國 立 交 通 大 學

電信工程研究所

碩 士 論 文

下鏈 LTE-A 蜂巢式系統中協調式多點傳輸方法之研究

A Study on Coordinated Multi-Point Transmission (CoMP) Scheme for the Downlink LTE-A Cellular System

研 究 生: 楊為守

指導教授: 黃家齊 博士

中華民國 一百零一 年 七 月

下鏈 LTE-A 蜂巢式系統中協調式多點傳輸方法之研究

A Study on Coordinated Multi-Point Transmission (CoMP) Scheme for the Downlink LTE-A Cellular System

研 究 生：楊為守　　Student: Wei-Shou Yang

指導教授：黃家齊 博士　　Advisor: Dr. Chia-Chi Huang

國 立 交 通 大 學

電信工程研究所

碩 士 論 文

A Thesis
Submitted to Institute of Communication Engineering
College of Electrical and Computer Engineering
National Chiao Tung University
in Partial Fulfillment of the Requirements
for the Degree of
Master of Science
in
Communication Engineering
July 2012
Hsinchu, Taiwan, Republic of China

中華民國　一百零一 年 七 月

下鏈 LTE-A 蜂巢式系統中協調式多點傳輸方法之研究

研究生：楊為守　　指導教授：黃家齊 博士

國立交通大學電信工程研究所 碩士班

摘　　要

協調式多點傳輸 (Coordinated Multi-Point Transmission，CoMP) 是一種有效降低基地台間干擾的方法。其主要概念是挑選數個基地台彼此合作以消除干擾，這衍生出一個問題：哪些基地台該合作並形成一個協調式多點傳輸叢集 (CoMP Cluster)？針對下鏈傳輸，我們提出一個動態建立叢集的方法。為了降低複雜度，我們使用一種基於區塊對角化 (Block Diagonalization) 的線性前置編碼器。模擬結果顯示出此動態方法優於另一靜態方法。接下來，我們提出一最佳的功率分配方式，使得總傳輸功率最低並同時滿足誤碼率 (Bit Error Rate)、資料傳輸率與天線傳輸功率限制。我們使用 Lagrange 對偶分解 (Dual Decomposition) 來解決此非凸 (Non-Convex) 的最佳化問題。和一固定的功率分配方式比較後，模擬結果顯示此最佳化方法能提供較佳的效能，此外，各天線上的傳輸功率也較少超出其限制。最後，我們提出了一個降低峰均值功率比 (Peak-to-Average Power Ratio，PAPR) 的方法，其使用疊代的方式來改變星座點以降低 PAPR。模擬結果顯示僅需兩次的疊代就可獲得很好的 PAPR 降低效果。

A Study on Coordinated Multi-Point Transmission (CoMP) Scheme for the Downlink LTE-A Cellular System

Student: Wei-Shou Yang Advisor: Dr. Chia-Chi Huang

Institute of Communication Engineering
National Chiao Tung University

ABSTRACT

Coordinated multi-point transmission (CoMP) is a promising way to suppress inter-base station (BS) interference. The main idea of CoMP is to select several BSs which could cooperate together to mitigate interference, which raises an intrinsic problem of which BSs should form a CoMP cluster. We propose a dynamic clustering method for downlink transmission, which forms CoMP clusters adaptively. To reduce the complexity, a linear precoder based on block diagonalization (BD) is used throughout this thesis. Simulation results show that our dynamic scheme outperforms another static method. Next, we design an optimal power allocation method that minimizes the total transmit power while satisfying bit error rate (BER), user rate requirement and per-antenna power constraints. Lagrange dual decomposition is used to solve this non-convex optimization problem. The numerical results reveal the great performance gain against fixed power allocation, and the transmit power on each antenna seldom exceeds the power limit. Finally, we propose a peak-to-average power ratio (PAPR) reduction method, which reduces signal peak by altering signal constellations. The simulation results show large PAPR reduction using an iterative procedure with only two iterations.

誌　謝

首先感謝黃家齊教授這兩年來，對於我的研究、課業與生活上的指導與勉勵，以及對於論文內容的建議，使我得以完成碩士學位。同時感謝口試委員高銘盛教授、陳紹基教授與古孟霖教授給予寶貴的意見與指教，使得本論文更加完整。

特別感謝馬峻楹學長、蕭煒翰學長與溫紹閔學長在我研究過程中給予的指導，讓我的觀念更加清楚、紮實。感謝實驗室的同學俊諺、伯謙、永勝、明佳、凱偉以及學弟妹健瑋、冠銘、日翔、雅涵與駿逸的砥礪與照顧，並帶給實驗室許多歡樂。也感謝梓瑄這段時間的支持，在我遇到挫折時鼓勵我、陪伴我。

感謝我的家人給予我的關心，你們是我最強大的支柱，讓我能無後顧之憂的完成學業。最後，再次感謝所有人，讓我有個非常精彩的碩士生涯。

TABLE OF CONTENTS

中文摘要 i

ABSTRACT ii

誌謝 iii

LIST OF FIGURES vii

LIST OF TABLES viii

CHAPTER 1 INTRODUCTION 1

1.1 CoMP Concept 2

1.2 Resource Allocation 4

1.3 PAPR Reduction 5

1.4 Organization of this thesis 6

CHAPTER 2 CLUSTERING TECHNIQUES FOR COMP 8

2.1 System Model and Transmission Schemes 8

2.1.1 Single Cell Processing 9

2.1.2 Base Station Cooperation 12

2.2 Clustering Algorithms 14

2.2.1 Static Clustering 14

2.2.2 Dynamic Clustering 16

2.3 Simulation Results 19

CHAPTER 3 ADAPTIVE RESOURCE ALLOCATION 27

3.1 System Model and Transmission Schemes 27

3.2 Problem Formulation 28

3.3 Low Complexity Solution for Power Minimization 30

3.3.1 Optimization Based on Dual Decomposition 30

3.3.2 Convergence Behavior Control 35

3.4 Simulation Results 38

CHAPTER 4 PAPR REDUCTION FOR MU-MIMO SYSTEMS **44**

4.1 System Model and Transmission Schemes 44

4.2 Suboptimal Power Minimization Algorithm 45

4.3 Multiuser Active Constellation Extension Method 46

4.3.1 PAPR Definition and ACE Concept 46

4.3.2 Problem Formulation 48

4.3.3 An Efficient Algorithm for ACE 50

4.3.4 Modifications of ACE 54

4.4 Simulation Results 55

CHAPTER 5 CONCLUSION **61**

APPENDIX A DERIVATIONS FOR CHAPTER 3 **63**

A.1 Derivation of Competitive Water Filling Solution 63

A.2 Derivation of Supergradient 65

REFERENCES **67**

LIST OF FIGURES

Figure **Page**

1.1 Illustration for CoMP-JT mode. The solid arrows stand for the signal links. 2

1.2 Illustration for CoMP-CB mode. The solid arrows stand for the signal links and the dashed ones are the interfering links. 3

2.1 A clustering example with $B = 2$ and $N_b = 4$. 9

2.2 Proposed static clustering table when cluster size $B = 3$. 16

2.3 A snapshot of the dynamic clustering. 18

2.4 Average rate versus the edge-SNR. 20

2.5 The average rate versus different pilot SINR threshold. 22

2.6 The CDF of the CoMP requesting users. 24

2.7 The CDF of the CoMP included users. 24

2.8 The CDF of the last five percent users. 25

2.9 The average rate versus different CoMP cluster size. 26

2.10 The CDF of the CoMP requesting users using weighted sum utility versus different CoMP cluster size. 26

3.1 Block diagram for downlink multi-user MIMO-OFDM. 27

3.2 Flowchart of our resource allocation algorithm. 37

3.3 Illustration for three-sectorized CoMP. 38

3.4 Edge-SNR versus different target rate. 40

3.5 Edge-SNR versus different number of antennas. 40

3.6 User rate convergence behavior. 42

3.7 Sum-power convergence behavior. 42

3.8 Normalized transmit power on antenna 1 over 100 channel realizations when $p_a^{\text{con}} \cong 8.019$ W. 43

3.9 Normalized transmit power on antenna 1 over 100 channel realizations when $p_a^{\text{con}} \cong 1.271$ W. 43

4.1 Illustration for ACE with QPSK modulation. 47

4.2 Illustration for ACE with 16-QAM modulation. 48
4.3 Illustration of block-IDFT. 49
4.4 The SNR versus different rate requirements. 56
4.5 A snapshot of the data symbols with QPSK modulation after ACE. 56
4.6 A snapshot of the data symbols with 16-QAM modulation after ACE. 57
4.7 The CCDF of the PAPR after ACE. 58
4.8 The CCDF of the PAPR after ACE with different oversampling factors. 58
4.9 The CCDF of the PAPR after ACE with different PAPR targets. 59
4.10 The CCDF of the PAPR after ACE with different subcarrier sizes. 59

LIST OF TABLES

Table **Page**

2.1 The CoMP request ratio and the actual CoMP included ratio against different pilot SINR threshold. 21

CHAPTER 1

INTRODUCTION

With the demand of high data rate in the future wireless communication systems, multiple-input multiple-output (MIMO) techniques have been proposed to improve the system throughput. Besides, in a cellular network, the frequency band can be reused in a one cell fashion to maximize the spectral efficiency. However, one cell frequency reuse limits the performance of MIMO systems due to severe other-cell interference (OCI) [1]. Although the receiver can eliminate the interference by applying techniques like successive interference cancellation (SIC), but in the downlink, this burdens the user equipment (UE) with high computational complexity. Recently, a promising way called CoMP has been proposed by the long-term evolution advanced (LTE-A) to suppress OCI. The idea of CoMP is to gather a group of BSs which share channel state information (CSI) and/or user data via high speed backhaul. In this way, the BSs can cooperate to lower the interference. On the other hand, since the channel varies with the user location, we can allocate power in different domains like frequency and space according to their channel quality. LTE-A prescribes orthogonal frequency division multiplexing (OFDM) for downlink transmission, such a multi-tone system suffers from high PAPR, which reduces the transmit power efficiency. In this thesis, we discuss the CoMP, power allocation and PAPR reduction issues, and some algorithms suitable for BS coordination scenario are proposed. It should be noted that ideal synchronization and perfect CSI estimation are assumed.

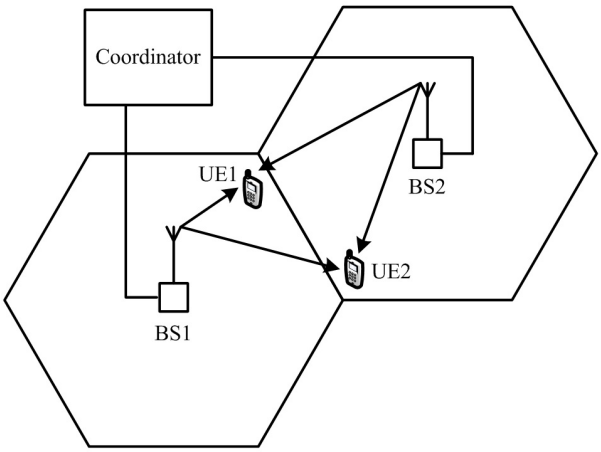

Figure 1.1: Illustration for CoMP-JT mode. The solid arrows stand for the signal links.

1.1 CoMP Concept

In LTE-A downlink, there are two classes of CoMP schemes named joint transmission (JT) and coordinated beamforming (CB). One can distinguish these two modes by the type of information sharing. The former requires both CSI and data exchange and the latter needs only CSI exchange.

The concept of CoMP-JT can be illustrated using Figure 1.1. Conceptually, the cellular network which deploys CoMP-JT is equivalent to multi-user MIMO (MU-MIMO) with some distinctions that the transmit antennas now belong to distributed BSs and the channels to different users experience independent pathloss and shadowing. Joint transmission means the signal intended for any user is jointly pre-processed and transmitted from all the BSs. Therefore, the interfering links (a link is the channel from a BS to a user) are transformed into useful links. The drawback of this mode is that the data for every user and the CSI of all the links need to be shared among the BSs, which increases the backhaul overhead.

In the case of CoMP-CB (see Figure 1.2), as conventional single cell scheme, each BS transmits the data to its users. But the BSs can cooperate to mitigate the interference to the UE in

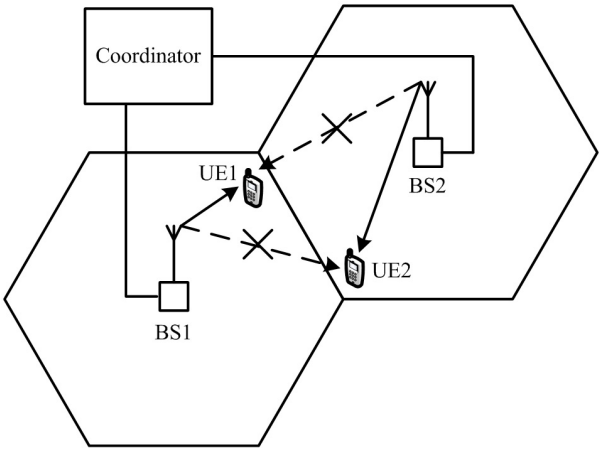

Figure 1.2: Illustration for CoMP-CB mode. The solid arrows stand for the signal links and the dashed ones are the interfering links.

the other cell by acquiring the CSI of the interfering links. In this mode, there is no data sharing, which reduces the backhaul traffic. However, the design target is to eliminate the generated interference but not to make use of it, which leads limited performance gain compared to CoMP-JT. In both cases, the users need to feedback the CSI from all the links including the interfering ones. In time division duplex (TDD) mode, the job of channel estimation can be placed in either BS or user side, while in frequency division duplex (FDD) mode, user has to estimate the CSI and feedback through uplink channels.

Plentiful research works on CoMP-JT can be found in the recent years [2]-[6]. [2] gives an overview including the currently known techniques for CoMP-JT (also a part of CoMP-CB), practical issues related to system complexity and main challenges for future CoMP design. A straightforward way to implement CoMP-JT is to build a central coordinator (CC) which collects all the CSI and then computes the precoding weights for all the users. The performance analysis to this way, using nonlinear (dirty paper coding, DPC) and linear (zero-forcing, ZF and minimum min square error, MMSE) precoding, was studied in [3] and [4], respectively. Such a structure would be more complicated and infeasible as the size of the network increases. A way

to overcome this problem is dividing the BSs into several clusters [5], [6].

As for CoMP-CB, [7] proposed a beamforming algorithm which aims to find the interfering users with similar channels. Another way to suppress interference using distributed resource allocation is discussed in [8]. In this thesis, we mainly focus on CoMP-JT.

In order to reduce the overhead, there are only a limited number of BSs can be included in a CoMP cluster. This leads to the question which BSs should form the clusters to maximize the system performance at manageable complexity. Static [9] and dynamic [10] clustering are two types of forming CoMP clusters. Static clustering can be performed in advance based on field measurements or geographical relations. Whereas dynamic clustering exploits CSI and changes the cluster groups over time. Algorithms for these two types can be found in Chapter 2.

1.2 Resource Allocation

After the BSs in the cellular network are grouped into several CoMP clusters, there are more jobs we can do in each cluster to improve the system performance. OFDM is chosen to be the modulation scheme for the downlink LTE-A, one advantage of the multicarrier system is that the power and rate can be allocated over different tones according to their variety if the transmitter has the CSI. In general, there are two objectives of resource allocation: the sum-rate maximization and power minimization [11], [12]. We pay our attention on the latter in this thesis for power saving.

Traditionally, resource allocation problems are formulated under sum-power constraint on the transmit antennas. However, per-antenna power constraint would be more realistic since each antenna has independent power amplifier. Besides, for avoiding inter-user interference, many systems apply orthogonal frequency division multiple access (OFDMA) technique which allows only one user on each subcarrier [11]. But the spectrum efficiency can be further im-

proved by separating the users in the spatial domain when the transmitter equips multiple antennas [13], in this way, multiple users can share the same subcarrier. As for the user terminals, different users may have individual quality of service (QoS) requirements such as minimal data rate and acceptable bit error rate (BER).

In Chapter 3, the power minimization problem under user rate and BER constraints as in [12] is considered. To make this problem more general, we add additional per-antenna power constraint and extend the system to a multi-cell scenario. The problem is non-convex since it aims to find the optimal set among different subcarrier and user combinations, and the complexity increases exponentially with the number of subcarriers and users. Therefore, efficient solutions must be found to make the complexity feasible. Although the original problem is not convex, it can be transformed to another problem based on Lagrange dual decomposition [12]. The transformed problem is always concave regardless of the convexity of the original problem. In this way, conventional convex optimization techniques can be applied to solve this problem efficiently.

