MEE10:118

Master's Thesis
Electrical Engineering with emphasis on Radio Communications

TDD Linear Precoding Methods for Next Generation Mobile Communication Systems

Author: Yuan Ding
Email: dingyuan.hit@gmail.com

Author: YiBing Wu
Email: yibingwu1986@gmail.com

Blekinge Institute of Technology
March 2011

Blekinge Institute of Technology
School of Engineering
Department of Electrical Engineering
Supervisor: Professor Abbas Mohammed
Examiner: Professor Abbas Mohammed

Acknowledgment

We are most grateful to our supervisor Professor Abbas Mohammed, who made this thesis possible and helped us with valuable discussions and feedback.

To our fellow student colleagues and friends at BTH, we will never forget this unique moment in our life, sharing different ideas from all around the world and learning from each other throughout the whole master program study.

We are especially thankful to our families for their relentless efforts to support us, both morally and financially, during the two years.

Finally, we give our best regards to BTH faculty and the lovely people in Karlskrona, who are friendly and helpful to our foreign students and make the study and living here really comfortable.

Abstract

The mobile communication has stepped into the era of the third generation (3G) worldwide, and plenty of new products and services for mobile communications are emerging every day, for giving people better audio and video experience. Recently, the fourth generation (4G) is starting to be deployed in some countries. However, people cannot help to wonder what will the next generation mobile communication system look like? In this thesis, we give our prediction about the next mobile communication system, which extends the capacity in spatial domain by using a large number of antennas and still keeps the compatibility with the former generation mobile communication systems. However, for the incompatible part—precoding—in such new system, we present two TDD linear precoding methods in this thesis. Through simulations, we can see that these two precoding methods are feasible for the new system and help to increase its throughput performance.

Keyword: linear precoding, next generation, training, TDD

TABLE OF CONTENTS

1 INTRODUCTION 5

1.1 RESEARCH QUESTIONS 5
1.2 HYPOTHESIS 5
1.3 MAIN CONTRIBUTION 5
1.4 ORGANIZATION 6

2 NEXT GENERATION MOBILE COMMUNICATION SYSTEMS AND CORRESPONDING PRECODING METHODS 7

2.1 INTRODUCTION 7
2.2 RETROSPECT THE HISTORY OF THE DEVELOPMENT OF THE MOBILE COMMUNICATION SYSTEMS 7
2.3 NEXT GENERATION MOBILE COMMUNICATION SYSTEMS 8
2.4 PRECODING METHODS FOR NEXT GENERATION MOBILE COMMUNICATION SYSTEMS 8
2.5 SUMMARY 9

3 GENERALIZED ZERO-FORCING PRECODING BASED ON THE UPLINK TRAINING 11

3.1 INTRODUCTION 11
3.2 SYSTEM MODEL AND UPLINK TRAINING 12
3.2.1 System model 12
3.2.2 Channel estimation and uplink training 12
3.3 GENERALIZED ZERO-FORCING PRECODING 13
3.4 ACHIEVABLE SUM DATA RATE 14
3.5 OPTIMIZATION OF PRECODING MATRIX 17
3.6 SCHEDULING STRATEGY 19
3.6.1 Homogeneous users 19
3.6.2 Heterogeneous Users 21
3.7 OPTIMAL TRAINING LENGTH 21
3.8 SUMMARY 22

4 SVH PRECODING METHOD AND TRAINING ON BOTH UPLINK AND DOWNLINK 23

4.1 INTRODUCTION 23
4.2 SVH PRECODING METHOD 23
4.3 MODIFIED SVH PRECODING METHOD 24
4.4 DOWNLINK TRAINING 26
4.5 ACHIEVABLE THROUGHPUT 26
4.5.1 Lower Bound of the Sum Rate 26
4.5.2 Upper Bound of the Sum Rate 27
4.6 SUMMARY 27

5 SIMULATION 28

5.1 INTRODUCTION 28
5.2 SIMULATION ON THE GENERALIZED ZERO-FORCING PRECODING METHOD 29
5.2.1 Homogeneous Users 29
5.2.2 Heterogeneous Users 31
5.3 SIMULATION ON THE SVH PRECODING METHOD 33
5.3.1 Simulation on the SVH precoding with perfect knowledge of the channel 33

5.3.2 Simulation on the modified SVH precoding based on the estimation of the channel and the statistics of the error *34*
5.3.3 Simulation on the modified SVH2 precoding based only on the estimation of the channel *35*
5.4 SIMULATION ON THE LARGER COHERENCE TIME INTERVAL CASE 37
5.5 SIMULATION ON THE DIFFERENT ANTENNA ARRAY CASE 39
5.6 SUMMARY 40
6 CONCLUSIONS 41
REFERENCES 43

LIST OF FIGURES

Figure 3-1 Three-phase transmission scheme over one coherence interval 11
Figure 4-1 Four-phase transmission scheme over one coherence interval 23
Figure 5-1 The net sum rate performance after using the generalized ZF precoding without scheduling. 29
Figure 5-2 The comparison of the net sum rate performance after the generalized ZF precoding with and without scheduling. 30
Figure 5-3 The net sum rate curve with the change of the downlink SNR 31
Figure 5-4 The net sum rate performance of the heterogeneous setting after using generalized ZF precoding without scheduling. 32
Figure 5-5 The net sum rate performance of the heterogeneous setting after using the generalized ZF precoding with scheduling. 32
Figure 5-6 The comparison of the net sum rate performance after the generalized ZF precoding and after the SVH precoding. 33
Figure 5-7 The comparison of the net sum rate performance after the generalized ZF precoding, after the SVH precoding and after the modified SVH precoding. 35
Figure 5-8 The comparison of the net sum rate performance after the generalized ZF precoding, after the SVH precoding, after the modified SVH1 precoding and after the modified SVH2 precoding. 36
Figure 5-9 The net sum rate performance with different iteration length. 37
Figure 5-10 The comparison of the net sum rate performance after the modified SVH2 precoding when the coherence time interval $T = 30$ and $T = 50$ 38
Figure 5-11 The comparison of the net sum rate performance after the generalized ZF precoding, after the SVH precoding and after the modified SVH2 precoding when the coherence time interval $T = 50$ 38
Figure 5-12 The comparison of the net sum rate after the modified SHV2 precoding when the number of antenna $M = 100$, $M = 150$ and $M = 16$ 39

LIST OF ABBREVIATIONS AND MATHEMATICAL NOTATION

TDD	Time Division Duplex
FDD	Frequency Division Duplex
CSI	Channel State Information
LTE	Long Term Evolution
CDMA	Code Division Multiple Access
WCDMA	Wideband Code Division Multiple Access
TD-SCDMA	Time Division-Synchronous Code Division Multiple Access
MIMO	Multiple Input and Multiple Output
OFDM	Orthogonal Frequency Division Multiplexing
SNR	Signal to Noise Ratio
LMSE	Least Mean Square Error
ZF	Zero-Forcing
SVH	Stojnic Vikalo and Hassibi
$(\bullet)^T$	the transpose
$(\bullet)^*$	the conjugate
$(\bullet)^H$	the Hermitian of vectors and matrices
$\mathrm{Tr}(\mathbf{A})$	the trace of the matrix $\mathbf{A}$
$\mathbf{A}^{-1}$	the inverse of the matrix $\mathbf{A}$
$diag(a)$	diagonal matrix with diagonal entries equal to the components of a
$\mathrm{E}[\bullet]$	the expectation
$\mathrm{var}[\bullet]$	the variance

1 INTRODUCTION

Guglielmo Marconi invented the wireless telegraph in 1896. In 1901, he transmitted the telegraphic signals across the Atlantic Ocean. In the 1960s, communication satellites were first launched, which could only handle 240 voice circuits [1]. In the 1980s, the first generation (1G) analog communication systems were born. In the middle of 1990s, 1G communication systems were replaced by the second generation (2G) digital communication systems. At the beginning of 21st Century, some countries started to deploy 3G mobile communications, using WCDMA, CDMA2000 and TD-SCDMA standards [2]. However, when the third generation (3G) communications are just accepted by the users, engineers are busy in figuring out the fourth generation (4G) standards. Though it is said that 4G LTE is a long term evolution, we cannot help thinking about what next generation mobile communication systems will look like?

1.1 Research questions

1. What do next generation mobile communication systems (after 4G) look like?

2. How to implement the efficient precoding schemes for such systems?

1.2 Hypothesis

1. Next generation mobile communication systems will mainly extend their communication capacity in the spatial dimension, which means a large antenna array will be used in such systems. Also, the next generation mobile communication systems will be compatible with the techniques in the former generations, like TDD, CDMA, MIMO, and OFDM.

2. Using training sequences and TDD linear procoding will give us low cost but good throughput performance solutions for next generation mobile communication systems.

1.3 Main contribution

1. Give a prediction of next generation mobile communication systems.

2. Analyse the performance of precoding schemes, using training squences and TDD linear procoding methods for next generation mobile communication systems.

3. Implement MATLAB simulation to test and compare different precoding schemes and figure out the practical one.

1.4 Organization

Chapter 2 -- presents our prediction about next generation mobile communication systems and the corresponding precoding methods for such systems.

Chapter 3 -- presents the theory part of the generalized zero-forcing precoding method. After that, we analyze the sum rate performance, optimization of the precoding matrix, scheduling strategy (homogeneous users and heterogeneous users) and optimal training length.

Chapter 4 -- mainly presents the theory part of the SVH precoding method and two modified forms.

Chapter 5 -- shows the MATLAB simulation results of the two precoding methods in Chapter 3 and 4, then compares and analyzes the results.

Chapter 6 -- the conclusion and future work.

2 NEXT GENERATION MOBILE COMMUNICATION SYSTEMS AND CORRESPONDING PRECODING METHODS

2.1 Introduction

Next generation mobile communication system will be mainly built in a large antenna array system. Thus, the feature techniques in such generation mobile communication system must be suitable for such large antenna array case, which means that some technologies originally fit for the small antenna array systems (including 2 to 4 antennas) in 4G may not be suitable anymore and new approaches will be developed for the new mobile communication systems.

In this chapter, we first look back on the development of current and former mobile communication systems, from 1G to 4G. With the help of the retrospect, we will clue our prediction of the next generation mobile communication systems. After that, we give the reasons why we have to choose TDD linear precoding methods for next generation mobile communication systems.

2.2 Retrospect the history of the development of the mobile communication systems

In order to explain why we predict next generation mobile communication systems that have a large antenna array at the base station, it might be a good idea to retrospect the history about the development of the mobile communication systems (from 1G to 4G). 1G are the analog systems, which were used for public voice service with the speed up to 2.4kbps. The frequency domain is the main dimension for the mobile communication systems to implement the duplexing and multi-user accessing to realize the mobile communication, where FDD and FDMA are the main feature techniques. Compared with 1G, 2G systems are based on the digital technology and the data rate is up to 300kbps. The time domain is the added dimension for the systems to increase the capacity of the mobile communications, where TDD and TDMA are the main feature technologies beyond the original ones. When it comes for the 3G systems, the code domain is the new dimension for the systems to increase the capacity of the mobile communication. Although there are three standards -- WCDMA, CDMA2000 and TD-SCDMA, generally speaking, CDMA is the main feature technique for 3G, which helps to increase the data rate up to 2Mbps. Now, for the 4G systems, the spatial domain is involved

for increasing the data rate performance of the mobile communication systems, where 2 to 4 antennas are used and the main feature techniques are MIMO and OFDM [3]. The representative standards are LTE and WiMAX, and the data rate is expected to be 1Gbps for the low speed users in the near future.

2.3 Next generation mobile communication systems

From the retrospect above, there is an interesting observation that when a new standard is set up for the new generation mobile communication systems, a new dimension will be introduced for increasing the capacity of the new systems at the same time. In the similar way, it is expected that another new dimension should be added in future generation mobile communication systems. However, it might be impossible to find such new dimension, since it seems that all the domains that have the gift to increase the mobile communication quality are used. But if we take a close look at these four dimensions, we can find that the spatial domain still have the potential to increase the data rate of the mobile communication systems. As we know, the frequency resources are limited and also charged by the governments for the commercial use, which means we are not free to use the frequency bandwidth as wide as possible. For the time domain, due to the technique and practical limitation, we cannot divide the time slot as short as we can. After turbo codes were introduced, it seems that the codes give us a great balance for achieving Shannon limitation and practical implementation. Although we still have the chance to find even better codes, the cost paid for the hunt and the numerical complexity of the codes will probably be rather high. However, when we turn to the spatial domain, it gives us a heuristic hint. Only 2 to 4 antennas are used in 4G. We can improve the data rate performance of next generation mobile systems by increasing the number of antennas at the base station and the cost is also relatively low. In addition, it is fairly easy to be compatible with 4G systems by simply scheduling the large array into some small sub-arrays that contains 2 to 4 antennas just like it in 4G. MIMO and OFDM, for example, are still useful for the new systems. Thus, we predict the large antenna array and the corresponding technologies for such systems are the key features of next generation mobile communication systems.

2.4 Precoding methods for next generation mobile communication systems

The channel model for small array is not suitable for next generation mobile communication systems and should be modified. Accordingly, for precoding part, precoding algorithms for FDD systems and non-linear precoding algorithms like dirty paper coding [4]

are not helpful for large antenna array systems anymore, because the overhead from the feedback and complicity of the devices are increasing greatly with the increase of the number of antennas. Thus, TDD linear precoding algorithms for multiple users are the techniques that are applicable for large antenna array case.

The research about precoding strategies has been well developed in recent decade. A lot of literature, however, has focused on the precoding algorithms in the FDD systems [5]-[9], while the interest in precoding methods in TDD systems has grown only in recent years [11]-[13], [15]-[18]. Although it seems that FDD and TDD is interchangeable schemes for mobile communication systems, there are some fundamental differences that make the corresponding precoding algorithms quite different with each other. One of the primary differences between TDD and FDD systems is the different ways to obtain the CSI (channel state information) at the transmitter. In TDD systems, the time lag between the uplink and downlink is relatively small compared to the channel coherence time, thus the channel reciprocity principle can be used for obtaining the CSI at the base station by implementing the training in the uplink. In FDD systems, due to the large frequency offset between the uplink and downlink (normally 5% of the carrier frequency), channel reciprocity generally does not hold, the closed-loop methods using feedback from the receiver of the downlink are more applicable [10]. However, in next generation mobile communication systems, where the base station is made of an antenna array with a large number of antennas, the overhead from the feedback in FDD systems will be rather large, which is often neglected in small antenna array setting. Thus, the TDD precoding algorithm becomes an attractive approach for next generation mobile communication systems.

