x}\|^2\right] = \sum_{i=1}^{2} P_i = P_t$. In the analysis, we assume $P_1 = P_2 = P_d$ and denote $\gamma_d = P_d/\sigma_d^2$.

As already mentioned, in practical scenarios, perfect CSI is difficult to obtain and the CSI at the BSs is obtained through feedback from the users. We are particularly interested in the channel model where we can express the downlink channel between the user k in cell j and BS i as

$$\mathbf{h}_{k,j,i} = \sqrt{\phi_{k,j,i}}\widehat{\mathbf{h}}_{k,j,i} + \widetilde{\mathbf{h}}_{k,j,i}, \tag{7.1}$$

where $\widehat{\mathbf{h}}_{k,j,i}$ represents the channel estimate, and $\widetilde{\mathbf{h}}_{k,j,i}$ the channel uncertainty or estimation error. The channel estimates are used by the BSs to construct the precoder.

The transmitted signal, precoder and SINR for each user for each cooperation scheme will be presented in the following subsections.

7.2.1 MCP

As previously mentioned, in the MCP, both BSs share the channel information and data symbols for all users in the network. Therefore, we may consider the network as a broadcast channel with $2N$ transmit antennas and $2K$ single antenna users. The BSs construct the precoding matrix using their channel estimates. In this work, we consider the RCI precoding, for which the precoding or beamforming vector for user k in cell j, $\mathbf{w}_{kj}$, can be written as [69]

$$\mathbf{w}_{kj} = c\left(\widehat{\mathbf{H}}^H\widehat{\mathbf{H}} + \alpha\mathbf{I}_{2N}\right)^{-1}\widehat{\mathbf{h}}_{k,j}^H,$$

where $\widehat{\mathbf{h}}_{k,j} = [\widehat{\mathbf{h}}_{k,j,1}\ \widehat{\mathbf{h}}_{k,j,2}]$ and $\widehat{\mathbf{H}} = [\widehat{\mathbf{h}}_{1,1}^H\ \widehat{\mathbf{h}}_{2,1}^H \cdots\ \widehat{\mathbf{h}}_{K,1}^H\ \widehat{\mathbf{h}}_{1,2}^H\ \widehat{\mathbf{h}}_{2,2}^H\ \cdots\ \widehat{\mathbf{h}}_{K,2}^H]^H$. Let $\hat{\mathbf{w}} = \mathbf{w}/c$. The transmitted data vector can be expressed as

$$\mathbf{x} = c\sum_{j=1}^{2}\sum_{k=1}^{K}\hat{\mathbf{w}}_{kj}s_{kj},$$

where $s_{kj} \sim \mathcal{CN}(0,1)$ denotes the symbol to be transmitted to user k in cell j. It is also assumed that the data symbols across the users are independent, i.e., $\mathbb{E}[\mathbf{s}\mathbf{s}^H] = \mathbf{I}_{2K}$ with $\mathbf{s} = [s_{11}\ \ldots\ s_{K1}\ s_{12}\ \ldots\ s_{K2}]^T$. c is a scaling factor ensuring the total power constraint is met with equality ($\mathbb{E}[\mathbf{x}\mathbf{x}^H] = P_t$):

$$c^2 = \frac{P_t}{\mathrm{Tr}\left\{\left(\widehat{\mathbf{H}}^H\widehat{\mathbf{H}} + \alpha\mathbf{I}_{2N}\right)^{-2}\widehat{\mathbf{H}}^H\widehat{\mathbf{H}}\right\}}.$$

The received signal at user k in cell j can be written as

$$\begin{aligned} y_{kj} &= \mathbf{h}_{k,j}\mathbf{x} + n_{k,j} \\ &= c\mathbf{h}_{k,j}\left(\widehat{\mathbf{H}}^H\widehat{\mathbf{H}} + \alpha\mathbf{I}_{2N}\right)^{-1}\widehat{\mathbf{H}}^H\mathbf{s} + n_{k,j} \\ &= c\mathbf{h}_{k,j}\left(\widehat{\mathbf{H}}^H\widehat{\mathbf{H}} + \alpha\mathbf{I}_{2N}\right)^{-1}\widehat{\mathbf{h}}_{k,j}^H s_{k,j} + c\mathbf{h}_{k,j}\left(\widehat{\mathbf{H}}^H\widehat{\mathbf{H}} + \alpha\mathbf{I}_{2N}\right)^{-1}\widehat{\mathbf{H}}_{k,j}^H\mathbf{s}_{k,j} + n_{k,j}\ , \end{aligned}$$

where $\mathbf{h}_{k,j}$ follows the channel model (7.1) with $\widetilde{\mathbf{h}}_{k,j} = [\widetilde{\mathbf{h}}_{k,j,1}\ \widetilde{\mathbf{h}}_{k,j,2}]$. The term $\widehat{\mathbf{H}}_{k,j}$ and $\mathbf{s}_{k,j}$ are obtained from $\widehat{\mathbf{H}}$ and $\mathbf{s}$ by removing the row corresponding to user k in cell j

respectively. Hence, the SINR for user k in cell j can be expressed as

$$\text{SINR}_{k,j} = \frac{c^2 \left| \mathbf{h}_{k,j} \left(\widehat{\mathbf{H}}^H \widehat{\mathbf{H}} + \alpha \mathbf{I}_{2N} \right)^{-1} \widehat{\mathbf{h}}_{k,j}^H \right|^2}{c^2 \mathbf{h}_{k,j} \left(\widehat{\mathbf{H}}^H \widehat{\mathbf{H}} + \alpha \mathbf{I}_{2N} \right)^{-1} \widehat{\mathbf{H}}_{k,j}^H \widehat{\mathbf{H}}_{k,j} \left(\widehat{\mathbf{H}}^H \widehat{\mathbf{H}} + \alpha \mathbf{I}_{2N} \right)^{-1} \mathbf{h}_{k,j}^H + \sigma_d^2}. \tag{7.2}$$

7.2.2 Coordinated Beamforming

In this scheme, the base stations only share the channel information, so that, for cell j, $\mathbf{x}_j$ can be expressed as

$$\mathbf{x}_j = c_j \sum_{k=1}^{K} \hat{\mathbf{w}}_{kj} s_{kj},$$

where $s_{kj} \sim \mathcal{CN}(0,1)$ denotes the data symbol for user k in cell j; $c_j^2 = \frac{P_j}{\sum_{k=1}^{K} \|\hat{\mathbf{w}}_{kj}\|^2}$. We let

$$\hat{\mathbf{w}}_{kj} = \left(\alpha \mathbf{I}_N + \sum_{(l,m) \neq (k,j)} \widehat{\mathbf{h}}_{l,m,j}^H \widehat{\mathbf{h}}_{l,m,j} \right)^{-1} \widehat{\mathbf{h}}_{k,j,j},$$

which is an extension of regularized zero-forcing to the coordinated beamforming setup [109]. Note that designing the precoding matrix at BS j requires *local* CSI only (the $\widehat{\mathbf{h}}_{k,i,j}$ from BS j to all users, but not the channels from the other BS to the users). The SINR of user k in cell j can be expressed as

$$\text{SINR}_{k,j} = \frac{c_j^2 |\mathbf{h}_{k,j,j} \hat{\mathbf{w}}_{kj}|^2}{\sum_{(k',j') \neq (k,j)} c_{j'}^2 |\mathbf{h}_{k,j,j'} \hat{\mathbf{w}}_{k'j'}|^2 + \sigma_d^2}, \tag{7.3}$$

where, once again, $\mathbf{h}_{k,j,j}$ and $\mathbf{h}_{k,j,j'}$ follow (7.1).

7.2.3 Analog Feedback through AWGN Channel

In the *analog feedback* scheme, proposed in [54], each user feeds back the CSI to the base stations using the linear analog modulation. Since we skip quantizing and coding the channel information, we can convey this information very rapidly [54]. We also consider a simple uplink channel model, an AWGN channel. A more realistic multiple access

(MAC) uplink channel model could be a subject for future investigation. Each user in cell j feeds back its CSI $\mathbf{h}_{k,j}$ orthogonally (in time). Since each user has to transmit $2N$ symbols (its channel coefficients), it needs $2\kappa N$ channel uses to feed back the CSI, where $\kappa \geq 1$. User k in cell j sends

$$\mathbf{h}_{k,j}\mathbf{\Lambda}_j^{\frac{1}{2}}, \tag{7.4}$$

where $\mathbf{\Lambda}_j$ is a diagonal matrix such that the first N diagonal entries are equal to λ_{j1} and the remaining diagonal entries are equal to λ_{j2}, with $\lambda_{jj} = 2\nu\kappa P_u$, $\lambda_{j\bar{j}} = 2\epsilon^{-1}(1-\nu)\kappa P_u$ and P_u is the user's average transmit power per channel use. Equation (7.4) satisfies the uplink power constraint $\mathbb{E}[\|\mathbf{h}_{k,j}\mathbf{\Lambda}_j^{\frac{1}{2}}\|^2] = 2\kappa N P_u$. Thus, the power allocated to feedback the direct and interfering channel is controlled by $\nu \in [0,1]$. We should note that in (7.4), it is assumed that κ is an integer. If κN is an integer, we can modulate the signal (7.4) with $2N \times 2\kappa N$ spreading matrix [8,54] and the analysis presented below still holds.

Now, let b_ℓ, $\ell = 1, 2, \cdots, 2N$, be the ℓth element of $\mathbf{h}_{k,j}$, λ_ℓ be the corresponding element on the diagonal of $\mathbf{\Lambda}$, and $\epsilon_\ell = \mathbb{E}[b_\ell b_\ell^*]$. When this channel coefficient is transmitted, the signal received by the coordinating BSs is

$$\mathbf{y}_\ell = \sqrt{\lambda_\ell}\begin{bmatrix} \mathbf{1}_N \\ \sqrt{\epsilon}\mathbf{1}_N \end{bmatrix} b_\ell + \mathbf{n}_u = \sqrt{\lambda_\ell}\mathbf{p}b_\ell + \mathbf{n}_u,$$

where $\mathbf{n}_u \in \mathbb{C}^{2N\times 1} \sim \mathcal{CN}(\mathbf{0}, \sigma_u^2\mathbf{I}_{2N})$ is the noise vector at the coordinating BSs and $\mathbf{1}_N$ is a column vector of length N with all 1 entries. Using the fact that the path-gain from the users in cell j to BS $\bar{j}$ is ϵ, the MMSE estimate of b_ℓ becomes

$$\hat{b}_\ell = \sqrt{\lambda_\ell}\epsilon_\ell \mathbf{p}^T\left[\lambda_\ell\epsilon_\ell\mathbf{p}\mathbf{p}^T + \sigma_u^2\mathbf{I}_{2N}\right]^{-1}\mathbf{y}_\ell,$$

and its MMSE is $\sigma_{b_\ell}^2 = \epsilon_\ell - \lambda_\ell\epsilon_\ell^2\mathbf{p}^T\left[\lambda_\ell\epsilon_\ell\mathbf{p}\mathbf{p}^T + \sigma_u^2\mathbf{I}_{2N}\right]^{-1}\mathbf{p}$. We should note that $\{\hat{b}_\ell\}$ are mutually independent. By using the property of MMSE estimation, we can express $\mathbf{h}_{k,j,i}$ as

$$\mathbf{h}_{k,j,i} = \widehat{\mathbf{h}}_{k,j,i} + \widetilde{\mathbf{h}}_{k,j,i}, \tag{7.5}$$

where $\widehat{\mathbf{h}}_{k,j,i}$ represents the channel estimate, and $\widetilde{\mathbf{h}}_{k,j,i}$ the channel uncertainty or estimation error. Note that the entries of each vector $\widehat{\mathbf{h}}_{k,i,j}$ and $\widetilde{\mathbf{h}}_{k,i,j}$ are independent and

identically distributed (i.i.d.) and distributed according to $\mathcal{CN}(0, \omega_{ji})$ and $\mathcal{CN}(0, \delta_{ji})$, respectively, where

$$\delta_{ji} = \begin{cases} \frac{1}{1+\nu\bar{\gamma}_u}, & j = i \\ \frac{\epsilon}{1+(1-\nu)\bar{\gamma}_u}, & j \neq i, \end{cases}, \quad \omega_{ji} = \begin{cases} \frac{\nu\bar{\gamma}_u}{1+\nu\bar{\gamma}_u}, & j = i \\ \frac{\epsilon(1-\nu)\bar{\gamma}_u}{1+(1-\nu)\bar{\gamma}_u}, & j \neq i, \end{cases} \tag{7.6}$$

and $\bar{\gamma}_u = 2\gamma_u\kappa(1+\epsilon)$ with $\gamma_u = NP_u/\sigma_u^2$. The channel estimates are used by the BSs to construct the precoder. Since each δ_{ij} and ω_{ij} are identical for all users, we denote $\delta_d = \delta_{jj}, \delta_c = \delta_{j\bar{j}}, \omega_d = \omega_{jj}$ and $\omega_c = \omega_{j\bar{j}}$. From (7.6), we have $\omega_d = 1 - \delta_d$ and $\omega_c = \epsilon - \delta_c$.

7.2.4 Quantized Feedback via RVQ

In the digital feedback case, user k in cell j uses $B_{k,j,j}$ and $B_{k,j,\bar{j}}$ bits to quantize or feedback the direct and interfering channels, respectively. The total number of feedback bits is assumed to be fixed. It is also assumed that each user has different codebooks: $\mathcal{U}_{k,j,j}$ with size $2^{B_{k,j,j}}$ and $\mathcal{U}_{k,j,\bar{j}}$ with size $2^{B_{k,j,\bar{j}}}$, to quantize the direct and interfering channel, respectively. Moreover, these codebooks are different for each user. In this work, $B_{k,j,j}$ is the same for all users and $B_{k,j,j} = B_d, \forall k, j = 1, 2$. Similarly, $B_{k,j,\bar{j}} = B_c, \forall k, j = 1, 2$. The total number of feedback bits is denoted by B_t, where $B_t = B_d + B_c$.

Since the optimal codebook design for the quantized feedback is not known yet, therefore in this work, for analytical tractability, we consider the well known RVQ scheme. As suggested by its name, RVQ uses a random vector quantization codebook where the quantization vectors in the codebook are independently chosen from the isotropic distribution on the N-dimensional unit sphere [36, 77]. The codebook is known by the base station and the user. The user quantizes its channel by finding the quantization vector in the codebook which is closest to its channel vector and feedbacks the index of the quantization vector to the BSs. We should note that only the channel direction is quantized. Most of the works that employ RVQ for the feedback model assume that only channel direction information is sent to the BSs. As mentioned in [36], the channel norm information can also be used for some problems that need channel quality information (CQI) such as power allocation across the channel and users scheduling [73].

The user k in cell j finds its quantization vector for the channel $\mathbf{h}_{k,j,i}$ according to

$$\hat{\mathbf{u}}_{k,j,i} = \arg \max_{\mathbf{u}_{k,j,i} \in \mathcal{U}_{k,j,i}} \frac{|\mathbf{h}_{k,j,i}\mathbf{u}_{k,j,i}^H|}{\|\mathbf{h}_{k,j,i}\|}.$$

The quantization error or distortion $\tau_{k,j,i}^2$ is defined as

$$\tau_{k,j,i}^2 = 1 - \frac{\|\mathbf{h}_{k,j,i}\hat{\mathbf{u}}_{k,j,i}\|^2}{\|\mathbf{h}_{k,j,i}\|^2} = \sin^2\left(\angle\left(\frac{\mathbf{h}_{k,j,i}}{\|\mathbf{h}_{k,j,i}\|}, \hat{\mathbf{u}}_{k,j,i}\right)\right).$$

It is a random variable whose distribution is equivalent to the minimum of $2^{B_{k,j,i}}$ beta random variables with parameters $N-1$ and 1 (see [2, 36]). Each realization of $\tau_{k,j,i}$ is different for each user even though the users have the same amount of feedback bits.

Having obtained $\hat{\mathbf{u}}_{k,j,i}$, each user sends its index in the code book and also the channel magnitude $\|\mathbf{h}_{k,j,i}\|$ (see also [73]). By assuming that the BSs can receive the information perfectly, the channel estimate at the BS can be written as

$$\widehat{\mathbf{h}}_{k,j,i} = \|\mathbf{h}_{k,j,i}\|\hat{\mathbf{u}}_{k,j,i}. \tag{7.7}$$

Note that $\widehat{\mathbf{h}}_{k,j,i}$ has the same statistical distribution as $\mathbf{h}_{k,j,i}$ i.e., $\widehat{\mathbf{h}}_{k,j,i} \sim \mathcal{CN}(0, \epsilon_{ji}\mathbf{I}_N)$, where $\epsilon_{ji} = 1$ when $i = j$ and otherwise, $\epsilon_{ji} = \epsilon$.

From [36, 37], we can model $\mathbf{h}_{k,j,i}$ as follows

$$\mathbf{h}_{k,j,i} = \sqrt{1-\tau_{k,j,i}^2}\widehat{\mathbf{h}}_{k,j,i} + \tau_{k,j,i}\|\mathbf{h}_{k,j,i}\|\mathbf{z}_{k,j,i}, \tag{7.8}$$

where $\mathbf{z}_{k,j,i}$ is isotropically distributed in the null-space of $\hat{\mathbf{u}}_{k,j,i}$ and independent of $\tau_{k,j,i}$. Moreover, it can be written as follows

$$\mathbf{z}_{k,j,i} = \frac{\mathbf{v}_{k,j,i}\mathbf{\Pi}^{\perp}_{\widehat{\mathbf{h}}_{k,j,i}}}{\|\mathbf{v}_{k,j,i}\mathbf{\Pi}^{\perp}_{\widehat{\mathbf{h}}_{k,j,i}}\|},$$

where $\mathbf{\Pi}_{\widehat{\mathbf{h}}_{k,j,i}}$ is the projection matrix in the column space of $\widehat{\mathbf{h}}_{k,j,i}$, $\mathbf{\Pi}^{\perp}_{\widehat{\mathbf{h}}_{k,j,i}} = \mathbf{I}_N - \frac{\widehat{\mathbf{h}}_{k,j,i}^H\widehat{\mathbf{h}}_{k,j,i}}{\|\widehat{\mathbf{h}}_{k,j,i}\|^2}$ and $\mathbf{v}_{k,j,i} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$ is independent of $\widehat{\mathbf{h}}_{k,j,i}$. It is clear that the channel model (7.8) has the same structure as (7.1) with $\phi_{k,j,i} = 1 - \tau_{k,j,i}^2$ and $\widetilde{\mathbf{h}} = \tau_{k,j,i}\|\mathbf{h}_{k,j,i}\|\mathbf{z}_{k,j,i}$.

7.2.5 Achievable and Limiting Sum Rate

Besides $\text{SINR}_{k,j}$, another relevant performance measure is the achievable rate. For the user k at cell j, it is defined as

$$R_{k,j} = \log_2(1 + \text{SINR}_{k,j}). \tag{7.9}$$

It is obtained by treating the interference as noise or equivalently performing single-user decoding at the receiver. Observing (7.9), it is obvious that there is a one-to-one continuous mapping between the SINR and the achievable rate (see also [91]). The total sum rate, or just the sum rate, can then be defined as follows

$$R_{\text{sum}} = \sum_{j=1}^{2} \sum_{k=1}^{K} R_{kj}. \tag{7.10}$$

As shown later in Section 7.3 and 7.4, as $K, N \to \infty$, we have

$$\text{SINR}_{kj} - \text{SINR}^{\infty} \to 0, \tag{7.11}$$

where SINR^{∞} is a deterministic quantity and also called the limiting SINR. It is also shown that the limiting SINR is the same for all users. By using the result (7.11) and based on continuous mapping theorem [96], the following holds (see also [101])

$$\frac{1}{2N}\mathbb{E}\left[R_{\text{sum}}\right] - R_{\text{sum}}^{\infty} \to 0,$$

where the limiting achievable sum rate can be expressed as $R_{\text{sum}}^{\infty} = \beta \log_2(1 + \text{SINR}^{\infty})$. For the numerical simulations, we also introduce the normalized sum rate difference, defined as

$$\Delta R_{\text{sum}} = \frac{\frac{1}{2N}\mathbb{E}\left[R_{\text{sum}}\right] - R_{\text{sum}}^{\infty}}{\frac{1}{2N}\mathbb{E}\left[R_{\text{sum}}\right]} \tag{7.12}$$

that quantifies the sum rate difference, $\frac{1}{2N}\mathbb{E}\left[R_{\text{sum}}\right] - R_{\text{sum}}^{\infty}$, compared to the (actual) finite-size system average sum rate.

7.3 MCP and CBf with Noisy Analog Feedback

In this section, we will discuss the large system results and feedback optimization for the MCP and CBf by using the analog feedback model discussed in Section 7.2.3. First, the large system limit expression for the SINR is derived. Then, the corresponding optimal regularization parameter that maximizes the limiting SINR is investigated. Finally, the optimal ν that maximizes the limiting SINR that already incorporates the optimal regularization parameter will be discussed.

7.3.1 MCP

We start with the theorem that states the large system limit of the SINR (7.2).

Theorem 7.1. *Let* $\rho_{M,AF} = (\omega_d+\omega_c)^{-1}\alpha/N$ *and* $g(\beta,\rho)$ *be the solution of* $g(\beta,\rho) = \left(\rho + \frac{\beta}{1+g(\beta,\rho)}\right)^{-1}$. *In the large system limit, the SINR of MCP given in* (7.2) *converges in probability to a deterministic quantity given by*

$$SINR^{\infty}_{MCP,AF} = \gamma_e g(\beta,\rho_{M,AF}) \frac{1+\frac{\rho_{M,AF}}{\beta}(1+g(\beta,\rho_{M,AF}))^2}{\gamma_e + (1+g(\beta,\rho_{M,AF}))^2}, \tag{7.13}$$

where the effective SNR γ_e *is expressed as*

$$\gamma_e = \frac{\omega_d+\omega_c}{\delta_d+\delta_c+\frac{1}{\gamma_d}} = \frac{1-\delta_d+\epsilon-\delta_c}{\delta_d+\delta_c+\frac{1}{\gamma_d}}. \tag{7.14}$$

Proof. See Appendix 7.7.1.1 □

It is obvious from above that the limiting SINR is the same for all users in both cells. This is due to the channel statistics of all users in both cells are the same. The channel uncertainty, captured by $\omega_\bullet$ and $\delta_\bullet$, affects the system performance (limiting SINR) via the effective SNR and regularization parameter $\rho_{\text{M,AF}}$.

As discussed previously, the (effective) regularization parameter $\rho_{\text{M,AF}}$ controls the amount of interference introduced to the users and provides the trade-off between suppressing the inter-user interference and increasing desired signal energy. The optimal choice of $\rho_{\text{M,AF}}$ that maximizes (7.13) is given in the following.