1.3 PAPR Reduction

Although the average transmit power can be minimized via the techniques introduced in Chapter 3, however, the power consumption of the power amplifier (PA) is dominated by the peak power rather than the average power. One of the main drawbacks of OFDM is the large PAPR, which makes the efficiency of the PA very poor (the definition of PAPR is left in section 4.3.1. In order to transmit a signal with wide power rang, an expensive PA is needed. For improving the power efficiency, several classes of PAPR reduction techniques have been proposed [14]-[16].

Tone reservation method [14] is one of the most popular PAPR reduction methods. It re-

serves some unused subcarriers and inserts signals to reduce the PAPR. This method causes no distortion to the data-carrying subcarriers due to the orthogonality of subcarriers. However, there exists a tradeoff between the PAPR reduction performance and the number of the reserved subcarriers. Reserving more subcarriers yields better performance, but sacrifices the available bandwidth for transmitting information data.

Another class called *active constellation extension* (ACE) tries to reduce PAPR by altering the constellation of the data [15], [16]. Without reserving any subcarrier, this scheme maps the original constellation to a constrained space which produces lower PAPR.

In Chapter 4, we first propose a suboptimal power allocation algorithm to reduce the average transmit power. As stated above, the power consumption depends mainly on the peak power. Hence we introduce a PAPR reduction scheme which combines tone reservation and constellation extension. This scheme is based on the idea of [15]. Additionally, we made some modifications so that it is suitable for CoMP-JT systems. As a note, this approach is designed for the signal which employs quadrature amplitude modulation (QAM).

1.4 Organization of this thesis

The rest of this thesis is organized as follows. In Chapter 2, we start with the CoMP clustering concept. How does a cellular network implement CoMP and which BSs should form a CoMP cluster will be illustrated. In Chapter 3, assuming all the CoMP clusters have been planned, we design an optimal power allocation method for each cluster. For power saving, the objective is to minimize the total transmit power. Chapter 4 deals with the problem of PAPR. A practical PAPR reduction algorithm suitable for the CoMP system is proposed. Finally, Chapter 5 gives the conclusion of this thesis. It should be noted that Chapter 2 - 4 have their own system models and simulation results.

Throughout this thesis, we adopt some notations: Matrix and vectors are denoted by uppercase and lower case boldface letters; $\mathbf{I}_N$ is the $N \times N$ identity matrix; $\mathbf{A}[i,j]$ indicates the element in the row i and column j of the matrix $\mathbf{A}$; $\|\cdot\|_F^2$ and $\|\cdot\|_\infty^2$ are the Frobenius norm and infinity norm; $(\cdot)^*$, $(\cdot)^T$ and $(\cdot)^H$ represent the conjugate, transpose and conjugate transpose operators.

CHAPTER 2

CLUSTERING TECHNIQUES FOR COMP

2.1 System Model and Transmission Schemes

Consider a downlink cellular network consists of N_b BSs with N_t antennas each and K_b active users in the b^{th} cell with N_r antennas each. All the BSs operate on the same carrier frequency so they will cause interference to each other. Assume that the BSs in the network are divided into several clusters, each contains B BSs, where $B \leq N_b$ (see Figure 2.1). The clusters are all non-overlapping groups. In other words, if one BS has been involved in a cluster, it cannot join the others. Let $\mathcal{G}$ be one of the set of the selected BSs in a cluster. The received signal of the k^{th} user in the b^{th} cell of the set $\mathcal{G}$ can be written as

$$\mathbf{y}_{k,b}^{\mathcal{G}} = \mathbf{H}_{k,b}^{b}\mathbf{x}_{k,b} + \sum_{i\neq k, i\leftrightarrow b} \mathbf{H}_{k,b}^{b}\mathbf{x}_{i,b} + \sum_{\bar{b}\in\mathcal{G},\bar{b}\neq b} \sum_{j\leftrightarrow\bar{b}} \mathbf{H}_{k,b}^{\bar{b}}\mathbf{x}_{j,\bar{b}} + \sum_{\widetilde{b}\notin\mathcal{G}} \sum_{l\leftrightarrow\widetilde{b}} \mathbf{H}_{k,b}^{\widetilde{b}}\mathbf{x}_{l,\widetilde{b}} + \mathbf{n}_{k,b}, \qquad (2.1)$$

where $\mathbf{H}_{k,b}^{\bar{b}}$ is the $N_r \times N_t$ MIMO channel from the BS $\bar{b}$ to the user k served by the BS b, $\mathbf{x}_{k,b}$ is the $N_t \times 1$ transmitted signal and $\mathbf{n}_{k,b}$ is the corresponding $N_r \times 1$ received noise vector, in which each element is a zero-mean complex Gaussian random variable with variance N_0. The first term of (2.1) is the desired signal, the second is the inter-user interference (IUI) in the b^{th} cell where $i \leftrightarrow b$ means the user i is served by the BS b, the third is the intra-cluster interference (ICI) from the BSs in the cluster $\mathcal{G}$ except the BS b, and the last term is the outer-cluster interference (OCI) from the BSs outside the cluster $\mathcal{G}$.

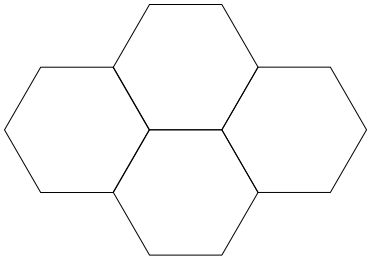

Figure 2.1: A clustering example with $B = 2$ and $N_b = 4$.

2.1.1 Single Cell Processing

In this scenario, each BS serves its users without coordinating with the other BSs, i.e., $B = 1$. Assume the b^{th} BS is considered. The transmit precoding matrices for the user k in the BS b are designed in the following two cases respectively.

Case 1) Eigenmode Precoding for $K_b = 1$

When there is only one user per BS, the received signal of the user k in the cell b can be written as

$$\mathbf{y}_{k,b} = \mathbf{H}^b_{k,b}\mathbf{x}_{k,b} + \sum_{\bar{b}\neq b}\sum_{i\leftrightarrow\bar{b}} \mathbf{H}^{\bar{b}}_{k,b}\mathbf{x}_{i,\bar{b}} + \mathbf{n}_{k,b}. \tag{2.2}$$

Note that the index for cluster is ignored here since there is no clustering concept when $B = 1$. When each BS has only the CSI of its served user, the precoding matrix can be designed by performing singular value decomposition (SVD) on $\mathbf{H}^b_{k,b}$:

$$\mathbf{H}^b_{k,b} = \mathbf{U}^b_{k,b}\mathbf{S}^b_{k,b}\left(\mathbf{V}^b_{k,b}\right)^H, \tag{2.3}$$

where $\mathbf{S}^b_{k,b} \in \mathbb{C}^{N_r\times N_t}$ is the matrix which contains the singular values, $\mathbf{U}^b_{k,b} \in \mathbb{C}^{N_r\times N_r}$ and $\mathbf{V}^b_{k,b} \in \mathbb{C}^{N_t\times N_t}$ collects the left and right singular vectors, respectively. Let $\mathbf{P}^b_{k,b}$ be the power allocation matrix for the user k in the BS b. Here we assume a simple equal power allocation, that is

$$\mathbf{P}^b_{k,b} = \left(\sqrt{p^{\text{con}}/N_r}\right)\mathbf{I}_{N_r}, \tag{2.4}$$

where p^{con} is the sum power constraint per BS. The precoding matrix can be chosen as

$$\mathbf{F}_{k,b}^{b} = \mathbf{V}_{k,b}^{b} \tag{2.5}$$

and the receive equalization matrix is

$$\mathbf{Q}_{k,b}^{b} = \left(\mathbf{U}_{k,b}^{b}\right)^{H}. \tag{2.6}$$

Having the precoding matrix, the pre-processing can be done as

$$\mathbf{x}_{k,b}^{b} = \mathbf{F}_{k,b}^{b}\mathbf{P}_{k,b}^{b}\mathbf{d}_{k,b}^{b}, \tag{2.7}$$

where $\mathbf{d}_{k,b}^{b} \in \mathbb{C}^{N_r \times 1}$ represents the information data. After receive equalization, the signal becomes

$$\begin{aligned}
\mathbf{r}_{k,b} &= \mathbf{Q}_{k,b}^{b}\mathbf{y}_{k,b} \\
&= \mathbf{Q}_{k,b}^{b}\left(\mathbf{H}_{k,b}^{b}\mathbf{x}_{k,b} + \sum_{\bar{b}\neq b}\sum_{i\leftrightarrow\bar{b}}\mathbf{H}_{k,b}^{\bar{b}}\mathbf{x}_{i,\bar{b}} + \mathbf{n}_{k,b}\right) \\
&= \left(\mathbf{U}_{k,b}^{b}\right)^{H}\mathbf{U}_{k,b}^{b}\mathbf{S}_{k,b}^{b}\left(\mathbf{V}_{k,b}^{b}\right)^{H}\mathbf{V}_{k,b}^{b}\mathbf{P}_{k,b}^{b}\mathbf{d}_{k,b}^{b} + \mathbf{Q}_{k,b}^{b}\sum_{\bar{b}\neq b}\sum_{i\leftrightarrow\bar{b}}\mathbf{H}_{k,b}^{\bar{b}}\mathbf{x}_{i,\bar{b}} + \mathbf{Q}_{k,b}^{b}\mathbf{n}_{k,b} \\
&= \mathbf{S}_{k,b}^{b}\mathbf{P}_{k,b}^{b}\mathbf{d}_{k,b}^{b} + \sum_{\bar{b}\neq b}\mathbf{i}_{k,b}^{\bar{b}} + \widetilde{\mathbf{n}}_{k,b},
\end{aligned} \tag{2.8}$$

where $\mathbf{i}_{k,b}^{\bar{b}} = \sum_{i\leftrightarrow\bar{b}}\mathbf{Q}_{k,b}^{b}\mathbf{H}_{k,b}^{\bar{b}}\mathbf{x}_{i,\bar{b}}$ is the equivalent interference from BS $\bar{b}$. Therefore the spatial streams can be extracted as

$$r_{k,b,l} = s_{k,b,l}^{b}\sqrt{p_{k,b,l}^{b}}d_{k,b,l}^{b} + i_{k,b,l}^{\bar{b}} + \widetilde{n}_{k,b,l},\ 1 \leq l \leq N_r, \tag{2.9}$$

where $d_{k,b,l}^{b}$ is the l^{th} element of $\mathbf{d}_{k,b}^{b}$, $s_{k,b,l}^{b}$ and $p_{k,b,l}^{b}$ are the corresponding singular value and the power loading weight, respectively, $i_{k,b,l}^{\bar{b}}$ is the interference on the l^{th} stream and $\widetilde{n}_{k,b,l}$ is the complex Gaussian noise with variance N_0. As a result, the $N_r \times N_t$ MIMO channel $\mathbf{H}_{k,b}^{b}$ is decoupled into N_r parallel single-input single-output (SISO) channels due to the effect of eigenmode precoding and equalization. In this case, the

signal-to-interference plus noise ratio (SINR) of the l^{th} stream for the user k in the BS b is

$$SINR_{k,b,l}^{\text{eigen}} = \frac{\left(s_{k,b,l}^{b}\right)^2 p_{k,b,l}^{b}}{E\left[\sum_{\bar{b}\neq b} i_{k,b,l}^{\bar{b}}\left(i_{k,b,l}^{\bar{b}}\right)^*\right] + N_0} \tag{2.10}$$

and the capacity is

$$C_{k,b}^{\text{eigen}} = \sum_{l=1}^{N_r} \log_2\left(1 + SINR_{k,b,l}^{eigen}\right). \tag{2.11}$$

Case 2) Block Diagonalization (BD) [13] for $K_b > 1$

In this case, the interference seen by the user k in the BS b can come from the other users in the same BS and those in the other BSs. The received signal can be written as

$$\mathbf{y}_{k,b} = \mathbf{H}_{k,b}^{b}\mathbf{x}_{k,b} + \sum_{i\neq k, i\leftrightarrow b} \mathbf{H}_{k,b}^{b}\mathbf{x}_{i,b} + \sum_{\bar{b}\neq b}\sum_{j\leftrightarrow\bar{b}} \mathbf{H}_{k,b}^{\bar{b}}\mathbf{x}_{j,\bar{b}} + \mathbf{n}_{k,b}. \tag{2.12}$$

Having only the CSI of the users in its scope, each BS can apply BD to eliminate the IUI, i.e., the second term of (2.12). There is a restriction on BD, which is $N_t \geq K_b N_r$, that is, the number of the transmit antennas must be larger than or equal to the sum of the number of the receive antennas. We introduce the procedure of designing the BD precoder below.

For simplicity, the index for BS is ignored here. The composite channel of all the users in the cluster can be written as $\mathbf{H} = \left[\mathbf{H}_1^T, \ldots, \mathbf{H}_K^T\right]^T \in \mathbb{C}^{KN_r \times N_t}$. Then we collect the interfering channels to the user k and apply SVD as

$$\left[\mathbf{H}_1^T, \ldots, \mathbf{H}_{k-1}^T, \mathbf{H}_{k+1}^T, \ldots, \mathbf{H}_K^T\right]^T = \mathbf{U}_k\mathbf{S}_k\left[\mathbf{V}_k\overline{\mathbf{V}}_k\right]^H, \tag{2.13}$$

where $\overline{\mathbf{V}}_k \in \mathbb{C}^{N_t \times (N_t-(K-1)N_r)}$ is the matrix which contains the right singular vectors that correspond to the zero singular values and therefore is the null space. The equivalent channel to the user k after orthogonalization is given by

$$\widetilde{\mathbf{H}}_k = \mathbf{H}_k\overline{\mathbf{V}}_k. \tag{2.14}$$

Apply SVD again on $\widetilde{\mathbf{H}}_k$ and get

$$\widetilde{\mathbf{H}}_k = \widetilde{\mathbf{U}}_k \widetilde{\mathbf{S}}_k \widetilde{\mathbf{V}}_k. \tag{2.15}$$

Similarly, we can use eigenmode precoder to decompose the equivalent MIMO channel $\widetilde{\mathbf{H}}_k$. The transmit precoding matrix can be designed as

$$\mathbf{F}_k = \overline{\mathbf{V}}_k \widetilde{\mathbf{V}}_k \tag{2.16}$$

and the receive equalization matrix is

$$\mathbf{Q}_k = \widetilde{\mathbf{U}}_k^H. \tag{2.17}$$

Hence, the users in the BD deployed BS will not observe interference from each other because their channels are mutually orthogonal, although the interference from the other BSs still remains. Since there are K users now, the power allocated to each user should be divided by K. Thus, the power allocation matrix should be

$$\mathbf{P}_k = \left(\sqrt{p^{\text{con}} / K N_r} \right) \mathbf{I}_{N_r}. \tag{2.18}$$

Since the inter-user interference is eliminated, the capacity for the user k can be represented in the same form as 2.11, the difference is that each BS now applies BD but not eigenmode precoding.