In addition, the linear precoding algorithms are more applicable than the non-linear ones in next generation mobile communication systems. Take the dirty paper coding for example, this non-linear procoding algorithm can help the systems to achieve the channel capacity with perfect CSI. However, if the CSI is not perfect, which is often the case in practice, there will be a significant loss in the data rate performance. Furthermore, even with perfect CSI, it is also not easy to implement such kind of non-linear precoding algorithms in practice. Moreover, the practical mobile devices are usually simple and low-cost. We cannot assume that they have the ability to cancel the interference. Therefore, the linear precoding algorithms in TDD would be the first choice for next generation mobile communication systems [11].

2.5 Summary

In this chapter, we gave our prediction on next generation mobile communication systems, the main feature of which is the large number of antennas setting at the base station. In such

systems, many of the techniques from former generation mobile communication systems will still be useful except for the precoding part. The TDD linear precoding methods could be a practical choice for next generation mobile communication systems.

3 GENERALIZED ZERO-FORCING PRECODING BASED ON THE UPLINK TRAINING

3.1 Introduction

In TDD systems, the channel reciprocity principle can be used for obtaining the CSI at the base station by using the training sequences in the uplink. Thus, we consider a transmission scheme that divides the whole coherence interval into three phases, as shown in Figure 3-1. There are the training phase, computing phase and data transmission phase. In the training phase, the users send the training sequences to the base station by the uplink channel. Then, in the computing phase, after receiving these training signals, the base station calculates the LMSE (linear least mean square error) estimation of the channel, which is regarded as an estimation of the instantaneous channel state. The base station gives the precoding matrix based on this channel estimation. We assume that it costs one symbol delay to do this computation. In practice, this delay may be different with different systems. In the data transmission phase, the base station sends data to the users.

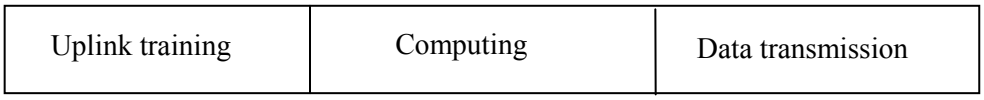

Figure 3-1 Three-phase transmission scheme over one coherence interval

Since we consider the setting of a large array with many antennas at the base station, the precoding method should take advantage of this antenna setting. In addition, the transmission period in one coherence interval is relatively short in the TDD systems and the CSI is imperfect at the base station. Therefore, the precoding method should be simple for saving more time for the data transmission phase but robust for the imperfect CSI situation. We implement the zero-forcing precoding, a well-studied simple linear algorithm in the perfect CSI setting, into our setting with the LMSE channel estimation, which we call the general zero-forcing precoding algorithm [11].

In this chapter, we obtain the channel estimation from uplink training, and then present the generalized zero-forcing precoding based on the channel estimation. After that, we analyse the sum rate performance, optimization of the precoding matrix, scheduling strategy (homogeneous users and heterogeneous users) and optimal training length.

3.2 System Model and Uplink Training

3.2.1 System model

The system model consists of a base station with M antennas and K single-antenna users. The downlink channel is characterized by the matrix $\mathbf{H}_{K\times M}$. For simplicity, we assume that the frequency selective fading is resolved by using the OFDM technique, and so we only pay our attention to the flat fading in this chapter. Here, the Rayleigh block fading model is implemented as the channel model, where the entries of the channel matrix $\mathbf{H}$ are independent and identically distributed (i.i.d) zero-mean, circularly symmetric complex Gaussian $\mathbb{C}N(0,1)$ random variables.

Let the downlink and uplink SNR of the k^{th} user be ρ_k^d and ρ_k^u, respectively. The $M\times 1$ signal vector is $\mathbf{s}^d$. The additive noise $\mathbf{z}^d$ is i.i.d $\mathbb{C}N(0,1)$. Then, the vector $\mathbf{x}^d$ received at the users is

$$\mathbf{x}^d = \mathbf{E}^d\mathbf{H}\mathbf{s}^d + \mathbf{z}^d \tag{3-1}$$

where $\mathbf{E}^d = diag\left\{\left[\sqrt{\rho_1^d},\sqrt{\rho_1^d},\cdots,\sqrt{\rho_K^d}\right]^T\right\}$.

Similarly, on the uplink, the vector $\mathbf{x}^u$ received at the base station is

$$\mathbf{x}^u = \mathbf{H}^T\mathbf{E}^u\mathbf{s}^u + \mathbf{z}^u \tag{3-2}$$

where $\mathbf{E}^u = diag\left\{\left[\sqrt{\rho_1^u},\sqrt{\rho_1^u},\cdots,\sqrt{\rho_K^u}\right]^T\right\}$.

3.2.2 Channel estimation and uplink training

Channel reciprocity is one of the key properties of TDD systems, since the base station exploits to obtain the channel estimation by the training on the uplink. Each user transmits a sequence of training signals of τ^u symbols duration in every coherence interval. The k^{th} user transmits the training sequence vector $\sqrt{\tau^u}\psi_k^H$. We use orthonormal sequences, i.e., $\psi_i^H\psi_j = \delta_{ij}$, where δ_{ij} is the Kronecker delta. In order to cover all the channels between the users and the base station, the minimum length of the training sequence is K, i.e., $\tau^u \geq K$.

The training signal matrix received at the base station is

$$\mathbf{Y} = \sqrt{\tau^u}\mathbf{H}^T\mathbf{E}^u\boldsymbol{\Psi}^H + \mathbf{V}^u \tag{3-3}$$

where $\boldsymbol{\Psi} = [\psi_1\psi_2\cdots\psi_K]$ $(\boldsymbol{\Psi}^H\boldsymbol{\Psi} = \mathbf{I})$ and the elements of $\mathbf{V}^u$ are i.i.d. $\mathbb{C}N(0,1)$.

The base station obtains the LMSE estimate of the channel

$$\hat{\mathbf{H}} = diag\left\{\left[\frac{\sqrt{\rho_1^u\tau^u}}{1+\rho_1^u\tau^u}, \frac{\sqrt{\rho_2^u\tau^u}}{1+\rho_2^u\tau^u}, \cdots, \frac{\sqrt{\rho_K^u\tau^u}}{1+\rho_K^u\tau^u}\right]^T\right\}\boldsymbol{\Psi}^T\mathbf{Y}^T \tag{3-4}$$

where $\hat{\mathbf{H}}$ is the LMSE estimation of the channel. By the properties of condition mean and joint Gaussian distribution, the estimation $\hat{\mathbf{H}}$ is independent of the estimation error $\tilde{\mathbf{H}} = \mathbf{H} - \hat{\mathbf{H}}$. The components of $\tilde{\mathbf{H}}$ and $\hat{\mathbf{H}}$ are independent, and the elements of their k^{th} row are $\mathbb{C}N\left(0, \frac{\rho_k^u\tau^u}{1+\rho_k^u\tau^u}\right)$ and $\mathbb{C}N\left(0, \frac{1}{1+\rho_k^u\tau^u}\right)$, respectively.

3.3 Generalized Zero-Forcing Precoding

This generalized zero-forcing precoding algorithm can be implemented in both homogeneous and heterogeneous users' settings. In the heterogeneous users setting, this precoding method is performed in two steps: (1) select users and (2) precoding optimization for the selected users. We denote the scheduling algorithm that selects the users as $S(\hat{\mathbf{H}}) = \{S_1, S_2, \cdots, S_N\} \subseteq \{1,2,\cdots,K\}$, i.e., the scheduling algorithm selects users $S_1, S_2, \cdots, S_N$ based on the channel estimation $\hat{\mathbf{H}}$. Next, denote $p_1, p_2, \cdots, p_K$ as the optimization of the precoding matrix, which are some positive constants. Let

$$\mathbf{D}_S = diag\left\{\left[p_{S_1}^{-\frac{1}{2}}, p_{S_2}^{-\frac{1}{2}}, \cdots, p_{S_N}^{-\frac{1}{2}}\right]^T\right\}. \tag{3-5}$$

$\hat{\mathbf{H}}_S$ is the matrix formed by the rows in set $S(\hat{\mathbf{H}})$ of matrix $\hat{\mathbf{H}}$. Similarly, $\mathbf{H}_S$ be the matrix formed by the rows in set $S(\hat{\mathbf{H}})$ of matrix $\mathbf{H}$, and $\tilde{\mathbf{H}}_S$ be the matrix formed by the rows in set $S(\hat{\mathbf{H}})$ of matrix $\tilde{\mathbf{H}}$, respectively.

Let $\hat{\mathbf{H}}_{DS} = \mathbf{D}_S\hat{\mathbf{H}}_S$. The generalized zero-forcing precoding matrix is

$$\mathbf{A}_{DS} = \frac{\hat{\mathbf{H}}_{DS}^H(\hat{\mathbf{H}}_{DS}\hat{\mathbf{H}}_{DS}^H)^{-1}}{\sqrt{\mathrm{Tr}\left[\left((\hat{\mathbf{H}}_{DS}\hat{\mathbf{H}}_{DS}^H)^{-1}\right)\right]}}. \tag{3-6}$$

This precoding matrix is normalized so that

$$\mathrm{Tr}(\mathbf{A}_{DS}^{H}\mathbf{A}_{DS}) = 1 \tag{3-7}$$

For this linear precoding algorithm, the transmission signal-vector for the i^{th} user of these selected users is given by

$$\mathbf{s}_i^d = \mathbf{A}_{DS}\mathbf{q}_i \tag{3-8}$$

where $\mathbf{s}_i^d$ is the transmission signal of the i^{th} user in the downlink and $\mathbf{q}_i$ is the input information sequence of the i^{th} user.

The constraint power of the transmitter at the base station is satisfied by the condition $\mathrm{E}\left[\left\|\mathbf{q}_i\right\|\right] = 1, \ \forall i \in \{1,2,\cdots,N\}$

From (3-5) and (3-6), we can see that this generalized zero-forcing precoding algorithm needs a scheduling algorithm and the values of p_i, which will be explained later in this chapter. In the next section, we will give the achievable sum data rate in the downlink by using this precoding method.

3.4 Achievable Sum Data Rate

Recall that M is the number of antennas of the array at the base station and K is the number of users. ρ_k^d is the downlink SNR associated with the k^{th} user and ρ_k^u is the uplink SNR of the k^{th} user. Let w_k be the weight of the k^{th} user and γ_k be the probability of selecting the k^{th} user in a scheduling algorithm. The achievable throughput in the downlink transmission is:

$$R_{\Sigma} = \sum_{k=1}^{K} \gamma_k w_k \log_2 \left(1 + \frac{\rho_k^d p_k \, \mathrm{E}^2(\chi)}{1 + \rho_k^d \left(\frac{1}{1+\rho_k^u \tau^u} + p_k \, \mathrm{var}\{\chi\} \right)} \right) \tag{3-9}$$

where χ is the scalar random variable given by

$$\chi = \left(\mathrm{Tr}\left[\left(\hat{\mathbf{H}}_{DS} \hat{\mathbf{H}}_{DS}^{H} \right)^{-1} \right] \right)^{-\frac{1}{2}} \tag{3-10}$$

Given a scheduling algorithm, when the k^{th} user is selected, the probability of selecting the k^{th} user γ_k is 1. At this time, the weighted sum rate in the downlink is:

$$R_{\Sigma} = \sum_{k=1}^{K} w_k \log_2 \left(1 + \frac{\rho_k^d p_k \mathrm{E}^2(\chi)}{1 + \rho_k^d \left(\frac{1}{1+\rho_k^u \tau^u} + p_k \operatorname{var}\{\chi\} \right)} \right) \tag{3-11}$$

Proof:

The signal-vector $\mathbf{x}^d$ received at the selected users:

$$\mathbf{x}^d = \mathbf{E}_S^d \mathbf{H}_S \mathbf{A}_{DS} \mathbf{q} + \mathbf{z}^d \tag{3-12}$$

where $\mathbf{E}_S^d = diag\left\{ \left[\sqrt{\rho_{S_1}^d}, \sqrt{\rho_{S_2}^d}, \cdots, \sqrt{\rho_{S_N}^d} \right]^T \right\}$ and the additive noise $\mathbf{z}^d$ is i.i.d. $\mathbb{C}N(0,1)$.