Corollary 7.1. *The optimal $\rho_{M,AF}$ that maximizes $SINR^{\infty}_{MCP,AF}$ is*

$$\rho^{*}_{M,AF} = \frac{\beta}{\gamma_e}, \tag{7.15}$$

and the corresponding limiting SINR is

$$SINR^{*,\infty}_{MCP,AF} = g(\beta, \rho^{*}_{M,AF}). \tag{7.16}$$

Proof. It is obvious that the limiting SINR (7.13) has the same structure as (3.16). Hence, the proof follows Appendix 3.6.1. □

It is interesting to see that the limiting SINR expression with $\rho^{*}_{\text{M,AF}}$ becomes simpler and it depends only the cell-loading (β) and the effective SNR. Clearly from (7.14), γ_e is a function of the total MSE, $\delta_t = \delta_d + \delta_c$, that can be considered as a reasonable measure of the CSIT quality. Thus, $\rho^{*}_{\text{M,AF}}$ adjusts its value as δ_t changes. Now, from (7.14), it is obvious that γ_e is a decreasing function of δ_t. As a result, $\rho^{*}_{\text{M,AF}}$ is increasing with δ_t. In other words, if the total quality of CSIT improves then $\rho^{*}_{\text{M,AF}}$ gets smaller. In the perfect CSIT case, i.e., $\delta_t = 0$, and in the high SNR, $\rho^{*}_{\text{M,AF}}$ goes to zero and we have the ZF precoder.

Now, we will investigate how to allocate ν to maximize the limiting SINR (7.16), or equivalently $g(\beta, \rho^{*}_{\text{M,AF}})$. ν is captured by γ_e (or $\rho^{*}_{\text{M,AF}}$) via δ_d. It can be shown that g is decreasing (increasing) in $\rho_{\text{M,AF}}$ (γ_e). Then, for a fixed β the limiting SINR is maximized by solving the following optimization problem

$$\max_{\nu\in[0,1]} \quad \gamma_e = \frac{\epsilon - \delta_c + 1 - \delta_d}{(\delta_d + \delta_c) + \frac{1}{\gamma_d}}.$$

As mentioned earlier, γ_e is a decreasing function of δ_t. Thus, the optimization problem above can be rewritten as

$$\min_{\nu\in[0,1]} \quad \delta_t = \delta_d + \delta_c = \frac{1}{\nu\bar{\gamma}_u + 1} + \frac{\epsilon}{(1-\nu)\bar{\gamma}_u + 1}. \tag{7.17}$$

From the above, it is very interesting to note that *the optimal ν that maximizes $SINR^{*,\infty}_{MCP}$ is the same as the one that minimizes the total MSE, δ_t.*

It is easy to check that the optimization problem above is a convex program and the optimal ν, denoted by ν^*, can be expressed as follows

$$\nu^* = \begin{cases} 0, & \sqrt{\epsilon} \geq \bar{\gamma}_u + 1 \\ 1, & \sqrt{\epsilon} \leq \frac{1}{\bar{\gamma}_u + 1} \\ \frac{1+\frac{1}{\bar{\gamma}_u}(1-\sqrt{\epsilon})}{1+\sqrt{\epsilon}}, & \text{otherwise.} \end{cases} \tag{7.18}$$

As a result, for $\sqrt{\epsilon} \leq \frac{1}{\bar{\gamma}_u+1}$, the BSs should not waste resources trying to learn about the "interfering" channel states. In this situation, *the coordination breaks down* and the base stations perform SCP. The completely opposite scenario, in which the BSs should not learn the "direct" channels, occurs when $\sqrt{\epsilon} \geq \bar{\gamma}_u + 1$. Clearly, this can only happen if $\epsilon > 1$. When $\sqrt{\epsilon} \geq \bar{\gamma}_u + 1$, the BSs also perform SCP but each BS transmits to the users in the neighboring cell.

We end this subsection by characterizing the behavior of γ_e (equivalently $\text{SINR}_{\text{MCP}}^{*,\infty}$), after optimal feedback power allocation, as the cross channel gain ϵ varies. This also implicitly shows how the total MSE, δ_t, affects the limiting SINR. Let $\breve{\gamma}_u = \frac{\bar{\gamma}_u}{(1+\epsilon)}$. We analyze the different cases in (7.18) separately.

1) Case $\sqrt{\epsilon} \leq \frac{1}{\bar{\gamma}_u+1}$ *:* This is the case when the BSs perform SCP for the users in their own cell. For fixed $\breve{\gamma}_u$, this inequality is equivalent to $\epsilon \leq \epsilon_{\text{max}}^{\text{SCP}}$, where $\epsilon_{\text{max}}^{\text{SCP}} \geq 0$ satisfies $\sqrt{\epsilon_{\text{max}}^{\text{SCP}}} = \frac{1}{\breve{\gamma}_u(1+\epsilon_{\text{max}}^{\text{SCP}})+1}$. Now, by taking the first derivative $\frac{\partial \gamma_e}{\partial \epsilon}$ and setting it to zero, the (unique) stationary point is given by

$$\epsilon_{\text{AF}}^{\text{SCP}} = \frac{1}{\sqrt{\gamma_d \breve{\gamma}_u}} - 1.$$

If $\sqrt{\epsilon_{\text{AF}}^{\text{SCP}}} \in [0, \sqrt{\epsilon_{\text{max}}^{\text{SCP}}}]$, it is easy to check that the limiting SINR is increasing until $\epsilon = \epsilon_{\text{AF}}^{\text{SCP}}$ and then decreasing. If $\sqrt{\gamma_d \breve{\gamma}_u} > 1$ then $\epsilon_{\text{AF}}^{\text{SCP}} < 0$, or equivalently, $\frac{\partial \gamma_e}{\partial \epsilon} < 0$. Consequently, for this case, the limiting SINR is decreasing in ϵ. Moreover, $\sqrt{\epsilon_{\text{AF}}^{\text{SCP}}} \geq \sqrt{\epsilon_{\text{max}}^{\text{SCP}}}$ if the following condition holds

$$\sqrt{\gamma_d \breve{\gamma}_u}(2 - 2\gamma_d - \breve{\gamma}_u) \geq (2\gamma_d \breve{\gamma}_u - \gamma_d - \breve{\gamma}_u), \tag{7.19}$$

in which case $\frac{\partial \gamma_e}{\partial \epsilon} > 0$, which implies that the limiting SINR always increases over ϵ.

This behavior of γ_e as a function of ϵ can be intuitively explained as follows. When $\nu = 1$, the total MSE is $\delta_t = \frac{1}{(1+\epsilon)\breve{\gamma}_u+1} + \epsilon$, where the first and second terms are δ_d and δ_c, respectively. As ϵ increases, δ_d decreases whereas δ_c increases. This shows that there is a trade-off between the quality of the direct channel and the strength of the interference. The trade-off is also influenced by parameters γ_d and $\breve{\gamma}_u$. As shown in the analysis, when $\sqrt{\gamma_d \breve{\gamma}_u} > 1$, the effect of cross channel to the limiting SINR dominates. In contrast, if the condition in (7.19) is satisfied, the effect of the quality of the direct channel (δ_t) becomes dominant. If the aforementioned conditions do not hold, δ_t causes the SINR to increase until $\epsilon_{\text{AF}}^{\text{SCP}}$ and after that the interference from the cross channel takes over as the dominant factor, thereby reducing the limiting SINR.

2) Case $\bar{\gamma}_u + 1 \geq \sqrt{\epsilon} \geq \frac{1}{\bar{\gamma}_u+1}$ *:* Here, the BSs perform MCP. By taking $\frac{\partial \gamma_e}{\partial \epsilon}$ in that interval of ϵ, it can be shown that we have a unique stationary which we denote as $\sqrt{\epsilon_{\text{AF}}^{\text{M}}}$. We can also show that γ_e is a convex function for $\epsilon \in [0, 1]$ and is increasing for $\epsilon \geq 1$. Thus, if $\frac{1}{\bar{\gamma}_u+1} \leq \sqrt{\epsilon_{\text{AF}}^{\text{M}}} \leq \bar{\gamma}_u + 1$, the limiting SINR will decrease for $\sqrt{\epsilon} \in [\frac{1}{\breve{\gamma}_u(1+\epsilon)+1}, \sqrt{\epsilon_{\text{AF}}^{\text{M}}}]$ and increase after that; Otherwise, the limiting SINR increases in the region. Here, for $\sqrt{\epsilon} \in [\frac{1}{\bar{\gamma}_u+1}, 1]$, we still can see the effect of the trade-off within δ_t to the limiting SINR as ϵ changes. In that interval, the quality of the direct channel becomes better as ϵ increases; However, that of the cross channel decreases and this affects the SINR badly until $\epsilon_{\text{AF}}^{\text{M}}$. After this point, the improvement in the quality of the direct channel will outweigh the deterioration of that of the cross channel, causing the SINR to increase.

3) Case $\sqrt{\epsilon} \geq \bar{\gamma}_u + 1$ *:* In this case, each BS performs SCP, but serves the other cell's users. We can establish that $\frac{\partial \gamma_e}{\partial \epsilon} > 0$. Hence, for this case, the limiting SINR is increasing in ϵ.

7.3.2 Coordinated Beamforming

Theorem 7.2. *Let $\rho_{C,AF} = \frac{\alpha}{N}$, and let Γ_A be the solution of the following cubic equation*

$$\Gamma_A = \frac{1}{\rho_{C,AF} + \frac{\beta\omega_c}{1+\omega_c\Gamma_A} + \frac{\beta\omega_d}{1+\omega_d\Gamma_A}}. \tag{7.20}$$

In the large system limit, the SINR of the coordinated beamforming given in (7.3) converges almost surely to a deterministic quantity given by

$$SINR^{\infty}_{CBf,AF} = \frac{\frac{\omega_d}{\beta}\Gamma_A\left[\rho_{C,AF} + \frac{\beta\omega_c}{(1+\omega_c\Gamma_A)^2} + \frac{\beta\omega_d}{(1+\omega_d\Gamma_A)^2}\right]}{\left(\frac{1}{\gamma_d} + \delta_d + \delta_c + \frac{\omega_d}{(1+\omega_d\Gamma_A)^2} + \frac{\omega_c}{(1+\omega_c\Gamma_A)^2}\right)}. \tag{7.21}$$

Proof. See Appendix 7.7.2.1 □

Similar to the MCP case, the limiting SINR expression (7.21) is the same for all users. Comparing (7.15) and (7.22), we can see that $\rho_{\text{C,AF}} = \rho_{\text{M,AF}}$ for a given α. For $\nu \in [0, 1]$, the optimal $\rho_{\text{C,AF}}$ that maximizes the limiting SINR (7.21) is given in the following.

Corollary 7.2. *The limiting SINR (7.21) is maximized by choosing the regularization parameter according to*

$$\rho^*_{C,AF} = \beta\left(\frac{1}{\gamma_d} + \delta_d + \delta_c\right) \tag{7.22}$$

and the corresponding limiting SINR is

$$SINR^{*,\infty}_{CBf,AF} = \omega_d\Gamma^*_A, \tag{7.23}$$

where Γ^*_A is Γ_A with $\rho_{C,AF} = \rho^*_{C,AF}$.

Proof. Let $\gamma_{eff} = \beta\left(\gamma_d^{-1} + \delta_d + \delta_c\right)$ and $\Psi = \frac{\beta\omega_d}{(1+\omega_d\Gamma_A)^2} + \frac{\beta\omega_c}{(1+\omega_c\Gamma_A)^2}$. It is easy to show that

$$\frac{\partial \text{SINR}^{\infty}_{\text{CBf,AF}}}{\partial\rho_{\text{C,AF}}} = \omega_d\frac{\gamma_{eff} - \rho_{\text{C,AF}}}{[\gamma_{eff} + \Psi]^2}\frac{\partial\Psi}{\partial\rho_{\text{C,AF}}}, \tag{7.24}$$

where $\frac{\partial\Psi}{\partial\rho_{\text{C,AF}}} = -2\beta\frac{\partial\Gamma_A}{\partial\rho_{\text{C,AF}}}\left(\frac{\omega_d^2}{(1+\omega_d\Gamma_A)^3} + \frac{\omega_c^2}{(1+\omega_c\Gamma_A)^3}\right) > 0$ with $\frac{\partial\Gamma_A}{\partial\rho_{\text{C,AF}}} < 0$ is given by (7.56). Thus, it follows that $\rho^*_{\text{C,AF}} = \gamma_{eff}$ is the unique stationary point and the global optimizer. Plugging back $\rho_{\text{C,AF}}$ into (7.21) yields (7.23). □

Similar to the MCP case, the corollary above shows that the optimal regularization parameter adapts to the changes of CSIT quality and it is a decreasing function of δ_t.

Finding ν that maximizes the limiting SINR of the CBf is more complicated than in the MCP case. It is equivalent to maximizing $\omega_d\Gamma$, such that $\nu \in [0,1]$: this is a non-convex program. The maximizer ν^* is the boundaries of the feasible set ($\nu = \{0,1\}$) or the stationary point, denoted by ν°, which is the solution of

$$\nu^\circ = -\frac{\Gamma_A^*}{\frac{\partial \Gamma_A^*}{\partial \rho_{\text{C,AF}}^*}(1+\nu^\circ \bar{\gamma}_u)}. \tag{7.25}$$

The point $\nu = 0$ can be eliminated from the feasible set since the derivative of the limiting SINR with respect to ν at this point is always positive.

7.3.3 Numerical Results

Since propagation channels fluctuate, the SINR expressions in (7.2) and (7.3) are random quantities. Consequently, the average sum rates are also random. Figure 7.2 illustrates how the random average sum rates approach the limiting sum rates as the dimensions of the system increase. This is quantified by the normalized sum rate difference which is defined in (7.12). The average sum rate is obtained by averaging the sum rates over 1000 channel realizations. The optimal regularization parameter and power splitting obtained in the large system analysis are used in computing the limiting and average sum rates. We can see that as the system size increases, the normalized sum rate difference becomes smaller and this hints that the approximation of the average sum rate by the limiting sum rate becomes more accurate. The difference is already about 1.3% and 0.5% for the MCP and CBf respectively for $N = 60, K = 36$.

Figure 7.3 describes the applicability of the large system results into finite-size systems. We choose a reasonable system-size in practice, i.e., $N = 10, K = 6$. Then, 250 channel realizations are generated. For each channel realization, with a fixed regularization parameter of the precoder, the optimal ν, denoted by ν_{FS}^*, is computed. Then the resulting average sum rate is compared to the average sum rate that using ν^* from the large system analysis, i.e., (7.18) and (7.25), for different values of ϵ. We can see that

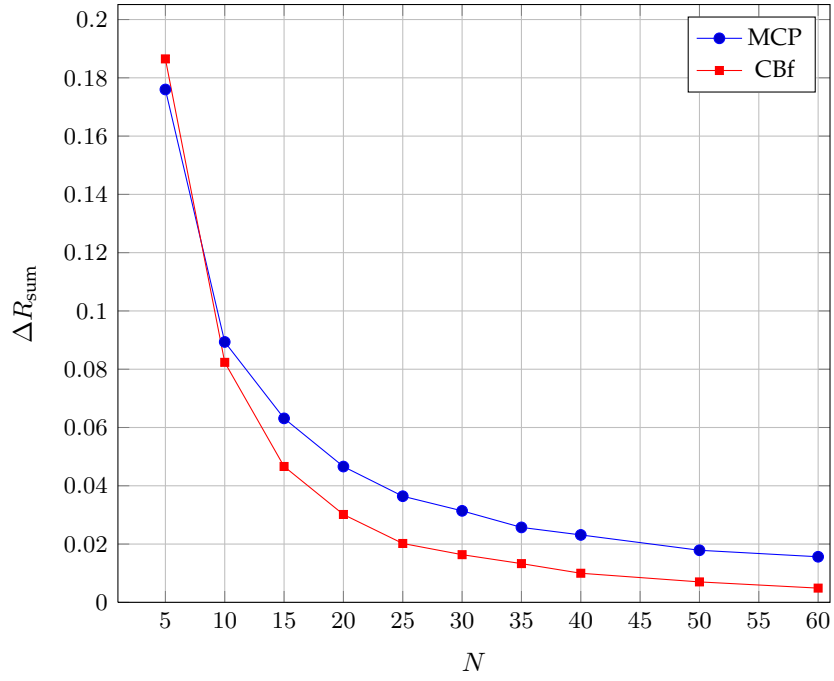


Figure 7.2: The normalized sum-rate difference for different system dimensions with $\beta = 0.6$, $\epsilon = 0.5$ $\gamma_d = 10$ dB and $\gamma_u = 0$ dB.

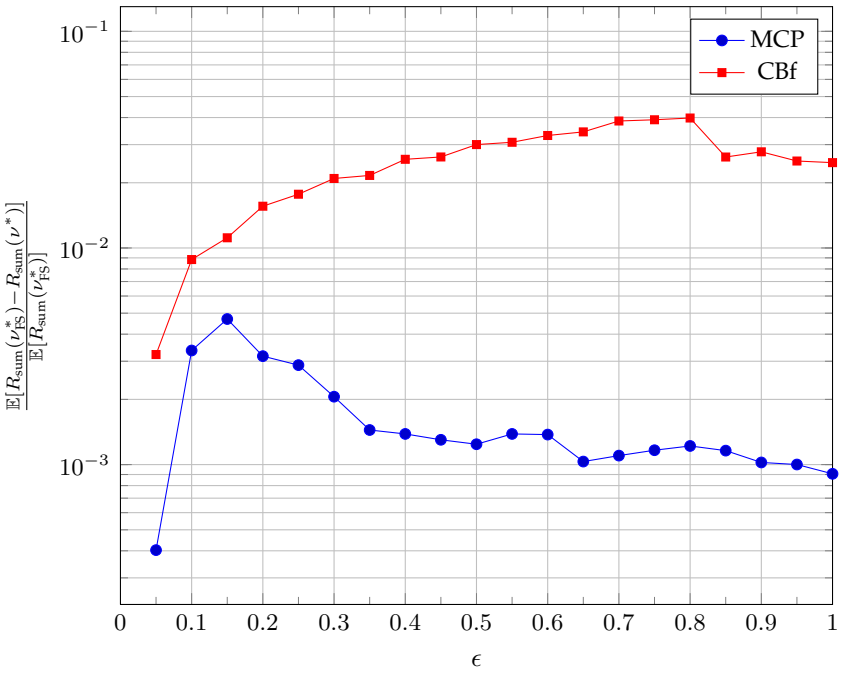


Figure 7.3: The normalized average sum-rate difference of the finite-size system by using the ν_{FS} and ν^* with $N = 10, \beta = 0.6$, $\gamma_d = 10$ dB and $\gamma_u = 0$ dB.

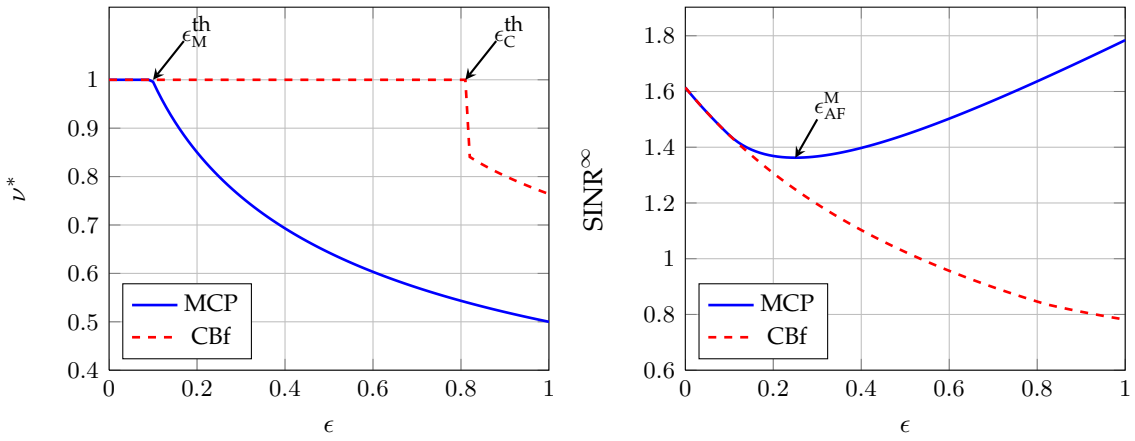


Figure 7.4: (a) The optimal ν^* and (b) the limiting SINR for the MCP and CBf scheme as ϵ varies in $[0, 1]$ with $\beta = 0.6$, $\gamma_d = 10$ dB, $\gamma_u = 0$ dB.

the normalized average sum rate difference, i.e., $\frac{\mathbb{E}[|R_{\mathrm{sum}}(\nu_{\mathrm{FS}}) - R_{\mathrm{sum}}(\nu^*)|]}{R_{\mathrm{sum}}(\nu_{\mathrm{FS}})}$, for CBf has a peak around 4% that can be considered as a reasonable value for the chosen system size. For MCP, it is less than 0.47% which is about ten times smaller compared to that for CBf. To this end, our simulation results indicate that the large system results discussed earlier approximate the finite-system quite well.

In the following, we present some numerical simulations that visualize the characteristics of the optimal ν^* (in the large system limit) and the corresponding limiting SINR for each cooperation scheme. We are primarily interested in their characteristics when the interfering channel gain ϵ varies, as depicted in Figure 7.4. In general, we can see that for the same system parameters, the CBf scheme allocates more power to feed back the direct channel compared to the MCP. From Figure 7.4(a), we can see that for values of ϵ ranging from 0 up to a certain threshold (denoted by $\epsilon_{\mathrm{M}}^{\mathrm{th}} = \epsilon_{\max}^{\mathrm{SCP}}$ and $\epsilon_{\mathrm{C}}^{\mathrm{th}}$ for MCP and CBf respectively), the optimal ν is 1: in other words, it is optimal in this range for the BSs not to try to get information about the cross channels and to construct the precoder based on the direct channel information only. Effectively, the two schemes reduce to the SCP scheme when $\nu^* = 1$: as a result, the same limiting SINR is achieved by both schemes.

We also observe a peculiar behavior of the limiting SINR of MCP, which we already highlighted in the analysis of Section 7.3.1. When $\sqrt{\epsilon} \leq \frac{1}{\bar{\gamma}_u + 1}$, i.e. when $\nu^* = 1$, the

SINR is decreasing as ϵ increases. After that the SINR is still decreasing until ϵ reaches $\epsilon_{\text{AF}}^{\text{M}}$ and then increasing: this reflects the trade-off between δ_c and δ_d. Note that this initial decrease does not occur in the perfect CSI case where the SINR is strictly increasing in ϵ for MCP. Similar to the MCP case, we can see that the limiting SINR of CBf is decreasing in ϵ when $\nu^* = 1$ (SCP). Moreover, it is still decreasing when both BSs perform CBf.

7.4 Quantized Feedback via Random Vector Quantization (RVQ)

In this section, we will derive the approximations of the SINR for the MCP (7.2) and CBf (7.3) by analyzing them in the large system limit. We use these approximations to optimize the feedback bit allocation and regularization parameter for maximizing the limiting SINR. This joint optimization problem can be split into two steps. First, we derive the optimal bit allocation for the direct and cross links, i.e., the optimal $\bar{B}_d = \frac{B_d}{N}$ and $\bar{B}_c = \frac{B_c}{N}$, respectively. Plugging the optimal bit allocation back into the limiting SINR expression, we can then proceed to the second step where we obtain the optimal regularization parameter. At the end of the section, some comparisons of the limiting SINR and bit allocation values for the two schemes are illustrated.