2.1.2 Base Station Cooperation

In this section, we show where the BD algorithm should be modified when it is applied to a multi-cell scenario. Assume a CoMP clustering algorithm is applied and some CoMP clusters are formed. In JT mode, all the users' information data and channels are shared among the BSs in the same cluster. Therefore, the grouped BSs can be seen as a huge BS with additional antennas. In this scenario, each CoMP cluster applies BD to decouple the channels of all the users in its

scope. The multi-cell BD can be done in a way similar to the Case 2 of single cell processing with some little differences that the transmit antenna N_t becomes BN_t and the sum power constraint p^{con} changes to Bp^{con}. Like single cell BD, the channels of all the users in the same cluster will become orthogonal to each other after multi-cell BD. Therefore, the interference only comes from the BSs outside the cluster, and the received signal of the user k in the BS b of the cluster $\mathcal{G}$ can be written as

$$\mathbf{y}_{k,b}^{\mathcal{G}} = \mathbf{H}_{k,b}^{\mathcal{G}}\mathbf{x}_{k,b}^{\mathcal{G}} + \sum_{\bar{b}\notin\mathcal{G}}\sum_{i\leftrightarrow\bar{b}} \mathbf{H}_{k,b}^{\bar{b}}\mathbf{x}_{i,\bar{b}} + \mathbf{n}_{k,b}^{\mathcal{G}}, \tag{2.19}$$

where $\mathbf{H}_{k,b}^{\mathcal{G}}$ is the $N_r \times BN_t$ multi-cell MIMO channel and $\mathbf{x}_{k,b}^{\mathcal{G}}$ is the $BN_t \times 1$ transmitted signal. The joint BD precoding matrix is

$$\mathbf{F}_{k,b}^{\mathcal{G}} = \overline{\mathbf{V}}_{k,b}^{\mathcal{G}}\widetilde{\mathbf{V}}_{k,b}^{\mathcal{G}}, \tag{2.20}$$

where $\overline{\mathbf{V}}_{k,b}^{\mathcal{G}}$ is the null space of the other users in $\mathcal{G}$ except the user k and $\widetilde{\mathbf{V}}_{k,b}^{\mathcal{G}}$ is the matrix which collects the first N_r right singular vectors of the projected channel $\widetilde{\mathbf{H}}_{k,b}^{\mathcal{G}} = \mathbf{H}_{k,b}^{\mathcal{G}}\overline{\mathbf{V}}_{k,b}^{\mathcal{G}}$. The jointly pre-processed signal can be represented as

$$\mathbf{x}_{k,b}^{\mathcal{G}} = \mathbf{F}_{k,b}^{\mathcal{G}}\mathbf{P}_{k,b}^{\mathcal{G}}\mathbf{d}_{k,b}^{\mathcal{G}}, \tag{2.21}$$

where the power allocation matrix is

$$\mathbf{P}_{k,b}^{\mathcal{G}} = \left(\sqrt{Bp^{\text{con}}/\sum_{b\in\mathcal{G}} K_b N_r}\right)\mathbf{I}_{N_r}, \tag{2.22}$$

(assume every BS has the same transmit power constraint p^{con}) and $\mathbf{d}_{k,b}^{\mathcal{G}}$ is the information data. If the user k chooses the equalization matrix to be

$$\mathbf{Q}_{k,b}^{\mathcal{G}} = \left(\widetilde{\mathbf{U}}_{k,b}^{\mathcal{G}}\right)^{H}, \tag{2.23}$$

where $\widetilde{\mathbf{U}}_{k,b}^{\mathcal{G}}$ contains the left singular vectors of the projected channel $\widetilde{\mathbf{H}}_{k,b}^{\mathcal{G}}$. The equalized signal at the receiver side can be derived following the similar steps of (2.8) and given by

$$\begin{aligned}
\mathbf{r}_{k,b}^{\mathcal{G}} &= \mathbf{Q}_{k,b}^{\mathcal{G}} \left(\mathbf{H}_{k,b}^{\mathcal{G}} \mathbf{x}_{k,b}^{\mathcal{G}} + \sum_{\bar{b} \notin \mathcal{G}} \sum_{i \leftrightarrow \bar{b}} \mathbf{H}_{k,b}^{\bar{b}} \mathbf{x}_{i,\bar{b}} + \mathbf{n}_{k,b}^{\mathcal{G}} \right) \\
&= \mathbf{Q}_{k,b}^{\mathcal{G}} \mathbf{H}_{k,b}^{\mathcal{G}} \overline{\mathbf{V}}_{k,b}^{\mathcal{G}} \widetilde{\mathbf{V}}_{k,b}^{\mathcal{G}} \mathbf{P}_{k,b}^{\mathcal{G}} \mathbf{d}_{k,b}^{\mathcal{G}} + \sum_{\bar{b} \notin \mathcal{G}} \sum_{i \leftrightarrow \bar{b}} \mathbf{Q}_{k,b}^{\mathcal{G}} \mathbf{H}_{k,b}^{\bar{b}} \mathbf{x}_{i,\bar{b}} + \mathbf{Q}_{k,b}^{\mathcal{G}} \mathbf{n}_{k,b}^{\mathcal{G}} \\
&= \left(\widetilde{\mathbf{U}}_{k,b}^{\mathcal{G}} \right)^H \widetilde{\mathbf{U}}_{k,b}^{\mathcal{G}} \widetilde{\mathbf{S}}_{k,b}^{\mathcal{G}} \left(\widetilde{\mathbf{V}}_{k,b}^{\mathcal{G}} \right)^H \widetilde{\mathbf{V}}_{k,b}^{\mathcal{G}} \mathbf{P}_{k,b}^{\mathcal{G}} \mathbf{d}_{k,b}^{\mathcal{G}} + \sum_{\bar{b} \notin \mathcal{G}} \mathbf{i}_{k,b}^{\bar{b}} + \widetilde{\mathbf{n}}_{k,b}^{\mathcal{G}} \\
&= \widetilde{\mathbf{S}}_{k,b}^{\mathcal{G}} \mathbf{P}_{k,b}^{\mathcal{G}} \mathbf{d}_{k,b}^{\mathcal{G}} + \sum_{\bar{b} \notin \mathcal{G}} \mathbf{i}_{k,b}^{\bar{b}} + \widetilde{\mathbf{n}}_{k,b}^{\mathcal{G}}. \qquad (2.24)
\end{aligned}$$

Being similar to (2.9), the multi-cell MIMO channel is decoupled and the l^{th} spatial stream is

$$r_{k,b,l}^{\mathcal{G}} = s_{k,b,l}^{\mathcal{G}} \sqrt{p_{k,b,l}^{\mathcal{G}}} d_{k,b,l}^{\mathcal{G}} + i_{k,b,l}^{\bar{b}} + \widetilde{n}_{k,b,l}^{\mathcal{G}}, \; 1 \leq l \leq N_r. \qquad (2.25)$$

The SINR of the l^{th} stream for the user k in the BS b of the cluster $\mathcal{G}$ is

$$SINR_{k,b,l}^{\text{CoMP}} = \frac{\left(s_{k,b,l}^{\mathcal{G}} \right)^2 p_{k,b,l}^{\mathcal{G}}}{E\left[\sum_{\bar{b} \notin \mathcal{G}} i_{k,b,l}^{\bar{b}} \left(i_{k,b,l}^{\bar{b}} \right)^* \right] + N_0} \qquad (2.26)$$

and the capacity is

$$C_{k,b}^{\text{CoMP}} = \sum_{l=1}^{N_r} \log_2 \left(1 + SINR_{k,b,l}^{CoMP} \right). \qquad (2.27)$$

In BS cooperation scenario, the signal from different BSs will experience different delay. Therefore, synchronization is an important topic. Nevertheless, to simplify our system, we assume the synchronization is always perfect.

2.2 Clustering Algorithms

2.2.1 Static Clustering

Static clustering is a feasible way to form the clusters in the cellular network. It can be designed offline based on geographic or averaged channel characteristics. Once the planning is

determined, it will not change over time. The advantage of this clustering type is that routing CSI and user data to a central coordinator (CC) is unnecessary. Instead, it requires a distributed coordinator (DC) per cluster which controls the BSs. The cooperation only takes place in each cluster and different clusters do not communicate with each other, which reduces the overheads. Here we propose a static clustering algorithm as follow. In each cell, we assume there is only one scheduled user, so the index for user can be ignored and the new notation $\mathbf{H}_b^b$ stands for the channel from the BS b to its user.

Algorithm Static Clustering Algorithm.

1: Specify the CoMP cluster size B;

2: Each user measures channel gains and calculates his pilot SINR by

$$SINR_b^{\text{pilot}} = \left\|\mathbf{H}_b^b\right\|^2 / \left(\sum_{\bar{b}\neq b, \bar{b}\in\mathcal{I}_b} \left\|\mathbf{H}_b^{\bar{b}}\right\|^2 + N_0\right), \forall b, \tag{2.28}$$

where $\mathcal{I}_b$ is set of the six first tier interfering BSs around BS b, i.e., only the pilots from the neighboring BSs are regarded as the valid interference. If $SINR_b^{\text{pilot}} < \gamma$, where γ is the threshold, the user requests CoMP service to the BS b through uplink. Then BS b sends the request to its DC;

3: DC finds the remaining $B-1$ BSs which should cooperate with the BS b based on the pre-defined clustering table (see Figure 2.2 for the case of $B = 3$), and makes them to form a cluster;

4: Go back to step3 until all the CoMP needed users are satisfied;

The clusters are formed by neighboring BSs here. Since in average, they cause stronger interference compared with those farther BSs. Although static clustering reduces the inter-cluster communication overhead, it inherits the fairness problem from single cell scenario that the users located at static cluster-edge still suffer from severe inter-cluster interference.

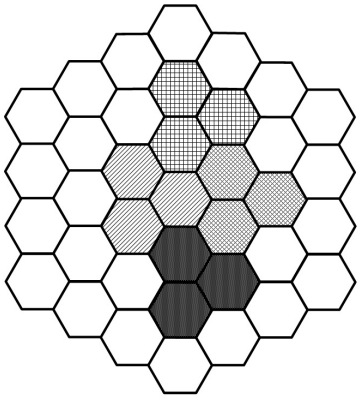

Figure 2.2: Proposed static clustering table when cluster size $B = 3$.

2.2.2 Dynamic Clustering

Static clustering has very limited performance gain since the variation of the channel condition is not fully exploited. As mentioned in the previous section, we select the neighboring BSs to form static clusters. Since on average, they are the ones which cause strong interference. Nevertheless, due to the effect of shadowing, a user might experience a better channel to a farther BS, in other words, interference does not always come from near BSs. Therefore, it is not flexible to form fixed clusters by grouping BSs which are close to each other. Besides, users at the edge of the static cluster experience much more interference from the neighboring clusters than the ones located around the center of the cluster, which causes fairness problem. In order to overcome the aforementioned problems, the idea of dynamic clustering has been introduced. Geographical relation is not the main concern anymore. Instead, we try to group the BSs which cause the strictest interference before any cooperation. The proposed greedy algorithm is illustrated in the following steps. Again, only one user per BS is assumed.

Algorithm Dynamic Clustering Algorithm.

1: Specify the CoMP cluster size B;

2: Each user calculates his $SINR_b^{\text{pilot}}$ as defined in (2.28). If $SINR_b^{\text{pilot}} < \gamma$, the user sends CoMP request to the serving BS b;

3: The CC collects all the requests and chooses a CoMP needed user who has not been chosen so far uniformly;

4: Find the remaining $B-1$ BSs which maximize the utility function $J\left(C_1^{\text{CoMP}}, \ldots, C_B^{\text{CoMP}}\right)$ with the user chosen in step 3, where C_b^{CoMP} is the capacity given by (2.27). We let only the first tier BSs around the selected user to be the candidates, i.e., $b \in \mathcal{I}_b$. If the available BSs is less than $B-1$, the user cannot acquire CoMP service at this time slot, then CC drops this user and picks another one uniformly;

5: Go back to step 3 until all the CoMP needed users find their partners;

We provide three choices of the utility function in this dynamic algorithm:

- Sum-rate (SR) utility:

$$J_1 = \frac{1}{B}\sum_{b=1}^{B} C_b^{\text{CoMP}} \tag{2.29}$$

- Proportional fair (PF) utility:

$$J_2 = \left(\prod_{b=1}^{B} C_b^{\text{CoMP}}\right)^{1/B} \tag{2.30}$$

- Weighted sum (WS) utility:

$$J_3 = \sum_{b=1}^{B} w_b C_b^{\text{CoMP}} \tag{2.31}$$

The purpose of the weight is to find the users who really need CoMP, i.e., the users with low SINR. So the weight is set to the reciprocal of the SINR, which is

$$w_b = c/q_b, \tag{2.32}$$

where

$$q_b = \log_2\left(1 + SINR_b^{\text{pilot}}\right) \tag{2.33}$$

is the transformed SINR which approximates capacity and

$$c = \frac{1}{\sum_{b=1}^{B} q_b^{-1}} \tag{2.34}$$

is a normalization factor such that $\sum_{b=1}^{B} w_b = 1$. Note that $SINR_b^{\text{pilot}}$ is the SINR experienced by the users before CoMP. The user with larger $SINR_b^{\text{pilot}}$ will get lower weight, since the performance gain is little when apply CoMP to the users with high SINR.

Figure 2.3 shows a snapshot of the dynamic clustering result. No matter which utility function is chosen, the CoMP clusters can be formed adaptively according to the change of channel conditions. Hence there are no constant cluster edges and therefore no users will suffer from more interference. However, there must exits a CC to run the dynamic clustering algorithm. Besides, the overhead of routing the CSI and user data is higher than the static scheme.

Since the available BSs for selection are reducing through the dynamic algorithm, it benefits the user that is chosen earlier in Step 3. To circumvent this fairness issue, we have to choose the user uniformly. Therefore, on average, everyone obtains close performance gain.

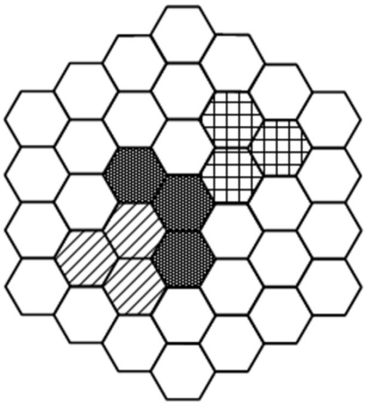

Figure 2.3: A snapshot of the dynamic clustering.

2.3 Simulation Results

We consider a downlink network consists of thirty-seven cells overall ($N_b = 37$), i.e., the first three tiers of cells, each cell has one BS at its center. Every BS has two omnidirectional antennas ($N_t = 2$) and each user has two antennas ($N_r = 2$). The cell radius is set to 1 Km. The MIMO channel from the BS $\bar{b}$ to the user k served by the BS b is

$$\mathbf{H}_{k,b}^{\bar{b}} = \mathbf{R}_{k,b}^{\bar{b}} / \sqrt{\beta \left(d_{k,b}^{\bar{b}} \right)^{\alpha} s_{k,b}^{\bar{b}}}, \tag{2.35}$$

where $\mathbf{R}_{k,b}^{\bar{b}}$ is the $N_r \times N_t$ Rayleigh fading channel, in which the elements are all i.i.d. complex Gaussian random variables, $s_{k,b}^{\bar{b}}$ is log-normal distributed with 8 dB standard deviation which models the shadowing effect and $d_{k,b}^{\bar{b}}$ is the corresponding distance in Km. For the pathloss, the 3GPP LTE pathloss model [10] is used, where $\beta = 10^{14.81}$ and $\alpha = 3.76$. The noise power spectral density is -174 dBm/Hz. Suppose that all of the BSs transmit on the same subcarrier with 15 KHz subcarrier spacing, so the noise power is $N_0 \cong -162.2391$ dBW. One user is generated uniformly in each cell ($K_b = 1, \forall b$) to simulate the round-robin scheduling, and we assume the user has not been handed off to another cell, i.e., the user generated in the cell b is served by the BS b, although the strongest signal may come from another BS. The cluster size B is fixed to three unless otherwise stated. We only observe the performance of the user in the central cell in the network, whereas all of the users in the central cell and its first tier BSs (total seven users) are allowed to request CoMP service. For our observed user in the central cell, only the signals from the six first tier BSs are treated as interference. The six BSs could be CoMP or non-CoMP BSs, which depends on the result of our clustering algorithm. If it is a CoMP BS, it deploys multi-cell BD (see section 2.1.2) with its partners. Otherwise, it applies single cell eigenmode precoding introduced in section 2.1.1.

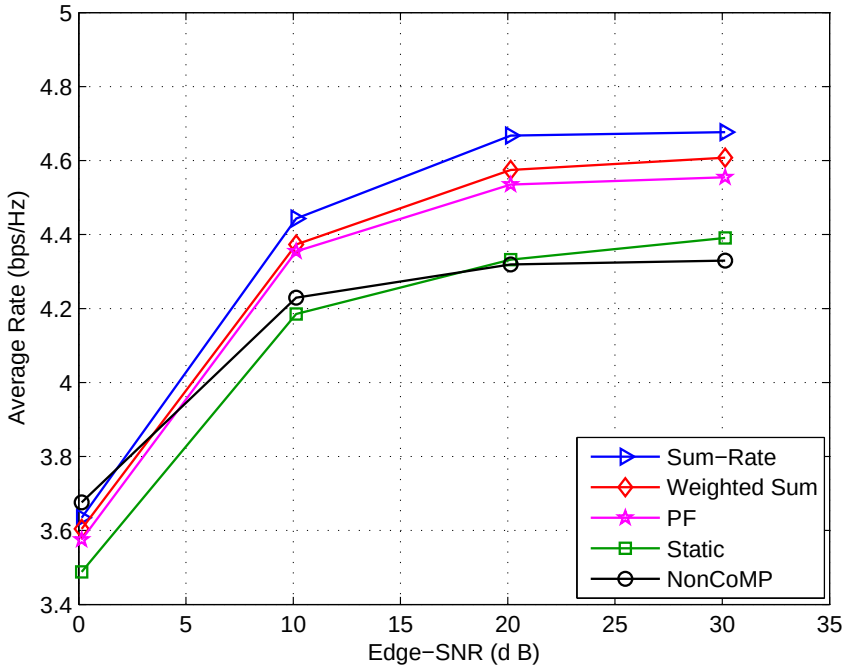

Figure 2.4: Average rate versus the edge-SNR.