Denote $\mathbf{G}^d$ as the effective downlink channel:

$$\begin{aligned}
\mathbf{G}^d &= \mathbf{E}_S^d \mathbf{H}_S \mathbf{A}_{DS} \\
&= \mathbf{E}_S^d \left(\hat{\mathbf{H}}_D + \tilde{\mathbf{H}}_S \right) \mathbf{A}_{DS} \\
&= \mathbf{E}_S^d \left(D_S^{-1} \hat{\mathbf{H}}_{DS} + \tilde{\mathbf{H}}_S \right) \mathbf{A}_{DS} \\
&= \mathbf{E}_S^d D_S^{-1} \hat{\mathbf{H}}_{DS} \mathbf{A}_{DS} + \mathbf{E}_S^d \tilde{\mathbf{H}}_S \mathbf{A}_{DS} \\
&= \mathbf{E}_S^d D_S^{-1} \hat{\mathbf{H}}_{DS} \frac{\hat{\mathbf{H}}_{DS}^H (\hat{\mathbf{H}}_{DS} \hat{\mathbf{H}}_{DS}^H)^{-1}}{\sqrt{\mathrm{Tr}\left[\left((\hat{\mathbf{H}}_{DS} \hat{\mathbf{H}}_{DS}^H)^{-1} \right) \right]}} + \mathbf{E}_S^d \tilde{\mathbf{H}}_S \mathbf{A}_{DS} \\
&= \mathbf{E}_S^d D_S^{-1} \hat{\mathbf{H}}_{DS} \hat{\mathbf{H}}_{DS}^H (\hat{\mathbf{H}}_{DS} \hat{\mathbf{H}}_{DS}^H)^{-1} \chi + \mathbf{E}_S^d \tilde{\mathbf{H}}_S \mathbf{A}_{DS} \\
&= \mathbf{E}_S^d D_S^{-1} \chi + \mathbf{E}_S^d \tilde{\mathbf{H}}_S \mathbf{A}_{DS}
\end{aligned} \tag{3-13}$$

The signal received at the k^{th} user is

$$x_k^d = \mathbf{g}_k^T \mathbf{q} + z_k^d \tag{3-14}$$

where $\mathbf{g}_k^T$ is the k^{th} row in matrix $\mathbf{G}$, corresponding to the k^{th} user. From (3-13), $\mathbf{g}_k^T$ is:

$$\mathbf{g}_k^T = \sqrt{\rho_k^d p_k} \chi \mathbf{e}_k^T + \sqrt{\rho_k^d} \tilde{\mathbf{h}}_k^T \mathbf{A}_{DS} \tag{3-15}$$

where $\tilde{\mathbf{h}}_k^T$ is the k^{th} row in matrix $\tilde{\mathbf{H}}_S$. $\mathbf{e}_k$ is the $N \times 1$ column vector, where only the k^{th} element is 1, and all other elements equal to 0. Substitute (3-15) in (3-14), we get

$$\begin{aligned}
x_k^d &= \left(\sqrt{\rho_k^d p_k} \chi \mathbf{e}_k^T + \sqrt{\rho_k^d} \tilde{\mathbf{h}}_k^T \mathbf{A}_{DS} \right) \mathbf{q} + z_k^d \\
&= \sqrt{\rho_k^d p_k} \chi \mathbf{e}_k^T \mathbf{q} + \sqrt{\rho_k^d} \tilde{\mathbf{h}}_k^T \mathbf{A}_{DS} \mathbf{q} + z_k^d \\
&= \sqrt{\rho_k^d p_k} \chi q_k + \sqrt{\rho_k^d} \tilde{\mathbf{h}}_k^T \mathbf{A}_{DS} \mathbf{q} + z_k^d
\end{aligned} \tag{3-16}$$

In (3-16), we add and subtract $\mathrm{E}[\chi]$, to obtain

$$\begin{aligned} x_k^d &= \sqrt{\rho_k^d p_k}\,(\chi + \mathrm{E}[\chi] - \mathrm{E}[\chi])\,q_k + \sqrt{\rho_k^d}\,\tilde{\mathbf{h}}_k^T \mathbf{A}_{DS}\mathbf{q} + z_k^d \\ &= \sqrt{\rho_k^d p_k}\,\mathrm{E}[\chi] q_k + \sqrt{\rho_k^d p_k}\,(\chi - \mathrm{E}[\chi])\,q_k + \sqrt{\rho_k^d}\,\tilde{\mathbf{h}}_k^T \mathbf{A}_{DS}\mathbf{q} + z_k^d \\ &= \sqrt{\rho_k^d p_k}\,\mathrm{E}[\chi] q_k + \tilde{z}_k^d \end{aligned} \tag{3-17}$$

where we denote the effective noise $\tilde{z}_k^d$ as :

$$\tilde{z}_k^d = \sqrt{\rho_k^d p_k}\,(\chi - \mathrm{E}[\chi])\,q_k + \sqrt{\rho_k^d}\,\tilde{\mathbf{h}}_k^T \mathbf{A}_{DS}\mathbf{q} + z_k^d . \tag{3-18}$$

Note that both the channel variation around the mean value and the imperfect knowledge at the base station contributes to the effective noise. z_k^d and $\tilde{\mathbf{h}}_k^T$ are independent of all the other terms, $E\left[z_k^T\right] = 0$ and $E\left[\tilde{\mathbf{h}}_k^T\right] = 0$, so the mean of the effective noise is $\mathrm{E}\left[\tilde{z}_k^d\right] = 0$. In addition, we also note that

$$\mathrm{E}\left[z_k^d \mid \mathbf{q}\right] = 0,\ \mathrm{E}\left[z_k^d \mid \mathbf{q}, \hat{\mathbf{H}}\right] = 0, \text{ and } \mathrm{E}\left[\tilde{\mathbf{h}}_k^T \mid \mathbf{q}, \hat{\mathbf{H}}\right] = 0.$$

$$\mathrm{E}\left[q_k q_k^* (\chi - \mathrm{E}[\chi])\right] = \mathrm{E}\left[q_k q_k^*\right](\mathrm{E}[\chi] - \mathrm{E}[\chi]) = 0$$

$$\mathrm{E}\left[q_k \left(\tilde{\mathbf{h}}_k^T \mathbf{A}_{DS}\mathbf{q}\right)^*\right] = \mathrm{E}\left[q_k \mathbf{q}^H \mathbf{A}_{DS}^H \tilde{\mathbf{h}}_k^*\right] = 0$$

$$\mathrm{E}\left[(\chi - \mathrm{E}[\chi])\, q_k \left(\tilde{\mathbf{h}}_k^T \mathbf{A}_{DS}\mathbf{q}\right)^*\right] = \mathrm{E}\left[(\chi - \mathrm{E}[\chi])\right]\mathrm{E}\left[q_k \left(\tilde{\mathbf{h}}_k^T \mathbf{A}_{DS}\mathbf{q}\right)^*\right] = 0$$

Hence, the variance of the effective noise is:

$$\begin{aligned} &\mathrm{var}\left\{\tilde{z}_k^d\right\} = \mathrm{E}\left[\left(\tilde{z}_k^d - \mathrm{E}\left[\tilde{z}_k^d\right]\right)\left(\tilde{z}_k^d - \mathrm{E}\left[\tilde{z}_k^d\right]\right)^*\right] = \mathrm{E}\left[\tilde{z}_k^d \tilde{z}_k^{d*}\right] \\ &= \mathrm{E}\left[\left(\sqrt{\rho_k^d p_k}\,(\chi - \mathrm{E}[\chi])\,q_k + \sqrt{\rho_k^d}\,\tilde{\mathbf{h}}_k^T \mathbf{A}_{DS}\mathbf{q} + z_k^d\right)\left(\sqrt{\rho_k^d p_k}\,(\chi - \mathrm{E}[\chi])\,q_k + \sqrt{\rho_k^d}\,\tilde{\mathbf{h}}_k^T \mathbf{A}_{DS}\mathbf{q} + z_k^d\right)^*\right] \\ &= \mathrm{E}\left[\rho_k^d p_k (\chi - \mathrm{E}[\chi])\, q_k (\chi - \mathrm{E}[\chi])\, q_k^*\right] + \mathrm{E}\left[\sqrt{\rho_k^d p_k}\,(\chi - \mathrm{E}[\chi])\, q_k \left(\sqrt{\rho_k^d}\,\tilde{\mathbf{h}}_k^T \mathbf{A}_{DS}\mathbf{q}\right)^*\right] \\ &+ \mathrm{E}\left[\sqrt{\rho_k^d p_k}\,(\chi - \mathrm{E}[\chi])\, q_k z_k^{d*}\right] + \mathrm{E}\left[\sqrt{\rho_k^d}\,\tilde{\mathbf{h}}_k^T \mathbf{A}_{DS}\mathbf{q}\left(\sqrt{\rho_k^d p_k}\,(\chi - \mathrm{E}[\chi])\, q_k\right)^*\right] \\ &+ \mathrm{E}\left[\sqrt{\rho_k^d}\,\tilde{\mathbf{h}}_k^T \mathbf{A}_{DS}\mathbf{q}\left(\sqrt{\rho_k^d}\,\tilde{\mathbf{h}}_k^T \mathbf{A}_{DS}\mathbf{q}\right)^*\right] + \mathrm{E}\left[\sqrt{\rho_k^d}\,\tilde{\mathbf{h}}_k^T \mathbf{A}_{DS}\mathbf{q} z_k^{d*}\right] \\ &+ \mathrm{E}\left[z_k^d \left(\sqrt{\rho_k^d p_k}\,(\chi - \mathrm{E}[\chi])\, q_k\right)^*\right] + \mathrm{E}\left[z_k^d \left(\sqrt{\rho_k^d}\,\tilde{\mathbf{h}}_k^T \mathbf{A}_{DS}\mathbf{q}\right)^*\right] + \mathrm{E}\left[z_k^d z_k^{d*}\right] \end{aligned}$$

$$= \mathrm{E}\left[\rho_k^d p_k \left(\chi - \mathrm{E}[\chi]\right) q_k \left(\chi - \mathrm{E}[\chi]\right) q_k^*\right] + \mathrm{E}\left[\sqrt{\rho_k^d}\tilde{\mathbf{h}}_k^T \mathbf{A}_{DS}\mathbf{q}\left(\sqrt{\rho_k^d}\tilde{\mathbf{h}}_k^T \mathbf{A}_{DS}\mathbf{q}\right)^*\right] + \mathrm{E}\left[z_k^d z_k^{d*}\right]$$

$$= \rho_k^d p_k \,\mathrm{var}\{\chi\} + \rho_k^d \left(\frac{1}{1+\rho_k^u \tau^u}\right) + 1$$

Therefore, the throughput of the k^{th} user is

$$R_k = \log_2\left(1 + (SNR)_k\right) = \log_2\left(1 + \frac{\left(\sqrt{\rho_k^d p_k}\,\mathrm{E}[\chi] q_k\right)^2}{\mathrm{var}\{\tilde{z}_k^d\}}\right)$$

$$= \log_2\left(1 + \frac{\rho_k^d p_k \,\mathrm{E}^2(\chi)}{1 + \rho_k^d\left(\frac{1}{1+\rho_k^u \tau^u} + p_k \,\mathrm{var}\{\chi\}\right)}\right)$$

Thus, the weighted sum rate in the downlink is:

$$R_\Sigma = \sum_{k=1}^{K} w_k \log_2\left(1 + \frac{\rho_k^d p_k \,\mathrm{E}^2(\chi)}{1 + \rho_k^d\left(\frac{1}{1+\rho_k^u \tau^u} + p_k \,\mathrm{var}\{\chi\}\right)}\right)$$

Proof ends.

3.5 Optimization of Precoding Matrix

The parameters $p_1, p_2, \cdots, p_k$ are introduced in the generalized zero-forcing precoding algorithm to handle the heterogeneous users, where the users are associated with different weights, downlink SNRs and uplink SNRs. In this section, we obtain the parameters $p_1, p_2, \cdots, p_k$ as a function of the weights, downlink SNRs and uplink SNRs. In order to make the problem mathematically tractable, we assume that $M \gg K$, which is fit for our large number of antennas setting [12].

By the weak law of large numbers, we know that $\lim_{\frac{M}{K}\to\infty} \frac{1}{M}\mathbf{Z}\mathbf{Z}^H = \mathbf{I}_K$, where $\mathbf{Z}$ is the $K \times M$ random matrix whose elements are i.i.d. $\mathbb{C}N(0,1)$ [13]. Thus, $\mathbf{Z}\mathbf{Z}^H$ is approximately $M\mathbf{I}_K$. Hence, the random variable χ in (3-10) can be approximated as:

$$\chi \approx \sqrt{\frac{M}{\sum_{i=1}^{K} a_i p_i}} \tag{3-19}$$

where $a_i = \left(\frac{\rho_i^u \tau^u}{1+\rho_i^u \tau^u} \right)^{-1}$. Substitute (3-19) in (3-11), we obtain :

$$R_\Sigma \approx J(\mathbf{p}) = \sum_{j=1}^{K} w_j \log_2 \left(1 + \frac{b_j p_j}{\sum_{i=1}^{K} a_i p_i} \right) \tag{3-20}$$

where $b_j = \frac{M \rho_j^d}{1+\rho_j^d \left(1+\rho_j^u \tau^u\right)^{-1}}$. We want to find the optimal values for the parameters $p_1, p_2, \cdots, p_k$ to maximize $J(\mathbf{p})$ in (3-20).

An optimal solution to maximize the function $J(\mathbf{p})$ is $\mathbf{p}'$, i.e., $\max J(\mathbf{p}) = J(\mathbf{p}')$. $\mathbf{p}'$ is in form of $c\overline{\mathbf{p}}'$, where c is a positive real number and $\overline{\mathbf{p}}' = \{\overline{p}_1', \overline{p}_2', \cdots, \overline{p}_K'\}$. Assume $c = 1$, then $\mathbf{p}'$ is given by

$$\overline{p}_i' = \max\left\{0, \left(\frac{w_i}{v' a_i} - \frac{1}{b_i} \right)\right\} \tag{3-21}$$

$$\sum_{i=1}^{K} a_i \max\left\{0, \left(\frac{w_i}{v' a_i} - \frac{1}{b_i} \right)\right\} = 1 \tag{3-22}$$

where v' is a positive real number and unique.

Proof: the idea of the proof is to obtain a convex optimization by introducing an additional constraint. It is shown that the introduction of the additional constraint does not affect the optimal value of the problem.

Note that $w_j > 0$, $a_i > 0$, and $b_j > 0$. Let $\mathbf{a} = [a_1, a_2, \cdots, a_K]^T$. We consider the optimization problem:

$$\text{maximize } J(\mathbf{p}) \tag{3-23}$$

$$\text{subject to } \mathbf{p} \succeq 0 \tag{3-24}$$

Since $J(\mathbf{p}) = J(c\mathbf{p})$ for any $c > 0$ and $\mathbf{p}' \neq 0$, $\mathbf{a}^T\mathbf{p}' = c$ is an optimal solution to (3-23), if and only if $\overline{\mathbf{p}}' = (1/c)\mathbf{p}'$ is an optimal solution to the convex optimization problem :

$$\text{minimize} \quad -\sum_{i=1}^{K} w_i \log_2(1 + b_i \overline{p}_i) \tag{3-25}$$

$$\text{subject to} \quad \overline{\mathbf{p}} \succeq 0, \ \mathbf{a}^T\overline{\mathbf{p}} = 1 \tag{3-26}$$

In order to solve (3-25), we introduce Lagrange multiplier $\boldsymbol{\lambda} \in \mathbb{R}^K$ for the inequality constraints $\overline{\mathbf{p}} \succeq 0$ and $v \in \mathbb{R}$ for the inequality constraint $\mathbf{a}^T\overline{\mathbf{p}} = 1$. The necessary and sufficient conditions for optimality are given by Karush-Kuhn-Tucker (KKT) condition [14]. These conditions are

$$\overline{\mathbf{p}}' \succeq 0, \ \mathbf{a}^T\overline{\mathbf{p}}' = 1, \ \boldsymbol{\lambda}' \succeq 0$$

$$\lambda_i'\overline{p}_i' = 0, \ -\frac{w_i b_i}{1 + b_i\overline{p}_i'} - \lambda_i' + v'a_i = 0 \qquad i = 1, 2, \cdots, K$$

This set of equations can be simplified to

$$\overline{p}_i' = \max\left\{0, \left(\frac{w_i}{v'a_i} - \frac{1}{b_i}\right)\right\}$$

$$\sum_{i=1}^{K} a_i \max\left\{0, \left(\frac{w_i}{v'a_i} - \frac{1}{b_i}\right)\right\} = 1$$

Since $\left(\frac{w_i}{v'a_i} - \frac{1}{b_i}\right)$ is an increasing function about $\frac{1}{v'}$, the solution v' is unique, which can be solved by numerical methods.