7.4.1 MCP

Theorem 7.3. *Let $\rho_{M,Q} = (1+\epsilon)^{-1}\alpha/N$ and $g(\beta,\rho)$ be the solution of $g(\beta,\rho) = \left(\rho + \frac{\beta}{1+g(\beta,\rho)}\right)^{-1}$. In the large system limit, the SINR in (7.2) converges in probability to a deterministic quantity given by*

$$SINR_{MCP,Q}^{\infty} = \gamma_e g(\beta,\rho_{M,Q}) \frac{1 + \frac{\rho_{M,Q}}{\beta}(1+g(\beta,\rho_{M,Q}))^2}{\gamma_e + (1+g(\beta,\rho_{M,Q}))^2}, \tag{7.26}$$

where

$$\gamma_e = \frac{d^2}{1-d^2+\frac{1}{\gamma_d(1+\epsilon)}} \tag{7.27}$$

is defined as the effective SNR and

$$d = \frac{\sqrt{1-2^{-\bar{B}_d}} + \epsilon\sqrt{1-2^{-\bar{B}_c}}}{1+\epsilon}. \tag{7.28}$$

Proof. Refer to Appendix 7.7.1.2. □

Theorem 7.3 shows that the limiting SINR is the same for all users in both cells. This is not surprising given the symmetry in their channel statistics and feedback mechanisms. Moreover, the only dependence of the limiting SINR on the bit allocation is via γ_e, which itself is a function of d: d can be interpreted as a measure of the total quality of the channel estimates; In fact, given that $\bar{B}_d$ and $\bar{B}_c$ are constrained to sum up to $\bar{B}_t$, d in (7.28) highlights a trade-off between increasing feedback bits for direct channel and cross channel. Comparing (7.13) and (7.26), we can immediately recognize an identical structure between them. The effective SNR expressions (7.14) and (7.27) also share a similar construction, where $(1+\epsilon)d^2$ in (7.28) can be thought to be equivalent to $\omega_d + \omega_c$.

Now, we move tho the first step of the joint optimization i.e., determining the optimal bit allocation that maximizes (7.26). It is clear from (7.26) that $\bar{B}_d$ and $\bar{B}_c$ contributes to the limiting SINR through d. It is easy to check that the limiting SINR is an increasing and a convex function of d. Thus, maximizing $\text{SINR}^{\infty}_{\text{MCP,Q}}$ is equivalent to maximizing d, i.e. solving (cf. Eq. (7.28))

$$\max_{x_d \in [X_t, 1]} \sqrt{1 - x_d} + \epsilon \sqrt{1 - \frac{X_t}{x_d}}, \tag{7.29}$$

where $X_t = 2^{-\bar{B}_t}$, $\bar{B}_t = \frac{B_t}{N}$ and $x_d = 2^{-\bar{B}_d}$. The solution of (7.29) is presented in the following theorem.

Theorem 7.4. *$\text{SINR}^{\infty}_{\text{MCP,Q}}$ is maximized by allocating $\bar{B}_d = -\log_2(x_d^*)$ bits to feed back the direct channel information, and $\bar{B}_c = \bar{B}_t - \bar{B}_d$ to feed back the interfering channel information, where x_d^* is the positive (real) solution of the following quartic equation*

$$x_d^4 - X_t x_d^3 + (\epsilon X_t)^2 (x_d - 1) = 0. \tag{7.30}$$

Proof. The first derivative of the objective function over x_d is given by

$$(1+\epsilon)\frac{\partial \mathbb{E}[d]}{\partial x_d} = \frac{1}{2}\left(-\frac{1}{\sqrt{1 - x_d}} + \frac{1}{x_d^2} \frac{\epsilon X_t}{\sqrt{1 - \frac{X_t}{x_d}}} \right) \tag{7.31}$$

and $\lim_{x_d \to X_t} \frac{\partial \mathbb{E}[d]}{\partial x_d} = \infty, \lim_{x_d \to 1} \frac{\partial \mathbb{E}[d]}{\partial x_d} = -\infty$. Moreover, the objective function is a concave function in x_d since

$$(1+\epsilon)\frac{\partial^2 \mathbb{E}[d]}{\partial x_d^2} = \frac{1}{2}\left(-\frac{1}{2}(1-x_d)^{-3/2} - \frac{2}{x_d^3}\frac{\epsilon X_t}{\sqrt{1-\frac{X_t}{x_d}}} - \frac{1}{2}\frac{\epsilon X_t}{x_d^4}\left(1-\frac{X_t}{x_d}\right)^{-3/2}\right) < 0,$$

for $x_d \in [X_t, 1]$. The stationary point, x_d^*, is obtained by setting the derivative equal to 0 and it is the non-negative (real) solution of

$$x_d^4 - X_t x_d^3 + (\epsilon X_t)^2(x_d - 1) = 0.$$

Since the objective function is concave over x_d, x_d^* is the global optimizer. □

Now, let us discuss how the optimal bit allocation vary with ϵ. Since $x_d = x_d^*$ satisfies (7.30), then by taking the (implicit) derivative of (7.30) w.r.t. ϵ, we have

$$\frac{\partial x_d^*}{\partial \epsilon} = \frac{2\epsilon X_t^2(1-x_d)}{4x_d^3 - 3X_t x_d^2 + (\epsilon X_t)^2} > 0, \quad \text{for } X_t \le x_d^* \le 1.$$

This implies that as ϵ increases, x_d^* ($\bar{B}_d^*$) increases (decreases). This is consistent with the intuition that for higher ϵ, more resources would be allocated to quantize the cross channel information. At one of the extremes, i.e., $\epsilon = 0$, $x_d^* = X_t$, or $\bar{B}_d = \bar{B}_t$. If $\epsilon = 0$, $x_d^* = X_t$, so that when there is no interference from the neighboring BS, all feedback bits are used to convey the direct channel states, as expected. At the other extreme, when $\epsilon \to \infty$, $x_d^* \to 1$ or $\bar{B}_d \to 0$. This can be shown by setting the derivative (7.31) equal to zero and we have

$$\frac{1}{\epsilon} = \frac{X_t\sqrt{1-x_d}}{x_d^2\sqrt{1-\frac{X_t}{x_d}}}.$$

As $\epsilon \to \infty$, the left hand side goes to zero and the stationarity is achieved by setting $x_d = 1$.

It is also interesting to see how d, after optimal bit allocation, behaves as the cross channel gain varies. Let d^* is d evaluated at $x_d = x_d^*$. By taking $\frac{\partial d^*}{\partial \epsilon}$, we can show the following property.

Proposition 7.1. *For $\epsilon \leq 1$, d^* is decreasing in ϵ and increasing for $\epsilon \geq 1$. Consequently, d^* is minimum at $\epsilon = 1$.*

As mentioned previously, x_d^* increases and consequently $1 - x_d^*$ decreases as ϵ increases. On the other side, $\epsilon\sqrt{1 - X_t/x_d^*}$ is getting larger. So, from the calculation we can conclude that d^* is mostly affected by $\sqrt{1 - x_d^*}$ for $\epsilon \leq 1$, while for the other values of ϵ, the other term takes over.

We now proceed to find the optimal $\rho_{\text{M,Q}}$ that maximizes $\text{SINR}_{\text{MCP,Q}}^{\infty}$.

Theorem 7.5. *Let γ_e^* be γ_e evaluated at $d = d^*$. The optimal ρ_M that maximizes $SINR_{MCP}^{\infty}(d^*)$ is*

$$\rho_{M,Q}^* = \frac{\beta}{\gamma_e^*}. \tag{7.32}$$

The corresponding limiting SINR is given by

$$SINR_{MCP}^{*,\infty} = g\left(\beta, \rho_{M,Q}^*\right).$$

Proof. The equation (7.26) has the same structure as (7.13) and thus, (7.32) follows. □

From Theorem 7.5, d^* affects the regularization parameter and the limiting SINR via effective SNR γ_e^*. The latter grows with d^* (cf. (7.27)). Thus, $\rho_{\text{M,Q}}^*$ declines as the CSIT quality, d^*, increases and this behavior is also observed for the cooperation schemes with the analog feedback.

In Proposition 7.1, we established how d^* changes with ϵ. We can show that γ_e^* has a similar behavior but reaches its minimum at a different value of ϵ due to the last term in the denominator in (7.27). For $\text{SINR}_{\text{MCP}}^{*,\infty}$, it attains its minimum at $\epsilon = \epsilon_{\text{Q}}^{\text{M}}$, as described in the next proposition.

Proposition 7.2. *Suppose that $\epsilon = \epsilon_{\text{Q}}^{M}$ satisfies*

$$(x_d^*)^2 = \frac{\gamma_d(1+\epsilon) - \frac{1}{2}}{\epsilon X_t \left[\gamma_d(1+\epsilon) + 1 + \frac{\epsilon}{2}\right]}. \tag{7.33}$$

Then, $SINR_{MCP,Q}^{*,\infty}$ decreasing for $\epsilon \leq \epsilon_{\text{Q}}^{M}$ and increasing for $\epsilon \geq \epsilon_{\text{Q}}^{M}$.

The characterization of $\text{SINR}^{*,\infty}_{\text{MCP,Q}}$ above reminds us a similar behavior of $\text{SINR}^{*,\infty}_{\text{MCP,AF}}$ after optimal power allocation. We can conclude that *the limiting SINR of MCP under both feedback schemes has a common behavior as ϵ varies*.

7.4.2 Coordinated Beamforming

Theorem 7.6. *Let $\rho_{\text{C,Q}} = \alpha/N$ and Γ_Q be the solution of the following cubic equation*

$$\Gamma_Q = \frac{1}{\rho_{\text{C,Q}} + \frac{\beta}{1+\Gamma_Q} + \frac{\beta\epsilon}{1+\epsilon\Gamma_Q}}. \tag{7.34}$$

Let $\phi_d = 1 - 2^{-\bar{B}_d}$, $\phi_c = 1 - 2^{-\bar{B}_c}$, $\delta_d = 2^{-\bar{B}_d}$ and $\delta_c = \epsilon 2^{-\bar{B}_c}$. In the large system limit, the SINR (7.3) for the quantized feedback via RVQ converges weakly to a deterministic quantity given by

$$SINR^{\infty}_{\text{CBf,Q}} = -\frac{\phi_d\Gamma_Q^2}{\beta\left(\frac{1}{\gamma_d} + \frac{\phi_d}{(1+\Gamma_Q)^2} + \frac{\phi_c\epsilon}{(1+\epsilon\Gamma_Q)^2} + \delta_d + \delta_c\right)\frac{\partial\Gamma_Q}{\partial\rho}}, \tag{7.35}$$

where

$$-\frac{\partial\Gamma_Q}{\partial\rho_{\text{C,Q}}} = \frac{\Gamma_Q}{\rho_{\text{C,Q}} + \frac{\beta\epsilon}{(1+\epsilon\Gamma_Q)^2} + \frac{\beta}{(1+\Gamma_Q)^2}}.$$

Proof. See Appendix 7.7.2.2 □

As in Theorem 7.3, Theorem 7.6 shows that that the limiting SINR is the same for all users. The quantization error variance of estimating the direct channel, δ_d, affects both the signal strength (via ϕ_d) and the interference energy, in which it captures the effect of the *intra-cell* interference. δ_c, on the other hand, only contributes to the interference term: It represents the quality of the cross channel and determines the strength of the *inter-cell* interference. Since $\bar{B}_t$ is fixed, increasing $\bar{B}_d$, or equivalently reducing $\bar{B}_c$, will strengthen the desired signal and reduce the intra-cell interference: it does so, however, at the expense of strengthening the inter-cell interference. Thus, feedback bits' allocation is important in order to improve the performance of the system.

To solve the joint optimization problem, it is useful to write (7.35) as follows

$$\text{SINR}_{\text{CBf,Q}}^{\infty} = G_1 \frac{1 - x_d}{\frac{1}{\gamma_d} + (1 - x_d)(G_2 - 1) + \epsilon \left(1 - \frac{X_t}{x_d}\right)(G_3 - 1) + 1 + \epsilon},$$

where x_d and X_t are defined as in the previous subsection. For brevity, we denote: $G_1 = -\Gamma_Q^2 \left(\beta \frac{\partial \Gamma_Q}{\rho_{\text{C,Q}}}\right)^{-1}$, $G_2 = (1 + \Gamma_Q)^{-2}$ and $G_3 = (1 + \epsilon\Gamma_Q)^{-2}$. The optimal bit allocation can be found by solving the following optimization problem

$$\max_{x_d \in [X_t, 1]} \quad \text{SINR}_{\text{CBf,Q}}^{\infty}. \tag{7.36}$$

The solution of (7.36) is summarized in the following theorem.

Theorem 7.7. *For a fixed $\bar{B}_t$, the optimal bit allocation, in terms of $x_d = 2^{-\bar{B}_d}$, that maximizes $SINR_{CBf,Q}^{\infty}$ is given by*

$$x_d^* = \begin{cases} X_t, & \epsilon \leq \frac{X_t(\frac{1}{\gamma_d}+1)}{1-G_3-X_t(2-G_3)} = \epsilon_{th} \\ X_d = \frac{\epsilon X_t(G_3-1)+\sqrt{\epsilon^2 X_t^2(G_3-1)^2-\epsilon X_t\left(\frac{1}{\gamma_d}+1+\epsilon G_3\right)(G_3-1)}}{\frac{1}{\gamma_d}+1+\epsilon G_3}, & \textit{otherwise.} \end{cases} \tag{7.37}$$

Proof. Differentiating the objective function (7.36), we get

$$\frac{\partial \text{SINR}_{\text{CBf,Q}}^{\infty}}{\partial x_d} = G_1 \frac{-x_d^2(\frac{1}{\gamma_d} + 1 + \epsilon G_3) + \epsilon(G_3 - 1)(2X_t x_d - X_t)}{x_d^2 \left(\frac{1}{\gamma_d} + (1 - x_d)(G_2 - 1) + \epsilon \left(1 - \frac{X_t}{x_d}\right)(G_3 - 1) + 1 + \epsilon\right)^2}$$

and the stationary is given by

$$x_d^{\circ} = \frac{\epsilon X_t(G_3 - 1) + \sqrt{\epsilon^2 X_t^2(G_3 - 1)^2 - \epsilon X_t(\frac{1}{\gamma_d} + 1 + \epsilon G_3)(G_3 - 1)}}{\frac{1}{\gamma_d} + 1 + \epsilon G_3}. \tag{7.38}$$

Now let us consider the term $Z = -x_d^2(\frac{1}{\gamma_d} + 1 + \epsilon G_3) + \epsilon(G_3 - 1)(2X_t x_d - X_t)$ in the numerator. It can be verified that the sign of Z is the same as the sign of $\frac{\partial \text{SINR}_{\text{CBf,Q}}^{\infty}}{\partial x_d}$. Thus, $X_d = x_d^{\circ}$ will be the unique positive solution of the quadratic equation $Z = 0$.

It can be also checked that $\frac{\partial Z}{\partial x_d} = -2x_d(\frac{1}{\gamma_d} + 1 + \epsilon G_3) + \epsilon(G_3 - 1)(2X_t) < 0$ and thus,

Z is decreasing in x_d. Since at $x_d = 1$, $Z < 0$, we should never allocate $x_d^* = 1$. We will allocate $x_d = X_t$ if $Z \leq 0$ at $x_d = X_t$ (this condition is satisfied whenever $\epsilon \leq \epsilon_{\text{th}}$). □

Unlike the MCP case where $x_d^* = X_t$ *only when* $\epsilon = 0$, in the CBf, it is optimal for a user to allocate all B_t to the direct channel when $0 \leq \epsilon \leq \epsilon_{\text{th}}$. Note that $x_d^* = X_t$ does not imply that the cooperation breaks down or that both BSs perform single-cell processing. It is easy to check that ϵ_{th} increases when $\bar{B}_t$ or γ_d is decreased. This suggests that when the resource for the feedback bits is scarce or the received SNR is low then it is preferable for the user to allocate all the feedback bits to quantize the direct channel. So, in this situation, quantizing the cross channel does more harm to the performance the system. However, as ϵ increases beyond ϵ_{th}, quantizing the cross channel will improve the SINR. We can show that x_d^*, particularly X_d, is increasing in ϵ. In doing that, we need to take the derivative of X_d over ϵ. It is easy to show that Γ_Q is decreasing in ϵ. Then, it follows that G_3 is decreasing in ϵ. Using this fact, we can then show $\frac{\partial X_d}{\partial \epsilon} > 0$. So, as in the case of MCP, this suggests that more resources are allocated to feedback the cross-channel when ϵ increases.

Once we have the optimal bit allocation, we can find the optimal $\rho_{\text{C,Q}}$, as we did for the MCP. For that purpose, we can rewrite (7.37) w.r.t $\rho_{\text{C,Q}}$ as follows

$$x_d^* = \begin{cases} X_t, & \rho_{\text{C,Q}} \geq \rho_{\text{th}} \\ X_d, & \text{otherwise,} \end{cases}$$

where for given X_t, ϵ and γ_d, the threshold ρ_{th} satisfies $\epsilon = \epsilon_{\text{th}}$. So, we have $\text{SINR}_{\text{CBf,Q}}^{\infty}(X_d)$ for $\rho_{\text{C,Q}} < \rho_{\text{th}}$ and $\text{SINR}_{\text{CBf,Q}}^{\infty}(X_t)$ for other values of $\rho_{\text{C,Q}}$.

Now, let us investigate the optimal $\rho_{\text{C,Q}}$ when $x_d^* = X_d$. By evaluating $\frac{\partial \text{SINR}_{\text{CBf,Q}}^{\infty}(X_d)}{\partial \rho_{\text{C,Q}}} = 0$, we can determine the stationary point, which is given by

$$\rho_{X_d}^{\circ} = \Bigg\{ (1 - X_d) \left((G_2' + \epsilon G_3') \left[\frac{1}{\gamma_d} + X_d + \epsilon \frac{X_t}{X_d} \right] + \epsilon (G_2 G_3' - G_3 G_2') \left[-X_d + \frac{X_t}{X_d} \right] \right) - X_d' (G_2 + \epsilon G_3) \gamma_e \Bigg\} \frac{\beta}{X_d' \gamma_e + (1 - X_d) \left((1 - X_d) G_2' + \epsilon \left(1 - \frac{X_t}{X_d} \right) G_3' \right)},$$

where $\gamma_e = \frac{1}{\gamma_d} + 1 + \epsilon + \epsilon(G_3 - 1)\left(1 - \frac{2X_t}{X_d} + \frac{X_t}{X_d^2}\right)$ and $G_2' = \frac{\partial G_2}{\partial \rho_{\text{C,Q}}}$ and $G_3' = \frac{\partial G_3}{\partial \rho_{\text{C,Q}}}$.

We can show that the derivative is positive for $\rho_{\text{C,Q}} \in [0, \rho^{\circ}_{X_d})$ and negative for $\rho_{\text{C,Q}} \in (\rho^{\circ}_{X_d}, \infty)$. Since $\text{SINR}^{\infty}_{\text{CBf,Q}}(X_d)$ is defined for $\rho_{\text{C,Q}} \leq \rho_{\text{th}}$, if $\rho^{\circ}_{X_d} < \rho_{\text{th}}$ then $\text{SINR}^{\infty}_{\text{CBf,Q}}(X_d)$ is increasing for $\rho_{\text{C,Q}} \in [0, \rho^{\circ}_{X_d}]$ and decreasing for $\rho_{\text{C,Q}} \in [\rho^{\circ}_{X_d}, \rho_{\text{th}})$. If $\rho^{\circ}_{X_d} \geq \rho_{\text{th}}$ then $\text{SINR}^{\infty}_{\text{CBf,Q}}(X_d)$ is increasing for $\rho_{\text{C,Q}} \in [0, \rho_{\text{th}})$.

Then, we move to the case when $x_d^* = X_t$. By setting $\frac{\partial \text{SINR}^{\infty}_{\text{CBf,Q}}(X_t)}{\partial \rho_{\text{C,Q}}} = 0$, the stationary point is then given by

$$\rho^{\circ}_{X_t} = \beta \frac{(G_2' + \epsilon G_3')(X_t + 1/\gamma_d + \epsilon) + \epsilon(1 - X_t)(G_2 G_3' - G_3 G_2')}{(1 - X_t) G_2'}.$$

We can also show that the derivative is positive for $\rho_{\text{C,Q}} \in [0, \rho^{\circ}_{X_t})$ and negative for $\rho_{\text{C,Q}} \in (\rho^{\circ}_{X_t}, \infty)$. Since $\text{SINR}^{\infty}_{\text{CBf,Q}}(X_t)$ is defined for $\rho_{\text{C,Q}} \geq \rho_{\text{th}}$, if $\rho^{\circ}_{X_t} > \rho_{\text{th}}$ then $\text{SINR}^{\infty}_{\text{CBf,Q}}(X_t)$ is increasing for $\rho_{\text{C,Q}} \in [\rho_{\text{th}}, \rho^{\circ}_{X_t}]$ and decreasing for $\rho_{\text{C,Q}} \in [\rho^{\circ}_{X_t}, \infty)$. If $\rho^{\circ}_{X_t} \leq \rho_{\text{th}}$ then $\text{SINR}^{\infty}_{\text{CBf,Q}}(X_t)$ is decreasing for $\rho_{\text{C,Q}} \in [\rho_{\text{th}}, \infty)$.

In what follows, by knowing the stationary point in both regions of ρ, we will investigate how to obtain the optimal $\rho_{\text{C,Q}}$, denoted by $\rho^*_{\text{C,Q}}$, for $\rho_{\text{C,Q}} \in [0, \infty)$. By inspecting $\partial \text{SINR}^{\infty}_{\text{CBf,Q}}(X_d)/\partial \rho_{\text{C,Q}}$ and $\partial \text{SINR}^{\infty}_{\text{CBf,Q}}(X_t)/\partial \rho_{\text{C,Q}}$ we can see that that $\text{SINR}^{\infty}_{\text{CBf,Q}}(x_d^*)$ is continuously differentiable for the region, $\rho_{\text{C,Q}} \in [0, \rho_{\text{th}})$ and $\rho_{\text{C,Q}} \in [\rho_{\text{th}}, \infty)$, respectively. To show $\text{SINR}^{\infty}_{\text{CBf,Q}}(x_d^*)$ is continuously differentiable for $\rho_{\text{C,Q}} \in [0, \infty)$ we need to establish $\partial \text{SINR}^{\infty}_{\text{CBf,Q}}(x_d^*)/\partial \rho_{\text{C,Q}}$ to be continuous at $\rho_{\text{C,Q}} = \rho_{\text{th}}$, or equivalently

$$\lim_{\rho_{\text{C,Q}} \to \rho_{\text{th}}^-} \frac{\partial \text{SINR}^{\infty}_{\text{CBf,Q}}(X_d)}{\partial \rho_{\text{C,Q}}} = \lim_{\rho_{\text{C,Q}} \to \rho_{\text{th}}^+} \frac{\partial \text{SINR}^{\infty}_{\text{CBf,Q}}(X_t)}{\partial \rho_{\text{C,Q}}} = \left.\frac{\partial \text{SINR}^{\infty}_{\text{CBf,Q}}(X_t)}{\partial \rho_{\text{C,Q}}}\right|_{\rho_{\text{C,Q}} = \rho_{\text{th}}}. \tag{7.39}$$

When $\rho_{\text{C,Q}} \to \rho_{\text{th}}^-$, $X_d \to X_t$ and therefore the denominator of $\partial \text{SINR}^{\infty}_{\text{CBf,Q}}(X_d)/\partial \rho_{\text{C,Q}}$ and $\partial \text{SINR}^{\infty}_{\text{CBf,Q}}(X_t)/\partial \rho_{\text{C,Q}}$ are equal. Let $\mathcal{N}(f)$ denote the numerator of f. As $X_d \to X_t$, we have

$$\begin{aligned}\lim_{\rho_{\text{C,Q}} \to \rho_{\text{th}}^-} \mathcal{N}\left(\frac{\partial \text{SINR}^{\infty}_{\text{CBf,Q}}(X_d)}{\partial \rho_{\text{C,Q}}}\right) = \big[&\beta(G_2' + \epsilon G_3')(1/\gamma_d + 1 + \epsilon - (1 - X_t)) \\ &+ \beta\epsilon(1 - X_t)(G_2 G_3' - G_3 G_2') - \rho_{\text{th}}(1 - X_t) G_2'\big] \Gamma_Q (1 - X_t) \\ &- \lim_{\rho_{\text{C,Q}} \to \rho_{\text{th}}^-} X_d'(\beta G_2 + \beta\epsilon G_3 + \rho)\gamma_e\end{aligned}$$

$$= \mathcal{N}\left(\left.\frac{\partial \text{SINR}^{\infty}_{\text{CBf,Q}}(X_t)}{\partial \rho_{\text{C,Q}}}\right|_{\rho_{\text{C,Q}}=\rho_{\text{th}}}\right) - \lim_{\rho_{\text{C,Q}}\to\rho_{\text{th}}^{-}} X_d'(\beta G_2 + \beta\epsilon G_3 + \rho)\gamma_e,$$

where $X_d' = \partial X_d/\partial\rho_{\text{C,Q}}$. We should note that $\lim_{\rho_{\text{C,Q}}\to\rho_{\text{th}}^{-}} X_d' = -\frac{1}{2}\epsilon G_3' \frac{1-X_t}{\frac{1}{\gamma_d}+1+\epsilon} \neq 0$. This shows that x_d^* is not continuously differentiable over $\rho_{\text{C,Q}}$. It can be verified that the following holds

$$\begin{aligned}\lim_{\rho_{\text{C,Q}}\to\rho_{\text{th}}^{-}} \gamma_e &= \frac{1}{\gamma_d} + 1 + \epsilon + \epsilon(G_3 - 1)\left(-1 + \frac{1}{X_t}\right) \\ &= \frac{1}{X_t}\left[\left(\frac{1}{\gamma_d}+1\right)X_t + \epsilon(2X_t - 1) - \epsilon G_3(X_t - 1)\right] = 0,\end{aligned}$$

since as $\rho_{\text{C,Q}} \to \rho_{\text{th}}^{-}$, from the (equivalent) condition $\epsilon = \epsilon_{\text{th}}$, the term in the bracket becomes 0. This concludes (7.39) and therefore $\text{SINR}^{\infty}_{\text{CBf,Q}}(x_d^*)$ is continuously differentiable for $\rho_{\text{C,Q}} \in [0, \infty)$.