Figure 2.4 shows the average rate as a function of the edge-SNR when the pilot SINR threshold $\gamma = -8.3$ dB. The edge-SNR is defined as the SNR measured by the user located at cell-edge without shadowing and interference, which is given by

$$SNR^{\text{edge}}\,(\text{dB}) = p^{\text{con}}\,(\text{dBW}) - 10\log_{10}(\beta) - 10\alpha\log_{10}(r_{\text{cell}}) - N_0\,(\text{dBW})\,, \tag{2.36}$$

where r_{cell} is the cell radius. In (2.36), we count the transmit power from only one BS, but each user receives multiple signal from all the BSs in the same CoMP cluster in our simulation. We can see that all the clustering schemes outperform the non-CoMP one when $SNR^{\text{edge}} > 20$ dB since the interference is reduced. The three dynamic approaches provide better average rate gain against the static clustering because they exploit the instantaneous CSI. Due to the target of the utility function, using sum-rate utility gets the best average rate and weighted sum is better than proportional fair.

Table 2.1: The CoMP request ratio and the actual CoMP included ratio against different pilot SINR threshold.

γ (dB)	Request ratio	SR included ratio	PF included ratio	WS included ratio	Static included ratio
-20.8	2.88%	8.64%	9.12%	8.06%	8.78%
-18.3	4.74%	14.06%	13.58%	13.60%	13.26%
-15.8	6.88%	19.94%	20.74%	19.48%	20.00%
-13.3	10.90%	29.28%	28.96%	27.72%	29.04%
-10.8	15.58%	38.32%	38.62%	37.90%	38.34%
-8.3	20.42%	48.84%	45.52%	47.52%	49.36%
-5.8	27.44%	60.98%	60.84%	59.48%	61.12%
-3.3	36.16%	70.64%	71.98%	68.72%	73.10%

Table 2.1 gives the probability of the CoMP requesting users and the actual CoMP included users with different pilot SINR threshold when $SNR^{\text{edge}} \cong 30$ dB. In both of the static and dynamic algorithms, there are two types of users named the CoMP requesting users and the CoMP included users. The CoMP requesting user is the one whose pilot SINR is lower than the threshold, in other words, is the user who sends CoMP request. The CoMP included users are the CoMP requesting user plus the $B-1$ users who are forced to get CoMP service (see the clustering algorithm in either section 2.2.1 or 2.2.2. For instance, if $B=3$ and the user in the BS 1 sends a CoMP request. The coordinator makes the users in the BS 2 and 3 to be the CoMP partners of the user in the BS 1. Then the user served by the BS 1 is called the CoMP requesting user and all of them are called CoMP included users. This table tells us the percentage of these two types of users with different threshold. It is interesting to notice that the percentage of the

static CoMP included users is higher than the three dynamic ones when the threshold is high. Because when the threshold is getting larger, more and more users request CoMP. As a result, some CoMP requesting users in the dynamic clustering algorithm cannot find sufficient BSs to form a cluster.

Figure 2.5 plots the average rate versus different pilot SINR threshold when $SNR^{\text{edge}} \cong 30$ dB, which corresponds to $p^{\text{con}} = 16$ dBW. As stated in Table 2.1, with the increasing of the threshold, more users can request CoMP and therefore the average rate is getting better with the price of higher network overhead. The average rates saturate when the threshold is high enough (about 15 dB above), since almost 100 % users are included in CoMP areas now. The static scheme obtains relatively low performance gain compared to the three dynamic ones because the variety of channel does not been fully utilized.

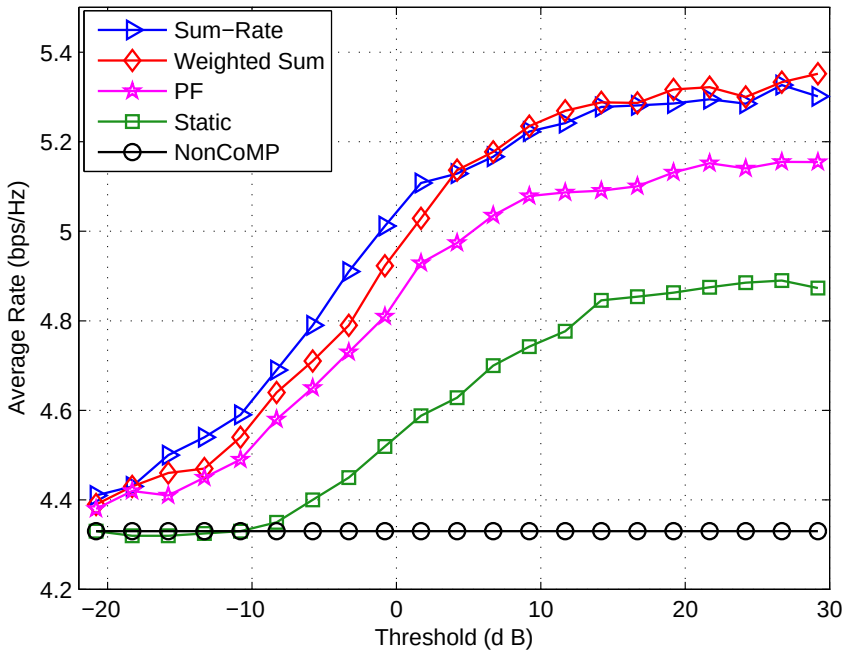

Figure 2.5: The average rate versus different pilot SINR threshold.

In Figure 2.6, the cumulative distribution function (CDF) of the capacity for the CoMP requesting users is plotted. The setup here is $p^{\text{con}} = 16$ dBW and $\gamma = -8.3$dB. The weighted sum dynamic scheme outperforms the others since it makes the users with low SINR to form the CoMP cluster and hence reduces the most interference. The black line with circle marker is the users whose $SINR^{\text{pilot}} < \gamma$ but there is no CoMP service available in the network, which is very poor compared with the other CoMP approaches.

Figure 2.7 is the CDF of the CoMP included users when the setup is the same as the one in Figure 2.6. Seemingly, the sum-rate and the proportional fair schemes are better than the weighted one. However, the reason is that they find the high SINR users to be the partners of the CoMP requesting users. Note that if the SINR is already high, the effect of CoMP is limited since the purpose of BS cooperation is to suppress interference. Therefore it is inefficient to include users with high SINR in the CoMP cluster, even though it seems that the average rate of the CoMP included users is increased.

Figure 2.8 plots the CDF of the last five percent users using the same setup as the one in Figure 2.6. As mentioned above, we only observe the performance of the user in the central cell. However, we can generate different samples of the user locations. For each sample, the user capacity after running our BS clustering algorithms is calculated, and we sort the capacity for all the samples and observe the ones with the lowest 5 % capacity. In general, the users with such low capacity are located near cell-edge. So this figure shows the approximate performance of the cell-edge users. It can be seen that the dynamic clustering scheme improves significant fairness. The sum-rate approach provides less performance gain since it puts resource to the users with good channel quality and sacrifices the cell-edge users.

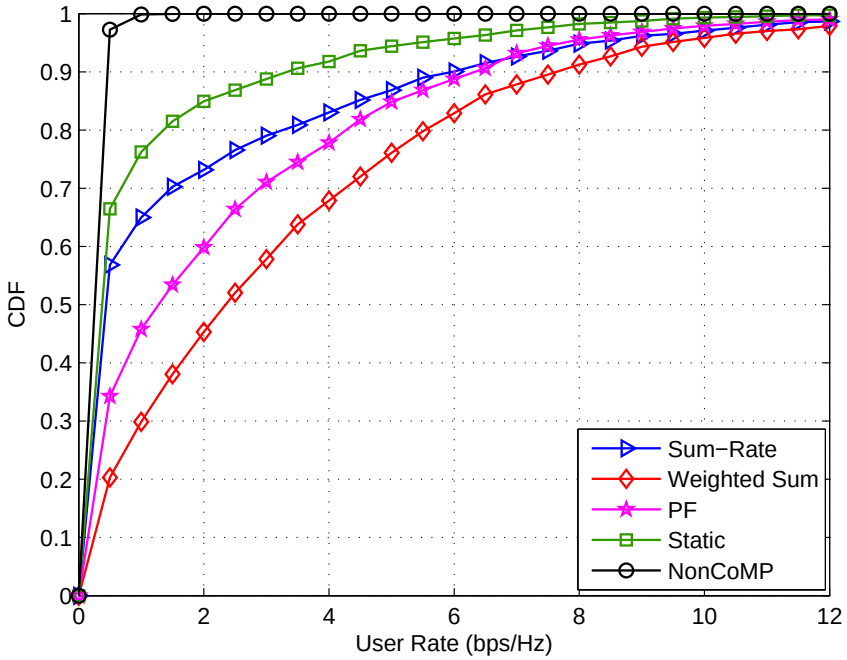

Figure 2.6: The CDF of the CoMP requesting users.

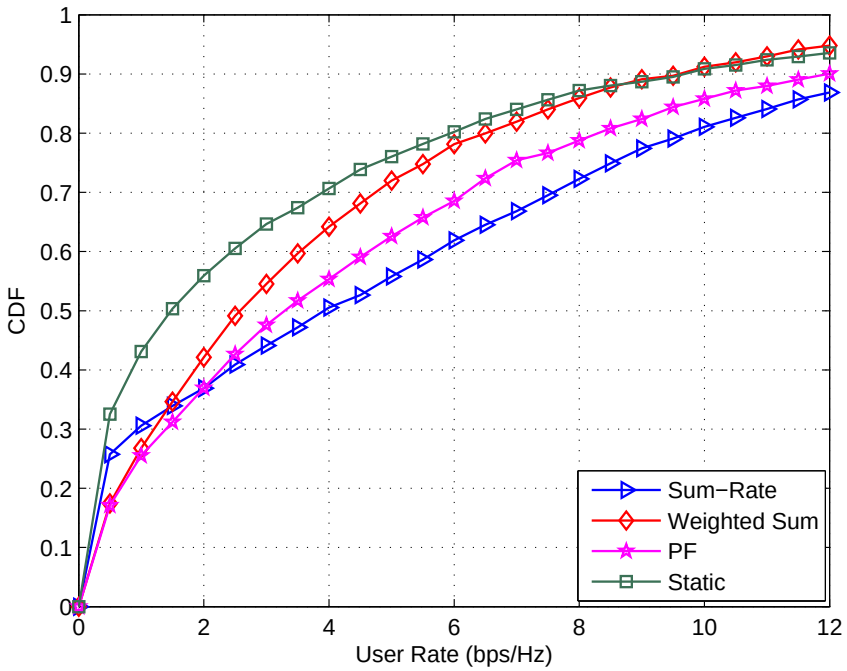

Figure 2.7: The CDF of the CoMP included users.

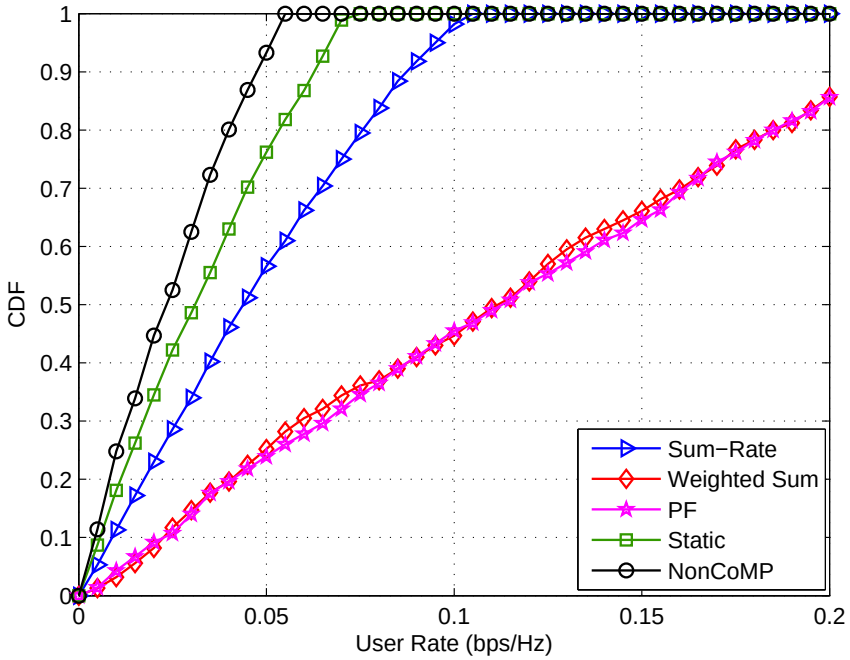

Figure 2.8: The CDF of the last five percent users.

Figure 2.9 illustrates the effect of increasing the CoMP cluster size when the setup is the same as the one in Figure 2.6. As expected, the average rate increases with the cluster size since more interference is eliminated. The drawback is the raise of network overhead.

In Figure 2.10, the CDF of the weighted sum scheme versus different CoMP cluster size is plotted when the setup is the same as the one in Figure 2.6. Only the performance of the CoMP requesting users is observed. In addition to sum rate benefit, increasing the cluster size provides significant gain to the low SINR users (the CoMP requesting users).

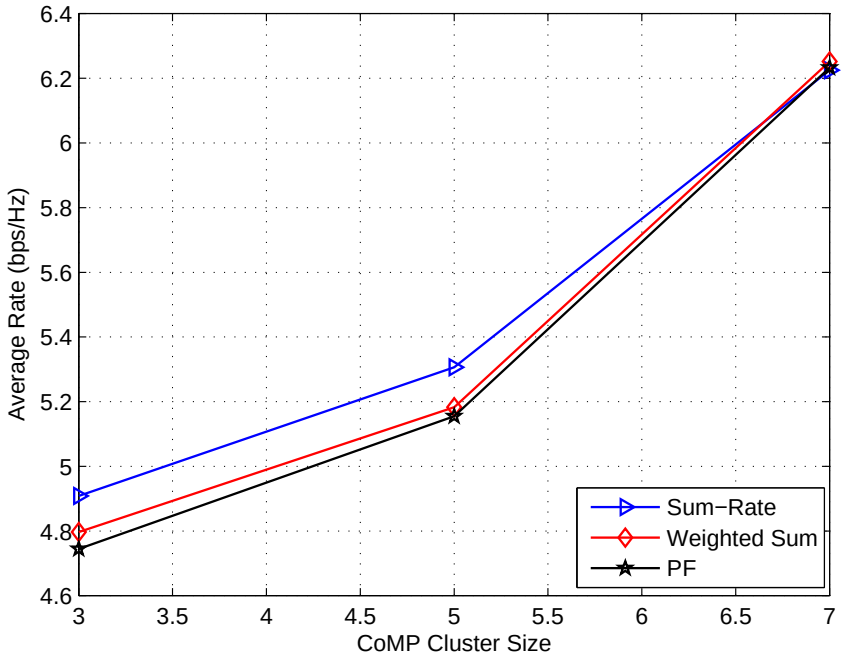

Figure 2.9: The average rate versus different CoMP cluster size.

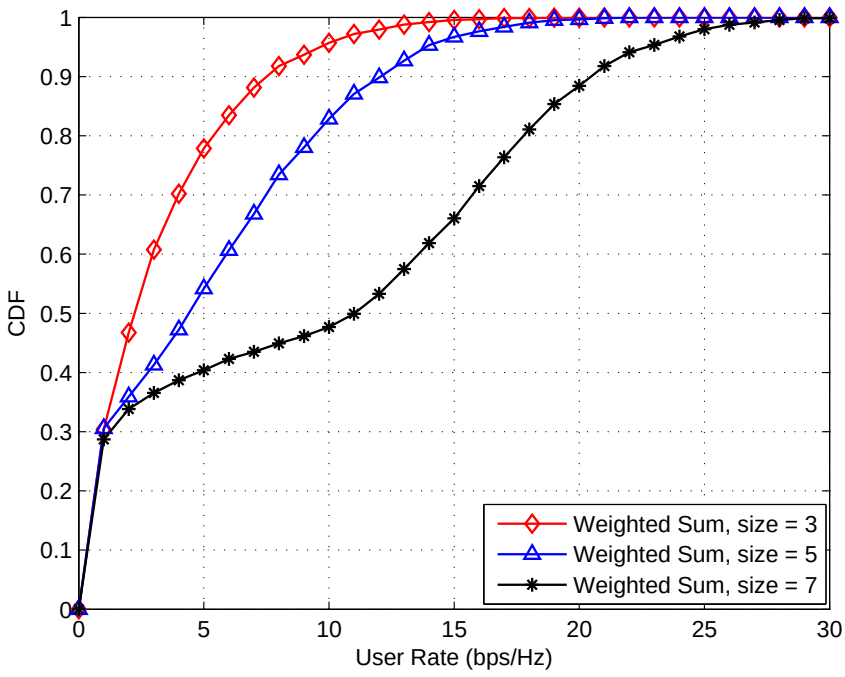

Figure 2.10: The CDF of the CoMP requesting users using weighted sum utility versus different CoMP cluster size.