The proof completes.

3.6 Scheduling Strategy

The scheduling strategy is to find the optimal scheduling of the users based on the scaled estimated channel gains [15].

3.6.1 Homogeneous users

In the homogeneous setting, all the downlink SNRs (ρ^d) from the base station to the users

are equal and all the uplink SNRs (ρ^u) from the users to the base station are equal too. In addition, the weights assigned to the users are the same, which we assume them as unity, i.e., $w_k = 1 \ \ (k = 1,2,\cdots,K)$. Given the perfect knowledge at the base station $\left(\hat{\mathbf{H}} = \mathbf{H}\right)$ and without scheduling $(N = K)$, the zero-forcing precoding diagonalizes the effective downlink channel and all users obtain the same effective channel gain.

We use a simple scheduling strategy here at the base station. In each coherence interval, the base station selects N users with largest estimated channel gains. Let $\hat{\mathbf{h}}_{(1)}^T, \hat{\mathbf{h}}_{(2)}^T, \cdots, \hat{\mathbf{h}}_{(K)}^T$ be the norm-ordered rows (from large to small) of the estimated channel $\hat{\mathbf{H}}$. After scheduling, the channel $\hat{\mathbf{H}}_S$ is $\hat{\mathbf{H}}_S = \left[\hat{\mathbf{h}}_{(1)}^T, \hat{\mathbf{h}}_{(2)}^T, \cdots, \hat{\mathbf{h}}_{(N)}^T\right]^T$. Thus, the achievable sum rate under this scheduling is

$$R_\Sigma = N \log_2 \left(1 + \frac{\rho^d \dfrac{\rho^u \tau^u}{1+\rho^u \tau^u} \mathrm{E}^2(\eta)}{1 + \rho^d \left(\dfrac{1}{1+\rho^u \tau^u} + \dfrac{\rho^u \tau^u}{1+\rho^u \tau^u} \mathrm{var}\{\eta\} \right)} \right) \tag{3-27}$$

Here, the random variable $\eta = \left(\mathrm{Tr}\left[\left(\mathbf{U}\mathbf{U}^H \right)^{-1} \right] \right)^{-\frac{1}{2}}$, where $\mathbf{U}$ is the $N \times M$ matrix formed by the N rows with largest norms of a $N \times M$ random matrix $\mathbf{Z}$ whose elements are i.i.d. $\mathbb{C}N(0,1)$.

With the change of the values of N, the sum rate is different. Here, we denote N_{opt} for N that maximize (3-27), i.e.,

$$R_{\Sigma(N_{opt})} = \max_N \{ N \log_2 \left(1 + \frac{\rho^d \dfrac{\rho^u \tau^u}{1+\rho^u \tau^u} \mathrm{E}^2(\eta)}{1 + \rho^d \left(\dfrac{1}{1+\rho^u \tau^u} + \dfrac{\rho^u \tau^u}{1+\rho^u \tau^u} \mathrm{var}\{\eta\} \right)} \right) \} \tag{3-28}$$

Net achievable sum rate accounts for the achievable sum rate in the data transmission phase, excluding those parts in the training and computation phase. We assume that in every coherence interval of T symbols. The first τ^u symbols are used for uplink training, one symbol is used for computation and the rest $\left(T - \tau^u - 1\right)$ symbols are used for data transmission. Moreover, the training length τ^u can be optimized to maximize the net throughput of the system. Therefore, the net achievable sum rate is:

$$R_{net} = \max_{\tau^u, N} \frac{T - \tau^u - 1}{T} R_{\Sigma} \tag{3-29}$$

subject to $K \le \tau^u \le T - 1$, $N \le K$.

3.6.2 Heterogeneous Users

Let $\mathbf{z}_1^T, \mathbf{z}_2^T, \cdots, \mathbf{z}_K^T$ be the rows of the matrix

$$\mathbf{Z} = diag\left\{ \left[\frac{\sqrt{1+\rho_1^u \tau^u}}{\rho_1^u \tau^u}, \frac{\sqrt{1+\rho_2^u \tau^u}}{\rho_2^u \tau^u}, \cdots, \frac{\sqrt{1+\rho_K^u \tau^u}}{\rho_K^u \tau^u} \right]^T \right\} \hat{\mathbf{H}} \tag{3-30}$$

where $\hat{\mathbf{H}}$ is the LMSE estimated channel given by (3-4). Note that $\mathbf{Z}$ is normalized such that elements are independent and identically distributed. In our heterogeneous users setting, the scheduling strategy is as follows. In each coherence interval, the users are ordered such that $\overline{p}'_{(1)} \left\| \mathbf{z}_{(1)}^T \right\|^2 \ge \overline{p}'_{(2)} \left\| \mathbf{z}_{(2)}^T \right\|^2 \ge \cdots \ge \overline{p}'_{(K)} \left\| \mathbf{z}_{(K)}^T \right\|^2$ and the first N users under this ordering are selected. Here, we denote N_{opt} for N that maximize (3-29), i.e.

$$R_{\Sigma(N_{opt})} = \max_{N} \{ \sum_{k=1}^{N} w_k \log_2 \left(1 + \frac{\rho_k^d p_k \, \mathrm{E}^2(\upsilon)}{1 + \rho_k^d \left(\frac{1}{1+\rho_k^u \tau^u} + p_k \frac{\rho_k^u \tau^u}{1+\rho_k^u \tau^u} \mathrm{var}\{\upsilon\} \right)} \right) \} \tag{3-31}$$

where $\upsilon = \left(\mathrm{Tr}\left[\left(\mathbf{Z}\mathbf{Z}^H \right)^{-1} \right] \right)^{-\frac{1}{2}}$.

Similarly, we define the net achievable weighted-sum rate as

$$R_{net} = \max_{\tau^u, N} \frac{T - \tau^u - 1}{T} R_{\Sigma} \tag{3-32}$$

subject to $K \le \tau^u \le T - 1$, $N \le K$.

3.7 Optimal Training Length

To maximize the net achievable sum rate in (3-29) and (3-32), given the values of M, K, T, ρ^d and ρ^u, it is still difficult to obtain a closed-form expression for the optimal

training length. Therefore, we look at the limiting case $\rho^u \to 0$ and $\rho^u \to \infty$ to learn the behavior of the optimal training length with uplink SNR.

In the limit $\rho^u \to 0$, we can approximate the net rate (3-29) as

$$R_{net} \approx \frac{T-\tau^u-1}{T} N \log_2\left(1+\frac{\rho^d \rho^u \tau^u}{1+\rho^d} \mathrm{E}^2(\eta)\right) \tag{3-33}$$

We use the fact that $\log(1+x) \approx x$ as $x \to 0$ to obtain:

$$R_{net} \approx c_1 \frac{T-\tau^u-1}{T} \tau^u \tag{3-34}$$

where c_1 is a positive constant. It is clear that (3-34) is maximized when $\tau^u = \frac{T-1}{2}$ if we assume T is odd and $T > 2K$.

In the limit $\rho^u \to \infty$, we can approximate the net rate (3-29) as

$$R_{net} \approx c_2 \frac{T-\tau^u-1}{T} \tag{3-35}$$

where c_2 is a positive constant. It is clear that (3-35) is maximized when $\tau^u = K$.

When the uplink SNR is very low, nearly half of the coherence time should be spent for training. However, when the uplink SNR is very high, it costs minimum possible number of symbols (which is K symbols here) for training.

3.8 Summary

In this chapter, we first presented the channel model and uplink training based on the LMSE estimation of the channel. Then, a precoding method, the generalized zero-forcing precoding, was proposed based on this training. It consists of a scheduling component and an optimization component. A simple but practical scheduling strategy is used here and the values of the optimization are obtained by using a convex optimization problem from a relevant asymptotic of large number of the base station antennas. Finally, the optimal training length was analysed.

4 SVH PRECODING METHOD AND TRAINING ON BOTH UPLINK AND DOWNLINK

4.1 Introduction

SVH precoding method is proposed in [16] and named by the first letters of the three authors (Stojnic Vikalo and Hassibi), which is designed to find the precoding matrix for optimizing the sum rate in the downlink.

In the previous chapter, the users have no knowledge about the instantaneous channel. However, users can generally obtain the knowledge about their effective channel gain in the following two ways: (1) the base station sends the information of the effective channel gains to the users. (2) The base station sends forward pilots to the users so that the users can estimate the effective gains. Because of the large amount of the overhead when the base station sending the channel information to the users, we will choose the training method that sends pilots in the downlink, which is regarded as the downlink training. Therefore, it results in a transmission method consisting of four phases: uplink training phase, computation phase, downlink training phase and data transmission, as shown in Figure 4.1.

Uplink training	Computation	Downlink training	Data transmission

Figure 4-1 Four-phase transmission scheme over one coherence interval

In this chapter, we first study the SVH precoding method and its modified forms. Then we analyse the upper bound sum rate by using the precoding. Finally, we consider the transmission method that contains both uplink and downlink training.

4.2 SVH Precoding Method

As explained in chapter 3, the base station obtains the LMSE estimate $\hat{\mathbf{H}}$ of the channel by using the uplink training. Then from the estimate $\hat{\mathbf{H}}$, the base station forms a precoding matrix. Let $\mathbf{A}$ denote any precoding matrix which is a function of the channel estimate $\hat{\mathbf{H}}$, i.e., $\mathbf{A} = f\left(\hat{\mathbf{H}}\right)$. The precoding function $f(\cdot)$ usually depends on the system parameters

such as downlink SNR, uplink SNR and weights assigned to the users. The precoding matrix is normalized here, i.e., $\mathrm{Tr}(\mathbf{A}^H\mathbf{A})=1$.

Let $\mathbf{h}_i$ be the i^{th} row of the channel matrix $\mathbf{H}$ and $\mathbf{a}_j$ be the j^{th} column of the precoding matrix $\mathbf{A}$. The sum rate of the downlink channel can be written as:

$$R(\mathbf{H},\mathbf{A})=\sum_{i=1}^{K}\log_2\left(1+\frac{\left|\mathbf{h}_i\mathbf{a}_i\right|^2}{\sigma^2\,\mathrm{Tr}(\mathbf{A}\mathbf{A}^H)+\sum_{i\neq j}\left|\mathbf{h}_i\mathbf{a}_j\right|^2}\right) \tag{4-1}$$

where σ^2 is the noise variance.

Let $\mathbf{b}_i=\left|\mathbf{h}_i\mathbf{a}_i\right|^2$ and $\mathbf{c}_i=\sigma^2\,\mathrm{Tr}(\mathbf{A}^H\mathbf{A})+\sum_{i\neq j}\left|\mathbf{h}_i\mathbf{a}_j\right|^2$. Denote matrices $\mathbf{\Delta}$ and $\mathbf{\Omega}$ as:

$$\mathbf{\Delta}=diag\left\{\left[\frac{(\mathbf{HA})_{11}}{\mathbf{c}_1},\frac{(\mathbf{HA})_{22}}{\mathbf{c}_1},\cdots,\frac{(\mathbf{HA})_{KK}}{\mathbf{c}_1}\right]\right\} \tag{4-2}$$

$$\mathbf{\Omega}=diag\left\{\left[\frac{\mathbf{b}_1}{\mathbf{c}_1(\mathbf{b}_1+\mathbf{c}_1)},\frac{\mathbf{b}_2}{\mathbf{c}_2(\mathbf{b}_2+\mathbf{c}_2)},\cdots,\frac{\mathbf{b}_K}{\mathbf{c}_K(\mathbf{b}_K+\mathbf{c}_K)}\right]^T\right\} \tag{4-3}$$

To maximize the sum rate $R(\mathbf{H},\mathbf{A})$, it is to solve the equations $\frac{\partial R(\mathbf{H},\mathbf{A})}{\partial \mathbf{A}_{ij}}=0$.

$$\mathbf{A}=\left(\left(\sigma^2 Tr(\mathbf{\Omega})\right)\mathbf{I}_M+\mathbf{H}^H\mathbf{D}\mathbf{H}\right)^{-1}\mathbf{H}^H\mathbf{\Delta}. \tag{4-4}$$

We use the following iterative steps to determine an efficient $\mathbf{A}$:

(1) Assigning the initial values to matrices $\mathbf{\Delta}$ and $\mathbf{\Omega}$, for example $\mathbf{\Delta}=\mathbf{I}_K$, $\mathbf{\Omega}=\mathbf{I}_K$.

(2) Calculate $\mathbf{A}$ according to (4-4).

(3) Calculate $\mathbf{\Delta}$ and $\mathbf{\Omega}$ according to (4-2) and (4-3).

(4) Repeat step 2 and step 3 several times.