By using the property above and the facts that the $\text{SINR}^{\infty}_{\text{CBf,Q}}(X_d)$ and $\text{SINR}^{\infty}_{\text{CBf,Q}}(X_t)$ are quasi-concave (unimodal), we can determine the optimal $\rho_{\text{C,Q}}^*$ and x_d^* jointly as described in Algorithm 7.4.1. We can verify the steps 6-13 in the algorithm by using the following

Algorithm 7.4.1 Calculate $\rho_{\text{C,Q}}^*$ and x_d^*

1: Compute ρ_{th}
2: **if** $\rho_{\text{th}} \leq 0$ **then**
3: $\quad x_d^* = X_t$.
4: $\quad \rho_{\text{C,Q}}^* = \rho_{X_t}^{\circ}$
5: **else**
6: $\quad$ Compute $\rho_{X_t}^{\circ}$
7: $\quad$ **if** $\rho_{X_t}^{\circ} \geq \rho_{\text{th}}$ **then**
8: $\quad\quad x_d^* = X_t$
9: $\quad\quad \rho_{\text{C,Q}}^* = \rho_{X_t}^{\circ}$
10: $\quad$ **else**
11: $\quad\quad x_d^* = X_d$
12: $\quad\quad \rho_{\text{C,Q}}^* = \rho_{X_d}^{\circ}$
13: $\quad$ **end if**
14: **end if**

arguments: If $\rho_{X_t}^{\circ} > \rho_{\text{th}}$, then the derivate of $\text{SINR}^{\infty}_{\text{CBf,Q}}(X_t)$ is positive at $\rho_{\text{C,Q}} = \rho_{\text{th}}$ because $\text{SINR}^{\infty}_{\text{CBf,Q}}(X_t)$ is quasi-concave. Since the $\text{SINR}^{\infty}_{\text{CBf,Q}}(x_d^*)$ is continuously differentiable, then the derivative of $\text{SINR}^{\infty}_{\text{CBf,Q}}(X_d)$ is also positive when $\rho_{\text{C,Q}} \to \rho_{\text{th}}$. Since $\text{SINR}^{\infty}_{\text{CBf,Q}}(X_d)$ is also

quasi-concave, consequently $\text{SINR}^{\infty}_{\text{CBf,Q}}(X_d)$ is increasing for $\rho_{\text{C,Q}} \in [0, \rho_{\text{th}})$. This implies that $\rho^*_{\text{C,Q}} = \rho^{\circ}_{X_t}$. Similar types of arguments can be also used to verify that if $\rho^{\circ}_{X_t} < \rho_{\text{th}}$ then $\rho^*_{\text{C,Q}} = \rho^{\circ}_{X_d}$.

7.4.3 Numerical Results

The first two figures in this section are obtained by using a similar procedure to that followed in the analog feedback case. Figure 7.5 shows how well the limiting sum rate (equivalently the limiting SINR) approximates the finite-size system sum rate. The optimal regularization parameter and bit allocation are applied in computing the limiting and average sum rates. As N grows, the normalized sum rate difference become smaller. For $N = 60, K = 36$, it is arleady about 3.1% and 1.6% for MCP and CBf, respectively. Figure 7.6 shows the average sum rate difference, with a fixed regularization parameter, between the system that uses $B^*_{d,\text{FS}}$ and $\bar{B_d}^*$ to feed back the direct channel states. $B^*_{d,\text{FS}}$ denotes the optimal bit allocation of the finite-size system. For each channel realization, it is obtained by a grid search. With $N = 10, K = 6$, the maximum normalized average sum rate difference reaches 0.22% for MCP. It is about four-times bigger for CBf, which is approximately 0.86%. Thus, from those simulations, similar to the analog feedback case, the conclusions we can reach for the limiting regime are actually useful for the finite system case.

In the following, we present numerical simulations that show the behavior of the limiting SINR and optimal bit allocation for MCP and CBf as ϵ varies. The optimal bit allocation is illustrated in Figure 7.7(a). As shown in Section 7.4, the optimal B_d for MCP is decreasing in ϵ and $B^*_d = B_t$ when $\epsilon = 0$. For CBf, $B^*_d = B_t$ when $\epsilon \leq 0.19$, and after that decreases as ϵ grows. Overall, for given ϵ, B^*_d for CBf is larger than for MCP, implying the quality of the direct channel information is more important for CBf.

In Figure 7.7(b), the optimal values for the regularization parameter and bit allocation are used. From that figure, it is obvious that $\text{SINR}^{\infty}_{\text{CBf,Q}}$ decreases as ϵ increases. In the case of MCP, as predicted by the analysis, the limiting SINR is decreasing until $\epsilon^*_{\text{M,RVQ}} \approx 0.72$ and increasing after that point. By comparing the limiting SINR for both cooperation schemes, it is also interesting to see that for some values of ϵ, i.e., in the interval when

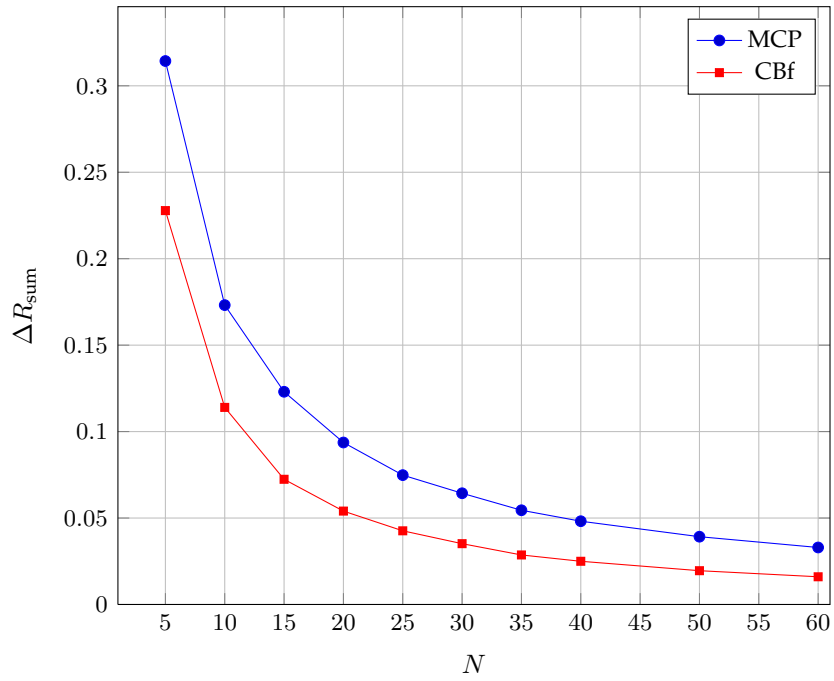


Figure 7.5: The total sum-rate difference for different system dimensions with $\beta = 0.6$, $\epsilon = 0.5$ $\gamma_d = 10$ dB and $\bar{B}_t = 4$.

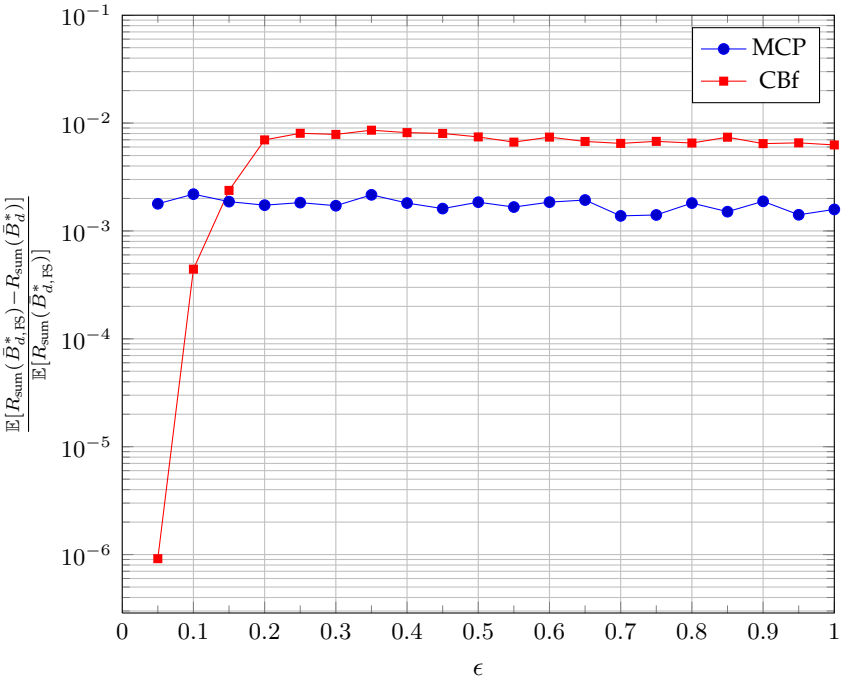


Figure 7.6: The (normalized) average sum-rate difference of the finite-size system by using the $\bar{B}^*_{d,\text{FS}}$ and $\bar{B}^*_d$ with $N = 10, \beta = 0.6$, $\gamma_d = 10$ dB and $\bar{B}_t = 4$.

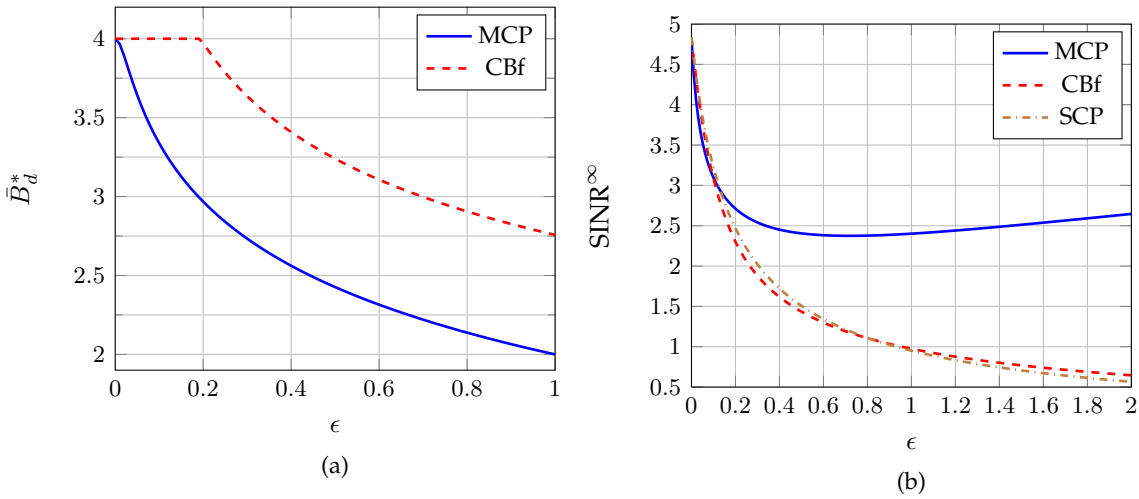


Figure 7.7: (a) Optimal bit allocation vs. ϵ, (b) Limiting SINR vs. ϵ. Parameters: γ_d=10 dB, $\bar{B}_t = 4$.

CBf has $\bar{B}_c^* = 0$, the CBf slightly outperforms MCP. We should note that within the current scheme, when $\bar{B}_c^* = 0$, CBf and MCP are not the same as single-cell processing (SCP): under RVQ, there is still a quantization vector in the codebook that is used to represent the cross channel (although it is uncorrelated with the actual channel vector being quantized).

Motivated by the above facts, we investigate whether SCP provides some advantages over MCP and CBf for some (low) values of ϵ. In SCP, we use $B_{k,j,j} = B_t$ bits ($\forall k, j$) to quantize the direct channel. The cross channels in the precoder are represented by vectors with zero entries. By following the steps in deriving Theorem 7.3 and 7.6, we can show that the limiting SINR is given by

$$\text{SINR}_{\text{SCP,Q}}^{\infty} = \gamma_e g(\beta, \rho_{\text{S}}) \frac{1 + \frac{\rho_{\text{S}}}{\beta}(1 + g(\beta, \rho_{\text{S}}))^2}{\gamma_e + (1 + g(\beta, \rho_{\text{S}}))^2},$$

where $\rho_{\text{S}} = N^{-1}\alpha$ and $\gamma_e = \frac{1-2^{-\bar{B}_t}}{2^{-\bar{B}_t}+\epsilon+\frac{1}{\gamma_d}}$. It follows that the optimal ρ_{S} maximizing $\text{SINR}_{\text{SCP,Q}}^{\infty}$ is $\rho_{\text{S}}^* = \frac{\beta}{\gamma_e}$ and the corresponding the limiting SINR is $\text{SINR}_{\text{SCP,Q}}^{*,\infty} = g(\beta, \rho_{\text{S}}^*)$.

From Figure 7.7, it is obvious that the SCP outperforms MCP and CBf for some values of ϵ. For $\epsilon \leq 0.13$, the SCP outperforms MCP and CBf. Surprisingly, the CBf is still beaten

by SCP until $\epsilon \approx 0.82$. This means that the SCP still gives advantages over the CBf even in a quite strong interference regime with this level of feedback.

7.5 Analog vs. Digital Feedback

In this section we will compare the performance of the analog and quantized feedback for each cooperation scheme. For the quantized feedback, we follow the approach in [8,45,78,111] that translates feedback bits to symbols for a fair comparison with the analog feedback. In this regard, there are two approaches [111]:

(i) By assuming that the feedback channel is error free and transmitted at the uplink rate (even though this assumption could be unrealistic in practice), we can write

$$\bar{B}_t = \frac{B_t}{N} = 2\kappa \log_2 \left(1 + (1+\epsilon)\gamma_u\right). \tag{7.40}$$

This approach is introduced in [8, 45]. (7.40) is obtained by assuming that each feedback bit is received by both base stations in different cells where the path-gains from a user to its own BS and other BS are different i.e, 1 and ϵ respectively. We can think the feedback transmission from a user to both BSs as a Single-Input Multi-Output (SIMO) system. The BSs linearly combine the feedback signal from the user and the corresponding maximum SNR is $(1+\epsilon)\gamma_u$ (see [46]). The pre-log factor $2\kappa N$ for B_t in (7.40) presents the channel uses (symbols) for transmitting the feedback bits which are the same as those for the analog feedback. κ follows the discussion in Section 7.2.3. Our approach is different from the approach in [111] in which the user k in cell j sends the feedbacks only to its own BS j. In that case, (7.40) becomes $\bar{B}_t = 2\kappa \log_2 (1+\gamma_u)$.

(ii) Following [78], the second approach translates the feedback bits to symbols based on the modulation scheme used in the feedback transmission. In the analog feedback, the feedback takes $2\kappa N$ channel uses per user. Let η be a conversion factor that links the bits and symbols and it depends on the modulation scheme. As an example, for the binary phase shift keying (BPSK), $\eta = 1$. Thus, we can write (see

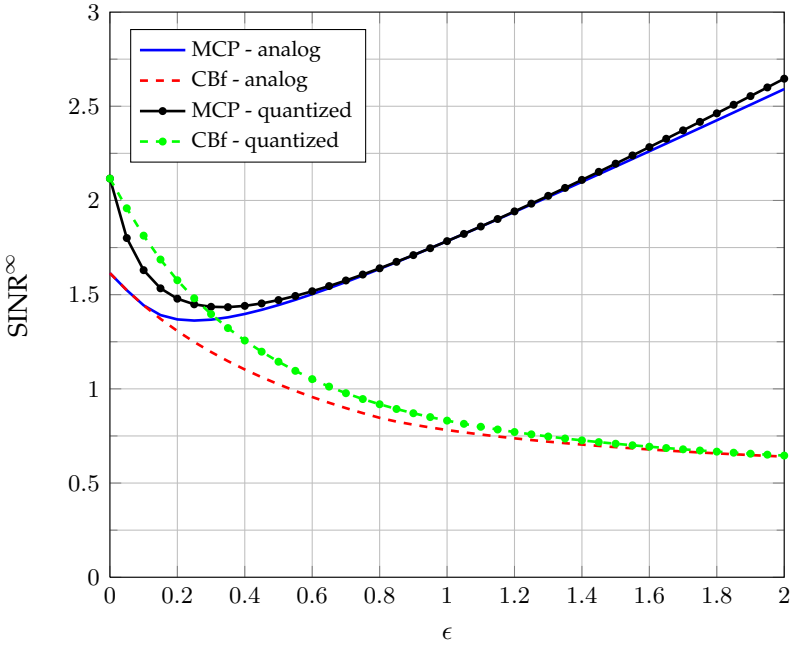
also [111])

$$\eta B_t = 2\kappa N. \tag{7.41}$$

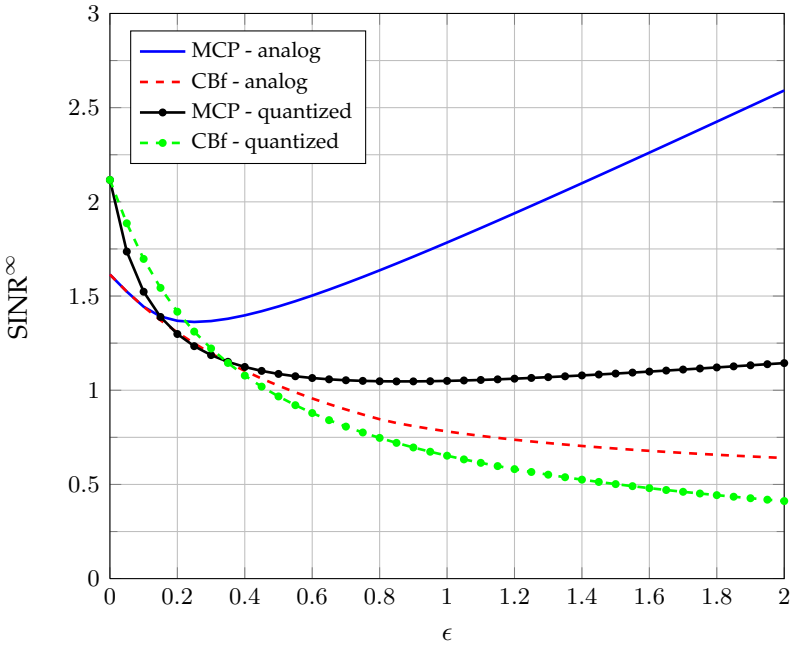
We should note that using this approach, for a fixed κ there is no link between $\bar{B}_t$ and γ_u as we can see in (7.40).

Let us assume that $\kappa = 1$. Thus, with the first approach, we have $X_t = 2^{-\bar{B}_t} = \frac{1}{(1+(1+\epsilon)\gamma_u)^2}$. The comparison of the limiting SINR based on the analog and quantized feedback for MCP and CBf can be seen in Figure 7.8(a). It shows that the quantized feedback beats the analog feedback in both MCP and CBf for ϵ less than about 1. A similar situation still occurs for CBf even for $\epsilon \in [0, 2]$. The opposite happens for MCP when ϵ is above 1.5. The comparison of the analog and quantized feedback with the second approach, also with $\kappa = 1$, is illustrated in Figure 7.8(b). Similar to the previous, one can see that the quantized feedback outperforms the analog feedback if ϵ is below a certain threshold. Otherwise, the analog feedback gives better performance. Those observations can be explained by verifying whether the feedback scheme that provides better CSIT will give a better performance. This is easier to check by looking at the MCP scheme because from our discussions in Section 7.3 and 7.4, its performance can be measured by the total CSIT quality, i.e., $\omega_c + \omega_d$ in the analog feedback and $(1 + \epsilon)d^2$ in the digital feedback. Plotting those over ϵ, not shown here, will give the same behaviors for the MCP as we observed in Figure 7.8. Thus, from our simulations above, the CSIT quality of the quantized feedback is better than that of analog feedback when the cross channel gain is below a certain threshold. The plots above also confirm that more feedback resources will increase the system performance: for a fixed $\gamma_u = 0$ dB, $\bar{B}_t$ in the left plot is larger than that in the right plot and hence gives a higher (limiting) SINR for the quantized feedback scheme.