CHAPTER 3

ADAPTIVE RESOURCE ALLOCATION

3.1 System Model and Transmission Schemes

In Chapter 2, we focus on a specific subcarrier and tackle the problem of BS clustering. In this chapter, we extend the scenario to a more general multi-carrier system. Suppose some CoMP clusters are formed in the cellular network based on a clustering algorithm. For the sake of illustration, only one cluster is taken into consideration. Assume there are B BSs and K users overall in this CoMP cluster, each BS equips N_t antennas and each user has N_r antennas. In CoMP-JT mode, user data and CSI are shared across all the BSs in the same cluster, therefore, a cluster is equivalent to a super BS with $N_T = BN_t$ antennas which serves K users simultaneously. Let there are M available subcarriers, the downlink transmission can be illustrated by Figure 3.1. In general, users are allocated to different subcarriers in order to avoid inter-user interference. However, we can apply BD to decouple the channels of different users on the same subcarrier. In this way, the spectral efficiency will be improved since multiple users can share the same bandwidth.

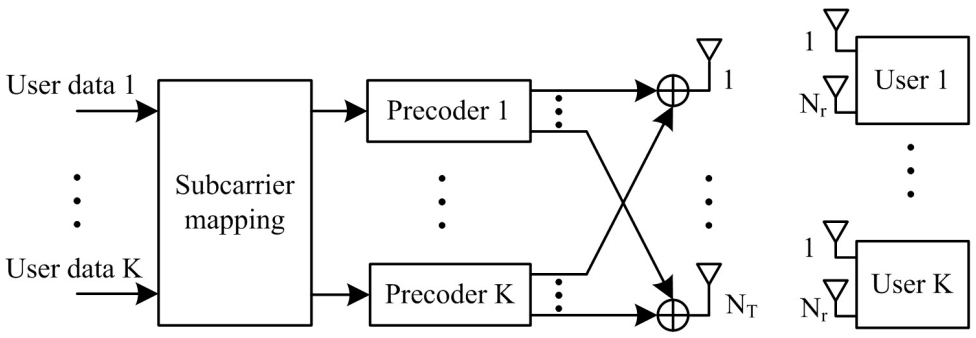

Figure 3.1: Block diagram for downlink multi-user MIMO-OFDM.

Assume we place K_m users on subcarrier m. After removing the cyclic prefix (CP) at the user side, the signal can be processed in a per-subcarrier MIMO fashion. That is, the input-output relation on every subcarrier can be represented in a MIMO structure. Ignoring the outer-cluster interference, the received signal of the user k on the subcarrier m can be represented as

$$\mathbf{y}_{k,m} = \mathbf{H}_{k,m}\mathbf{x}_{k,m} + \sum_{n=1,n\neq k}^{K_m} \mathbf{H}_{n,m}\mathbf{x}_{k,m} + \mathbf{n}_{k,m}, \tag{3.1}$$

where $\mathbf{H}_{k,m}$ is the $N_r \times N_T$ MIMO channel, $\mathbf{x}_{k,m}$ is the $N_T \times 1$ transmitted signal for the user k and $\mathbf{n}_{k,m}$ is the zero-mean complex Gaussian noise vector with covariance matrix $E\left[\mathbf{n}_{k,m}(\mathbf{n}_{k,m})^H\right] = N_0\mathbf{I}_{N_r}$. After BD (refer to section 2.1.1) pre-filtering, the transmitted signal can be represented as

$$\mathbf{x}_{k,m} = \mathbf{F}_{k,m}\mathbf{P}_{k,m}\mathbf{d}_{k,m} \tag{3.2}$$

where $\mathbf{P}_{k,m}$ is the $N_r \times N_r$ power allocation matrix and $\mathbf{d}_{k,m}$ is the $N_r \times 1$ data vector before pre-filtering. The BD matrix is $\mathbf{F}_{k,m} = \overline{\mathbf{V}}_{k,m}\widetilde{\mathbf{V}}_{k,m}$, where $\overline{\mathbf{V}}_{k,m}$ makes the channel of the user k to be orthogonal to the channels of the others, and $\widetilde{\mathbf{V}}_{k,m}$ decouples the orthogonalized MIMO channel of the user k into N_r parallel SISO channels. Following the derivation similar to (2.24) and (2.25), the l^{th} spatial stream for the user k is

$$r_{k,m,l} = s_{k,m,l}\sqrt{p_{k,m,l}}d_{k,m,l} + \widetilde{n}_{k,m,l},\ 1 \leq l \leq N_r, \tag{3.3}$$

where $s_{k,m,l}$ is the l^{th} singular value of the orthogonalized channel matrix of the user k on subcarrier m, $p_{k,m,l}$ is the power allocated on this stream, $d_{k,m,l}$ is the l^{th} element in $\mathbf{d}_{k,m}$ and $\widetilde{n}_{k,m,l}$ is a zero-mean complex Gaussian noise with variance N_0.

3.2 Problem Formulation

The mathematical problem of power minimization is formulated in this section. In order to save power in each CoMP cluster, we try to find a power allocation approach which minimizes

the transmit power while satisfies some constraints. In practice, users may have different QoS requirements such as target data rate and tolerable BER. On the other hand, each antenna has its own power amplifier and therefore has a unique transmit power constraint. Considering all the issues above, the power minimization problem can be formulated as

$$\begin{aligned}
\underset{p_{k,m,l}}{\text{minimize}} \quad & \sum_{k=1}^{K}\sum_{m=1}^{M}\sum_{l=1}^{N_r} p_{k,m,l} \\
\text{subject to} \quad & \sum_{m=1}^{M}\sum_{l=1}^{N_r} r_{k,m,l} \geq M r_k^{\text{tar}},\ 1 \leq k \leq K \\
& \sum_{k=1}^{K}\sum_{m=1}^{M}\sum_{l=1}^{N_r} |\mathbf{F}_{k,m}[a,l]|^2 p_{k,m,l} \leq p_a^{\text{con}},\ 1 \leq a \leq N_T \\
& p_{k,m,l} \geq 0,\ \forall k,\ m,\ l,
\end{aligned} \tag{3.4}$$

where $p_{k,m,l}$ and $r_{k,m,l}$ is the transmit power and rate allocated on the l^{th} stream of the user k on the subcarrier m, r_k^{tar} is the target data rate in bps/Hz, $\mathbf{F}_{k,m}[a,l]$ is the element in the a^{th} row and l^{th} column of the BD matrix $\mathbf{F}_{k,m}$, and p_a^{con} is the power constraint on transmit antenna a. Note that if the bandwidth of each subcarrier is β, the total rate requirement per user of one OFDM symbol is $M\beta r_k^{\text{tar}}$ (bps), therefore, Mr_k^{tar} bits are required during the period of one OFDM symbol. The rate $r_{k,m,l}$ can be written as

$$r_{k,m,l} = \log_2\left(1 + \frac{p_{k,m,l}s_{k,m,l}^2}{\tau N_0}\right), \tag{3.5}$$

where τ is the SNR gap given by

$$\tau = -\frac{\ln(5BER^{\text{tar}})}{1.5}, \tag{3.6}$$

where BER^{tar} is the target BER specified by the users. Although the capacity is hard to achieve in practice, it provides an upper bound that tells us how well we can do. The problem (3.4) implies a user selection problem on each subcarrier. If $p_{k,m} = \sum_{l=1}^{N_r} p_{k,m,l} = 0$, the user k is absent on the subcarrier m; otherwise, the user k is allocated on this subcarrier. Problem (3.4) is a

modified version of the one in [12], which considers no power constraint. However, as the user's target rate increases, the minimized power may exceed the transmit power limit. Therefore, we want to see how the results will be when the per-antenna power constraints are added to the power minimization problem.

3.3 Low Complexity Solution for Power Minimization

3.3.1 Optimization Based on Dual Decomposition

In this section, we propose a low complexity solution to (3.4) based on the Lagrange dual transformation. For the sake of interpretation, we rewrite (3.4) in a new form:

$$\begin{aligned} \underset{\mathbf{r}}{\text{minimize}} \quad & f(\mathbf{r}) \\ \text{subject to} \quad & \sum_{m=1}^{M} \mathbf{r}_m \succeq M\mathbf{r}^{\text{tar}} \\ & g(\mathbf{r}) \preceq \mathbf{p}^{\text{con}}, \end{aligned} \tag{3.7}$$

where $\mathbf{r} = \left[\mathbf{r}_1^T, \ldots, \mathbf{r}_m^T, \ldots, \mathbf{r}_M^T\right]^T$, with $\mathbf{r}_m = [r_{1,m}, \ldots, r_{k,m}, \ldots r_{K,m}]^T$, with $r_{k,m} = \sum_{l=1}^{N_r} r_{k,m,l}$ are the allocated rates, $\mathbf{r}^{\text{tar}} = [r_1^{\text{tar}}, \ldots, r_K^{\text{tar}}]^T$ are the target rates, $\mathbf{p}^{\text{con}} = \left[p_1^{\text{con}}, \ldots, p_{N_T}^{\text{con}}\right]^T$ are the per-antenna power constraints, $f(\cdot)$ and $g(\cdot)$ are the $\mathbb{R}^{MK} \rightarrow \mathbb{R}$ and $\mathbb{R}^{MK} \rightarrow \mathbb{R}^{N_T}$ mapping functions, respectively, and $\mathbf{a} \succeq \mathbf{b}$ means $a_i \geq b_i,\ \forall i$. The constraints $p_{k,m,l} \geq 0,\ \forall k, m, l$ in (3.4) are removed temporarily and will be considered afterwards (in Appendix A.1). Although the original objective function $f(\cdot)$ is not convex, it can be transformed into a dual function, which is always concave regardless of the convexity of $f(\cdot)$. Hence traditional convex optimization techniques can be used to solve the transformed problem. We start from the Lagrangian of (3.7), which is

$$\mathcal{L}(\mathbf{r}, \boldsymbol{\mu}, \boldsymbol{\kappa}) = f(\mathbf{r}) + \boldsymbol{\mu}^T \left(M\mathbf{r}^{\text{tar}} - \sum_{m=1}^{M} \mathbf{r}_m \right) + \boldsymbol{\kappa}^T \left(g(\mathbf{r}) - \mathbf{p}^{\text{con}} \right), \tag{3.8}$$

where $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_K]^T$ and $\boldsymbol{\kappa} = [\kappa_1, \ldots, \kappa_{N_T}]^T$ are the vectors of Lagrange multipliers correspond to the rate and power constraints in (3.7). The dual function is defined as

$$d(\boldsymbol{\mu}, \boldsymbol{\kappa}) = \mathcal{L}(\mathbf{r}^*, \boldsymbol{\mu}, \boldsymbol{\kappa}), \tag{3.9}$$

where $\mathbf{r}^* = \min_{\mathbf{r}} \mathcal{L}(\mathbf{r}, \boldsymbol{\mu}, \boldsymbol{\kappa})$. The dual problem can be formulated as

$$\begin{aligned} \underset{\boldsymbol{\mu},\boldsymbol{\kappa}}{\text{maximize}} \quad & d(\boldsymbol{\mu}, \boldsymbol{\kappa}) \\ \text{subject to} \quad & \boldsymbol{\mu} \succeq \mathbf{0} \\ & \boldsymbol{\kappa} \succeq \mathbf{0}. \end{aligned} \tag{3.10}$$

In another word, we let the Lagrange multipliers to be constants temporarily and find the $\mathbf{r}^*$ which minimizes the Lagrangian $\mathcal{L}(\mathbf{r}, \boldsymbol{\mu}, \boldsymbol{\kappa})$, this is the definition of the dual function $d(\boldsymbol{\mu}, \boldsymbol{\kappa})$. Next, we formulate the dual problem which aims to find the optimal Lagrange multipliers that maximize the dual function. In contrast with the *dual problem*, the original problem (3.7) is called the *primal problem*. The dual problem is equivalent to the primal problem if the original objective function $f(\cdot)$ is convex, otherwise, there exists a *duality gap* between these two problems [18]. In our case, $f(\cdot)$ is not convex since it is a pointwise minimum of several convex functions. Nevertheless, [19] shows that this gap can be reduced by increasing the subcarrier size M. In order to find the $\mathbf{r}^*$ in (3.9), we first express (3.8) in another form:

$$\begin{aligned} \widetilde{\mathcal{L}} = & \sum_{k=1}^{K}\sum_{m=1}^{M}\sum_{l=1}^{N_r} p_{k,m,l} + \sum_{k=1}^{K} \mu_k \left(M r_k^{\text{tar}} - \sum_{m=1}^{M}\sum_{l=1}^{N_r} r_{k,m,l} \right) \\ & + \sum_{a=1}^{N_T} \kappa_a \left(\sum_{k=1}^{K}\sum_{m=1}^{M}\sum_{l=1}^{N_r} |\mathbf{F}_{k,m}[a,l]|^2 p_{k,m,l} - p_a^{\text{con}} \right). \end{aligned} \tag{3.11}$$

Since data rate is a function of power, therefore, finding the $\mathbf{r}^*$ which minimizes $\mathcal{L}$ is equivalent to finding $p^*_{k,m,l},\ \forall k, m, l$ which minimize $\widetilde{\mathcal{L}}$, so we set the latter to be our new goal. After making some arrangements to (3.11), it becomes

$$\widetilde{\mathcal{L}} = \sum_{m=1}^{M}\sum_{k=1}^{K}\sum_{l=1}^{N_r} \left(\overline{\mathcal{L}}(m,k,l) \right) + \sum_{k=1}^{K} \mu_k M r_k^{\text{tar}} - \sum_{a=1}^{N_T} \kappa_a p_a^{\text{con}}, \tag{3.12}$$

where

$$\overline{\mathcal{L}}(m,k,l) = p_{k,m,l} - \mu_k r_{k,m,l} + \sum_{a=1}^{N_T} \kappa_a |\mathbf{F}_{k,m}[a,l]|^2 p_{k,m,l}. \tag{3.13}$$

In the dual function, μ_k and κ_a are treated as constants temporarily, so the last two terms in (3.12) are unrelated terms. Therefore, minimizing $\widetilde{\mathcal{L}}$ is equivalent to minimizing $\overline{\mathcal{L}}$. On each subcarrier, when the user selection has been determined, we can obtain the BD precoding matrices $\mathbf{F}_{k.m},\ \forall k,m$, the minimal power and rate allocated on each spatial stream can be given by

$$p_{k,m,l} = \max\left\{ \frac{\mu_k}{\ln(2)} \frac{1}{1 + \sum_{a=1}^{N_T} \kappa_a |\mathbf{F}_{k,m}[a,l]|^2} - \frac{\tau N_0}{s_{k,m,l}^2}, 0 \right\} \tag{3.14}$$

and

$$r_{k,m,l} = \log_2 \left(\max\left\{ \frac{\mu_k s_{k,m,l}^2}{\ln(2)\,\tau N_0 \left(1 + \sum_{a=1}^{N_T} \kappa_a |\mathbf{F}_{k,m}[a,l]|^2\right)}, 1 \right\} \right). \tag{3.15}$$

For brevity, the derivations of (3.14) and (3.15) are left in Appendix A.1. We call this solution the *competitive water-filling* solution. The reason for the name will be explained later.