4.3 Modified SVH Precoding Method

SVH precoding method requires the base station having the perfect knowledge of the channel, which is often impossible in practice. However, the approach can be extended for the case that only the estimation of the channel $\hat{\mathbf{H}}$ and the statistics of the estimation error $\tilde{\mathbf{H}}$ are

available. Here, we can use the LMSE estimation of the channel discussed in the previous chapter. To maximize the average sum rate, we get:

$$R(\mathbf{H},\mathbf{A}) = \mathrm{E}\left[R(\hat{\mathbf{H}} + \tilde{\mathbf{H}}, \mathbf{A}) \right] \tag{4-5}$$

Assume there are L samples of $\hat{\mathbf{H}}$, which are $\hat{\mathbf{H}}^{(1)}, \hat{\mathbf{H}}^{(2)}, \cdots, \hat{\mathbf{H}}^{(L)}$. Define $\mathbf{H}^{(i)} = \hat{\mathbf{H}} + \tilde{\mathbf{H}}^{(i)}$. The average sum rate will be

$$R(\hat{\mathbf{H}},\mathbf{A}) = \frac{1}{L}\sum_{k=1}^{L}\sum_{i=1}^{K}\log_2\left(1 + \frac{\left|\mathbf{h}_i^{(k)}\mathbf{a}_i\right|^2}{\sigma^2\,\mathrm{Tr}(\mathbf{A}\mathbf{A}^H) + \sum_{i\neq j}\left|\mathbf{h}_i^{(k)}\mathbf{a}_j\right|^2}\right). \tag{4-6}$$

We denote $\mathbf{\Delta}^{(i)}$ and $\mathbf{\Omega}^{(i)}$ as in (4-2) and (4-3) using $\mathbf{H}^{(i)}$ instead of $\mathbf{H}$. We solve the equations $\dfrac{\partial R\left(\hat{\mathbf{H}},\mathbf{A}\right)}{\partial \mathbf{A}_{ij}} = 0$ to obtain

$$\sum_{i=1}^{L}\left[\mathbf{H}^{(i)}\mathbf{\Delta}^{(i)} - \mathbf{H}^{(i)H}\mathbf{\Omega}^{(i)}\mathbf{H}^{(i)}\mathbf{A} - \sigma^2\,\mathrm{Tr}(\mathbf{\Omega}^{(i)})\mathbf{A}\right] = 0 \tag{4-7}$$

Let

$$\mathbf{W} = \sum_{i=1}^{L}\mathbf{H}^{(i)H}\mathbf{\Omega}^{(i)}\mathbf{H}^{(i)} - \sigma^2\,\mathrm{Tr}(\mathbf{\Omega}^{(i)})\mathbf{I}_M \quad \text{and} \quad \mathbf{S} = \sum_{i=1}^{L}\mathbf{H}^{(i)}\mathbf{\Delta}^{(i)}$$

We have

$$\mathbf{A} = \mathbf{W}^{-1}\mathbf{S} \tag{4-8}$$

We use the following iterative steps to determine $\mathbf{A}$:

(1) Assigning the initial values to matrices $\mathbf{\Delta}^{(i)}$ and $\mathbf{\Omega}^{(i)}$, for example $\mathbf{\Delta}^{(i)} = \mathbf{I}_K$, $\mathbf{\Omega}^{(i)} = \mathbf{I}_K$.

(2) Calculate $\mathbf{A}$ according to (4-8).

(3) Calculate $\mathbf{\Delta}^{(i)}$ and $\mathbf{\Omega}^{(i)}$ according to (4-2) and (4-3) using $\mathbf{H}^{(i)}$ instead of $\mathbf{H}$.

(4) Repeat step 2 and step 3 several times.

Similarly, when the statistics of the estimation error $\tilde{\mathbf{H}}$ is not available or it is time-consuming to get the samples $\tilde{\mathbf{H}}^{(i)}$, we can use $\hat{\mathbf{H}}$ to replace $\mathbf{H}$. After several iterative steps, we obtain a better precoding matrix $\mathbf{A}$.

(1) Assigning the initial values to matrices $\mathbf{\Delta}$ and $\mathbf{\Omega}$, for example $\mathbf{\Delta} = \mathbf{I}_K$, $\mathbf{\Omega} = \mathbf{I}_K$.

(2) Calculate $\mathbf{A}$ according to (4-4).

(3) Calculate $\mathbf{\Delta}$ and $\mathbf{\Omega}$ according to (4-2) and (4-3) , using $\hat{\mathbf{H}}$ instead of $\mathbf{H}$.

(4) Repeat step 2 and step 3 several times.

4.4 Downlink Training

The users can obtain the estimation of their channel gain by the downlink training, where the base station sends τ^d pilots to the users. Considered the short coherence interval in the TDD systems, the downlink training pilots should be much less than that in the uplink training. In the special case, the downlink training is reduced to only one pilot, which is fairly attractive for the TDD systems. Furthermore, orthogonal pilots are not restricted in the downlink training [17], [18]. For example, the downlink training vector is $\mathbf{q}^p = [1,1,\cdots 1]^T$. The signal received by the users is

$$\mathbf{x}^p = \mathbf{E}^d \mathbf{H}\mathbf{A}\mathbf{q}^p + \mathbf{z}^d \tag{4-9}$$

where the downlink gain $\mathbf{E}^d = diag\left\{\left[\sqrt{\rho_1^d}, \sqrt{\rho_2^d}, \cdots, \sqrt{\rho_K^d}\right]^T\right\}$ and $\mathbf{z}^d$ is the downlink additive white noise. We denote the effective downlink channel as $\mathbf{G} = \mathbf{E}^d \mathbf{H}\mathbf{A}$. We will get

$$x_i^p = \sqrt{P_i} g_{ii} + \sqrt{P_i} \sum_{i \neq j} g_{ij} + z_i \qquad i = 1,2,\cdots,K \tag{4-10}$$

where P_i is the power of the i^{th} element of the pilot signal-vector $\mathbf{q}^p$. If the precoding matrix $\mathbf{A}$ matches the channel $\mathbf{H}$ quite well, i.e., $\mathbf{H}\mathbf{A} \approx \mathbf{I}_K$, the expectation of the received signal will equal to the scaled downlink gains, i.e., $\mathrm{E}\left[x_i^p\right] = \sqrt{P_i}\sqrt{\rho_i^d}$.

4.5 Achievable Throughput

4.5.1 Lower Bound of the Sum Rate

We use similar techniques in the previous section to obtain the net achievable throughput for the four-phase transmission method with uplink and downlink training. From (4-9) and (4-10), a lower bound in the downlink weighted sum capacity during the transmission is given by

$$R_{\Sigma} = \sum_{k=1}^{K} w_k \, \mathrm{E}\left[\log_2 \left(1 + \frac{\left| \mathrm{E}[g_{kk}] \right|^2}{1 + \sum_{i \neq k} \mathrm{E}\left[\left| g_{ki} \right|^2 \right] + \mathrm{var}\{g_{kk}\}} \right) \right] \tag{4-11}$$

The net achievable weighted sum rate is

$$R_{net} = \max_{\tau^u} \frac{T - \tau^u - \tau^d - 1}{T} R_{\Sigma} \tag{4-12}$$

4.5.2 Upper Bound of the Sum Rate

We assume that the estimate $\hat{\mathbf{H}}$, the statistics of $\hat{\mathbf{H}}$ and $\tilde{\mathbf{H}}$, and the downlink ρ_k^d are available at the base station. The precoding matrix $\mathbf{A}$ is obtained from this information. The signal-vector received by users is

$$\mathbf{x} = \mathbf{E}^d \mathbf{H} \mathbf{A} \mathbf{q} + \mathbf{z} \tag{4-13}$$

Denote C as the sum capacity of the downlink channel. The upper bound of C is given by

$$C \leq \sum_{i=1}^{K} \log_2 \left(1 + \frac{\rho_i^d \left| \mathbf{h}_i \mathbf{a}_i \right|^2}{1 + \sum_{i \neq j} \rho_j^d \left| \mathbf{h}_i \mathbf{a}_j \right|^2} \right) \tag{4-14}$$

In practice, $\mathbf{H}$ is usually not available at the base station. We can use the approximate approach in (4-5), $R(\mathbf{H}, \mathbf{A}) = \mathrm{E}\left[R(\hat{\mathbf{H}} + \tilde{\mathbf{H}}, \mathbf{A}) \right]$, to obtain the local maxim. We can find a number of such local maxims, and then let the upper bound C be the maximum of these local maximums.

4.6 Summary

In this chapter, we proposed another precoding technique, SVH precoding, which takes advantage of the channel estimation to obtain a good precoding matrix for maximizing the sum rate in the downlink. Then the training on the downlink is introduced, which will help the users to estimate the effective channel gains. Finally, we analysed the achievable throughput performance of the precoding method and training.

5 SIMULATION

5.1 Introduction

The system investigated in this thesis consists of a base station with a large array of M antennas and K single antenna users. In [11], the authors set the number of antennas at the base station $M = 16$ in their simulations. However, we will hypothesize $M = 100$ in our simulations in this thesis, due to the large number of antennas prediction in next generation mobile communication systems. In the last simulation of the chapter, we will compare the different performance of the system, when M grows from 16 to 100. To simplify the simulation, we use linear wire antennas here, and the distance between antennas is 2λ, for making sure that there is no correlation. For TDD systems, the coherence time interval is often short. Assume that there is a carrier frequency of 1900 MHz and the maximum velocity of the mobile users is 240 km/h. Then the coherence time interval is about $400\mu s$. If the coherence bandwidth is $50-200$ kHz, the effective symbol rate will approximately be $5-20\mu s$. Thus, we will have $20-80$ symbols in one coherence time interval. In the simulations, we assume the coherence time $T = 30$ symbols. It implies that the users are moving in a relatively high speed. Let $K = 15,$ which means at least 15 symbols will be used for the training. Furthermore, the reciprocity principle is used here, i.e., the uplink channel between any user and the base station is a scaled version of the downlink channel. To simplify the system, we also assume that OFDM is employed to reduce the multi-path and remove frequency selective fading. The duration of the coherence interval is set in one of the parallel OFDM sub-bands. Thus, only flat fading is considered in the simulations, which is assumed as Rayleigh block fading. In Rayleigh fading, the entries of the channel matrix $\mathbf{H}$ are independent and i.i.d $\mathbb{C}N(0,1)$.

In this chapter, all the simulations are done in Matlab. First, we simulate and analyse the performance of the generalized zero-forcing precoding methods, which are presented in Chapter 3. Then, the SVH precoding methods presented in Chapter 4 will be simulated and compared. Here, the achievable throughput is the main performance that we are concentrated on.

5.2 Simulation on the Generalized Zero-Forcing Precoding methods

5.2.1 Homogeneous users case

A) Achievable throughput without scheduling

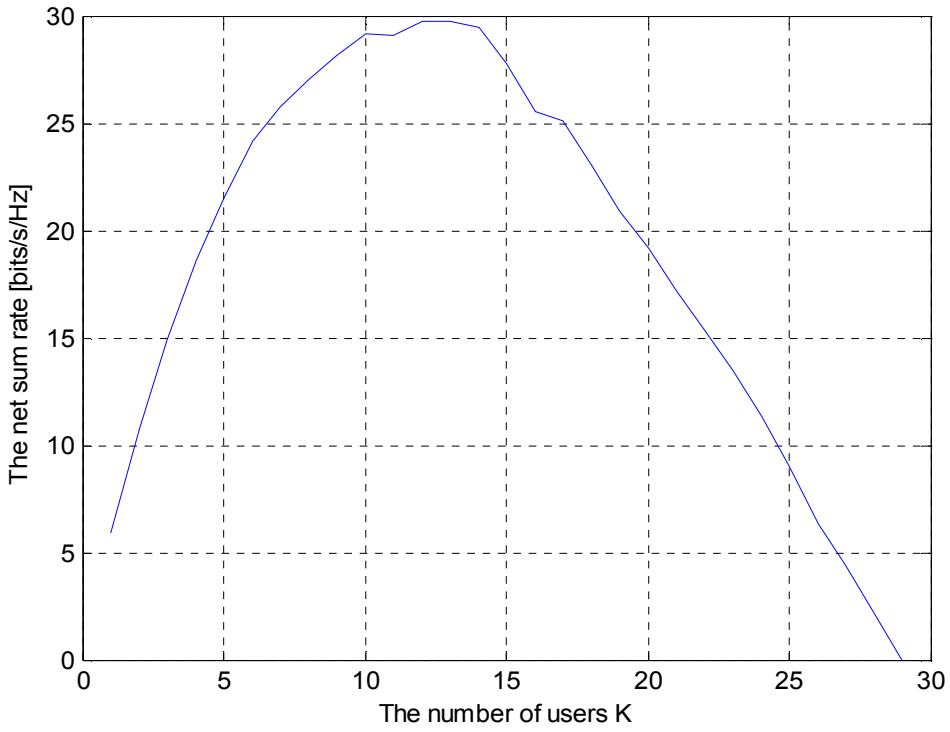

Figure 5-1 The net sum rate performance after using the generalized ZF precoding without scheduling.

In this simulation, the downlink and uplink SNR are 10dB and 0dB, respectively. The variable is the number of the users K, where $1 \le K \le T-1$. Let T be the duration of the coherence time interval, $T = 30$. The achievable throughput is calculated using (3-27) and (3-29).

As shown in Figure 5-1, we observe that the net sum rate firstly grows up with increasing the number of users K, but after the point $K = 13$, the net sum rate falls down as K increases. This can be explained from (3-29). When K is not large, it is obvious that the rates summed up by the parallel channels will increase as the number of users increase. However, we should remember that the time used for data transmission is only part of the whole coherence time interval and the other part of the interval is consumed for training and computation, which is shown clearly in (3-29). Thus, when the number of users K grows

beyond some point, say $K = 13$ in our simulation, the time used for data transmission falls down, which causes the decreasing of the net sum rate.

The highest net sum rate in the downlink in this simulation is about 30 bit/s/Hz. If the frequency bandwidth is 10 MHz, then the data rate will be about 300 Mbps, three times higher than 100 MHz, which is the data rate of mobile device moving in a relative high speed in 4G. Furthermore, if we use some higher order modulation techniques, such data rate is expected to be even higher.