Figure 7.9 depicts the limiting SINR of the analog and quantized feedback for different values of feedback rate. For the analog feedback, the values of feedback rate/bit is converted by using the previous approaches: $\kappa = \frac{\bar{B}_t}{2\log_2(1+(1+\epsilon)\gamma_u)}$ and $\kappa = \bar{B}_t/2$ respectively. For MCP, we can see that initially the analog feedback scheme outperforms the quantized feedback in both plots. However, after a certain value (threshold) of $\bar{B}_t$, the opposite happens. A similar observation also holds for the CBf scheme. The explana-

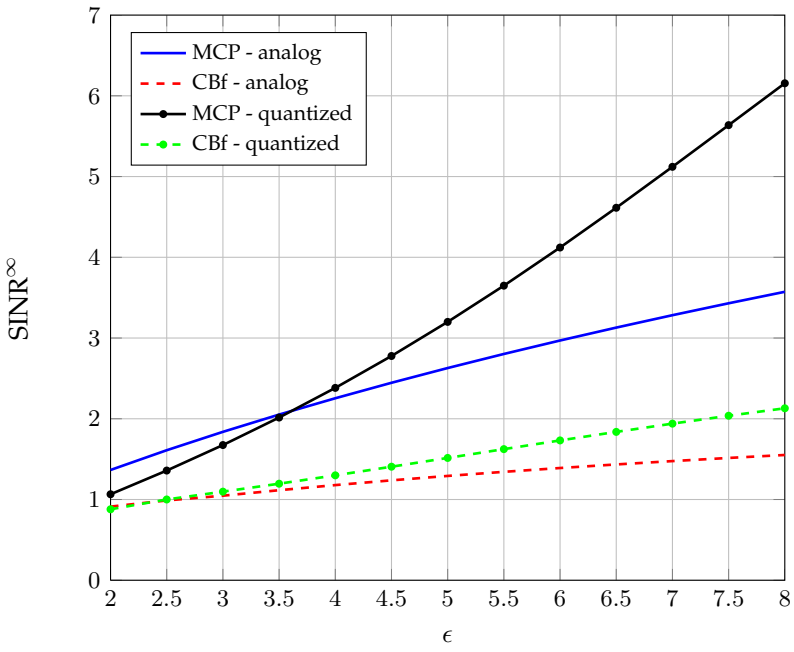


(a) $\bar{B}_t = 2\log_2(1 + (1+\epsilon)\gamma_u)$

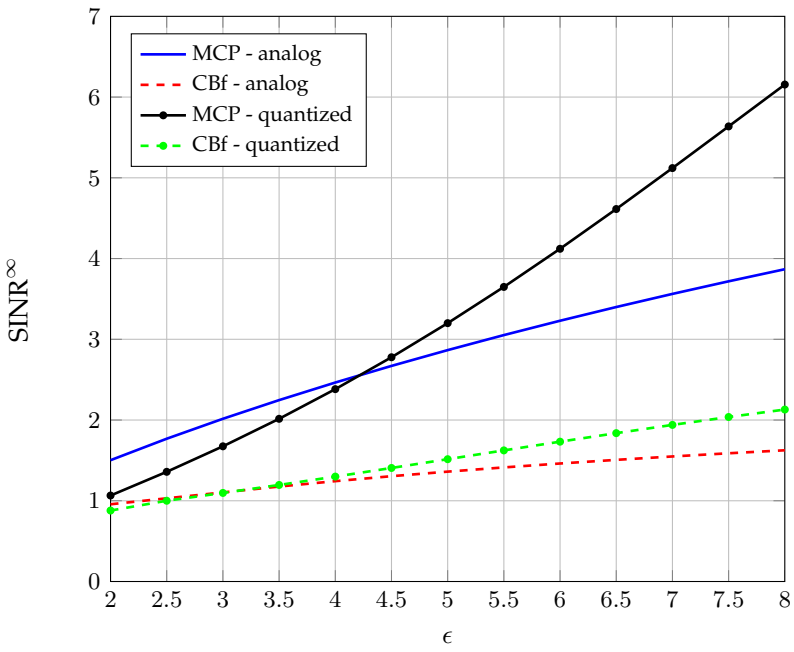


(b) $\bar{B}_t = 2$

Figure 7.8: The comparison of the limiting SINR of the analog and quantized feedback for different cooperation schemes. Parameters: $\beta = 0.6$, $\gamma_d = 10$ dB, $\gamma_u = 0$ dB.



(a) $\bar{B}_t = 2\kappa \log_2 (1 + (1 + \epsilon)\gamma_u)$



(b) $\bar{B}_t = 2\kappa$

Figure 7.9: The comparison of the limiting SINR of the analog and quantized feedback for different cooperation schemes vs. the feedback rates. Parameters: $\beta = 0.6$, $\epsilon = 0.6$,$\gamma_d = 10$ dB, $\gamma_u = 0$ dB.

tions for those phenomena follow the discussions for Figure 7.8. We should note that in generating the figures, the values for $\bar{B}_t$ are already determined. So, the limiting SINRs for the digital feedback are the same in both sub-figures. For the analog feedback, since κ with the approach (7.41) is larger (with $\gamma_u = 0$ dB) than that with the approach (7.40), then the training period in the former is longer and will result in a better CSIT. Thus, the limiting SINRs for the analog feedback in Figure 7.9(b) are larger compared to those in 7.9(a).

7.6 Conclusion

In this chapter, we perform feedback optimization for the analog and quantized feedback schemes in a symmetric two-cell network with different levels of cooperation between base stations. In both cooperation schemes, it is shown that more resources, uplink transmit power in the case of analog feedback or feedback bits in the case of quantized feedback, are allocated to feeding back the interfering channel information as the interfering channel gain increases. Moreover, if the interfering channel gain is below a certain threshold, the conventional network with no cooperation between base stations is preferable. Our analysis also shows that the limiting SINR for MCP, in both analog and quantized feedback, improves in ϵ if ϵ is above certain threshold. This also implies that above that threshold the (total) quality of the channel at the base stations is also getting better. Although our analysis is performed in the asymptotic regime, our numerical results hint to their validity in the finite-size system cases.

7.7 Appendix

7.7.1 Large System Results for the Network MIMO

First, we will expand the SINR expression (7.2). Let $\mathbf{\Phi}_{k,j} = \text{diag}\{\phi_{k,j,1}, \phi_{k,j,2}\}$. Based on (7.1) we can write $\mathbf{h}_{k,j} = \widehat{\mathbf{h}}_{k,j}\mathbf{\Phi}_{k,j}^{\frac{1}{2}} + \widetilde{\mathbf{h}}_{k,j}$. Consequently, the $\text{SINR}_{k,j}$ can be expressed as

$$\frac{c^2 \left| (\widehat{\mathbf{h}}_{k,j}\mathbf{\Phi}_{k,j}^{\frac{1}{2}} + \widetilde{\mathbf{h}}_{k,j}) \left(\widehat{\mathbf{H}}^H\widehat{\mathbf{H}} + \alpha\mathbf{I}_{2N}\right)^{-1} \widehat{\mathbf{h}}_{k,j}^H \right|^2}{c^2(\widehat{\mathbf{h}}_{k,j}\mathbf{\Phi}_{k,j}^{\frac{1}{2}} + \widetilde{\mathbf{h}}_{k,j}) \left(\widehat{\mathbf{H}}^H\widehat{\mathbf{H}} + \alpha\mathbf{I}_{2N}\right)^{-1} \widehat{\mathbf{H}}_{k,j}^H\widehat{\mathbf{H}}_{k,j} \left(\widehat{\mathbf{H}}^H\widehat{\mathbf{H}} + \alpha\mathbf{I}_{2N}\right)^{-1} (\widehat{\mathbf{h}}_{k,j}\mathbf{\Phi}_{k,j}^{\frac{1}{2}} + \widetilde{\mathbf{h}}_{k,j})^H + \sigma_d^2}.$$

By applying the matrix inversion lemma (MIL) to $\left(\widehat{\mathbf{H}}^H\widehat{\mathbf{H}} + \alpha\mathbf{I}_{2N}\right)^{-1}$, we can rewrite the SINR expression as follows

$$\text{SINR}_{k,j} = \frac{c^2 \left| \frac{\breve{A}_{k,j} + F_{k,j}}{1 + A_{k,j}} \right|^2}{c^2 \left(B_{k,j} + 2\Re[D_{k,j}] + E_{k,j}\right) + \sigma_d^2}, \tag{7.42}$$

where

$$\begin{aligned}
\breve{A}_{k,j} &= \frac{1}{N}\widehat{\mathbf{h}}_{k,j}\mathbf{\Phi}_{k,j}^{\frac{1}{2}} \left(\frac{1}{N}\widehat{\mathbf{H}}_{k,j}^H\widehat{\mathbf{H}}_{k,j} + \rho\mathbf{I}_{2N}\right)^{-1} \widehat{\mathbf{h}}_{k,j}^H \\
A_{k,j} &= \frac{1}{N}\widehat{\mathbf{h}}_{k,j} \left(\frac{1}{N}\widehat{\mathbf{H}}_{k,j}^H\widehat{\mathbf{H}}_{k,j} + \rho\mathbf{I}_{2N}\right)^{-1} \widehat{\mathbf{h}}_{k,j}^H \\
F_{k,j} &= \frac{1}{N}\widetilde{\mathbf{h}}_{k,j} \left(\frac{1}{N}\widehat{\mathbf{H}}_{k,j}^H\widehat{\mathbf{H}}_{k,j} + \rho\mathbf{I}_{2N}\right)^{-1} \widehat{\mathbf{h}}_{k,j}^H \\
B_{k,j} &= \frac{1}{N}\widehat{\mathbf{h}}_{k,j}\mathbf{\Phi}_{k,j}^{\frac{1}{2}}\mathbf{Z}_{k,j} \left(\frac{1}{N}\widehat{\mathbf{H}}_{k,j}^H\widehat{\mathbf{H}}_{k,j}\right) \mathbf{Z}_{k,j}\mathbf{\Phi}_{k,j}^{\frac{1}{2}}\widehat{\mathbf{h}}_{k,j}^H \\
D_{k,j} &= \frac{1}{N}\widehat{\mathbf{h}}_{k,j}\mathbf{\Phi}_{k,j}^{\frac{1}{2}}\mathbf{Z}_{k,j} \left(\frac{1}{N}\widehat{\mathbf{H}}_{k,j}^H\widehat{\mathbf{H}}_{k,j}\right) \mathbf{Z}_{k,j}\widetilde{\mathbf{h}}_{k,j}^H \\
E_{k,j} &= \frac{1}{N}\widetilde{\mathbf{h}}_{k,j}\mathbf{Z}_{k,j} \left(\frac{1}{N}\widehat{\mathbf{H}}_{k,j}^H\widehat{\mathbf{H}}_{k,j}\right) \mathbf{Z}_{k,j}\widetilde{\mathbf{h}}_{k,j}^H
\end{aligned}$$

with $\rho = \frac{\alpha}{N}$, $\mathbf{O}_{k,j} = \left(\frac{1}{N}\widehat{\mathbf{H}}_{k,j}^H\widehat{\mathbf{H}}_{k,j} + \rho\mathbf{I}_{2N}\right)^{-1}$ and $\mathbf{Z}_{k,j} = \left[\mathbf{O}_{k,j} - \frac{\mathbf{O}_{k,j}\left(\frac{1}{N}\widehat{\mathbf{h}}_{k,j}^H\widehat{\mathbf{h}}_{k,j}\right)\mathbf{O}_{k,j}}{1+A_{k,j}}\right]$. Note that for the analog feedback, $\mathbf{\Phi}_{k,j} = \mathbf{I}, \forall k, j$. In the following subsections, the large system limit of each term in the SINR (7.42), for the analog and quantized feedback cases, will be derived. For brevity in the presentation, we denote $\mathbf{Q}_{k,j} = \mathbf{O}_{k,j}\left(\frac{1}{N}\widehat{\mathbf{H}}_k^H\widehat{\mathbf{H}}_{k,j}\right)\mathbf{O}_{k,j}$.

7.7.1.1 Proof of Theorem 1: Analog Feedback case

In the analog feedback case, as previously mentioned in Section 7.3, $\widehat{\mathbf{h}}_{k,i,j} \sim \mathcal{CN}(0, \omega_{ij}\mathbf{I}_N)$ and $\widetilde{\mathbf{h}}_{k,i,j} \sim \mathcal{CN}(0, \delta_{ij}\mathbf{I}_N)$ are independent. We can rewrite those vectors as follows

$$\widehat{\mathbf{h}}_{k,j} = \mathbf{g}_{k,j}\mathsf{G}_{k,j}^{\frac{1}{2}} \text{ and } \widetilde{\mathbf{h}}_{k,j} = \mathbf{d}_{k,j}\mathsf{D}_{k,j}^{\frac{1}{2}},$$

where $\mathbf{g}_{k,j} \sim \mathcal{CN}(0, \mathbf{I}_{2N})$ and $\mathbf{d}_{k,j} \sim \mathcal{CN}(0, \mathbf{I}_{2N})$ are independent. The diagonal matrices $\mathsf{G}_{k,j}$ and $\mathsf{D}_{k,j}$ are given by $\mathsf{G}_{k,j} = \text{diag}\{\omega_{j1}\mathbf{I}_N, \omega_{j2}\mathbf{I}_N\}$ and $\mathsf{D}_{k,j} = \text{diag}\{\delta_{j1}\mathbf{I}_N, \delta_{j2}\mathbf{I}_N\}$, respectively.

In the analysis below, we heavily use Theorem 2.5. Therefore, it is useful to define the asymptotic variance profile for the matrix $\frac{1}{N}\widehat{\mathbf{H}}^H$ which is a $2N \times 2\beta N$ complex random matrix. Following Theorem 2.5, in our case, we have $x \in [0, 2]$, $y \in [0, 2\beta]$ and the asymptotic variance profile is given by

$$v(x,y) = \begin{cases} \omega_d & 0 \leq x < 1, 0 \leq y < \beta \\ \omega_c & 1 \leq x < 2, 0 \leq y < \beta \\ \omega_c & 0 \leq x < 1, \beta \leq y < 2\beta \\ \omega_d & 1 \leq x < 2, \beta \leq y < 2\beta. \end{cases}$$

In what follows, we will derive the large system limit for each term in (7.42).

1) $\breve{A}_{k,j}$ *:* It can rewritten as $\frac{1}{N}\mathbf{g}_{k,j}\mathsf{G}_{k,j}^{\frac{1}{2}}\mathbf{\Phi}_{k,j}^{\frac{1}{2}}\mathbf{O}_{k,j}\mathsf{G}_{k,j}^{\frac{1}{2}}\mathbf{g}_{k,j}^H$. Applying Lemma 2.2 yields

$$\breve{A}_{k,j} - \frac{1}{N}\text{Tr}\left(\mathsf{G}_{k,j}\mathbf{\Phi}_{k,j}^{\frac{1}{2}}\mathbf{O}_{k,j}\right) \xrightarrow{a.s.} 0.$$

The second term in the LHS can be written as

$$\frac{1}{N}\left[\sum_{i=1}^{N}\omega_{j1}\sqrt{\phi_{k,j,1}}[\mathbf{O}_{k,j}]_{ii} + \sum_{i=N+1}^{2N}\omega_{j2}\sqrt{\phi_{k,j,2}}[\mathbf{O}_{k,j}]_{ii}\right].$$

By Applying Theorem 2.5, it converges in probability to

$$\omega_{j1}\sqrt{\phi_{k,j,1}}\int_0^1 u(x,-\rho)\,dx+\omega_{j2}\sqrt{\phi_{k,j,2}}\int_1^2 u(x,-\rho)\,dx,$$

where for $0\leq x\leq 1$,

$$u(x,-\rho)=u_1=\frac{1}{\rho+\frac{\beta\omega_d}{1+u_1\omega_d+u_2\omega_c}+\frac{\beta\omega_c}{1+u_1\omega_c+u_2\omega_d}}$$

and for $1<x\leq 2$,

$$u(x,-\rho)=u_2=\frac{1}{\rho+\frac{\beta\omega_c}{1+u_1\omega_d+u_2\omega_c}+\frac{\beta\omega_d}{1+u_1\omega_c+u_2\omega_d}}.$$

The solution of the equations above is $u_1=u_2=u$ where u is the positive solution of

$$u=\frac{1}{\rho+\frac{\beta(\omega_d+\omega_c)}{1+u(\omega_d+\omega_c)}}.$$

Let $g(\beta,\rho)$ be the solution of $g(\beta,\rho)=\left(\rho+\frac{\beta}{1+g(\beta,\rho)}\right)^{-1}$. Then, we can express u in terms of $g(\beta,\rho)$ as

$$u=\frac{1}{\omega_d+\omega_c}g(\beta,\bar{\rho}),\qquad \bar{\rho}=\frac{\rho}{\omega_d+\omega_c}.$$

We should note that $\omega_{j1}+\omega_{j2}=\omega_d+\omega_c,\ \forall j$. Thus,

$$\breve{A}_{k,j}-\breve{A}^{\infty}_{k,j}\xrightarrow{i.p.}0,\ \text{with}\ \breve{A}^{\infty}_{k,j}=\frac{\sqrt{\phi_{k,j,1}}\omega_{j1}+\sqrt{\phi_{k,j,2}}\omega_{j2}}{\omega_d+\omega_c}g(\beta,\bar{\rho}).$$

Since in the current feedback scheme $\mathbf{\Phi}_{k,j}=\mathbf{I}_{2N}$ then $\breve{A}^{\infty}_{k,j}=g(\beta,\bar{\rho})$. Moreover, $\breve{A}^{\infty}_{k,j}=\breve{A}^{\infty}$ is the same for all users in both cells.

2) $A_{k,j}$ *:* This term is $\breve{A}_{k,j}$ with $\mathbf{\Phi}_{k,j}=\mathbf{I}_{2N}$. Thus, it follows that $A_{k,j}-g(\beta,\bar{\rho})\xrightarrow{i.p.}0$.

3) $F_{k,j}$ *:* It can be rewritten as $\frac{1}{N}\mathbf{d}_{k,j}\mathsf{D}_{k,j}^{\frac{1}{2}}\mathbf{O}_{k,j}\mathsf{G}_{k,j}^{\frac{1}{2}}\mathbf{g}_{k,j}^{H}$. Conditioning on $\widehat{\mathbf{H}}_{k,j}$, it is obvious that $\mathbf{d}_{k,j}$, $\mathbf{g}_{k,j}$ and $\mathsf{D}_{k,j}^{\frac{1}{2}}\mathbf{O}_{k,j}\mathsf{G}_{k,j}^{\frac{1}{2}}$ are independent of each other. By Lemma 2.3, it follows that $F_{k,j}\xrightarrow{a.s.}0$.

4) $D_{k,j}$ *:* Expanding $D_{k,j}$, we have

$$\begin{aligned} D_{k,j} &= \frac{1}{N}\widehat{\mathbf{h}}_{k,j}\mathbf{\Phi}^{\frac{1}{2}}\mathbf{Q}_{k,j}\widetilde{\mathbf{h}}_{k,j}^H + \frac{\breve{A}_{k,j}F_{k,j}^*\left(\frac{1}{N}\widehat{\mathbf{h}}_{k,j}\mathbf{Q}_{k,j}\widehat{\mathbf{h}}_k^H\right)}{(1+A_{k,j})^2} - \frac{F_{k,j}^*\left(\frac{1}{N}\widehat{\mathbf{h}}_{k,j}\mathbf{\Phi}^{\frac{1}{2}}\mathbf{Q}_{k,j}\widehat{\mathbf{h}}_{k,j}^H\right)}{1+A_{k,j}} \\ &\quad - \frac{\breve{A}_{k,j}\left(\frac{1}{N}\widehat{\mathbf{h}}_{k,j}\mathbf{Q}_{k,j}\widetilde{\mathbf{h}}_{k,j}^H\right)}{1+A_{k,j}} \\ &= D_{k,j}^{(1)} + D_{k,j}^{(2)} - D_{k,j}^{(3)} - D_{k,j}^{(4)}. \end{aligned} \tag{7.43}$$

Following the arguments in 3), it can be checked that $D_{k,j}^{(1)} \xrightarrow{a.s.} 0$. Similarly, $D_{k,j}^{(4)} \xrightarrow{a.s.} 0$. Since $F_{k,j} \xrightarrow{a.s.} 0$ then $D_{k,j}^{(2)} \xrightarrow{a.s.} 0$ and $D_{k,j}^{(3)} \xrightarrow{a.s.} 0$. Thus, it follows that $D_{k,j} \xrightarrow{a.s.} 0$.

5) $B_{k,j}$ *:* It can be rewritten as

$$\begin{aligned} B_{k,j} &= \frac{1}{N}\widehat{\mathbf{h}}_{k,j}\mathbf{\Phi}_{k,j}^{\frac{1}{2}}\mathbf{Q}_{k,j}\mathbf{\Phi}_{k,j}^{\frac{1}{2}}\widehat{\mathbf{h}}_{k,j}^H + \frac{|\breve{A}_{k,j}|^2\left(\frac{1}{N}\widehat{\mathbf{h}}_{k,j}\mathbf{Q}_{k,j}\widehat{\mathbf{h}}_{k,j}^H\right)}{(1+A_{k,j})^2} \\ &\quad - \frac{2\Re\left[\breve{A}_{k,j}^*\left(\frac{1}{N}\widehat{\mathbf{h}}_{k,j}\mathbf{\Phi}_{k,j}^{\frac{1}{2}}\mathbf{Q}_{k,j}\widehat{\mathbf{h}}_{k,j}^H\right)\right]}{1+A_{k,j}} \qquad (7.44) \\ &= B_{k,j}^{(1)} + \frac{|\breve{A}_{k,j}|^2 B_{k,j}^{(2)}}{(1+A_{k,j})^2} - \frac{2\Re[\breve{A}_{k,j}^* B_{k,j}^{(3)}]}{1+A_{k,j}}. \qquad (7.45) \end{aligned}$$

From Lemma 2.2, we can show $B_{k,j}^{(1)} - (N)^{-1}\mathrm{Tr}\left(\mathsf{G}_{k,j}\mathbf{\Phi}_{k,j}\mathbf{Q}_{k,j}\right) \xrightarrow{a.s.} 0$. Following (3.28), we have $\mathbf{Q}_{k,j} = \mathbf{O}_{k,j} + \rho\frac{\partial}{\partial\rho}\mathbf{O}_{k,j}$. By applying Theorem 2.5, we obtain

$$\frac{1}{N}\sum_{i=\lfloor aN\rfloor}^{\lfloor bN\rfloor}[\mathbf{Q}_{k,j}]_{ii} \xrightarrow{i.p.} \int_a^b u(x,-\rho)\,dx + \rho\frac{\partial}{\partial\rho}\int_a^b u(x,-\rho)\,dx\,.$$

Consequently, we can show $B_{k,j}^{(1)} - B_{k,j}^{(1),\infty} \xrightarrow{i.p.} 0$, where

$$B_{k,j}^{(1),\infty} = \frac{\omega_{j1}\phi_{k,j,1} + \omega_{j2}\phi_{k,j,2}}{\omega_d + \omega_c}\left[g(\beta,\bar{\rho}) + \bar{\rho}\frac{\partial}{\partial\bar{\rho}}g(\beta,\bar{\rho})\right]\,. \tag{7.46}$$

Similarly, we can also show that

$$B_{k,j}^{(2),\infty} = g(\beta,\bar{\rho}) + \bar{\rho}\frac{\partial}{\partial\bar{\rho}}g(\beta,\bar{\rho}),$$

$$B_{k,j}^{(3),\infty} = \frac{\omega_{j1}\sqrt{\phi_{k,j,1}} + \omega_{j2}\sqrt{\phi_{k,j,2}}}{\omega_d + \omega_c}\left[g(\beta,\bar{\rho}) + \bar{\rho}\frac{\partial}{\partial\bar{\rho}}g(\beta,\bar{\rho})\right].$$

Since $\mathbf{\Phi}_{k,j} = \mathbf{I}_{2N}$, it follows that $B_{k,j}^{(1),\infty} = B_{k,j}^{(2),\infty} = B_{k,j}^{(3),\infty}$. Thus, $B_{k,j} - B^{\infty} \xrightarrow{i.p.} 0$, where

$$B^{\infty} = \frac{1}{(1+g(\beta,\bar{\rho}))^2}\left[g(\beta,\bar{\rho}) + \bar{\rho}\frac{\partial}{\partial\bar{\rho}}g(\beta,\bar{\rho})\right].$$

6) $E_{k,j}$ *:* Expanding this term gives

$$\begin{aligned} E_{k,j} &= \frac{1}{N}\widetilde{\mathbf{h}}_{k,j}\mathbf{Q}_{k,j}\widetilde{\mathbf{h}}_{k,j}^H - 2\Re\left[\frac{F_{k,j}\left(\frac{1}{N}\widehat{\mathbf{h}}_{k,j}\mathbf{Q}_{k,j}\widetilde{\mathbf{h}}_{k,j}^H\right)}{1+A_{k,j}}\right] + \frac{|F_{k,j}|^2\,\frac{1}{N}\widehat{\mathbf{h}}_{k,j}\mathbf{Q}_{k,j}\widehat{\mathbf{h}}_{k,j}^H}{(1+A_{k,j})^2} \\ &= E_{k,j}^{(1)} - E_{k,j}^{(2)} + E_{k,j}^{(3)}. \end{aligned} \tag{7.47}$$

By using the previous results, we can show $E_{k,j}^{(2)} \xrightarrow{a.s.} 0$ and $E_{k,j}^{(3)} \xrightarrow{i.p.} 0$. From Lemma 2.2, $E_{k,j}^{(1)} - (N)^{-1}\mathrm{Tr}\left(\mathsf{D}_{k,j}\mathbf{Q}_{k,j}\right) \xrightarrow{a.s.} 0$. Following the steps in obtaining $B_{k,j}^{(1),\infty}$, it is straightforward to show that $E_{k,j}^{(1)} - E_{k,j}^{(1),\infty} \xrightarrow{i.p.} 0$ where

$$E_{k,j}^{(1),\infty} = \frac{\delta_{k,j,1} + \delta_{k,j,2}}{\omega_d + \omega_c}\left[g(\beta,\bar{\rho}) + \bar{\rho}\frac{\partial}{\partial\bar{\rho}}g(\beta,\bar{\rho})\right].$$

Thus, we have $E_{k,j} - E_{k,j}^{(1),\infty} \xrightarrow{i.p.} 0$.