Since BD can mitigate the inter-user interference on the same subcarrier, multiple users can share the same bandwidth. The optimal user selection on each subcarrier would be to search over 2^K user combinations and find the one that minimizes

$$\widehat{\mathcal{L}}(m) = \sum_{k=1}^{K} \sum_{l=1}^{N_r} \overline{\mathcal{L}}(m,k,l). \tag{3.16}$$

For the overall M subcarriers, there would be $M2^K$ choices. As mentioned above, the duality gap approaches zero when M goes to infinity. However, this will make user selection problem to be computational infeasible. The complexity could be reduced by a suboptimal greedy user selection introduced in [12]: For each subcarrier, allocate the user that minimizes $\widehat{\mathcal{L}}(m)$ on subcarrier m. Next, add another one from the remaining $K-1$ users if $\widehat{\mathcal{L}}(m)$ can be further reduced, and so on. Note that if $\widehat{\mathcal{L}}(m) \geq 0$, there is no user allocated on this subcarrier, since

positive $\widehat{\mathcal{L}}(m)$ will not minimize $\widetilde{\mathcal{L}}$. As the number of users on this subcarrier increases, the BD precoder will project each user's channel to a more restricted space (see section 2.1.1), which makes the channels weak. Hence, it is not always the best to put all the users on each subcarrier, even though they do not interfere to each other after BD. The suitable number of users that allocated on each subcarrier can be found by the greedy user selection algorithm above. In this way, the maximum combination of users over the total M subcarriers becomes $M\sum_{j=0}^{K-1}\binom{K-j}{1} = \frac{MK(K+1)}{2}$, which is small compared to $M2^K$ when K is large. As for the globally optimal solution, even if the per-antenna power constraints are ignored and the minimal power which satisfies the user rate constraint is obtained by the water-filling solutions (which are (3.14) and (3.15) after setting $\kappa_a = 0,\ \forall a$), it still needs a search over 2^{KM} possibilities to find the optimal solution, which is computationally prohibitive.

So far, we have found the $p^*_{k,m,l},\ \forall k, m, l$ that minimize $\widetilde{\mathcal{L}}$ and therefore the dual function $d(\boldsymbol{\mu}, \boldsymbol{\kappa})$ is obtained. Next, we need to find the optimal $\boldsymbol{\mu}$ and $\boldsymbol{\kappa}$ that maximize $d(\boldsymbol{\mu}, \boldsymbol{\kappa})$. Since $d(\boldsymbol{\mu}, \boldsymbol{\kappa})$ is concave, we can update $\boldsymbol{\mu}$ and $\boldsymbol{\kappa}$ along some directions to find the optimal point. We adopt a special searching direction named supergradient [20]. In general, the supergradient at a point $\boldsymbol{\alpha} \in \mathbb{R}^{n\times 1}$ is defined as a vector $\boldsymbol{\chi} \in \mathbb{R}^{n\times 1}$ which satisfies

$$d(\widetilde{\boldsymbol{\alpha}}) \le d(\boldsymbol{\alpha}) + \boldsymbol{\chi}^T(\widetilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}),\ \forall \widetilde{\boldsymbol{\alpha}} \ne \boldsymbol{\alpha}. \tag{3.17}$$

In our optimization problem (3.7), $\boldsymbol{\alpha}$ comprises the Lagrange multiplier vectors $\boldsymbol{\mu}$ and $\boldsymbol{\kappa}$, and $\boldsymbol{\chi}$ can be decomposed into two directions $\boldsymbol{\chi}_1$ and $\boldsymbol{\chi}_2$, which are given by

$$\boldsymbol{\chi}_1 = M\mathbf{r}^{\text{tar}} - \sum_{m=1}^{M}\mathbf{r}^*_m \tag{3.18}$$

and

$$\boldsymbol{\chi}_2 = g(\mathbf{r}^*) - \mathbf{p}^{\text{con}}, \tag{3.19}$$

where $g(\cdot)$ is the function defined in (3.7). The proofs are shown in Appendix A.2. Without

changing the direction, we divide the first supergradient by M, that is,

$$\overline{\boldsymbol{\chi}}_1 = \boldsymbol{\chi}_1/M = \left[\overline{\chi}_{1,1}, \ldots, \overline{\chi}_{1,k}, \ldots, \overline{\chi}_{1,K}\right]^T, \tag{3.20}$$

where

$$\overline{\chi}_{1,k} = r_k^{\text{tar}} - \frac{1}{M}\sum_{m=1}^{M}\sum_{l=1}^{N_r} r_{k,m,l}. \tag{3.21}$$

The second supergradient remains the same, which is

$$\boldsymbol{\chi}_2 = \left[\chi_{2,1}, \ldots, \chi_{2,a}, \ldots \chi_{2,N_T}\right]^T \tag{3.22}$$

where

$$\chi_{2,a} = \sum_{k=1}^{K}\sum_{m=1}^{M}\sum_{l=1}^{N_r} \left|\mathbf{F}_{k,m}\left[a,l\right]\right|^2 p_{k,m,l} - p_a^{\text{con}}. \tag{3.23}$$

We update the two Lagrange multiplier vectors in an iteration manner:

$$\mu_k^{i+1} = \max\left\{\mu_k^i + \delta_1^i \overline{\chi}_{1,k}, 0\right\} \tag{3.24}$$

and

$$\kappa_a^{i+1} = \max\left\{\kappa_a^i + \delta_2^i \chi_{2,a}, 0\right\}, \tag{3.25}$$

where i is the iteration index, δ_1^i and δ_2^i are the two positive step sizes for $\boldsymbol{\mu}$ and $\boldsymbol{\kappa}$, respectively. As we can see in (3.21), if the allocated rate to the user k exceeds the its target, the direction becomes negative and μ_k will be reduced in the next iteration. On the contrary, μ_k will increase if it falls below the target rate. The similar actions can be observed in (3.23). Since the rate (3.15) is directly proportional to μ_k but inversely proportional to ϕ_a, consider a case that a user requests so much rate that the allocated power goes beyond the per-antenna power constraints, then $\boldsymbol{\chi}_2$ becomes positive and hence increases $\boldsymbol{\kappa}$, as a consequence, the power will be dropped. However, it will be raised again since the target rates are not satisfied due to insufficient power. This causes a struggle situation in our iterative algorithm and that is why the allocation scheme is named by competitive water-filling solution. On the other hand, if the required rates are not that

much or the power constraints are set to very high, the allocated rate to each user will gradually converge to their targets from an initial point without struggling.

3.3.2 Convergence Behavior Control

In this section, we design the initial point and step size for a faster convergence. Although the concavity of the dual function promises that the power and rate will converge along the supergradient, we can boost the speed of convergence by finding the proper initial values. The choice of the step size also has great impacts. A small step size lengthens the convergence time, while a large step size leads to coarse convergence result. The main design principle in this section is based on [12]. In order to give the initial values of $\boldsymbol{\mu}$ and $\boldsymbol{\kappa}$, we let each subcarrier is occupied by all the K users and BD is deployed to separate the users. Even though this is not the optimal way, it gives a good starting point to our algorithm. Since the user selection has been fixed now, we can apply the competitive water-filling solution (3.14) and (3.15), where the initial values of $\kappa_a^0,\ \forall a$ are generated uniformly on an open interval (0, 1), and the initial values of $\mu_k^0,\ \forall k$ are chosen to the levels such that

$$\sum_{m=1}^{M}\sum_{l=1}^{N_r} r_{k,m,l} = M r_k^{\text{tar}},\ \forall k. \tag{3.26}$$

The initial step sizes for updating $\boldsymbol{\mu}$ and $\boldsymbol{\kappa}$ can be designed as

$$\delta_1^0 = \eta_1 \frac{\sum_{k=1}^{K} \mu_k^0}{\sum_{k=1}^{K} r_k^{\text{tar}}} \tag{3.27}$$

and

$$\delta_2^0 = \eta_1 \frac{\sum_{a=1}^{N_T} \kappa_a^0}{\sum_{a=1}^{N_T} p_a^{\text{con}}}, \tag{3.28}$$

where $\eta_1 > 0$ is a constant. There are two behavior when the algorithm is running, the first is that $\boldsymbol{\mu}$ and $\boldsymbol{\kappa}$ are changing in one direction, which can be expressed as

$$sign\left(\chi_{1,k}^{i}\right) == sign\left(\chi_{1,k}^{i-1}\right), \ \forall k \tag{3.29}$$

and

$$sign\left(\chi_{2,a}^{i}\right) == sign\left(\chi_{2,a}^{i-1}\right), \ \forall a, \tag{3.30}$$

where "==" is the equality judgment. Since the initial values may be far from the optimal results, this situation tells us the $\boldsymbol{\mu}$ and $\boldsymbol{\kappa}$ are approaching the optimal values. Therefore, we can increase the step sizes to boost the convergence by adjusting

$$\delta_1^{i+1} = \eta_2 \delta_1^i \tag{3.31}$$

and

$$\delta_2^{i+1} = \eta_2 \delta_2^i, \tag{3.32}$$

where $\eta_2 > 1$ is a constant. On the other hand, if $\boldsymbol{\mu}$ and $\boldsymbol{\kappa}$ are oscillating, which means that they are already close to the optimal values. The oscillation behavior can be represented as

$$\exists k \text{ such that } \left[sign\left(\chi_{1,k}^{i}\right) \neq sign\left(\chi_{1,k}^{i-1}\right)\right] \cap \left[sign\left(\chi_{1,k}^{i-1}\right) \neq sign\left(\chi_{1,k}^{i-2}\right)\right], \tag{3.33}$$

and the second direction is

$$\exists a \text{ such that } \left[sign\left(\chi_{2,a}^{i}\right) \neq sign\left(\chi_{2,a}^{i-1}\right)\right] \cap \left[sign\left(\chi_{2,a}^{i-1}\right) \neq sign\left(\chi_{2,a}^{i-2}\right)\right]. \tag{3.34}$$

In this case, we can stabilize them by changing

$$\delta_1^{i+1} = \delta_1^i / \eta_3 \tag{3.35}$$

and

$$\delta_2^{i+1} = \delta_2^i / \eta_3, \tag{3.36}$$

where $\eta_3 > 1$ is a constant. If the condition does not belong to anyone of these two, then δ_1^i and δ_2^i remain the same. Besides, in order to prevent the step size from going unboundedly. We set the upper and lower bounds for the step size, which are

$$\delta_1^{(\max)} = \eta^{(\max)}\delta_1^0,\ \delta_2^{(\max)} = \eta^{(\max)}\delta_2^0 \tag{3.37}$$

and

$$\delta_1^{(\min)} = \eta^{(\min)}\delta_1^0,\ \delta_2^{(\min)} = \eta^{(\min)}\delta_2^0. \tag{3.38}$$

where $\eta^{(\max)} > 1$ and $0 < \eta^{(\min)} < 1$ are constants.

It should be noted that although the behavior of $\boldsymbol{\mu}$ and $\boldsymbol{\kappa}$ have some relation, but they are not fully identical. In other words, if $\boldsymbol{\mu}$ is moving in one direction, it is possible that $\boldsymbol{\kappa}$ is oscillating. In short, our resource allocation algorithm can be illustrated by Figure 3.2.

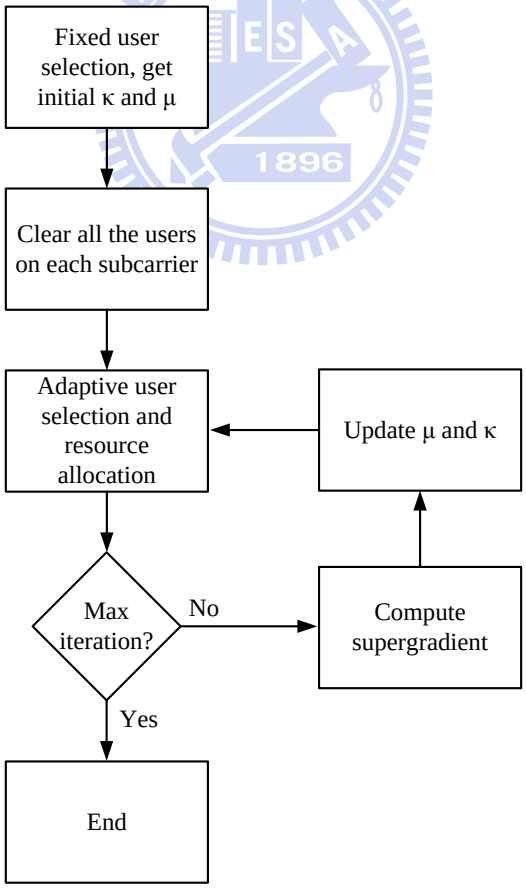

Figure 3.2: Flowchart of our resource allocation algorithm.

3.4 Simulation Results

In the simulation of this chapter, we consider a downlink CoMP cluster consists of three cells ($B = 3$), each cell is divided into three sectors and the BS cooperation takes place in the neighboring sectors. One user is dropped uniformly in each sector ($K = 3$). This scenario is shown in Figure 3.3. Each BS and user both have two antennas ($N_T = 6,\ N_r = 2$), and the target rate is set to $r_k^{\text{tar}} = 3$ bps/Hz, $\forall k$, unless otherwise stated. We let the SNR gap be 3 dB, which corresponds to the tolerable bit error rate $BER^{\text{tar}} \cong 10^{-2}$. There are sixty-four subcarriers in our system ($M = 64$) with 15 KHz bandwidth each. The noise power spectral density is -174 dBm/Hz, so the noise power within one subcarrier is $N_0 \cong -162.2391$ dBW. The path loss model is the same as the one in section 2.3. In order to show the effect of power allocation, we apply a seven taps channel with the power delay profile defined in the extended pedestrian-A ITU channel model [21], which is $[0, -1, -2, -3, -8, -17.2, -20.8]$ dB. Multiple users are allowed to occupy the same subcarrier, the number of user on each subcarrier is determined by the greedy user selection algorithm introduced in section 3.3.1. Based on [12], the parameters for the step size are set to $\eta_1 = 0.1,\ \eta_2 = 1.1,\ \eta_3 = 2,\ \eta^{(\max)} = 5$ and $\eta^{(\min)} = 0.1$.

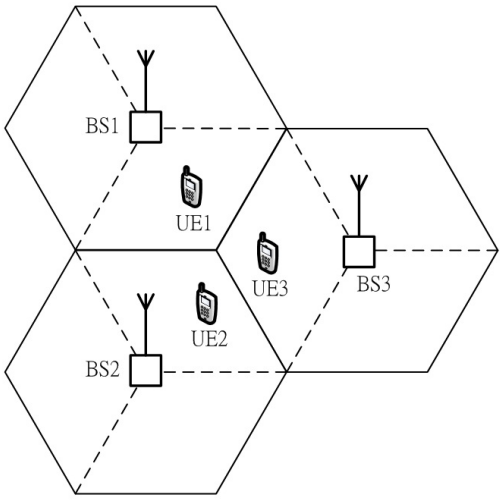

Figure 3.3: Illustration for three-sectorized CoMP.

Assume the power constraint is equivalent among all the transmit antennas. We first show the result of adaptive power allocation, for avoiding the competitive condition, we set the per-antenna power constraints to an extremely high value, which is $(N_T p_a^{\text{con}}) / (MN_0) = 180$ dB, this corresponds to $p_a^{\text{con}} \cong 637$ W, $\forall a$.

Figure 3.4 plots the edge-SNR versus different target rates. The edge-SNR here is defined as

$$SNR^{\text{edge}}\,(\text{dB}) = P_T\,(\text{dBW}) - 10\log_{10}(\beta) - 10\alpha\log_{10}(r_{\text{cell}}) - MN_0\,(\text{dBW}), \qquad (3.39)$$

where $\beta = 10^{14.81}$ is the pathloss constant, $\alpha = 3.76$ is the pathloss exponent and $r_{\text{cell}} = 1$ Km is the cell radius, note that in (3.39), $P_T = \sum_{a=1}^{N_T} P_a^t$, where P_a^t is the actual transmit power from antenna a after power allocation. We can see that the required edge-SNR increases with the target rate as expected. The "Fixed Alloc" is the initialization of our algorithm that each subcarrier is occupied by the maximum $K = 3$ users. Our adaptive allocation method outperforms the fixed one since it exploits the frequency-selectivity of the multipath channel and the degree of freedom of multiple antennas.

Figure 3.5 shows the edge-SNR as a function of the number of antennas when $p_a^{\text{con}} \cong 637$ W. In this figure, the number of antenna changes simultaneously at both the BS and user side. For example, one antenna means $N_r = N_t = 1$ so it is a 3×3 MU-MIMO scheme and so on. It is clear that multiple antennas can reduce the requirement of the edge-SNR under a given target rate.