B) Achievable throughput with scheduling

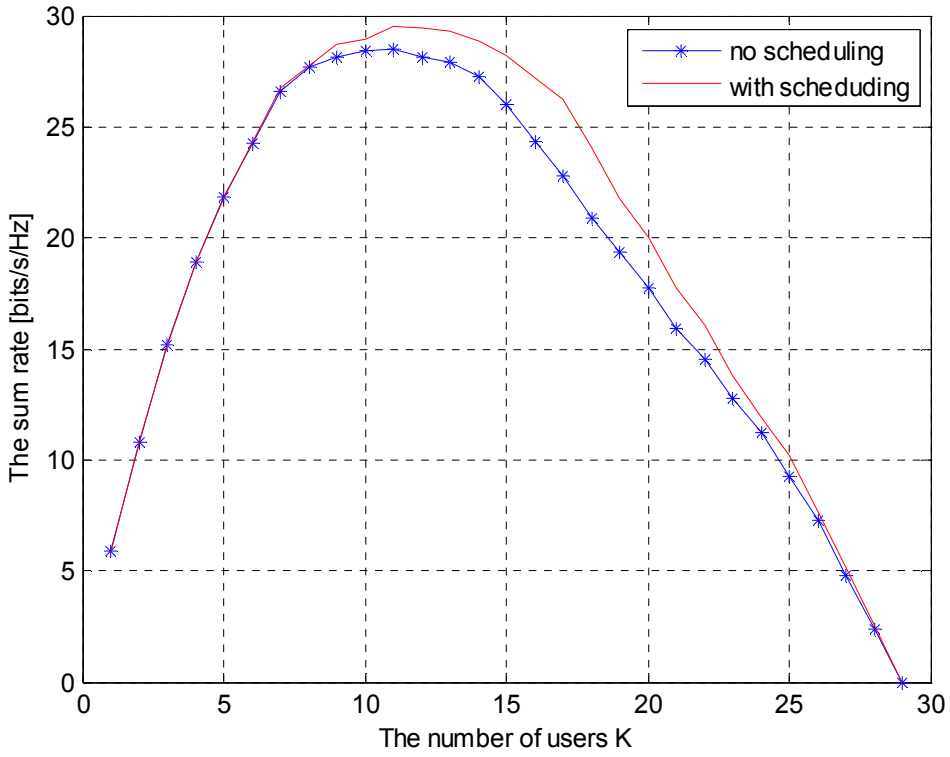

Figure 5-2 The comparison of the net sum rate performance after the generalized ZF precoding with and without scheduling.

From Figure 5-2, we can see that the net sum rate performs better after using the scheduling than that with no scheduling, especially when K is over the optimal values. This is because that the scheduling algorithm can help the base station to make some balance between the increase of the users and the decrease of the data transmission time slot in the whole coherence time interval. However, as K continues to grow, the scheduling algorithm fails as well.

C) The net sum rate with the change of downlink SNR

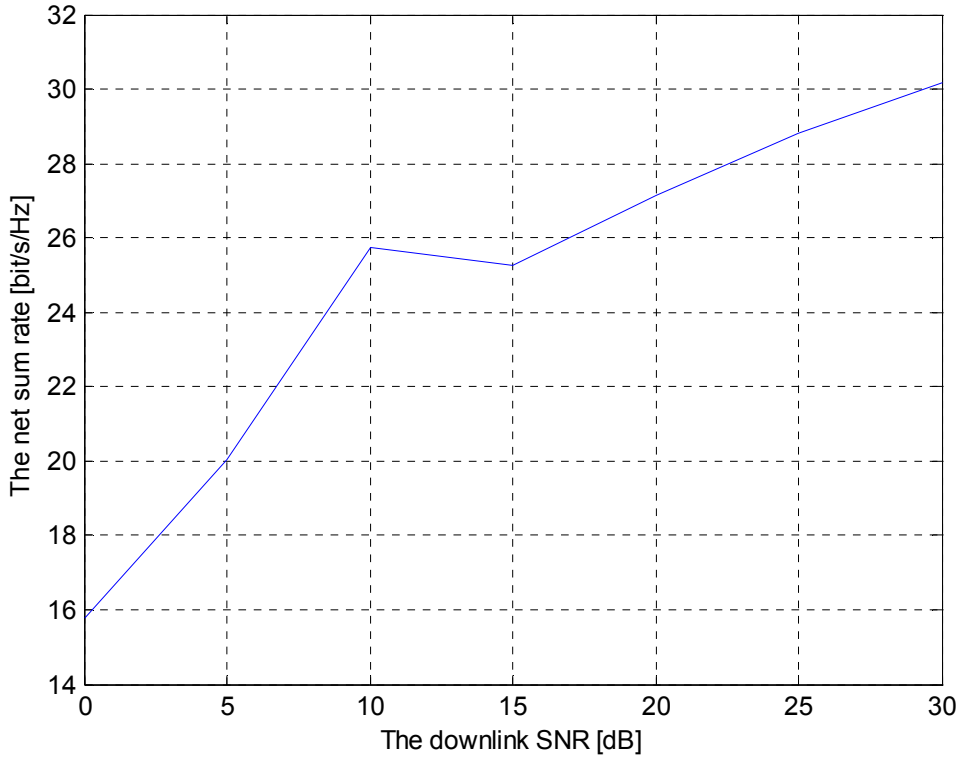

Figure 5-3 The net sum rate curve with the change of the downlink SNR

From Figure 5-3, we note that the net sum rate increases as we raise the downlink SNR. However, when the downlink SNR is higher, the increase of the throughput that brought by the downlink SNR is not as faster as that in the lower downlink SNR case.

5.2.2 Heterogeneous users case

A) Achievable throughput without scheduling

To simplify the system, we assume the downlink SNR and the weight values:

$\boldsymbol{\rho}^d = [0\ 0\ 0\ 5\ 5\ 5\ 5\ 5\ 5\ 10\ 10\ 10\ 12\ 12\ 12\ 0\ 0\ 0\ 5\ 5\ 5\ 5\ 5\ 5\ 10\ 10\ 10\ 12\ 12]^T$

$\mathbf{w} = [1\ 1\ 1\ 2\ 2\ 2\ 2\ 3\ 3\ 3\ 3\ 5\ 5\ 5\ 1\ 1\ 1\ 2\ 2\ 2\ 2\ 3\ 3\ 3\ 3\ 5\ 5]^T$. We use (3-31) and (3-32) to calculate the net sum rate.

As shown in Figure 5-4, the net sum rate performance in the heterogeneous users case has similar shape as that in homogeneous users setting, where the curve first increases to some peak point then falls down with increasing the number of users. But the peak net sum rate here is about 100 bit/s/Hz. Given 10 MHz bandwidth, the peak data rate will be 1 Gbps.

However, we cannot say the heterogeneous setting is better than the homogeneous setting, because the downlink SNRs and the weight value are different.

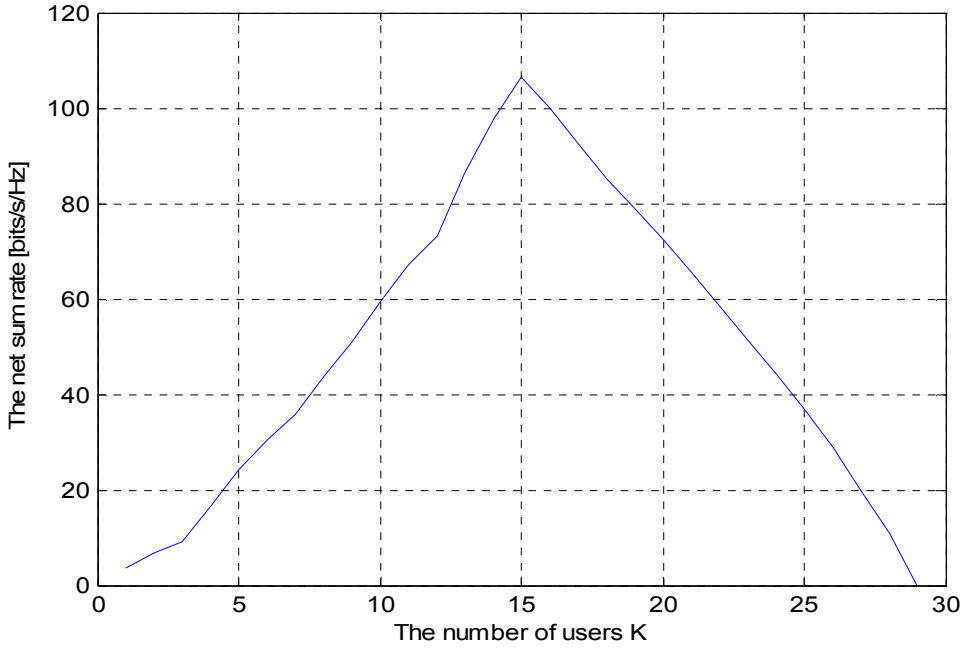

Figure 5-4 The net sum rate performance of the heterogeneous setting after using generalized ZF precoding without scheduling.

B) Achievable throughput with scheduling

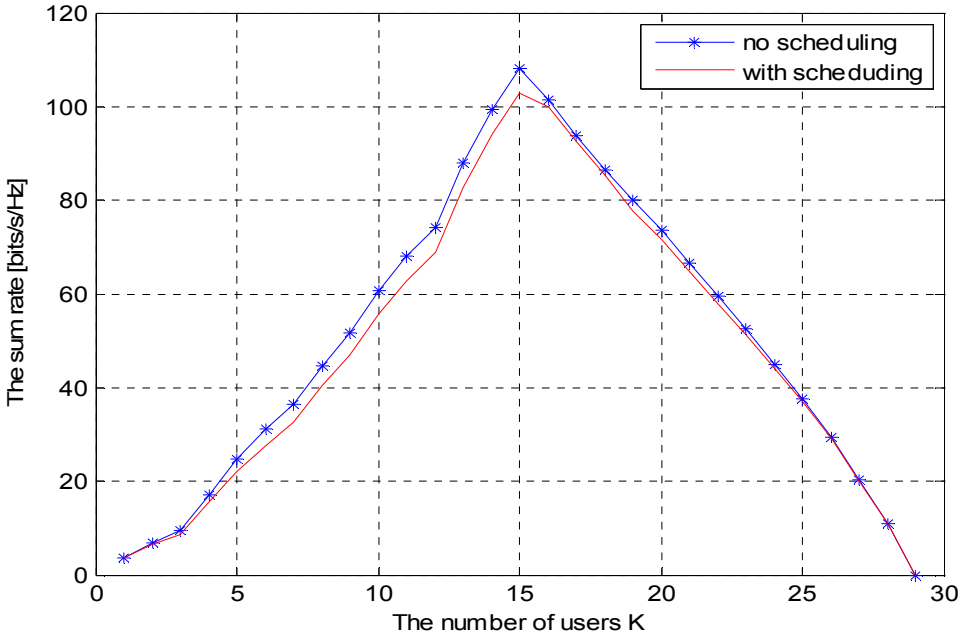

Figure 5-5 The net sum rate performance of the heterogeneous setting after using the generalized ZF precoding with scheduling.

In Figure 5-5, we see that the situation is different from that in the homogeneous case, for the throughput performance after scheduling becomes slightly worse than that without scheduling. This is because different users with different downlink SNRs and weight values seem to have been scheduled in a certain way already. If we continue using our scheduling strategy, it will cost some more time to do the calculation, which may cancel the benefits we get from scheduling. However, the difference with and without scheduling is small.

5.3 Simulation on the SVH precoding methods

In the previous section, we mainly consider the generalized ZF precoding methods, while in this section, we will focus on the other precoding methods -- SVH precoding, which are presented in Chapter 4. Because the homogeneous and heterogeneous users settings are relatively similar, we only assume the homogeneous users setting in this part of the simulations, which will simplify our system and ease the comparisons.

5.3.1 Simulation on the SVH precoding with perfect knowledge of the channel H

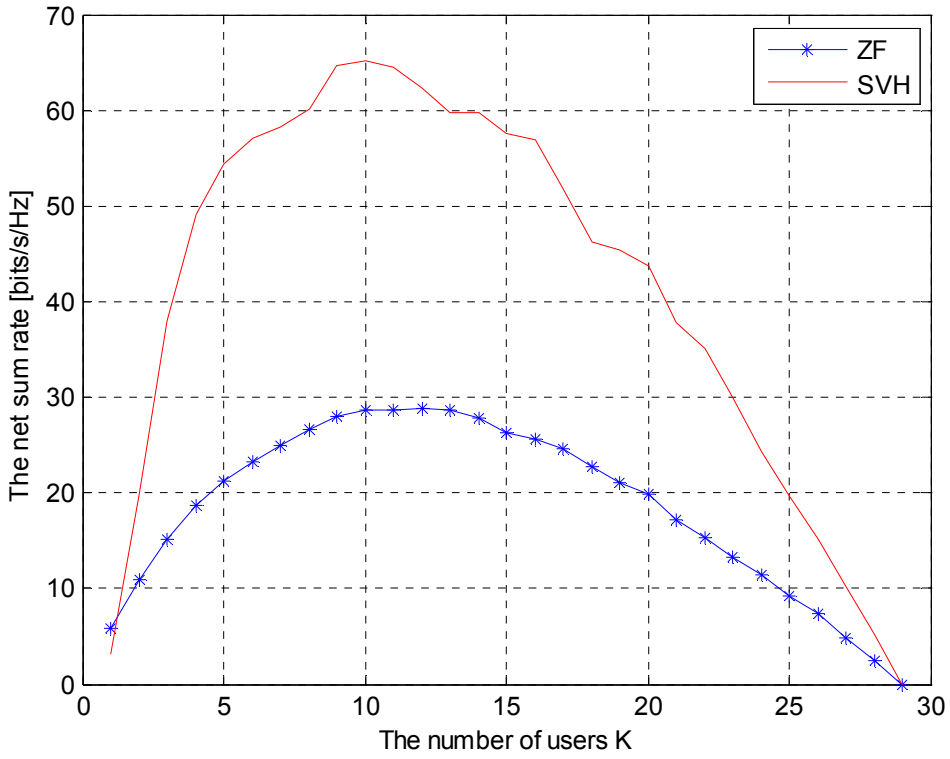

Figure 5-6 The comparison of the net sum rate performance after the generalized ZF precoding and after the SVH precoding.

In the simulations, let the downlink SNR and uplink SNR be 10dB and 0dB, respectively. The other parameters remain the same as before. The formulation for calculating the net sum rate is from equation (4-4).