7) c^2 *:* The denominator of c^2 can be written as follows

$$\frac{1}{N}\mathrm{Tr}\left\{\left(\frac{1}{N}\widehat{\mathbf{H}}^H\widehat{\mathbf{H}} + \rho\mathbf{I}_{2N}\right)^{-2}\frac{1}{N}\widehat{\mathbf{H}}^H\widehat{\mathbf{H}}\right\} = \int \frac{\lambda}{(\lambda+\rho)^2}\, dF_{\widehat{\mathbf{H}}^H\widehat{\mathbf{H}}}(\lambda),$$

where $F_{\widehat{\mathbf{H}}^H\widehat{\mathbf{H}}}$ is the empirical eigenvalue distribution of $\widehat{\mathbf{H}}^H\widehat{\mathbf{H}}$. From Theorem 2.5, $F_{\widehat{\mathbf{H}}^H\widehat{\mathbf{H}}}$ converges almost surely to a limiting distribution G^* whose Stieltjes transform

$$m(z) = \int_0^{\infty} \frac{1}{\lambda - z}\, dG^*(\lambda) = \int_0^1 u(x,z)\, dx.$$

Therefore,

$$\int \frac{\lambda}{(\lambda+\rho)^2}\, dF_{\widehat{\mathbf{H}}^H\widehat{\mathbf{H}}}(\lambda) = \int \frac{1}{\lambda+\rho} + \rho\frac{\partial}{\partial\rho}\frac{1}{\lambda+\rho}\, dF_{\widehat{\mathbf{H}}^H\widehat{\mathbf{H}}}(\lambda)$$
$$\xrightarrow{a.s.} m(-\rho) + \rho\frac{\partial}{\partial\rho}m(-\rho) = \int_0^1 u(x,-\rho) + \rho\frac{\partial}{\partial\rho}u(x,-\rho)\, dx.$$

Previously, we have shown that $\int_0^1 u(x,-\rho) = 2(\omega_d+\omega_c)^{-1}g(\beta,\bar{\rho})$ with $\bar{\rho} = 2\rho(\omega_d+\omega_c)^{-1}$. Hence, the last equation equals to $2(\omega_d+\omega_c)^{-1}\left(g(\beta,\bar{\rho}) + \bar{\rho}\frac{\partial}{\partial\bar{\rho}}g(\beta,\bar{\rho})\right)$ and we have

$$c^2 - \frac{\frac{1}{2}(\omega_d+\omega_c)P_t}{g(\beta,\bar{\rho}) + \bar{\rho}\frac{\partial}{\partial\bar{\rho}}g(\beta,\bar{\rho})} \xrightarrow{a.s.} 0.$$

The large system analysis in 1)-3) and 7) show that the signal strength, i.e, the numerator of (7.42), converges to

$$\frac{(\omega_d+\omega_c)P_d g^2(\beta,\bar{\rho})}{(1+g(\beta,\bar{\rho}))^2\left(g(\beta,\bar{\rho}) + \bar{\rho}\frac{\partial}{\partial\bar{\rho}}g(\beta,\bar{\rho})\right)}, \tag{7.48}$$

where we already substitute $P_d = \frac{1}{2}P_t$. Similarly, it follows that the interference (energy) converges to

$$\frac{P_d}{(1+g(\beta,\bar{\rho}))^2}\left(\omega_d+\omega_c - (\delta_d+\delta_c)\left(1+g(\beta,\bar{\rho})\right)^2\right). \tag{7.49}$$

By using (3.32), we can show

$$g(\beta,\bar{\rho}) + \bar{\rho}\frac{\partial}{\partial\bar{\rho}}g(\beta,\bar{\rho}) = \frac{\beta g(\beta,\bar{\rho})}{\beta+\bar{\rho}(1+g(\beta,\bar{\rho}))^2}.$$

By combining the large system results and denoting $\rho_{\text{M,AF}} = \bar{\rho}$, we can express the limiting SINR as in (7.13). This completes the proof.

7.7.1.2 Proof of Theorem 7.3: Quantized feedback (via RVQ) case

Observing (7.8), it is obvious that $\phi_{k,j,i}$ is random because it is the function of the quantization error. In opposite, it is not random in the analog feedback channel model. Therefore, the derivation of the limiting SINR for the limited feedback channel model is quite

different from the previous subsection. We start with the following notation. For a given a $2N \times 2N$ matrix $\mathbf{X}$, we can partition it as follows.

$$\mathbf{X}_{k,j} = \begin{bmatrix} \mathbf{X}_{k,j}^{11} & \mathbf{X}_{k,j}^{12} \\ \mathbf{X}_{k,j}^{21} & \mathbf{X}_{k,j}^{22} \end{bmatrix},$$

where $\mathbf{X}_{k,j}^{11} = [\mathbf{X}_{k,j}]_{lm}, l = 1, \cdots, N, m = 1, \cdots, N$, $\mathbf{X}_{k,j}^{12} = [\mathbf{X}_{k,j}]_{lm}, l = 1, \cdots, N, m = N+1, \cdots, 2N$, $\mathbf{X}_{k,j}^{21} = [\mathbf{X}_{k,j}]_{lm}, l = N+1, \cdots, 2N, m = 1, \cdots, N$, and $\mathbf{X}_{k,j}^{22} = [\mathbf{X}_{k,j}]_{lm}, l = N+1, \cdots, 2N, m = N+1, \cdots, 2N$.

In the following, the large system limit for each term in the SINR is derived.

1) $\breve{A}_{k,j}$ *:* We can write $\breve{A}_{k,j}$ as

$$\breve{A}_{k,j} = \frac{1}{N}\Big(\phi_{k,j,1}^{\frac{1}{2}}\widehat{\mathbf{h}}_{k,j,1}\mathbf{O}_{k,j}^{11}\widehat{\mathbf{h}}_{k,j,1}^{H} + \phi_{k,j,1}^{\frac{1}{2}}\widehat{\mathbf{h}}_{k,j,1}\mathbf{O}_{k,j}^{12}\widehat{\mathbf{h}}_{k,j,2}^{H} + \phi_{k,j,2}^{\frac{1}{2}}\widehat{\mathbf{h}}_{k,j,2}\mathbf{O}_{k,j}^{21}\widehat{\mathbf{h}}_{k,j,1}^{H} + \phi_{k,j,2}^{\frac{1}{2}}\widehat{\mathbf{h}}_{k,j,2}\mathbf{O}_{k,j}^{22}\widehat{\mathbf{h}}_{k,j,2}^{H}\Big). \tag{7.50}$$

Since $\widehat{\mathbf{h}}_{k,j,1}$ and $\widehat{\mathbf{h}}_{k,j,2}$ are independent, the second and third terms converge almost surely to 0. For the first term

$$\frac{1}{N}\widehat{\mathbf{h}}_{k,j,1}\mathbf{O}_{k,j}^{11}\widehat{\mathbf{h}}_{k,j,1}^{H} - \frac{\omega_{j1}}{N}\mathrm{Tr}\mathbf{O}_{k,j}^{11} \xrightarrow{a.s.} 0$$

or equivalently,

$$\frac{1}{N}\widehat{\mathbf{h}}_{k,j,1}\mathbf{O}_{k,j}^{11}\widehat{\mathbf{h}}_{k,j,1}^{H} - \frac{\omega_{j1}}{N}\sum_{i=1}^{N}[\mathbf{O}_{k,j}]_{ii} \xrightarrow{a.s.} 0.$$

By using Theorem 2.5, we have

$$\frac{1}{N}\sum_{i=1}^{N}[\mathbf{O}_{k,j}]_{ii} \xrightarrow{i.p.} \frac{1}{1+\epsilon}g(\beta, \bar{\rho}),$$

where $\bar{\rho} = \frac{\frac{1}{N}\alpha}{1+\epsilon}$. By using the same techniques as in [77], we can show that

$$\sqrt{\phi_{k,j,i}} = \sqrt{1-\tau_{k,j,i}^2} \xrightarrow{L_2} \begin{cases} \sqrt{1-2^{-\bar{B}_d}} & j=i \\ \sqrt{1-2^{-\bar{B}_c}} & \text{otherwise.} \end{cases}$$

We should also note that the convergence in mean square sense implies the convergence in probability. By doing the same steps for the last term of (7.50), we have

$$\breve{A}_{k,j} - \frac{\sqrt{1-2^{-\bar{B}_d}} + \epsilon\sqrt{1-2^{-\bar{B}_c}}}{1+\epsilon} g(\beta,\bar{\rho}) \xrightarrow{i.p.} 0.$$

2) $A_{k,j}$ *:* This term is $\breve{A}_{k,j}$ with $\mathbf{\Phi}_{k,j} = \mathbf{I}_{2N}$. Hence,

$$A_{k,j} - g(\beta,\bar{\rho}) \xrightarrow{i.p.} 0.$$

3) $F_{k,j}$ *:* We can expand the term as follows

$$F_{k,j} = \frac{1}{N}\left(\widetilde{\mathbf{h}}_{k,j,1}\mathbf{O}_{k,j}^{11}\widehat{\mathbf{h}}_{k,j,1}^H + \widetilde{\mathbf{h}}_{k,j,1}\mathbf{O}_{k,j}^{12}\widehat{\mathbf{h}}_{k,j,2}^H + \widetilde{\mathbf{h}}_{k,j,2}\mathbf{O}_{k,j}^{21}\widehat{\mathbf{h}}_{k,j,1}^H + \widetilde{\mathbf{h}}_{k,j,2}\mathbf{O}_{k,j}^{22}\widehat{\mathbf{h}}_{k,j,2}^H\right). \quad (7.51)$$

By using the same steps as in deriving (7.60), it follows that each term in the right hand side of the equation above converges in probability to 0. Hence, $F_{k,j} \xrightarrow{i.p.} 0$.

4) $B_{k,j}$ *:* By following the representation (7.45) for B_{kj} , we can express the first term in the right hand side as

$$B_{k,j}^{(1)} = \frac{1}{N}\Big(\phi_{k,j,1}\widehat{\mathbf{h}}_{k,j,1}\mathbf{Q}_{k,j}^{11}\widehat{\mathbf{h}}_{k,j,1}^H + \phi_{k,j,1}\widehat{\mathbf{h}}_{k,j,1}\mathbf{Q}_{k,j}^{12}\widehat{\mathbf{h}}_{k,j,2}^H + \phi_{k,j,2}\widehat{\mathbf{h}}_{k,j,2}\mathbf{Q}_{k,j}^{21}\widehat{\mathbf{h}}_{k,j,1}^H + \phi_{k,j,2}\widehat{\mathbf{h}}_{k,j,2}\mathbf{Q}_{k,j}^{22}\widehat{\mathbf{h}}_{k,j,2}^H\Big).$$

The second term and the third term converge (almost surely) to 0. For the first term,

$$\frac{1}{N}\widehat{\mathbf{h}}_{k,j,1}\mathbf{Q}_{k,j}^{11}\widehat{\mathbf{h}}_{k,j,1}^H - \frac{\omega_{j1}}{N}\mathrm{Tr}\mathbf{Q}_{k,j}^{11} \xrightarrow{a.s.} 0.$$

From (3.28), we have $\mathbf{Q}_{k,j} = \mathbf{O}_{k,j} + \rho\frac{\partial}{\partial\rho}\mathbf{O}_{k,j}$. Hence, we can show that

$$\frac{1}{N}\text{Tr}\mathbf{Q}_{k,j}^{11} = \frac{1}{N}\sum_{i=1}^{N}\left[\mathbf{Q}_{k,j}^{11}\right]_{ii} \xrightarrow{i.p.} \int_0^1 u(x,-\rho)\,dx + \rho\frac{\partial}{\partial\rho}\int_0^1 u(x,-\rho)\,dx$$
$$= \frac{1}{1+\epsilon}\left[g(\beta,\bar{\rho}) + \bar{\rho}\frac{\partial}{\partial\bar{\rho}}g(\beta,\bar{\rho})\right].$$

By doing the same steps for the last term of $B_{k,j}^{(1)}$, it follows that

$$B_{k,j}^{(1)} - \frac{1-2^{-\bar{B}_d}+\epsilon(1-2^{-\bar{B}_c})}{1+\epsilon}\left[g(\beta,\bar{\rho}) + \bar{\rho}\frac{\partial}{\partial\bar{\rho}}g(\beta,\bar{\rho})\right] \xrightarrow{i.p.} 0.$$

Similarly, we can also show that

$$B_{k,j}^{(2)} - g(\beta,\bar{\rho}) + \bar{\rho}\frac{\partial}{\partial\bar{\rho}}g(\beta,\bar{\rho}) \xrightarrow{i.p.} 0,$$
$$B_{k,j}^{(3)} - \frac{\sqrt{1-2^{-\bar{B}_d}}+\epsilon\sqrt{1-2^{-\bar{B}_c}}}{1+\epsilon}\left[g(\beta,\bar{\rho}) + \bar{\rho}\frac{\partial}{\partial\bar{\rho}}g(\beta,\bar{\rho})\right] \xrightarrow{i.p.} 0.$$

Combining the results together, we obtain

$$B_{k,j} - \left(\frac{1-2^{-\bar{B}_d}+\epsilon(1-2^{-\bar{B}_c})}{1+\epsilon} - \frac{d^2(2+g(\beta,\bar{\rho}))g(\beta,\bar{\rho})}{(1+g(\beta,\bar{\rho}))^2}\right)\left[g(\beta,\bar{\rho}) + \bar{\rho}\frac{\partial}{\partial\bar{\rho}}g(\beta,\bar{\rho})\right] \xrightarrow{i.p.} 0,$$

where

$$d = \frac{\sqrt{1-2^{-\bar{B}_d}}+\epsilon\sqrt{1-2^{-\bar{B}_c}}}{1+\epsilon}.$$

5) $D_{k,j}$ *:* We can expand $D_{k,j}$ as expressed in (7.43). Using the previous results, it follows that $D_{k,j}^{(2)}$ and $D_{k,j}^{(3)}$ converge to 0. We can expand the first term as follows

$$D_{k,j}^{(1)} = \frac{1}{N}\left(\phi_{k,j,1}^{\frac{1}{2}}\widehat{\mathbf{h}}_{k,j,1}\mathbf{Q}_{k,j}^{11}\widetilde{\mathbf{h}}_{k,j,1}^{H} + \phi_{k,j,1}^{\frac{1}{2}}\widehat{\mathbf{h}}_{k,j,1}\mathbf{Q}_{k,j}^{12}\widetilde{\mathbf{h}}_{k,j,2}^{H} + \phi_{k,j,2}^{\frac{1}{2}}\widehat{\mathbf{h}}_{k,j,2}\mathbf{Q}_{k,j}^{21}\widetilde{\mathbf{h}}_{k,j,1}^{H}\right.$$
$$\left. + \phi_{k,j,2}^{\frac{1}{2}}\widehat{\mathbf{h}}_{k,j,2}\mathbf{Q}_{k,j}^{22}\widetilde{\mathbf{h}}_{k,j,2}^{H}\right).$$

Again, by following the same steps as in deriving (7.60), each term converges in probability to 0 and hence $D_{k,j}^{(1)} \xrightarrow{i.p.} 0$. Similarly, $D_{k,j}^{(4)} \xrightarrow{i.p.} 0$. Therefore, in the end, $D_{k,j} \xrightarrow{i.p.} 0$.

6) $E_{k,j}$ *:* The expansion of E_{kj} follows (7.47). By applying the previous results, $E_{k,j}^{(2)}$ and $E_{k,j}^{(3)}$ converge in probability to 0. $E_{k,j}^{(1)}$ can rewritten as

$$E_{k,j}^{(1)} = \frac{1}{N}\left(\widetilde{\mathbf{h}}_{k,j,1}\mathbf{Q}_{k,j}^{11}\widetilde{\mathbf{h}}_{k,j,1}^{H} + \widetilde{\mathbf{h}}_{k,j,1}\mathbf{Q}_{k,j}^{12}\widetilde{\mathbf{h}}_{k,j,2}^{H} + \widetilde{\mathbf{h}}_{k,j,2}\mathbf{Q}_{k,j}^{21}\widetilde{\mathbf{h}}_{k,j,1}^{H} + \widetilde{\mathbf{h}}_{k,j,2}\mathbf{Q}_{k,j}^{22}\widetilde{\mathbf{h}}_{k,j,2}^{H}\right).$$

Since $\widetilde{\mathbf{h}}_{k,j,1}$ and $\widetilde{\mathbf{h}}_{k,j,2}$ are independent then the second and third term converge to 0. The first term in the equation above can be written as

$$\begin{aligned}
\frac{1}{N}\widetilde{\mathbf{h}}_{k,j,1}\mathbf{Q}_{k,j}^{11}\widetilde{\mathbf{h}}_{k,j,1}^{H} &= \frac{\tau_{k,j,1}^2\|\mathbf{h}_{k,j,1}\|^2}{N\|\mathbf{v}_{k,j,1}\mathbf{\Pi}_{\widehat{\mathbf{h}}_{k,j,1}}^{\perp}\|^2}\left(\mathbf{v}_{k,j,1} - \frac{(\mathbf{v}_{k,j,1}\widehat{\mathbf{h}}_{k,j,1}^{H})\widehat{\mathbf{h}}_{k,j,1}}{\|\widehat{\mathbf{h}}_{k,j,1}\|^2}\right)\mathbf{Q}_{k,j}^{11} \\
&\quad\times\left(\mathbf{v}_{k,j,1} - \frac{(\mathbf{v}_{k,j,1}\widehat{\mathbf{h}}_{k,j,1}^{H})\widehat{\mathbf{h}}_{k,j,1}}{\|\widehat{\mathbf{h}}_{k,j,1}\|^2}\right)^{H} \\
&= \frac{\tau_{k,j,1}^2\|\mathbf{h}_{k,j,1}\|^2}{\|\mathbf{v}_{k,j,1}\mathbf{\Pi}_{\widehat{\mathbf{h}}_{k,j,1}}^{\perp}\|^2}\left(\frac{1}{N}\mathbf{v}_{k,j,1}\mathbf{Q}_{k,j}^{11}\mathbf{v}_{k,j,1}^{H} + \frac{|\frac{1}{N}\mathbf{v}_{k,j,1}\widehat{\mathbf{h}}_{k,j,1}^{H}|^2\frac{1}{N}\widehat{\mathbf{h}}_{k,j,1}\mathbf{Q}_{k,j}^{11}\widehat{\mathbf{h}}_{k,j,1}^{H}}{\frac{1}{N^2}\|\widehat{\mathbf{h}}_{k,j,1}\|^4}\right. \\
&\quad\left.-2\Re\left[\frac{(\frac{1}{N}\mathbf{v}_{k,j,1}\widehat{\mathbf{h}}_{k,j,1}^{H})^{*}\frac{1}{N}\mathbf{v}_{k,j,1}\mathbf{Q}_{k,j}^{11}\widehat{\mathbf{h}}_{k,j,1}^{H}}{\frac{1}{N}\|\widehat{\mathbf{h}}_{k,j,1}\|^2}\right]\right).
\end{aligned}$$

Since $\mathbf{v}_{k,j,1}$ and $\widehat{\mathbf{h}}_{k,j,1}$ are independent then the second and third term in the bracket converge to 0. It can be shown that

$$\frac{1}{N}\mathbf{v}_{k,j,1}\mathbf{Q}_{k,j}^{11}\mathbf{v}_{k,j,1}^{H} - \frac{1}{1+\epsilon}\left[g(\beta,\bar{\rho}) + \bar{\rho}\frac{\partial}{\partial\bar{\rho}}g(\beta,\bar{\rho})\right] \xrightarrow{i.p.} 0.$$

The large system limit for the last term of $E_{k,j}$ can be done in the same way. Thus,

$$E_{k,j} - \frac{2^{-\bar{B}_d} + \epsilon 2^{-\bar{B}_c}}{1+\epsilon}\left[g(\beta,\bar{\rho}) + \bar{\rho}\frac{\partial}{\partial\bar{\rho}}g(\beta,\bar{\rho})\right] \xrightarrow{i.p.} 0.$$

7) c^2 *:* We can show that

$$c^2 - \frac{\frac{1}{2}P_t(1+\epsilon)}{g(\beta,\bar{\rho}) + \bar{\rho}\frac{\partial}{\partial\bar{\rho}}g(\beta,\bar{\rho})} \xrightarrow{a.s.} 0.$$

Since $\xrightarrow{a.s.}$ implies $\xrightarrow{i.p.}$ then c^2 also converges in probability to the same quantity as above.

Combining the results, it follows that the signal strength and the interference con-

verge to

$$\frac{P_d(1+\epsilon)d^2 g^2(\beta,\bar{\rho})}{(1+g(\beta,\bar{\rho}))^2\left(g(\beta,\bar{\rho})+\bar{\rho}\frac{\partial}{\partial\bar{\rho}}g(\beta,\bar{\rho})\right)} \tag{7.52}$$

and

$$P_d(1+\epsilon)\left(1-\frac{d^2 g(\beta,\bar{\rho})(2+g(\beta,\bar{\rho}))}{(1+g(\beta,\bar{\rho}))^2}\right), \tag{7.53}$$

respectively. Moreover, (7.26) also follows immediately with $\rho_{\text{M,Q}}=\bar{\rho}$.