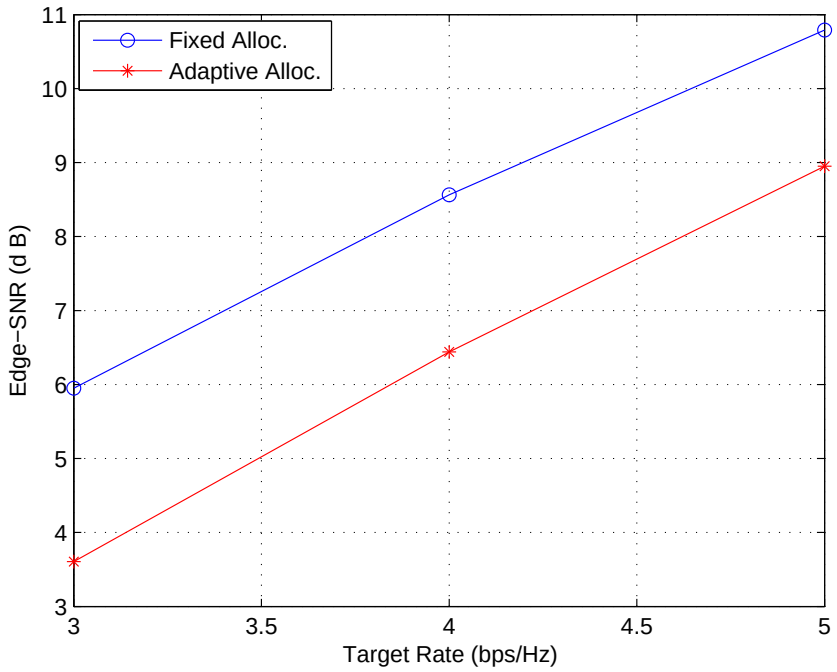

Figure 3.4: Edge-SNR versus different target rate.

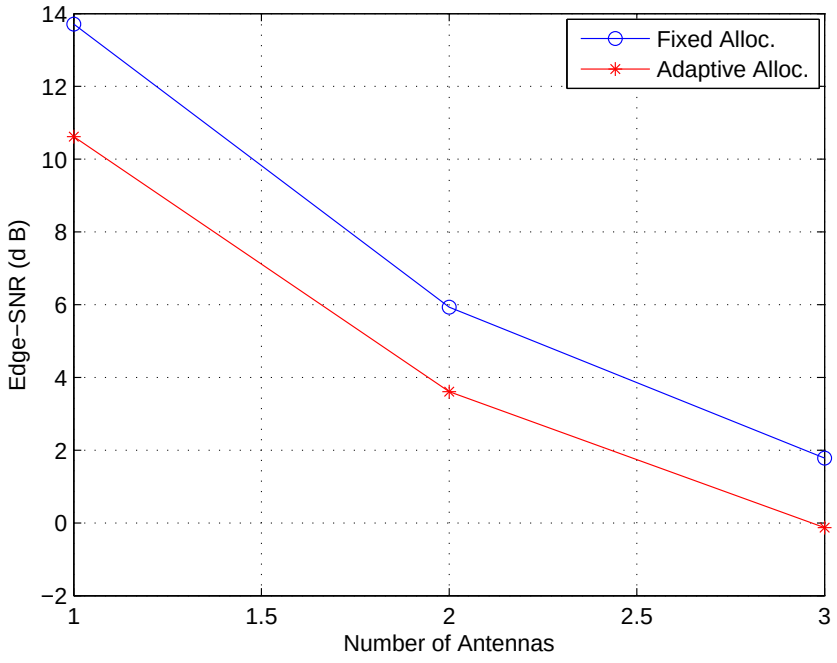

Figure 3.5: Edge-SNR versus different number of antennas.

Figure 3.6 and Figure 3.7 plot the user rate and sum-power versus the number of iterations under the same channel realization. The setups of power constraint of the two figures are both the same as the one in Figure 3.5. As stated in section 3.3.2, when the per-antenna power constraint is relatively large, the rates of all the users will gradually converge to their target ($r_k^{\text{tar}} = 3$ bps/Hz, $\forall k$, in our case), and there is no waste of power since only the power which satisfies the user rates are transmitted. Although the rates are fluctuating within the first 20 iterations, however, the step sizes will be adjusted adaptively according to the convergence condition, so they will become stable step by step.

Finally, we show the competitive condition. Figure 3.8 and 3.9 plots the transmit power on antenna 1 over 100 independent channel realizations with different setups.In both of the two figures, the target user rates are increased to $r_k^{\text{tar}} = 5$ bps/Hz, $\forall k$, the per-antenna power constraints are reduced to $(N_T p_a^{\text{con}}) \,/\, (MN_0) = 161$ dB, 153 dB, or equivalently, $p_a^{\text{con}} \cong 8.019,\ 1.271$ W, $\forall a$, respectively. The power is normalized over p_a^{con}, so 0 dB is equivalent to the value of the antenna power constraint. If the power constraints are removed, the goal is simply to find the minimum power allocation which satisfies the target rates. Although the rate requirements are achieved seemingly, however, the transmit power often exceeds the its constraint (up to about 8 and 16 dB in the two figures) so that the transmit power will be suppressed. Consequently, the actual transmitted rate is less than the expected one. Whereas the scheme which considers per-antenna power constraints tries to balance the transmit power on each antenna, we can see that the maximum power is reduced to about 4 and 7 dB, and it fluctuates slightly compared to the scheme without power constraints. The power on antenna 1 does not always falls below 0 dB, the reason is that the rate constraints in (3.4) tries to increase the transmit power, which leads to a competitive condition. It should be noted that the channel of the two scheme are independent to each other, i.e., they experience different channel in channel realization $i,\ 1 \leq i \leq 100$.

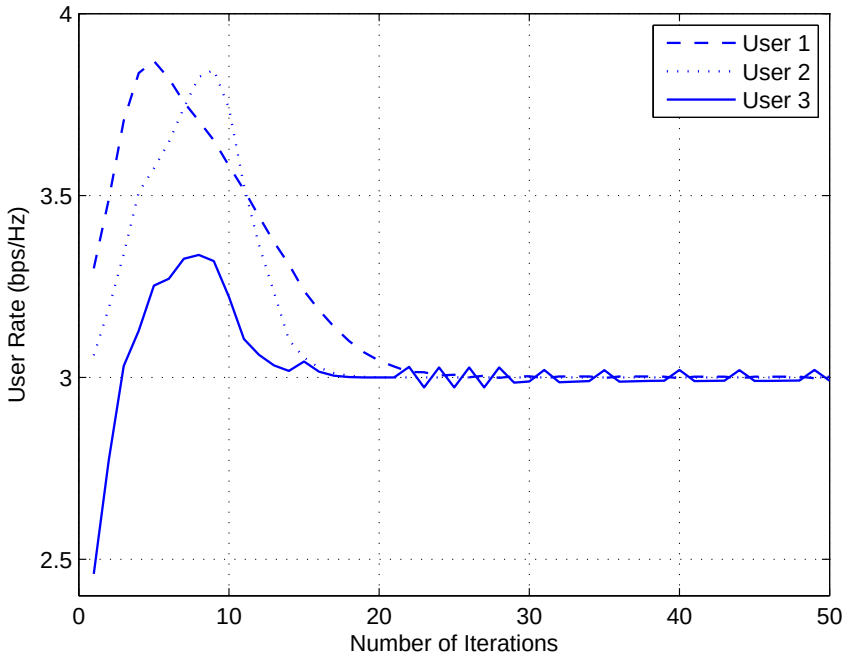

Figure 3.6: User rate convergence behavior.

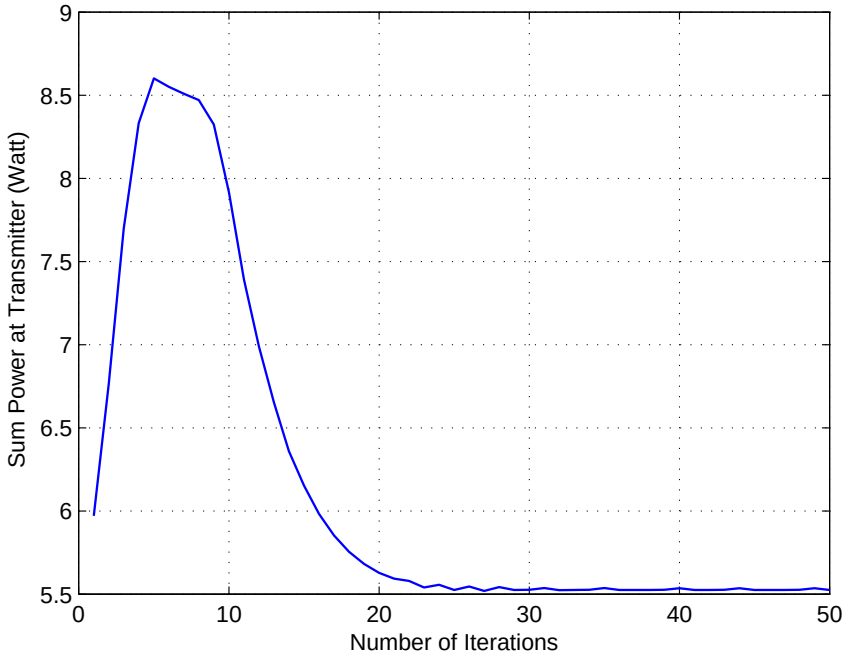

Figure 3.7: Sum-power convergence behavior.

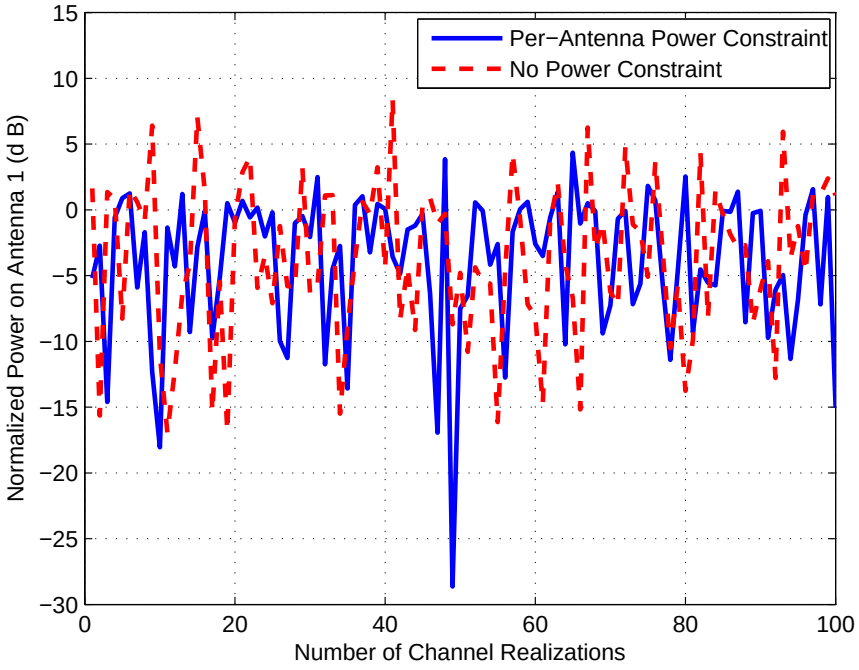

Figure 3.8: Normalized transmit power on antenna 1 over 100 channel realizations when $p_a^{\text{con}} \cong 8.019$ W.

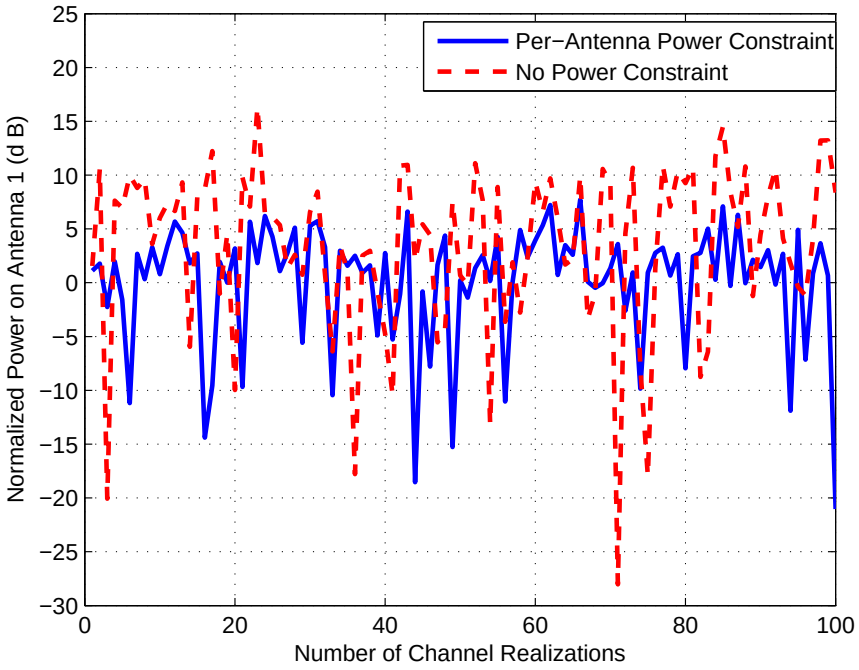

Figure 3.9: Normalized transmit power on antenna 1 over 100 channel realizations when $p_a^{\text{con}} \cong 1.271$ W.

CHAPTER 4

PAPR REDUCTION FOR MU-MIMO SYSTEMS

4.1 System Model and Transmission Schemes

In this chapter, we consider the downlink transmission in only one cell with small scale Rayleigh fading channels for brevity. This assumption is reasonable since in a CoMP cluster, the BSs which apply JT are equivalent to a super BS with additional antennas which serves multiple users. Therefore, the model in this chapter is a single cell MU-MIMO scheme, and it can be extended to a multi-cell scenario by adding large scale components to the channel matrix.

Assume there are M subcarriers and K users in our system. The BS equips N_t antennas and each user has N_r antennas. Ignoring the interference from the other BSs, the received signal of the user k on the subcarrier m can be represented as the same as (3.1). We rewrite this signal here for the sake of illustration:

$$\mathbf{y}_{k,m} = \mathbf{H}_{k,m}\mathbf{x}_{k,m} + \sum_{n=1,n\neq k}^{K_m} \mathbf{H}_{n,m}\mathbf{x}_{k,m} + \mathbf{n}_{k,m}, \tag{4.1}$$

where K_m is the number of user on the subcarrier m, $\mathbf{H}_{k,m}$ is the $N_r \times N_t$ MIMO channel, $\mathbf{x}_{k,m}$ is the $N_t \times 1$ transmitted signal for the user k and $\mathbf{n}_{k,m}$ is the zero-mean complex Gaussian noise vector with covariance matrix $E\left[\mathbf{n}_{k,m}(\mathbf{n}_{k,m})^H\right] = N_0\mathbf{I}_{N_r}$. After BD precoding and receive equalization (see section 3.1), the l^{th} spatial stream for the user k is

$$r_{k,m,l} = s_{k,m,l}\sqrt{p_{k,m,l}}d_{k,m,l} + \widetilde{n}_{k,m,l}, \tag{4.2}$$

where $s_{k,m,l}$ is the l^{th} singular value of the orthogonalized channel matrix of the user k on the subcarrier m, $p_{k,m,l}$ is the power allocated on this stream, $d_{k,m,l}$ is the l^{th} element in $\mathbf{d}_{k,m}$ and

$\widetilde{n}_{k,m,l}$ is the zero-mean complex Gaussian noise with variance N_0.