In Figure 5-6, we can clearly see that the net sum rate performance of the system has been greatly enhanced by using SVH precoding as compared with the generalized ZF precoding. The result bornes out the fact that the SVH precoding is designed to maximize the net sum rate performance. After the optimization, the peak net sum rate can reach as high as 65 bit/s/Hz, which is about twice higher than that obtained from the generalized ZF precoding. Given a bandwidth of 15 MHz, the corresponding peak throughput is approximately 1 Gbps. If we increase the downlink SNR or the number of antennas, the net sum rate performance will grow over 1 Gbps as well.

5.3.2 Simulation on the modified SVH precoding based on the estimation of the channel $\hat{\mathbf{H}}$ and the statistics of the error $\tilde{\mathbf{H}}$

It is often the case that the base station has no perfect knowledge of the whole transmission channel. Thus, the SVH precoding method cannot work in the same way that presented in the previous section. However, if we have estimation of the channel $\hat{\mathbf{H}}$ and the statistics of the error $\tilde{\mathbf{H}}$, the SVH precoding can be modified based on these information. In the simulations, $\hat{\mathbf{H}}$ and $\tilde{\mathbf{H}}$ remain the same as the previous sections, and so do the other parameters.

As shown in Figure 5-7, there are three curves that represent the net sum rate performance, where the performance after the modified SVH precoding is in the middle of those after the generalized ZF precoding and the SVH precoding. The peak net sum rate after the modified SVH precoding is about 42 bit/s/Hz, which is approximately 20 bit/s/Hz less than that in SVH precoding. This can be regarded as the loss in terms of throughput, which is due to the lack of perfect knowledge of the channel state information $\mathbf{H}$. However, because of the optimization provided by the modified SVH precoding, the peak net sum rate increases about 10 bit/s/Hz than that after the generalized ZF precoding.

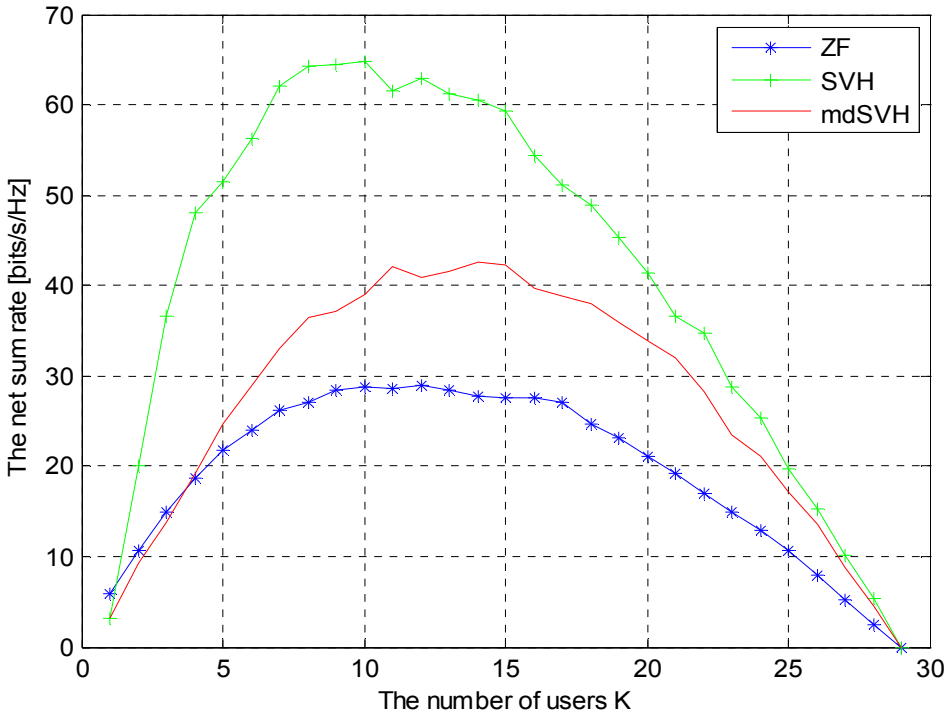

Figure 5-7 The comparison of the net sum rate performance after the generalized ZF precoding, after the SVH precoding and after the modified SVH precoding.

5.3.3 Simulation on the modified SVH2 precoding based only on the estimation of the channel $\hat{\mathbf{H}}$

In this section, we consider the case that only the estimation of the channel $\hat{\mathbf{H}}$ is available at the base station. We give this assumption based on two reasons. First, for the base station, it is often the case that not only the perfect knowledge of the channel state information $\mathbf{H}$ but also the statistics of the estimation error $\tilde{\mathbf{H}}$ is not available. Second, although for some cases, we have the statistics of the estimation error $\tilde{\mathbf{H}}$, it is a time-consuming process to do the calculation. Therefore, we try to simply the modified SVH precoding algorithm a little further, where we do the precoding only depending on the estimation of the channel $\hat{\mathbf{H}}$. Here, we denote this precoding method as the modified SVH2 precoding. Accordingly, the modified SVH precoding will be denoted as the modified SVH1 precoding.

A) Achievable throughput

In Figure 5-8, we see the net sum rate performance after the modified SVH2 precoding is still in the middle of those after the generalized ZF precoding and the SVH precoding, but very close to that after the modified SVH1 precoding. This means that our modified version

of SVH precoding based only on the estimation of the channel $\hat{\mathbf{H}}$ is feasible. Furthermore, the simplification brought by the modified SVH2 precoding algorithm will make this precoding method more practicable.

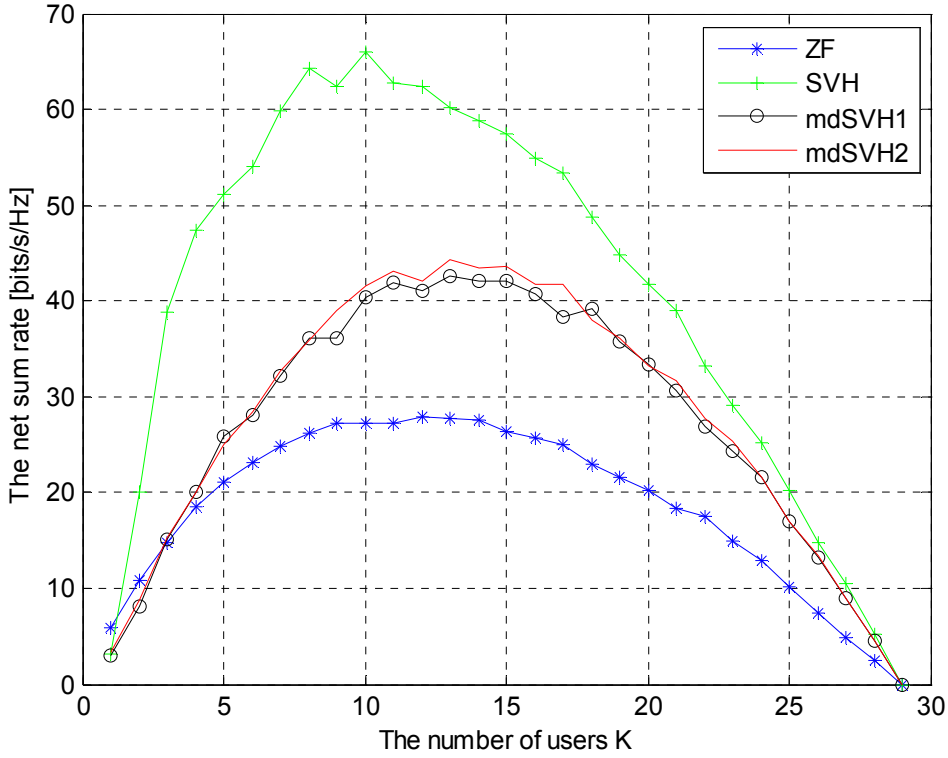

Figure 5-8 The comparison of the net sum rate performance after the generalized ZF precoding, after the SVH precoding, after the modified SVH1 precoding and after the modified SVH2 precoding.

B) The net sum rate performance with different iteration length

In the computation of the family of SVH precoding methods, we use iterative steps to obtain the optimal precoding matrix $\mathbf{A}$. In the following simulation, we want to test the net sum rate performance from two different iteration length, which are 2 and 20, respectively. Here, we implement the simulation on the modified SVH2 precoding method. However, the results from the other two SVH precoding algorithms will be similar.

From Figure 5-9, we can see that the change of the length of the iteration does increase the net sum rate performance of the system. However, this enhancement of the throughput performance is very limited. Considered about the calculation cost, the choice of the value of the iteration length should be system dependent. If the system has the powerful capability in computation, we can choose a relative larger iteration for the modified SVH2 precoding.

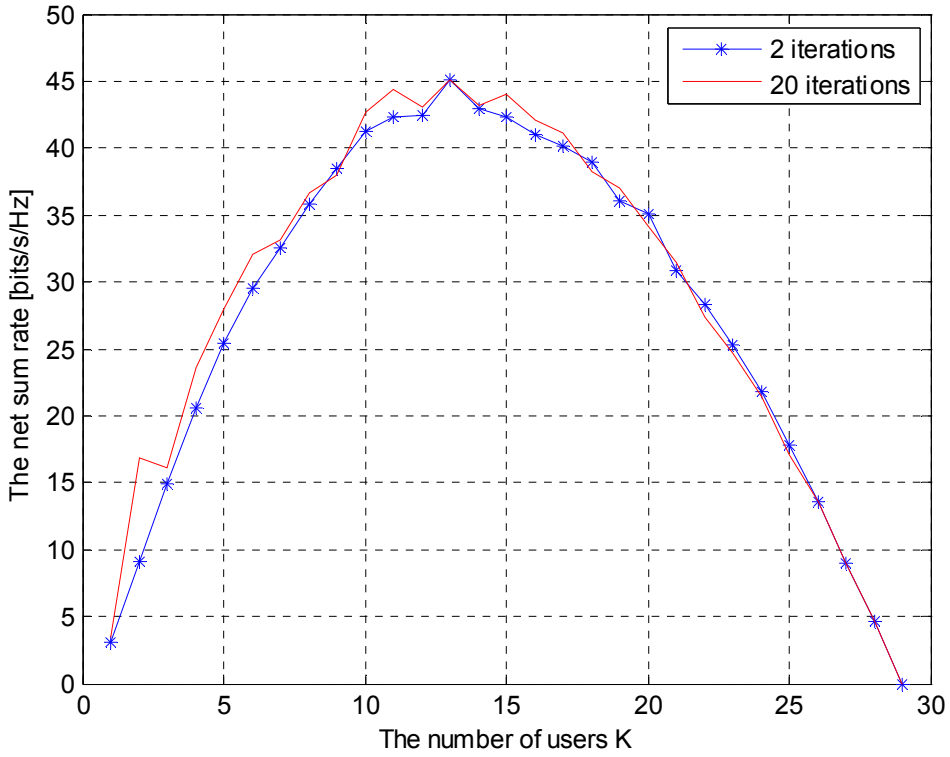

Figure 5-9 The net sum rate performance with different iteration length.

5.4 Simulation on the larger coherence time interval case

In the previous simulations, we assume that the coherence time interval is as small as $T = 30$ symbols and the corresponding number of users may be even less, considering the training. This assumption is based on users moving in a relatively high speed. However, if we loose the assumption a little bit. The mobile speed of the user is not fast, say $T = 50$. Then we redo the simulations on the net sum rate performance. In the simulations, we first compare the throughput performance after the modified SVH2 precoding when the coherence time interval $T = 30$ and $T = 50$, respectively. Then we compare the performance after the generalized ZF precoding, the SVH precoding and the modified SVH2 precoding, when the coherence time interval $T = 50$.

From Figure 5-10, we clearly see that the net sum rate of the system has been greatly improved. The peak net sum rate jumps from approximately 42 bit/s/Hz when $T = 30$ to about 65 bit/s/Hz when $T = 50$. Given the bandwidth 20 MHz, the peak data rate of the system can reach as high as 1.2 Gbps.

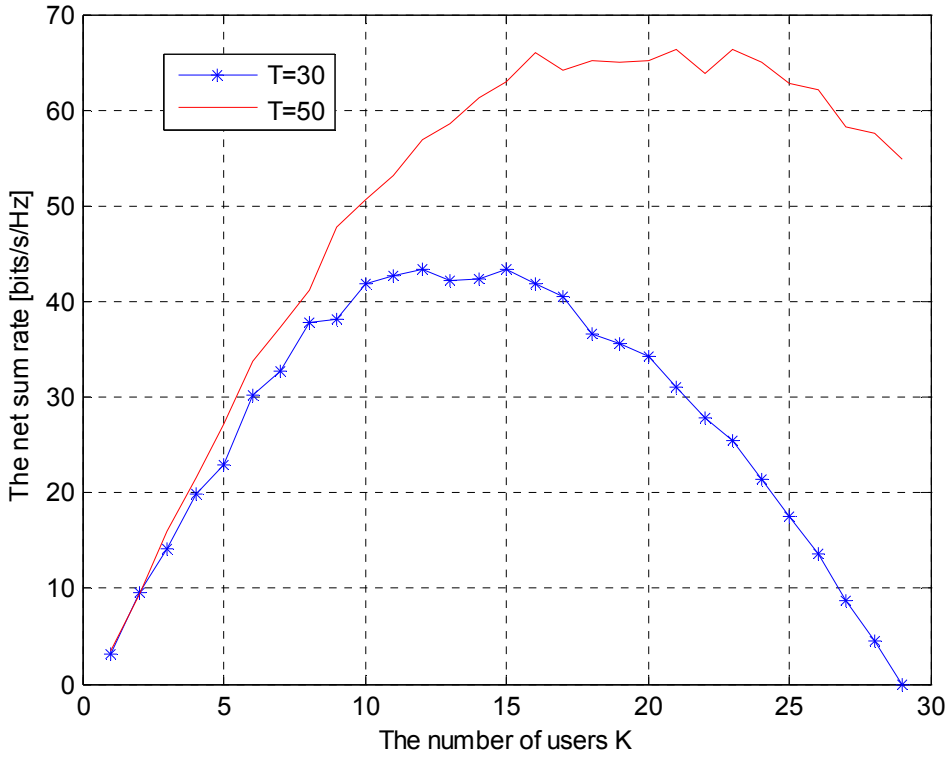

Figure 5-10 The comparison of the net sum rate performance after the modified SVH2 precoding when the coherence time interval $T = 30$ and $T = 50$.