7.7.2 Large System Results for the Coordinated Beamforming

For brevity in the proof, we define the following (see also [109])

$$\mathbf{A}_j=\left(\rho\mathbf{I}_N+\frac{1}{N}\sum_{m=1}^{2}\sum_{l=1}^{K}\widehat{\mathbf{h}}_{l,m,j}^H\widehat{\mathbf{h}}_{l,m,j}\right)^{-1}$$
$$\mathbf{A}_{kj}=\left(\rho\mathbf{I}_N+\frac{1}{N}\sum_{(l,m)\neq(k,j)}\widehat{\mathbf{h}}_{l,m,j}^H\widehat{\mathbf{h}}_{l,m,j}\right)^{-1}$$
$$\mathbf{A}_{kj,k'j',j}=\left(\rho\mathbf{I}_N+\frac{1}{N}\sum_{(l,m)\neq(k,j),(k',j')}\widehat{\mathbf{h}}_{l,m,j}^H\widehat{\mathbf{h}}_{l,m,j}\right)^{-1},$$

where $\rho=\frac{\alpha}{N}$. From the definitions above, we can write the numerator of the $\text{SINR}_{k,j}$ (7.3) excluding c_j^2, as follows

$$|\mathbf{h}_{k,j,j}\hat{\mathbf{w}}_{kj}|^2=\left|\frac{\sqrt{\phi_{k,j,j}}}{N}\widehat{\mathbf{h}}_{k,j,j}\mathbf{A}_{kj}\widehat{\mathbf{h}}_{k,j,j}^H\right|^2+\left|\frac{1}{N}\widetilde{\mathbf{h}}_{k,j,j}\mathbf{A}_{kj}\widehat{\mathbf{h}}_{k,j,j}^H\right|^2$$
$$+2\Re\left[\frac{1}{N^2}(\widetilde{\mathbf{h}}_{k,j,j}\mathbf{A}_{kj}\widehat{\mathbf{h}}_{k,j,j}^H)(\widehat{\mathbf{h}}_{k,j,j}\mathbf{A}_{kj}\widehat{\mathbf{h}}_{k,j,j}^H)\right]=\phi_{k,j,j}\left|S_{kj}^{(1)}\right|^2+|S_{kj}^{(2)}|^2+S_{kj}^{(3)}.$$

In the denominator, let us consider the term $|\mathbf{h}_{k,j,j'}\mathbf{w}_{k'j'}|^2$ that can be expanded as follows

$$|\mathbf{h}_{k,j,j'}\mathbf{w}_{k'j'}|^2=\frac{1}{N}\left(I_{kj,k'j'}^{(1)}+I_{kj,k'j'}^{(2)}-I_{kj,k'j'}^{(3)}\right)=\frac{1}{N}\mathcal{I}, \tag{7.54}$$

where

$$I^{(1)}_{kj,k'j'} = \frac{1}{N}\phi_{k,j,j'}\widehat{\mathbf{h}}_{k,j,j'}\mathbf{A}_{k'j'}\widehat{\mathbf{h}}^H_{k',j',j'}\widehat{\mathbf{h}}_{k',j',j'}\mathbf{A}_{k'j'}\widehat{\mathbf{h}}^H_{k,j,j'}$$
$$I^{(2)}_{kj,k'j'} = \frac{1}{N}\widetilde{\mathbf{h}}_{k,j,j'}\mathbf{A}_{k'j'}\widehat{\mathbf{h}}^H_{k',j',j'}\widehat{\mathbf{h}}_{k',j',j'}\mathbf{A}_{k'j'}\widetilde{\mathbf{h}}^H_{k,j,j'}$$
$$I^{(3)}_{kj,k'j'} = 2\Re\left[\frac{\sqrt{\phi_{k,j,j'}}}{N}\widetilde{\mathbf{h}}_{k,j,j'}\mathbf{A}_{k'j'}\widehat{\mathbf{h}}^H_{k',j',j'}\widehat{\mathbf{h}}_{k',j',j'}\mathbf{A}_{k'j'}\widehat{\mathbf{h}}^H_{k,j,j'}\right].$$

In what follows, we will derive the large system limit for each term in the numerator and denominator above. First, we are going to derive the large system limit for $\frac{1}{N}\mathrm{Tr}\mathbf{A}_j$ because it will be used frequently in this section. Let $\widehat{\mathsf{H}}_j = [\widehat{\mathbf{h}}_{1,1,j}\ \cdots\ \widehat{\mathbf{h}}_{K,1,j}\ \widehat{\mathbf{h}}_{1,2,j}\ \cdots\ \widehat{\mathbf{h}}_{K,2,j}]^T$ and $\widehat{\mathbf{h}}_{k,i,j} \sim \mathcal{CN}(0, \omega_{ij}\mathbf{I}_N)$. Then, $\mathbf{A}_j = \left(\rho\mathbf{I}_N + \frac{1}{N}\widehat{\mathsf{H}}^H_j\widehat{\mathsf{H}}_j\right)^{-1}$ and

$$\frac{1}{N}\mathrm{Tr}\mathbf{A}_j = \int \frac{1}{\lambda+\rho}\, dF_{\widehat{\mathsf{H}}^H_j\widehat{\mathsf{H}}_j},$$

where $F_{\widehat{\mathsf{H}}^H_j\widehat{\mathsf{H}}_j}$ is the empirical eigenvalue distribution of $\widehat{\mathsf{H}}^H_j\widehat{\mathsf{H}}_j$. From Theorem 2.5, this distribution converges almost surely to a limiting distribution F whose Stieltjes transform $m_{\mathsf{F}}(z)$. It can be shown that

$$\frac{1}{N}\mathrm{Tr}\mathbf{A}_j \xrightarrow{a.s.} m_{\mathsf{F}}(-\rho) = \int_0^1 u(x,-\rho)\, dx\ ,$$

where

$$u(x,-\rho) = u(-\rho) = \frac{1}{\rho + \frac{\beta\omega_{1j}}{1+\omega_{1j}u(-\rho)} + \frac{\beta\omega_{2j}}{1+\omega_{2j}u(-\rho)}} = \frac{1}{\rho + \frac{\beta\omega_d}{1+\omega_d u(-\rho)} + \frac{\beta\omega_c}{1+\omega_c u(-\rho)}}\ ,$$

for $0 \le x \le 1$. Let $\Gamma = u(-\rho)$, then $\frac{1}{N}\mathrm{Tr}\mathbf{A}_j \xrightarrow{a.s.} \Gamma$.

7.7.2.1 Analog Feedback

Based on the channel model (7.1), we have $\phi_{k,j,j} = \phi_{k,j,j'} = 1$. The definitions for other terms such as $\omega_\bullet$ and $\delta_\bullet$ can be seen in Section 7.2.3. Now, let us first derive the large system limit for the numerator of the SINR_{kj}. We start with the term $S^{(1)}_{kj}$. From Lemma

2.2 and by applying [48, Lemma 5.1], we can show that

$$\max_{j=1,2,k\leq K} \left| S_{kj}^{(1)} - \frac{\omega_d}{N} \mathrm{Tr}\mathbf{A}_{kj} \right| \xrightarrow{a.s.} 0.$$

By applying Lemma 3 in [109], we have

$$\max_{j=1,2,k\leq K} \left| S_{kj}^{(1)} - \frac{\omega_d}{N} \mathrm{Tr}\mathbf{A}_{j} \right| \xrightarrow{a.s.} 0,$$

where $\frac{1}{N}\mathrm{Tr}\mathbf{A}_j \xrightarrow{a.s.} \Gamma$.

Since $\widehat{\mathbf{h}}_{k,j,j}$, $\mathbf{A}_{kj}$ and $\widetilde{\mathbf{h}}_{k,j,j}$ are independent then it follows that

$$\max_{j=1,2,k\leq K} \left| \widetilde{\mathbf{h}}_{k,j,j} \mathbf{A}_{kj} \widehat{\mathbf{h}}_{k,j,j}^H \right| \xrightarrow{a.s.} 0.$$

Consequently,

$$\max_{j=1,2,k\leq K} \left| S_{kj}^{(2)} \right| \xrightarrow{a.s.} 0 \text{ and } \max_{j=1,2,k\leq K} \left| S_{kj}^{(3)} \right| \xrightarrow{a.s.} 0.$$

In summary,

$$\max_{j=1,2,k\leq K} \left| |\mathbf{h}_{k,j,j}\mathbf{w}_{kj}|^2 - \omega_d^2\Gamma^2 \right| \xrightarrow{a.s.} 0.$$

Now, let us move in analyzing the interference term. Using the matrix inversion lemma, we rewrite $I_{kj,k'j'}^{(1)}$ as

$$I_{kj,k'j'}^{(1)} = \frac{\frac{1}{N}\phi_{k,j,j'}\widehat{\mathbf{h}}_{k,j,j'}\mathbf{A}_{k'j',kj,j'}\widehat{\mathbf{h}}_{k',j',j'}^H\widehat{\mathbf{h}}_{k',j',j'}\mathbf{A}_{k'j',kj,j'}\widehat{\mathbf{h}}_{k,j,j'}^H}{\left(1+\frac{1}{N}\widehat{\mathbf{h}}_{k,j,j'}\mathbf{A}_{k'j',kj,j'}\widehat{\mathbf{h}}_{k,j,j'}^H\right)^2}.$$

By applying Lemma [48, Lemma 5.1] and [109, Lemma 3] twice, we can show

$$\max_{j,j'=1,2,k,k'\leq K,(k,j)\neq(k',j')} \left| \frac{1}{N}\widehat{\mathbf{h}}_{k,j,j'}\mathbf{A}_{k'j',kj,j'}\widehat{\mathbf{h}}_{k,j,j'}^H - \frac{\omega_{jj'}}{N}\mathrm{Tr}\mathbf{A}_{j'} \right| \xrightarrow{a.s.} 0.$$

Similarly,

$$\max_{j,j'=1,2,k,k'\leq K,(k,j)\neq(k',j')} \left| \frac{1}{N}\widehat{\mathbf{h}}_{k,j,j'}\mathbf{A}_{k'j',kj,j'}\widehat{\mathbf{h}}_{k',j',j'}^H\widehat{\mathbf{h}}_{k',j',j'}\mathbf{A}_{k'j',kj,j'}\widehat{\mathbf{h}}_{k,j,j'}^H - \frac{\omega_{jj'}\omega_d}{N}\mathrm{Tr}\mathbf{A}_{j'}^2 \right| \xrightarrow{a.s.} 0.$$

Since $\frac{1}{N}\text{Tr}\mathbf{A}_{j'} \to \Gamma$ and $\frac{1}{N}\text{Tr}\mathbf{A}_{j'}^2 \to -\frac{\partial}{\partial\rho}\Gamma$, we have

$$\max_{j,j'=1,2,k,k'\leq K,(k,j)\neq(k',j')} \left| I^{(1)}_{kj,k'j'} - \omega_d \left(-\frac{\omega_{jj'}}{(1+\omega_{jj'}\Gamma)^2}\frac{\partial\Gamma}{\partial\rho} \right) \right| \xrightarrow{a.s.} 0.$$

By following the same steps, we obtain

$$\max_{j,j'=1,2,k,k'\leq K,(k,j)\neq(k',j')} \left| I^{(2)}_{kj,k'j'} - \left(-\delta_{jj'}\omega_d \frac{\partial\Gamma}{\partial\rho} \right) \right| \xrightarrow{a.s.} 0 \,,$$

and

$$\max_{j,j'=1,2,k,k'\leq K,(k,j)\neq(k',j')} \left| I^{(3)}_{kj,k'j'} \right| \xrightarrow{a.s.} 0.$$

Combining the results, we have

$$\max_{j,j'=1,2,k,k'\leq K,(k,j)\neq(k',j')} \left| \mathcal{I} - \omega_d \left(-\frac{\omega_{jj'}}{(1+\omega_{jj'}\Gamma)^2} - \delta_{jj'} \right) \frac{\partial\Gamma}{\partial\rho} \right| \xrightarrow{a.s.} 0. \tag{7.55}$$

Using (7.55), the large system result for the interference term can be written as follows

$$\begin{aligned} \sum_{(k',j')\neq(k,j)} |\mathbf{h}_{k,j,j'}\mathbf{w}_{k'j'}|^2 &= \sum_{l=1,l\neq k}^{K} |\mathbf{h}_{k,j,j}\mathbf{w}_{lj}|^2 + \sum_{l=1}^{K} |\mathbf{h}_{k,j,\bar{j}}\mathbf{w}_{l\bar{j}}|^2 \\ &\xrightarrow{a.s.} -\beta\omega_d \left(\frac{\omega_d}{(1+\omega_d\Gamma)^2} + \frac{\omega_c}{(1+\omega_c\Gamma)^2} + \delta_d + \delta_c \right) \frac{\partial\Gamma}{\partial\rho}. \end{aligned}$$

Now, we just need to derive the large system limit for $c_j^2 = P_d \left(\sum_{k=1}^{K} \|\mathbf{w}_{kj}\|^2 \right)^{-1}$, where we can express $\|\mathbf{w}_{kj}\|^2 = \frac{1}{N^2}\widehat{\mathbf{h}}_{k,j,j}\mathbf{A}_{kj}^2\widehat{\mathbf{h}}_{k,j,j}^H$. We can show that

$$\max_{j=1,2,k\leq K} \left| \frac{1}{N^2}\widehat{\mathbf{h}}_{k,j,j}\mathbf{A}_{kj}^2\widehat{\mathbf{h}}_{k,j,j}^H - \frac{\omega_d}{N}\text{Tr}\mathbf{A}_j^2 \right| \xrightarrow{a.s.} 0.$$

Thus,

$$c_j^2 \xrightarrow{a.s.} \frac{P_d}{-\beta\omega_d\frac{\partial\Gamma}{\partial\rho}},$$

where we can show that

$$-\frac{\partial\Gamma}{\partial\rho} = -\Gamma' = \frac{\Gamma}{\rho + \frac{\beta\omega_c}{(1+\omega_c\Gamma)^2} + \frac{\beta\omega_d}{(1+\omega_d\Gamma)^2}}. \tag{7.56}$$

To sum up, from the analyses above, we can express the limiting signal energy as

$$\frac{1}{\beta}P_d\omega_d\Gamma\left(\rho+\frac{\beta\omega_c}{(1+\omega_c\Gamma)^2}+\frac{\beta\omega_d}{(1+\omega_d\Gamma)^2}\right), \tag{7.57}$$

and the limiting interference energy as

$$P_d\left(\frac{\omega_d}{(1+\omega_d\Gamma)^2}+\frac{\omega_c}{(1+\omega_c\Gamma)^2}+\delta_d+\delta_c\right). \tag{7.58}$$

Finally, the limiting SINR can be expressed as (7.21), with $\Gamma_A=\Gamma$ and $\rho_{\text{C,AF}}=\rho$.

7.7.2.2 Proof of Theorem 7.6: Quantized Feedback via RVQ

In the derivation of the large system limit SINR in this section, we use some of the results presented in the previous section. Here, we have $\omega_{jj}=\omega_d=1$ and $\omega_{j\bar{j}}=\omega_c=\epsilon$. From (7.1), we have $\phi_{k,j,i}=1-\tau^2_{k,j,i}$.

First, let us consider the numerator of the SINR. By using the result from previous section, we have

$$\max_{j=1,2,k\leq K}\left|S^{(1)}_{kj}-\frac{1}{N}\text{Tr}\mathbf{A}_j\right|\xrightarrow{a.s.}0,$$

where $\frac{1}{N}\text{Tr}\mathbf{A}_j\xrightarrow{a.s.}\Gamma$ and Γ is the solution of

$$\Gamma=\frac{1}{\rho+\frac{\beta}{1+\Gamma}+\frac{\beta\epsilon}{1+\epsilon\Gamma}}.$$

As stated in [77], we have,

$$\phi_{k,j,j}\xrightarrow{L_2}1-2^{-\bar{B}_d}, \tag{7.59}$$

where $\xrightarrow{L_2}$ denotes convergence in mean square sense. Since almost sure convergence and convergence in mean square imply the convergence in probability then

$$\phi_{k,j,j}|S^{(1)}_{kj}|^2-(1-2^{-\bar{B}_d})\Gamma^2\xrightarrow{i.p.}0.$$

By using (7.8), the term $\widetilde{\mathbf{h}}_{k,j,j}\mathbf{A}_{kj}\widehat{\mathbf{h}}_{k,j,j}^H$ in $S_{kj}^{(3)}$ can be rewritten as

$$\begin{aligned}\frac{1}{N}\widetilde{\mathbf{h}}_{k,j,j}\mathbf{A}_{kj}\widehat{\mathbf{h}}_{k,j,j}^H &= \frac{\tau_{k,j,j}\|\mathbf{h}_{k,j,j}\|}{\|\mathbf{v}_{k,j,j}\mathbf{\Pi}^{\perp}_{\widehat{\mathbf{h}}_{k,j,j}}\|}\left(\frac{1}{N}\mathbf{v}_{k,i,j}\mathbf{\Pi}^{\perp}_{\widehat{\mathbf{h}}_{k,i,j}}\widehat{\mathbf{h}}_{k,j,j}^H\right)\\ &= \frac{\tau_{k,j,j}\|\mathbf{h}_{k,j,j}\|}{\|\mathbf{v}_{k,j,j}\mathbf{\Pi}^{\perp}_{\widehat{\mathbf{h}}_{k,j,j}}\|}\left(\frac{1}{N}\mathbf{v}_{k,j,j}\mathbf{A}_{kj}\widehat{\mathbf{h}}_{k,j,j}^H - \frac{(\frac{1}{N}\mathbf{v}_{k,j,j}\widehat{\mathbf{h}}_{k,j,j}^H)\widehat{\mathbf{h}}_{k,j,j}\mathbf{A}_{kj}\widehat{\mathbf{h}}_{k,j,j}^H}{\|\widehat{\mathbf{h}}_{k,j,j}\|^2}\right).\end{aligned}$$

Since $\mathbf{v}_{k,j,j}$ and $\widehat{\mathbf{h}}_{k,j,j}^H$ are independent, then

$$\max_{j=1,2,k\leq K}\left|\frac{1}{N}\mathbf{v}_{k,j,j}\widehat{\mathbf{h}}_{k,j,j}^H\right| \xrightarrow{a.s.} 0, \text{ and } \max_{j=1,2,k\leq K}\left|\frac{1}{N}\mathbf{v}_{k,j,j}\mathbf{A}_{kj}\widehat{\mathbf{h}}_{k,j,j}^H\right| \xrightarrow{a.s.} 0.$$

It can also be shown that

$$\max_{j=1,2,k\leq K}\left|\frac{1}{N}\|\mathbf{h}_{k,j,j}\|^2 - 1\right| \xrightarrow{a.s.} 0, \text{ and } \max_{j=1,2,k\leq K}\left|\frac{1}{N}\|\mathbf{v}_{k,j,j}\mathbf{\Pi}^{\perp}_{\widehat{\mathbf{h}}_{k,j,j}}\|^2 - 1\right| \xrightarrow{a.s.} 0.$$

Hence,

$$\frac{1}{N}\widetilde{\mathbf{h}}_{k,j,j}\mathbf{A}_{kj}\widehat{\mathbf{h}}_{k,j,j}^H \xrightarrow{i.p.} 0, \tag{7.60}$$

and thus,

$$S_{kj}^{(2)} \xrightarrow{i.p.} 0, \text{ and } S_{kj}^{(3)} \xrightarrow{i.p.} 0.$$

Putting the results together, we have

$$|\mathbf{h}_{k,j,j}\mathbf{w}_{kj}|^2 - (1-2^{-\bar{B}_d})\Gamma^2 \xrightarrow{i.p.} 0.$$

For the interference terms, by using the same steps as in the previous section, we can show

$$\max_{j,j'=1,2,k,k'\leq K,(k,j)\neq(k',j')}\left|I_{kj,k'j'}^{(1)} - \left(-\frac{(1-2^{-\bar{B}_{jj'}})\omega_{jj'}}{(1+\omega_{jj'}\Gamma)^2}\frac{\partial\Gamma}{\partial\rho}\right)\right| \xrightarrow{i.p.} 0,$$

$$\max_{j,j'=1,2,k,k'\leq K,(k,j)\neq(k',j')}\left|I_{kj,k'j'}^{(2)} - \epsilon_{jj'}2^{-\bar{B}_{jj'}}\frac{\partial\Gamma}{\partial\rho}\right| \xrightarrow{i.p.} 0,$$

and

$$\max_{j,j'=1,2,k,k'\leq K,(k,j)\neq(k',j')}|I_{kj,k'j'}^{(3)}| \xrightarrow{i.p.} 0,$$

where $\bar{B}_{jj'} = \bar{B}_d$ when $j = j'$ and otherwise $\bar{B}_{jj'} = \bar{B}_c$.

Combining the results, we have

$$\max_{j,j'=1,2,k,k'\leq K,(k,j)\neq(k',j')} \left| \mathcal{I} - \left(-\frac{(1-2^{-\bar{B}_{jj'}})\omega_{jj'}}{(1+\omega_{jj'}\Gamma)^2} - \epsilon_{jj'}2^{-\bar{B}_{jj'}} \right) \frac{\partial \Gamma}{\partial \rho} \right| \xrightarrow{i.p.} 0. \tag{7.61}$$

Using (7.61), the large system result for the interference term can be written as follows

$$\begin{aligned} \sum_{(k',j')\neq(k,j)} |\mathbf{h}_{k,j,j'}\mathbf{w}_{k'j'}|^2 &= \sum_{l=1,l\neq k}^{K} |\mathbf{h}_{k,j,j}\mathbf{w}_{lj}|^2 + \sum_{l=1}^{K} |\mathbf{h}_{k,j,\bar{j}}\mathbf{w}_{l\bar{j}}|^2 \\ &\xrightarrow{i.p.} -\beta \left(\frac{1-2^{-\bar{B}_d}}{(1+\Gamma)^2} + \frac{\epsilon(1-2^{-\bar{B}_c})}{(1+\epsilon\Gamma)^2} + 2^{-\bar{B}_d} + \epsilon 2^{-\bar{B}_c} \right) \frac{\partial \Gamma}{\partial \rho}. \end{aligned}$$

By using the result from the previous results straightforwardly, we have

$$c_j^2 \xrightarrow{a.s.} \frac{P_t}{-\beta \frac{\partial \Gamma}{\partial \rho}}.$$

Putting all the large system results, we can show that the limiting signal strength is

$$\frac{1}{\beta} P_d \phi_d \Gamma \left(\rho + \frac{\beta\epsilon}{(1+\epsilon\Gamma)^2} + \frac{\beta}{(1+\Gamma)^2} \right), \tag{7.62}$$

and the limiting interference energy becomes

$$P_d \left(\frac{\phi_d}{(1+\Gamma)^2} + \frac{\epsilon\phi_c}{(1+\epsilon\Gamma)^2} + \delta_d + \delta_c \right). \tag{7.63}$$

Let $\rho_{\text{C,Q}} = \rho$ and $\Gamma_Q = \Gamma$. Then, we can obtain the limiting SINR given by (7.35) from (7.62) and (7.63) straightforwardly.

Chapter 8
Conclusion

IN this thesis, we consider system performance maximization in terms of SINR or achievable sum rate in single-cell and multicell MISO broadcast channels with the precoder having the RCI structure. We investigate optimal design parameters for the precoder for some channels and various CSI conditions available at the transmitter/base station. The results are obtained by conducting the analysis in the large system regime. The results give hints on the system performance and behaviors in the finite dimensional systems. Moreover, applying the optimal design parameters from the asymptotic analysis to the finite-size system only incurs a small loss, sometimes negligible, in system performance compared to the system optimizations in the finite-size system regimes. The latter could be computationally expensive.

A brief summary of the results in this work can be described as follows. In i.i.d. channel with perfect CSIT, the optimal regularization parameter of the precoder maximizing the limiting SINR turns out to be a ratio between the cell loading and received SNR. The same result is also obtained in the presence of spatial correlation at the transmitter side. It is a surprising result since the limiting SINR itself is affected by the correlation. In both cases, we also investigate the performance of the MPCI and SU precoders and observe that they are outperformed by the RCI precoder. In the context of TDD scheme, we consider the weighted sum rate maximization of the downlink and uplink data transmissions when the transmitter only has the imperfect CSI obtained from the channel training. We show that only a minimum training symbols per user, i.e., one symbol per user, is needed if the optimal power control between the training and uplink data transmission is performed. Allowing power allocation across the grouped users' data symbols where

the groups have different path-losses, we show that the maximum (limiting) sum rate is achieved by employing the water-filling strategy: a group with better path-loss gain receives more power. Under this optimal power policy, we also investigate the multi-mode transmission where we study the optimal number of groups communicating with the transmitter that maximizes the sum rate. Two cases are considered. First, all groups have the same cell-loading and the transmitter should decide whether to transmit to all user in the groups or to none of them. In the second case, the transmitter is allowed to send the data to only subsets of the users in the groups. In the last part of this thesis, we explore feedback optimization in a symmetric two-cell network for different levels of cooperation between the base stations. We take into account both the noisy analog and digital/quantized feedback schemes. A common behavior of the limiting SINR with respect to the interfering channel gain is observed. Below a certain threshold of that gain, no-cooperation between base stations is preferred. For the multicell processing scheme under some conditions, we observe a behavior where the limiting SINR is first decreasing until reaching a certain value of the interfering channel gain and then increasing.

The extensions of this thesis will be more toward design optimizations in multicell networks. Related to the last previous chapter, future works could consider a more general channel model such as the analog feedback through MAC channels and a more practical network models. Furthermore, feedback reduction problems in which the users or groups of users have different path-loss gains would be interesting to explore.

Bibliography

[1] T. Y. Al-Naffouri, M. Sharif, and B. Hassibi, "How Much Does Transmit Correlation Affect the Sum-Rate Scaling of MIMO Gaussian Broadcast Channels?" *IEEE Transactions on Communications*, vol. 57, no. 2, pp. 562–572, February 2009.

[2] C. K. Au-Yeung and D. J. Love, "On the performance of random vector quantization limited feedback beamforming in a MISO system," *IEEE Transactions on Wireless Communications*, vol. 6, no. 2, pp. 458–462, 2007.

[3] Z. D. Bai and J. W. Silverstein, "No eigenvalues outside the support of the limiting spectral distribution of large dimensional sample covariance matrices," *Annals of Probability*, vol. 26, no. 1, pp. 316–345, 1998.

[4] ——, *Spectral Analysis of Large Dimensional Random Matrices*, ser. Springer Series in Statistics. Springer, 2009.

[5] R. Bhagavatula and R. W. Heath, "Adaptive Limited Feedback for Sum-Rate Maximizing Beamforming in Cooperative Multicell Systems," *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 800–811, 2011.

[6] E. Biglieri, R. Calderbank, A. Costantinides, A. Goldsmith, and A. Paulraj, *MIMO Wireless Communications*. New York: Cambridge University Press, 2007.

[7] S. Boyd and L. Vandenbergehe, *Convex Optimization*. Cambridge University Press, 2004.

[8] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO Achievable Rates With Downlink Training and Channel State Feedback," *IEEE Transactions on Information Theory*, vol. 56, no. 6, pp. 2845–2866, 2010.

[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, 2003.

[10] M. Chiani, M. 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.

[11] D. Chizhik, G. Foschini, M. Gans, and R. Valenzuela, "Keyholes, correlations, and capacities of multielement transmit and receive antennas," *IEEE Transactions on Wireless Communications*, vol. 1, no. 2, pp. 361 –368, apr 2002.

[12] D. Chizhik, F. Rashid-Farrokhi, J. Ling, and A. Lozano, "Effect of antenna separation on the capacity of BLAST in correlated channels," *IEEE Communications Letters*, vol. 4, no. 11, pp. 337 –339, nov. 2000.

[13] C.-N. Chuah, D. Tse, J. M. Kahn, and R. A. Valenzuela, "Capacity scaling in MIMO wireless systems under correlated fading," *IEEE Transactions on Information Theory*, vol. 48, no. 3, pp. 637–650, March 2002.

[14] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications*. Cambridge University Press, 2011.

[15] X. Cui, Q. Zhang, and Z. Feng, "Generic procedure for tightly bounding the capacity of MIMO correlated Rician fading channels," *IEEE Transactions on Communications*, vol. 53, no. 5, pp. 890 – 898, may 2005.

[16] W. Dai, Y. Liu, B. Rider, and W. Gao, "How many users should be turned on in a multi-antenna broadcast channel?" *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 8, pp. 1526 –1535, october 2008.

[17] A. F. Dana, M. Sharif, and B. Hassibi, "On the capacity region of multi-antenna Gaussian broadcast channels with estimation error," in *Int. Symp. Information Theory (ISIT)*, Seattle, July 2006.

[18] G. Dimic and N. Sidiropoulos, "On downlink beamforming with greedy user selection: performance analysis and a simple new algorithm," *IEEE Transactions on Signal Processing*, vol. 53, no. 10, pp. 3857 – 3868, oct. 2005.

[19] U. Erez and S. ten Brink, "A close-to-capacity dirty paper coding scheme," *IEEE Transactions on Information Theory*, vol. 51, no. 10, pp. 3417–3432, 2005.

[20] J. Evans and D. N. C. Tse, "Large system performance of linear multiuser receivers in multipath fading channels," *IEEE Transactions on Information Theory*, vol. 46, no. 6, pp. 2059–2078, 2000.

[21] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, pp. 311–335, 1998.

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

[23] G. Foschini, K. Karakayali, and R. Valenzuela, "Coordinating multiple antenna cellular networks to achieve enormous spectral efficiency," *IEE Proceedings Communications*, vol. 153, no. 4, pp. 548–555, 2006.

[24] D. Gesbert, H. Bolcskei, D. Gore, and A. Paulraj, "Outdoor MIMO wireless channels: models and performance prediction," *IEEE Transactions on Communications*, vol. 50, no. 12, pp. 1926 – 1934, dec 2002.

[25] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 9, pp. 1380–1408, 2010.

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

[27] A. Goldsmith and P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Transactions on Information Theory*, vol. 43, no. 6, pp. 1986 –1992, nov 1997.

[28] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*, 7th ed., A. Jeffrey and D. Zwillinger, Eds. Elsevier, 2007.

[29] R. M. Gray, *Toeplitz and Circulant Matrices: a Review*, ser. Foundations and Trends in Communication and Information Theory. Now Publishers, 2005, vol. 2, no. 3.

[30] S. V. Hanly and D. N. C. Tse, "Resource pooling and effective bandwidths in CDMA networks with multiuser receivers and spatial diversity," *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1328–1351, 2001.

[31] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Transaction on Information Theory*, vol. 49, no. 4, 2003.

[32] B. Hochwald, C. Peel, and A. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part II: perturbation," *IEEE Transactions on Communications*, vol. 53, no. 3, pp. 537 – 544, march 2005.

[33] H. Huh, A. Tulino, and G. Caire, "Network MIMO With Linear Zero-Forcing Beamforming: Large System Analysis, Impact of Channel Estimation, and Reduced-Complexity Scheduling," *IEEE Transactions on Information Theory*, vol. 58, no. 5, pp. 2911 –2934, May 2012.

[34] S. Jafar and A. Goldsmith, "Isotropic Fading Vector Broadcast Channels:The Scalar Upper Bound and Loss in Degrees of Freedom," *IEEE Transactions on Information Theory*, vol. 51, no. 3, pp. 848 – 857, march 2005.

[35] X. Jin, Y. Yang, L. Tian, D. Pang, J. Shi, and E. Dutkiewicz, "QoS-aware optimal power allocation with channel inversion regularization precoding in MU-MIMO," in *Proc. IEEE International Conference on Communications*, June 2009.

[36] N. Jindal, "MIMO Broadcast Channels With Finite-Rate Feedback," *IEEE Transactions on Information Theory*, vol. 52, no. 11, pp. 5045–5060, 2006.

[37] ——, "Antenna combining for the MIMO downlink channel," *IEEE Transactions on Wireless Communications*, vol. 7, no. 10, pp. 3834–3844, 2008.

[38] S. Jing, D. N. C. Tse, J. B. Soriaga, J. Hou, J. E. Smee, and R. Adovani, "Multi-cell downlink capacity with coordinated processing," *EURASIP Journal on Wireless Communications and Networking*, vol. 2008, pp. 1–19, 2008.

[39] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot Contamination and Precoding in Multi-Cell TDD Systems," *IEEE Transactions on Wireless Communications*, vol. 10, no. 8, pp. 2640–2651, 2011.

[40] J. Jose, A. Ashikhmin, P. Whiting, and S. Vishwanath, "Channel Estimation and Linear Precoding in Multiuser Multiple-Antenna TDD Systems," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 5, pp. 2102 –2116, June 2011.

[41] J. Jose, A. Ashikhmint, P. Whiting, and S. Vishwanath, "Scheduling and pre-conditioning in multi-user MIMO TDD systems," in *IEEE International Conference on Communications*, 2008, pp. 4100–4105.

[42] M. Karakayali, G. Foschini, and R. Valenzuela, "Network coordination for spectrally efficient communications in cellular systems," *IEEE Wireless Communications*, vol. 13, no. 4, pp. 56 –61, aug. 2006.

[43] Kiran and D. N. C. Tse, "Effective interference and effective bandwidth of linear multiuser receivers in asynchronous cdma systems," *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1426–1447, 2000.

[44] M. Kobayashi, G. Caire, and N. Jindal, "How much training and feedback are needed in MIMO broadcast channels?" in *International Symposium on Information Theory (ISIT)*, Toronto, July 2008.

[45] M. Kobayashi, N. Jindal, and G. Caire, "Training and feedback optimization for multiuser MIMO downlink," *IEEE Transactions on Communications*, vol. 59, no. 8, pp. 2228–2240, 2011.

[46] E. Larsson and P. Stoica, *Space-Time Block Coding for Wireless Communications*. Cambridge University Press, 2003.

[47] J.-W. Lee, J.-Y. Ko, and H.-Y. Lee, "Effect of Transmit Correlation on the Sum-Rate Capacity of Two User Broadcast Channels," *IEEE Transactions on Communications*, vol. 57, no. 9, September 2009.

[48] Y.-C. Liang, G. Pan, and Z. D. Bai, "Asymptotic performance of MMSE receivers for large systems using random matrix theory," *IEEE Transactions on Information Theory*, vol. 53, no. 11, pp. 4173–4190, 2007.

[49] R. Louie, M. McKay, and I. Collings, "Impact of Correlation on the Capacity of Multiple Access and Broadcast Channels with MIMO-MRC," *IEEE Transactions on Wireless Communications*, vol. 7, no. 6, pp. 2397 –2407, june 2008.

[50] D. J. Love, R. W. Heath, V. K. N. Lau, D. Gesbert, B. D. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 8, pp. 1341–1365, 2008.

[51] D. Love, J. Heath, R.W., W. Santipach, and M. Honig, "What is the value of limited feedback for MIMO channels?" *IEEE Communications Magazine*, vol. 42, no. 10, pp. 54 – 59, Oct. 2004.

[52] 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, 2004.

[53] V. A. Marčenko and L. A. Pastur, "Distribution of eigenvalues for some sets of random matrices," *Math. USSR-Sbornik*, vol. 1, no. 4, pp. 457–483, 1967.

[54] T. L. Marzetta and B. M. Hochwald, "Fast transfer of channel state information in wireless systems," *IEEE Transactions on Signal Processing*, vol. 54, no. 4, pp. 1268–1278, 2006.

[55] M. McKay and I. Collings, "General Capacity Bounds for Spatially Correlated Rician MIMO Channels," *IEEE Transactions on Information Theory*, vol. 51, no. 9, pp. 3121 – 3145, sept. 2005.

[56] ——, "On the capacity of frequency-flat and frequency-selective Rician MIMO channels with single-ended correlation," *IEEE Transactions on Wireless Communications*, vol. 5, no. 8, pp. 2038 –2043, aug. 2006.

[57] X. Mestre, J. R. Fonollosa, and A. Pages-Zamora, "Capacity of MIMO channels: Asymptotic evaluation under correlated fading," *IEEE Journal on Selected Areas in Communication*, vol. 21, no. 5, pp. 829–838, June 2003.

[58] R. Muharar and J. Evans, "Downlink beamforming with transmit-side channel correlation: A large system analysis," in *Proc. IEEE International Conference on Communications (ICC)*, 2011, pp. 1–5.

[59] ——, "Optimal training for time-division duplexed systems with transmit beamforming," in *Proc. Australian Communications Theory Workshop (AusCTW)*, 2011, pp. 158–163.

[60] R. Muharar, R. Zakhour, and J. Evans, "Base Station Cooperation with Limited Feedback: A Large System Analysis," in *Proc. Int. Symp. Inform. Theory (ISIT)*, Cambridge, MA, July 2012.

[61] ——, "Base Station Cooperation with Noisy Analog Channel Feedback: A Large System Analysis," in *Proc. Int. Conf. Commun. (ICC)*, Ottawa, Canada, June 2012.

[62] V. K. Nguyen and J. S. Evans, "Multiuser Transmit Beamforming via Regularized Channel Inversion: A Large System Analysis," in *IEEE Global Telecommunications Conference (GLOBECOM)*, December 2008.

[63] V. K. Nguyen, R. Muharar, and J. Evans, "Multiuser transmit beamforming via regularized channel inversion: A large system analysis," *Unpublished*, November 2009. [Online]. Available: http://cubinlab.ee.unimelb.edu.au/rmuharar

[64] U. Niesen, D. Shah, and G. Wornell, "Adaptive alternating minimization algorithms," *IEEE Transactions on Information Theory*, vol. 55, no. 3, pp. 1423 – 1429, March 2009.

[65] O. Oyman, R. Nabar, H. Bolcskei, and A. Paulraj, "Characterizing the statistical properties of mutual information in MIMO channels," *IEEE Transactions on Signal Processing*, vol. 51, no. 11, pp. 2784 – 2795, nov 2003.

[66] S. Ozyurt and M. Torlak, "Performance analysis of optimum zero-forcing beamforming with greedy user selection," *IEEE Communications Letters*, vol. 16, no. 4, pp. 446 –449, april 2012.

[67] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge University Press, 2003.

[68] M. J. M. Peacock, "Random Matrix Theory Analysis of Fixed and Adaptive Linear Multiuser Receivers," Ph.D. dissertation, School of Electrical and Information Engineering, The University of Sydney, August 2005.

[69] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part I: channel inversion and regularization," *IEEE Transactions on Communications*, vol. 53, no. 1, pp. 195–202, 2005.

[70] V. Raghavan, J. Kotecha, and A. Sayeed, "Why Does the Kronecker Model Result in Misleading Capacity Estimates?" *IEEE Transactions on Information Theory*, vol. 56, no. 10, pp. 4843 –4864, oct. 2010.

[71] S. A. Ramprashad and G. Caire, "Cellular vs. network MIMO: A comparison including the channel state information overhead," in *Proc. IEEE 20th Int. Personal, Indoor and Mobile Radio Communications Symp*, 2009, pp. 878–884.

[72] S. A. Ramprashad, G. Caire, and H. C. Papadopoulos, "Cellular and network MIMO architectures: MU-MIMO spectral efficiency and costs of channel state information," in *Proc. The Forty-Third Asilomar Conf. Signals, Systems and Computers*, 2009, pp. 1811–1818.

[73] N. Ravindran and N. Jindal, "Multi-user diversity vs. accurate channel feedback for MIMO broadcast channels," in *Proc. IEEE Int. Conf. Communications (ICC)*, 2008, pp. 3684–3688.

[74] A. Razi, D. Ryan, I. Collings, and J. Yuan, "Sum rates, rate allocation, and user scheduling for multi-user MIMO vector perturbation precoding," *IEEE Transactions on Wireless Communications*, vol. 9, no. 1, pp. 356 –365, january 2010.

[75] F. Rusek, D. Persson, B. Lau, E. Larsson, T. Marzetta, O. Edfors, and F. Tufvesson, "Scaling Up MIMO: Opportunities and Challenges with Very Large Arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40 –60, jan. 2013.

[76] D. Samardzija and N. Mandayam, "Unquantized and uncoded channel state information feedback in multiple-antenna multiuser systems," *IEEE Transactions on Communications*, vol. 54, no. 7, pp. 1335–1345, 2006.

[77] W. Santipach and M. L. Honig, "Capacity of a Multiple-Antenna Fading Channel With a Quantized Precoding Matrix," *IEEE Transactions on Information Theory*, vol. 55, no. 3, pp. 1218–1234, 2009.

[78] ——, "Optimization of Training and Feedback Overhead for Beamforming Over Block Fading Channels," *IEEE Transactions on Information Theory*, vol. 56, no. 12, pp. 6103–6115, 2010.

[79] A. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Transactions on Signal Processing*, vol. 50, no. 10, pp. 2563 – 2579, oct 2002.

[80] S. Shamai and B. Zaidel, "Enhancing the cellular downlink capacity via co-processing at the transmitting end," in *53rd IEEE Vehicular Technology Conference*, vol. 3, 2001, pp. 1745–1749 vol.3.

[81] H. Shin and J. H. Lee, "Capacity of multiple-antenna fading channels: spatial fading correlation, double scattering, and keyhole," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2636 – 2647, oct. 2003.

[82] D.-S. Shiu, G. Foschini, M. Gans, and J. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Transactions on Communications*, vol. 48, no. 3, pp. 502 –513, mar 2000.

[83] J. W. Silverstein and Z. D. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," *Journal Multivariate Analysis*, vol. 54, pp. 175–192, 1995.

[84] O. Sjobergh, E. Jorswieck, and E. Larsson, "Greedy user selection for zero-forcing and MMSE multiuser beamforming with channel estimation errors," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 31 2008-april 4 2008, pp. 3137 –3140.

[85] O. Somekh, O. Simeone, Y. Bar-Ness, A. Haimovich, and S. Shamai, "Cooperative Multicell Zero-Forcing Beamforming in Cellular Downlink Channels," *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3206 –3219, July 2009.

[86] Q. H. Spencer, C. B. Peel, A. L. Swindlehurst, and M. Haardt, "An introduction to the multi-user MIMO downlink," *IEEE Communications Magazine*, vol. 42, no. 10, pp. 60–67, 2004.

[87] 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, 2004.

[88] T. Tao, *Topics in Random Matrix Theory*, ser. Graduate Studies in Mathematics. American Mathematical Society, 2012.

[89] E. Telatar, "Capacity of multi-antenna Gaussian channels," AT&T Bell Lab., Lucent Technology, Technical Report, 1995.

[90] ——, "Capacity of Multi-antenna Gaussian Channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.

[91] D. N. C. Tse and S. V. Hanly, "Linear multiuser receivers: effective interference, effective bandwidth and user capacity," *IEEE Transactions on Information Theory*, vol. 45, no. 2, pp. 641–657, 1999.

[92] D. N. C. Tse and O. Zeitouni, "Linear multiuser receivers in random environments," *IEEE Transactions on Information Theory*, vol. 46, no. 1, pp. 171–188, 2000.

[93] D. Tse and S. Verdu, "Optimum asymptotic multiuser efficiency of randomly spread cdma," *IEEE Transactions on Information Theory*, vol. 46, no. 7, pp. 2718 – 2722, nov 2000.

[94] A. M. Tulino, A. Lozano, and S. Verdu, "Impact of antenna correlation on the capacity of multiantenna channels," *IEEE Transactions on Information Theory*, vol. 51, no. 7, pp. 2491–2509, July 2005.

[95] A. M. Tulino and S. Verdu, "Random Matrix Theory and Wireless Communications," *Foundations and Trends in Communications and Information Theory*, vol. 1, no. 1, 2004.

[96] A. W. van der Vaart, *Asymptotic Statistics*, ser. Cambridge series in statistical and probablistic mathematics. Cambridge University Press, 1998.

[97] S. Verdu and S. Shamai, "Spectral efficiency of CDMA with random spreading," *IEEE Transactions on Information Theory*, vol. 45, no. 2, pp. 622 –640, mar 1999.

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

[99] 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, 2003.

[100] M. Vu and A. Paulraj, "MIMO Wireless Linear Precoding," *Signal Processing Magazine, IEEE*, vol. 24, no. 5, pp. 86 –105, sept. 2007.

[101] S. Wagner, R. Couillet, M. Debbah, and D. T. M. Slock, "Large System Analysis of Linear Precoding in Correlated MISO Broadcast Channels Under Limited Feedback," *IEEE Transactions on Information Theory*, vol. 58, no. 7, pp. 4509 –4537, July 2012.

[102] J. Wang, D. Love, and M. Zoltowski, "User Selection With Zero-Forcing Beamforming Achieves the Asymptotically Optimal Sum Rate," *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3713 –3726, aug. 2008.

[103] H. Weingarten, Y. Steinberg, and S. Shamai, "The Capacity Region of the Gaussian Multiple-Input Multiple-Output Broadcast Channel," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 3936–3964, 2006.

[104] J. Wishart, "The Generalised Product Moment Distribution in Samples from a Normal Multivariate Population," *Biometrika*, vol. 20A, no. 1/2, pp. 32–52, July 1928.

[105] Y. Xu and T. Le-Ngoc, "Optimal power allocation with Channel Inversion Regularization-based precoding for MIMO broadcast channels," *Eurasip Journal on Advances in Signal Processing*, vol. 2008, 2008.

[106] ——, "A Capacity-Achieving Precoding Scheme Based on Channel Inversion Regularization with Optimal Power Allocation for MIMO Broadcast Channels," in *IEEE Global Telecommunications Conference (GLOBECOM)*, Nov. 2007, pp. 3190 –3194.

[107] T. Yoo and A. Goldsmith, "On the Optimality of Multi-Antenna Broadcast Scheduling Using Zero-Forcing Beamforming," *IEEE Journal on Selected Areas in Communication*, vol. 24, no. 3, pp. 528–541, March 2006.

[108] 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, 2004.

[109] R. Zakhour and S. Hanly, "Base Station Cooperation on the Downlink: Large System Analysis," *IEEE Transactions on Information Theory*, vol. 58, no. 4, pp. 2079 –2106, April 2012.

[110] H. Zhang and H. Dai, "Cochannel interference mitigation and cooperative processing in downlink multicell multiuser MIMO networks," *Eurasip Journal on Wireless Communications and Networking*, vol. 2004, no. 2, pp. 222–235, December 2004.

[111] J. Zhang, J. G. Andrews, and K. B. Letaief, "Spatial Intercell Interference Cancellation with CSI Training and Feedback," *IEEE Transactions on Wireless Communications*, May 2011, submitted. [Online]. Available: http://arxiv.org/pdf/1105.4206

[112] J. Zhang, M. Kountouris, J. Andrews, and R. Heath, "Achievable throughput of multi-mode multiuser MIMO with imperfect CSI constraints," in *IEEE International Symposium on Information Theory (ISIT)*, 2009, pp. 2659 –2663.

[113] ——, "Multi-Mode Transmission for the MIMO Broadcast Channel with Imperfect Channel State Information," *IEEE Transactions on Communications*, vol. 59, no. 3, pp. 803 –814, March 2011.