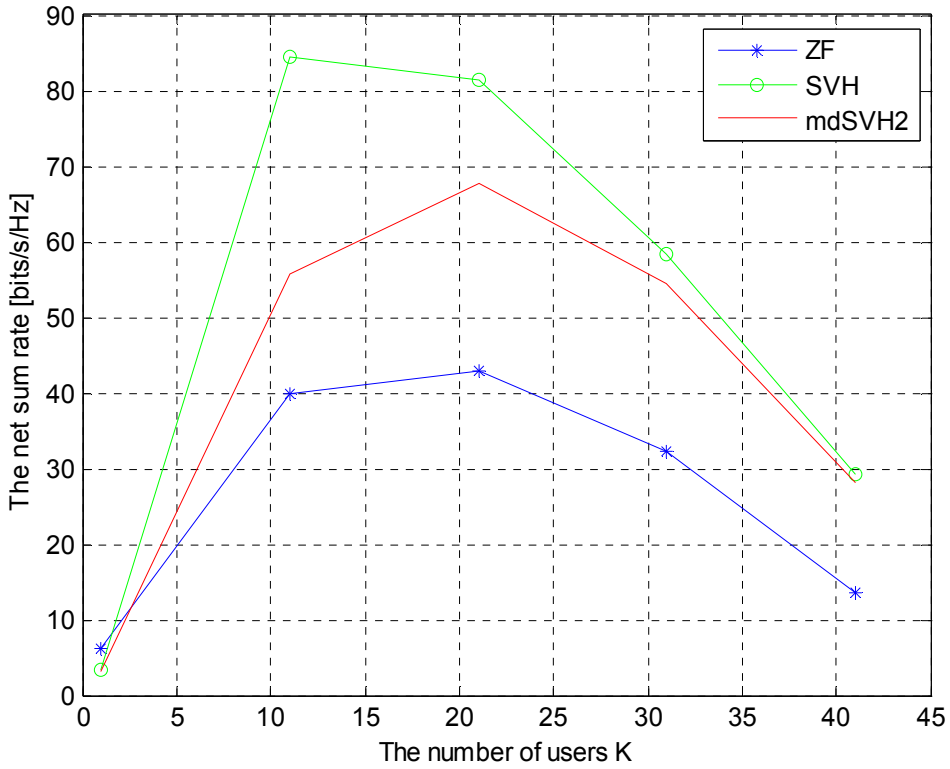

Figure 5-11 The comparison of the net sum rate performance after the generalized ZF precoding, after the SVH precoding and after the modified SVH2 precoding when the coherence time interval $T = 50$.

In Figure 5-11, we compare the throughput performance among different precoding algorithms. It is obvious that the net sum rates after using these three precoding methods are growing up, due to the increase of the coherence time interval. The net sum rate after the SVH precoding is the best among the three, followed by the modified SVH2 precoding, and the generalized ZF precoding, where the peak values are about 85 bit/s/Hz, 68 bit/s/Hz and 42 bit/s/Hz, respectively.

5.5 Simulation on the different antenna array case

In the previous sections, we fixed the number of the antenna array to $M = 100$. In this section, we try to enlarge the antenna array at the base station. We maintain the other parameters the same as before, using the modified SVH precoding algorithm, and compare the net sum rate performance when $M = 100$, $M = 150$ and $M = 16$, respectively.

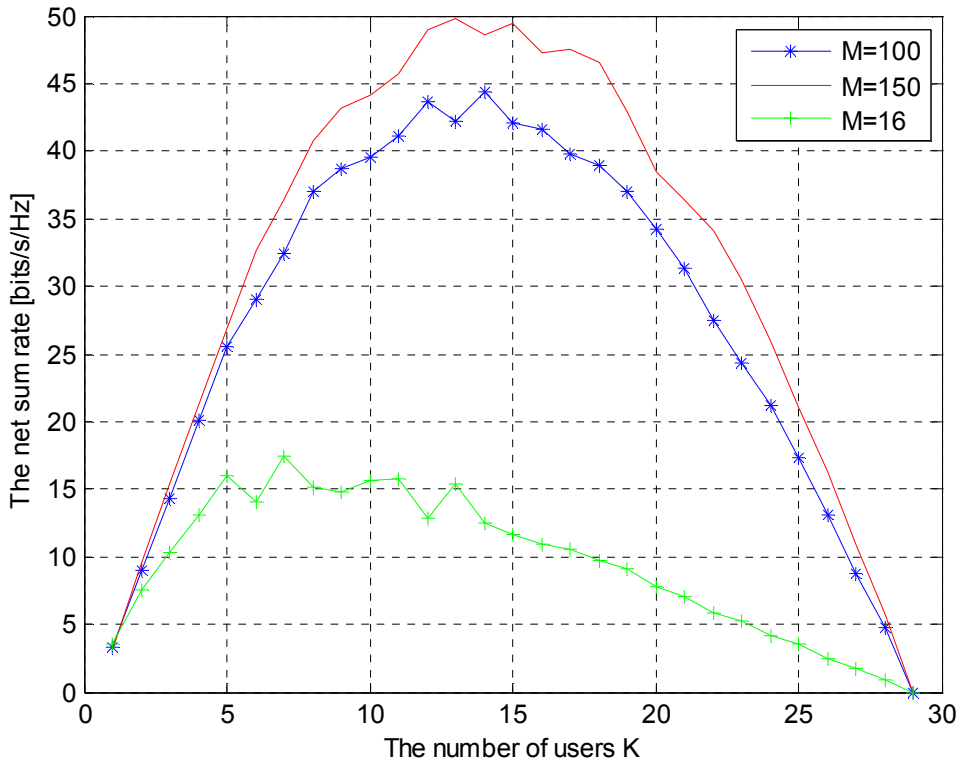

Figure 5-12 The comparison of the net sum rate after the modified SHV2 precoding when the number of antenna $M = 100$, $M = 150$ and $M = 16$.

In [11], the number of antenna used is $M = 16$ in the simulations. However, in the simulations of the thesis, we increase M from 16 to 100. From Figure 5-17, we see that the net sum rate performance is improved by using the larger antenna array. The peak net sum rate increases from approximately 15 bit/s/Hz to 45 bit/s/Hz, when M is growing from 16 to 100. It means that using the larger antenna array at the base station can help the system to

improve its throughput performance. From the next comparison, the peak net sum rate increases from approximately 45 bit/s/Hz to 50 bit/s/Hz, when M increases from 100 to 150. However, compared with the improvement brought by extending the coherence time interval in the last section, the improvement from using the larger antenna array is not as high as expected.

5.6 Summary

In this chapter, we focused on the simulations of the two precoding methods that presented in Chapter 3 and 4. Thus, the simulations can roughly be divided into two parts, one part is about the generalized zero-forcing precoding and the other part is about the SVH precoding. Here, we mainly examined the performance of the precoding methods, in terms of throughput. For the throughput performance, we want to see how much benefit in data rate we can obtain in the downlink channel by using these precoding methods. The results of these simulation shows that our setting of the large number of antennas at the base station can help to improve the whole capacity of the mobile communication system, which is the main assumption about next mobile communication systems proposed in Chapter 2.

We also did some comparisons in the simulation, which shows the differences of the two precoding methods. In the last two simulations, we tried to extend our initial setting about the coherence time interval length and the number of the antennas. Through the simulation results, we have seen different performance by using different coherence time interval length and number of the antennas at the base station.

6 CONCLUSIONS

We developed a general framework of next mobile communication systems in this thesis. The main feature of such systems is the large number of antennas setting, which can be the antenna array with hundreds of antennas at the base station, which is much larger than the 2-4 antennas setting in 4G systems. In this setting, the capacity of the new systems can be greatly increased with the help of the extension in the spatial domain. What is more, many of the techniques, like OFDM, MIMO, TDD, etc., which have been developed in current and former mobile communication systems, are still useful.

However, for the precoding part, the techniques that used in current systems are not suitable any longer. Thus, we proposed our solutions in this thesis. We focused our attention on the linear TDD precoding methods, accounting for the non-feedback and light in the respect of computation features, compared with the FDD and non-linear precoding algorithms. We presented two such precoding methods in Chapter 3 and 4, which are the generalized zero-forcing and SVH precoding methods. In order to estimate the channel state, we also proposed the uplink training method. Based on the reciprocity principle and the received signals of these training sequences, we can obtain the estimation of the downlink channel at the base station. Furthermore, we can form the precoding matrix of the system from the estimation. If the users want to obtain the estimation of their channel gain, they can use the downlink training method.

For the generalized zero-forcing precoding, we analysed the method further in two main situations, which are homogeneous users setting and heterogeneous setting. In both cases, we also considered their scheduling strategy. Here, we mainly paid our attention to the downlink throughput performance of the precoding method and presented their formulas, correspondingly.

For the SVH precoding, we analysed the method in the case that the base station has perfect knowledge of the channel state information. Thus, we presented two modified forms of the SVH precoding method, considering the practical case that the base station does not have perfect knowledge of the channel and only have the estimation of the channel. In addition, the modified SVH precoding also combines the generalized zero-forcing precoding and the SVH precoding together. The ways to calculate the downlink net sum rate of the system are also presented.

We presented the simulations of these two precoding methods in Chapter 5, which validate our theoritical analysis in previous chapters and also confirm the feasibility of our prediction

of next generation mobile communication systems. In the simulations, we mainly examined the performance of the precoding methods, which is the throughput performance. For the throughput performance, we want to see how the precoding methods help to increase the net sum rate of the system.

In the generalized zero-forcing precoding, for the throughput performance in our system settings of the simulation parameters, the peak net sum rate can approximately reach to 30 bit/s/Hz and 100 bit/s/Hz in the homogeneous users and heterogeneous users setting, respectively. Given a bandwidth of 10 MHz, the peak data rate will be about 300 Mbps and 1 Gbps, respectively, which are much faster than that in 4G systems.

For the SVH precoding, the throughput of the SVH, modified SVH1 and modified SVH2 precoding methods are about 65 bit/s/Hz, 42 bit/s/Hz and 42 bit/s/Hz, respectively, in our system settings of the simulation parameters. Compared with the throughput performance of the generalized zero-forcing method of the homogeneous users setting, there is an increase of approximately 20 bit/s/Hz by using the modified versions of SVH precoding method. Thus, in practice, the modified SVH2 precoding method will be a good choice, for it takes advantage of the two main precoding methods presented and ease of implementation.

For future work, if our prediction of the next generation mobile communication systems is widely accepted, there will be considerable works to do on setting the specifications of the systems and other details. Due to the large number of antennas case, new channel models and time-space codes should be found for such new system settings. For the precoding part, we have seen from the simulations that there is a big gap between the SVH precoding and modified SVH precodings in the net sum rate, which means that the precoding method is not perfect and still needs to develop to narrow the gap from the ideal one. Also, if we can find a better way to train and estimate the channel, the time for the training and computation will be compressed and the corresponding time taken for the data transmission will increase.

REFERENCES

[1] W. Stallings, *Wireless communications & networks*, 2nd ed. NJ: Prentice-Hall, 2005.

[2] T. Rappaport, *Wireless communications: principles and practice*, 2nd ed. NJ: Prentice-Hall, 2002.

[3] C. Oestges and B. Clerckx, *MIMO Wireless Communications: From Real-World Propagation to Space-Time Code Design*. Academic Press, 2007

[4] M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theroy*, vol. 29, pp. 439-441, May 1983.

[5] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channels with partial side information," *IEEE Trans. Inf Theory*, vol. 51, pp. 506-522, Feb. 2005.

[6] P. Ding, D. J. Love, and M. D. Zoltowski, "On the sum rate of channel subspace feedback for multi-antenna broadcast channels," in *Proc. IEEE Globecom*, Nov. 2005, pp. 2699-2703.

[7] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, pp. 5045-5060, Nov. 2006.

[8] T. Yoo, N. Jindal, and A. Goldsmith, "Multi-antenna broadcast channels with limited feedback and user selection," *IEEE J. Sel. Areas Commun.,* Vol. 25, pp. 1478-1491, 2007.

[9] K. Huang, R. W. Heath, Jr., and J. G. Andrews, "Space division multiple access with a sum feedback rate constraint," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3879-3891, 2007.

[10] E. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, A. Paulraj, and H. Vincent Poor, *MIMO Wireless Communications*. Cambridge University Press, New York, 2007.

[11] J. Jose, A. Ashikhmin, P. Whiting, and S. Vishwanath, "Precoding methods for multi-user TDD MIMO systems," in *Proc. 46th Asilomar Conf. on Communication, Control, and Computing*, Sept 2008, pp. 761-767.

[12] T. L. Marzetta, "How much training is required for multiuser MIMO?" in *Proc. Asilomar Conf. on Signals, Systems and Computers (ACSSC)*, Pacific Grove, CA, 2006, pp. 359-363.

[13] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, pp. 951-963, Apr. 2003.

[14] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004

[15] J. Jose, A. Ashikhmin, P. Whiting, and S. Vishwanath, "Scheduling and pre-conditioning in multi-user MIMO TDD systems," in *Proc. IEEE International Conf. On Communications (ICC)*, Beijing, China, May 2008, pp. 4100-4105.

[16] M. Stojnic, H. Vikalo, and B. Hassibi, "Rate Maximization in Multi- Antenna Broadcast Channels with Linear Preprocessing," *IEEE Trans. Wireless Commun*., Sept. 2006, vol. 5, pp. 2338-2342.

[17] K. S. Gomadam, H. C. Papadopoulos, and C. -E. W. Sundberg, "Techniques for multi-user MIMO with two-way training," in *Proc. IEEE International Conf. On Communications (ICC)*, Beijing, China, May 2008, PP. 3360-3366.

[18] Steger and A. Sabharwal, "Single-input two-way SIMO channel: diversity-multiplexing tradeoff with two-way training," *IEEE Trans. Wireless Commun*., vol. 7, no. 12, pp. 4877-4885, Dec. 2008.