4.2 Suboptimal Power Minimization Algorithm

In this section, we consider a power minimization problem similar to the one in section 3.2. For power saving, we try to find a way of power allocation that minimizes the total transmit power while satisfies some constraints. The problem can be formulated mathematically as

$$\begin{aligned} \underset{p_{k,m,l}}{\text{minimize}} \quad & \sum_{k=1}^{K}\sum_{m=1}^{M}\sum_{l=1}^{N_r} p_{k,m,l} \\ \text{subject to} \quad & \sum_{m=1}^{M}\sum_{l=1}^{N_r} b_{k,m,l} \geq M r_k^{\text{tar}},\ 1 \leq k \leq K \\ & b_{k,m,l} \in \mathcal{S},\ \forall k,\ m,\ l \\ & p_{k,m,l} \geq 0,\ \forall k,\ m,\ l, \end{aligned} \tag{4.3}$$

where $p_{k,m,l}$ and $b_{k,m,l}$ are the power and the bits allocated for the user k on the l^{th} stream of the subcarrier m, r_k^{tar} is the rate requirement in bps/Hz and $\mathcal{S}$ is a set of the available number of bits of different constellation, for example, if we use quadrature phase-shift keying (QPSK) and 16-QAM for modulation, then $S = \{0, 2, 4\}$. [22] shows that if QAM is deployed, then $p_{k,m,l}$ is given by

$$p_{k,m,l} = \frac{N_0}{\left(s_{k,m,l}\right)^2} \ln\left(\frac{1}{5BER^{\text{tar}}}\right) \frac{2^{b_{k,m,l}} - 1}{1.5}, \tag{4.4}$$

where BER^{tar} is the target BER. Based on (4.4), we propose a suboptimal greedy algorithm to solve (4.3) when only QPSK and 16-QAM are deployed, although it can be easily extended to a higher order modulation. For brevity, the user selection is fixed such that each subcarrier is occupied by all the K users, i.e., $K_m = K,\ \forall k$. The $\Delta p_{k,m,l}$ in Step 2 is defined as

$$\Delta p_{k,m,l} = \frac{N_0}{\left(s_{k,m,l}\right)^2} \ln\left(\frac{1}{5BER^{\text{tar}}}\right) \left(\frac{2^{b_{k,m,l}+2} - 1}{1.5} - \frac{2^{b_{k,m,l}} - 1}{1.5}\right). \tag{4.5}$$

We set the bit step size to two because that $S = \{0, 2, 4\}$ here.

Algorithm Greedy Power Allocation Algorithm

1: Assign zero bits to all streams;

2: Find the stream $\widehat{l}$ that requires the least $\Delta p_{k,m,l}$;

3: Add two bits to $\widehat{l}$ if the bits on $\widehat{l}$ does not exceed the maximum number (4 in our case); otherwise remove $\widehat{l}$ from the candidate list;

4: Terminate this algorithm when all the rate requirements are satisfied; otherwise go back to step2;

4.3 Multiuser Active Constellation Extension Method

4.3.1 PAPR Definition and ACE Concept

In order to understand why OFDM suffers from large PAPR, we start from the definition of the OFDM signal. An OFDM signal is the sum of M independent signals, which is

$$x_n = \frac{1}{M}\sum_{m=1}^{M} X_m e^{j2\pi mn/M},\ 1 \le n \le M \tag{4.6}$$

where X_m is the frequency doamin symbol (could be any type of modulation) on the subcarrier m and x_n is the n^{th} time domain symbol. The definition of the PAPR for an OFDM symbol is

$$PAPR\left(\mathbf{x}\right) = \frac{\| \mathbf{x} \|_{\infty}^{2}}{E\left[\| \mathbf{x} \|_{F}^{2}\right]/M}, \tag{4.7}$$

where $E\left[\cdot\right]$ is the expectation operator and $\mathbf{x} = \left[x_1, \ldots, x_M\right]^T$. Based on the central limit theorem, as M increases, x_n approaches a Gaussian distributed random variable. Therefore, it may become large under a certain probability and cause high PAPR. Since there is a saturation region in every power amplifier (PA), any signal exceeds the saturation threshold will be clipped. Hence we need to widen the power range of the PA if the transmitted signal has large PAPR, which is power inefficient.

Here we introduce the concept of active constellation extension (ACE) in [15]. ACE is a

method which tries to alter the constellation to reduce PAPR. The idea can be illustrated in Figure 4.1 with QPSK modulation. The shaded regions represent the legal extension regions (could be further extended). Suppose the received signal is corrupted with additive white Gaussian noise, the decision boundaries of the maximum likelihood detector are the real and imagine axes which divide the complex plane into four quadrants. Making any constellation point to be closer to the boundaries will increase the BER. By setting the extension regions properly, ACE can modify the constellation points without BER performance degradation .

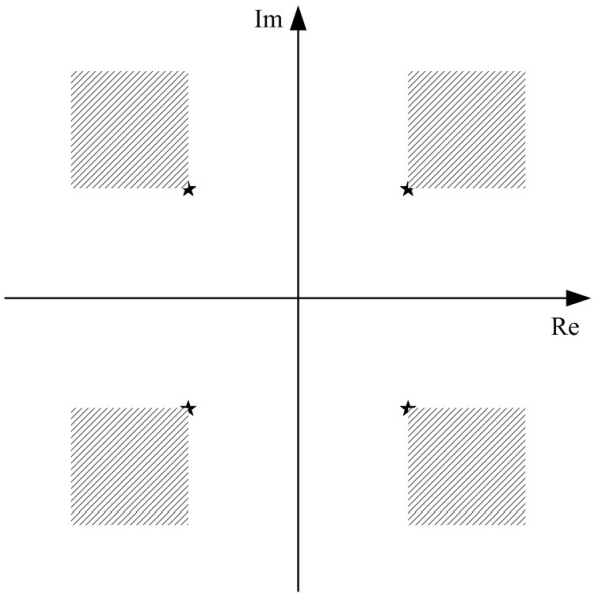

Figure 4.1: Illustration for ACE with QPSK modulation.

This idea can also be applied to QAM. For the example of 16-QAM as depicted in Figure 4.2, the constellation points can be classified into three types according to their position: the interior, exterior and corner points. Without increasing the BER, the interior points have to be fixed, the exterior points can only be moved in one direction, and the corner points can be moved inside the shaded regions. Hence, the higher the modulation order is, the less the flexibility of the ACE extension region will be. Our goal is to find a combination of these additional signals that can reduce time domain peaks.

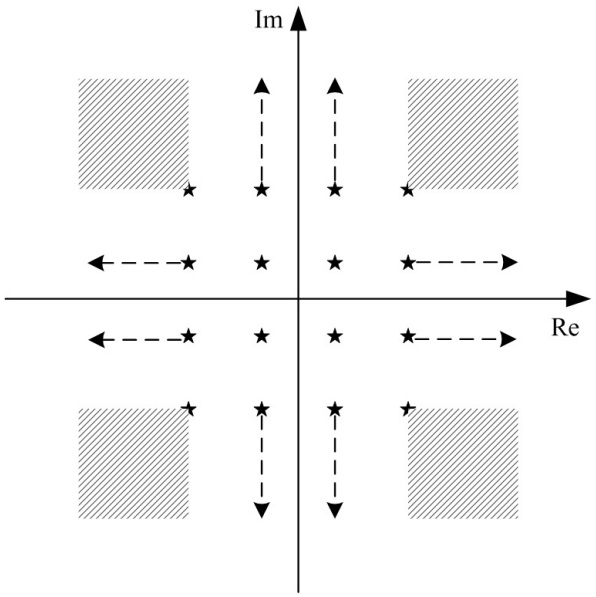

Figure 4.2: Illustration for ACE with 16-QAM.

4.3.2 Problem Formulation

In this section, we define the ACE PAPR reduction problem for MU-MIMO scheme. We first define the $N_t \times M$ frequency domain MIMO-OFDM signal as

$$\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_m, \ldots \mathbf{x}_M], \tag{4.8}$$

where

$$\begin{aligned}
\mathbf{x}_m &= \sum_{k=1}^{K_m} \mathbf{x}_{k,m} \\
&= \sum_{k=1}^{K_m} \mathbf{F}_{k,m}\mathbf{P}_{k,m}\mathbf{d}_{k,m} \\
&= \sum_{k=1}^{K_m} \mathbf{F}_{k,m}\widetilde{\mathbf{d}}_{k,m} \\
&= [\mathbf{F}_{1,m}, \ldots, \mathbf{F}_{K_m,m}] \left[\left(\widetilde{\mathbf{d}}_{1,m}\right)^H, \ldots, \left(\widetilde{\mathbf{d}}_{K_m,m}\right)^H \right]^H \\
&= \mathbf{F}_m \widetilde{\mathbf{d}}_m
\end{aligned} \tag{4.9}$$

is the column vector of $\mathbf{X}$ which sums the signals of all the K_m users on the subcarrier m. In (4.9), $\mathbf{F}_{k,m}$ is the BD precoding matrix, $\mathbf{P}_{k,m}$ and $\mathbf{d}_{k,m}$ are the power allocation matrix and

modulated symbol vector determined by the algorithm introduced in section 4.2. Then we obtain the $N_t \times M$ time domain signal by applying block-IDFT to each row of $\mathbf{X}$:

$$\begin{aligned} \mathbf{W} &= \text{Block} - \text{IDFT}\,(\mathbf{X}) \\ &= [\mathbf{w}_1, \ldots, \mathbf{w}_a, \ldots, \mathbf{w}_{N_T}]^T, \end{aligned} \tag{4.10}$$

where $\mathbf{w}_a^T$ is the a^{th} row vector of $\mathbf{W}$. The frequency-time relation via block-IDFT can be illustrated by Figure 4.3.

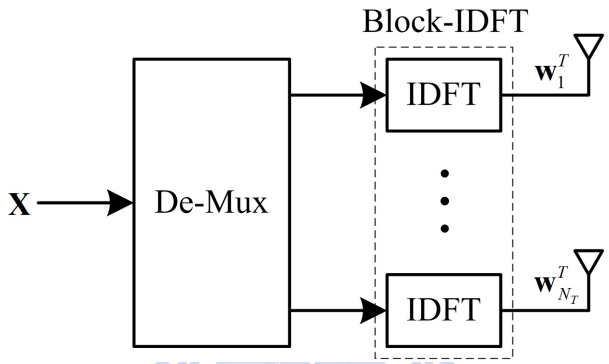

Figure 4.3: Illustration of block-IDFT.

We define the PAPR on the a^{th} antenna as

$$PAPR_a = \frac{\| \mathbf{w}_a^T \|_\infty^2}{E\left[\| \mathbf{w}_a^T \|_F^2\right]/M} \tag{4.11}$$

and the PAPR of a MIMO system as

$$PAPR^{\text{MIMO}} = \max_a \left(\text{PAPR}_a\right). \tag{4.12}$$

With the definitions above, the ACE PAPR minimization problem can be formulated as

$$\underset{\mathbf{C}\in\mathcal{C}}{\text{minimize}} \quad \overline{PAPR}^{\text{MIMO}}, \tag{4.13}$$

where $\mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_m, \ldots \mathbf{c}_M]$ is the extension matrix, $\mathcal{C}$ is the transformed legal ACE space and $\overline{PAPR}^{\text{MIMO}}$ is the new PAPR after altering the original symbol by $\overline{\mathbf{X}} = \mathbf{X} + \mathbf{C}$. When there are multiple users on each subcarrier, we can transform $\mathbf{C}$ back into the constellation domain by

$$\mathbf{e}_m = \text{inv}\left(\mathbf{F}_m\right)\mathbf{c}_m, \tag{4.14}$$

where $\mathbf{F_m}$ is the composite BD matrix defined in (4.9). Therefore, the signal on the subcarrier m after ACE processing can be given by

$$\begin{aligned}\overline{\mathbf{x}}_m &= \mathbf{x}_m + \mathbf{c}_m \\ &= \mathbf{F}_m\widetilde{\mathbf{d}}_m + \mathbf{F}_m\mathbf{e}_m. \end{aligned} \tag{4.15}$$

This equation tells us that finding $\mathbf{c_m}$ is equivalent to finding a constellation extension vector $\mathbf{e_m}$. Hence we can reformulate (4.13) in a new form:

$$\underset{\mathbf{E}\in\mathcal{E}}{\text{minimize}} \quad \overline{\text{PAPR}}^{\text{MIMO}}, \tag{4.16}$$

where $\mathbf{E} = [\mathbf{e}_1, \ldots, \mathbf{e}_m, \ldots \mathbf{e}_M]$ and $\mathcal{E}$ are the ACE extension matrix and legal space in the constellation domain, respectively. A low complexity algorithm for solving (4.16) will be introduced in section 4.3.3.

4.3.3 An Efficient Algorithm for ACE

In this section, we propose a suboptimal algorithm to implement ACE. The main idea of this algorithm is based on [15], we made some modifications so that it can be used in MU-MIMO systems. Based on the Step 4 of the approximate gradient-project method, we know that only the elements of $\mathbf{e}_m$ which corresponds to the data-carrying spatial streams would be altered due to ACE region constraints. The remaining empty spatial streams (usually have bad channel conditions) after power allocation can be fully used for PAPR reduction, which makes our algorithm has a tone reservation-like property.

Algorithm Approximate Gradient-Project Method

1: Obtain $\mathbf{X}$ by applying BD and the power allocation algorithm introduced in section 4.2. Determine the legal ACE extension regions according to the modulation results on each spatial stream. Deploy block-IDFT to get $\mathbf{W}$ and set the iteration index $i = 0$;

Algorithm Approximate Gradient-Project Method (continued)

2: Clip any signal which exceeds the threshold D:

$$\widetilde{\mathbf{W}}^i[a,n] = \begin{cases} \mathbf{W}^i[a,n] & , \left|\mathbf{W}^i[a,n]\right| \leq D \\ De^{j\theta[a,n]} & , \left|\mathbf{W}^i[a,n]\right| > D, \end{cases} \tag{4.17}$$

where

$$\mathbf{W}^i[a,n] = \left|\mathbf{W}^i[a,n]\right| e^{j\theta[a,n]} \tag{4.18}$$

is the element in the a^{th} row and n^{th} column of the matrix $\mathbf{W}^i$;

3: Compute the clipped portion:

$$\mathbf{W}_{\text{clip}} = \widetilde{\mathbf{W}}^i - \mathbf{W}^i \tag{4.19}$$

and transform it into frequency domain to obtain $\mathbf{C}$:

$$\mathbf{C} = \text{Block} - \text{DFT}\left(\mathbf{W}_{\text{clip}}\right); \tag{4.20}$$

4: For each subcarrier, compute $\mathbf{e}_m = \text{inv}\left(\mathbf{F}_m\right)\mathbf{c}_m$ and remove the elements (real and/or imaginary part) outside the legal ACE regions to obtain $\widetilde{\mathbf{e}}_m$;

5: Compute the time domain clipped portion which lies in legal region by

$$\widetilde{\mathbf{W}}_{\text{clip}} = \text{Block} - \text{IDFT}\left(\mathbf{F}_m\widetilde{\mathbf{e}}_m\right). \tag{4.21}$$

6: Compute

$$\mathbf{W}^{i+1} = \mathbf{W}^i + \delta\widetilde{\mathbf{W}}_{\text{clip}}, \tag{4.22}$$

where δ is the step size.

7: Stop this algorithm if the target level of $PAPR^{\text{MIMO}}$ or the maximum number of iteration has been reached. Otherwise, set $i = i + 1$ and go back to Step 2.

Next, we introduce an algorithm to find the δ in Step 6 of the approximate gradient-project method. The best choice of δ would be

$$\delta^* = \arg\min_{\delta} \frac{\| \left(\mathbf{w}^i_{a_{\max}}\right)^T + \delta\left(\widetilde{\mathbf{w}}^i_{\text{clip},a_{\max}}\right)^T \|^2_\infty}{E[\| \left(\mathbf{w}^i_{a_{\max}}\right)^T + \delta\left(\widetilde{\mathbf{w}}^i_{\text{clip},a_{\max}}\right)^T \|^2_F]/M}, \tag{4.23}$$

where

$$a_{\max} = \arg\min_{a} \left(PAPR_a\right) \tag{4.24}$$

and

$$\left(\mathbf{w}^i_{a_{\max}}\right)^T = \left[w^i_{a_{\max},1}, \ldots, w^i_{a_{\max},n}, \ldots w^i_{a_{\max},M}\right] \tag{4.25}$$

is the row $a_{\max}$ of $\mathbf{W}^i$. A heuristic way to find δ is to reduce the peak to another sample point:

$$\left|\mathbf{W}^i\left[a_{\max}, n_{\max}\right] + \delta\widetilde{\mathbf{W}}^i_{\text{clip}}\left[a_{\max}, n_{\max}\right]\right|^2 = \left|\mathbf{W}^i\left[a, n\right] + \delta\widetilde{\mathbf{W}}^i_{\text{clip}}\left[a, n\right]\right|^2, \tag{4.26}$$

where

$$n_{\max} = \arg\max_{n} \left(\left[\mid w^i_{a_{\max},1} \mid, \ldots, \mid w^i_{a_{\max},n} \mid, \ldots, \mid w^i_{a_{\max},M} \mid\right]\right). \tag{4.27}$$

Hence $\mathbf{W}\left[a_{\max}, n_{\max}\right]$ indicates peak location on antenna $a_{\max}$. This method needs a search over $MN_T - 1$ possibilities, i.e., it solves (4.26) for every sample $[a, n]$ except $[a_{\max}, n_{\max}]$ and finds the best δ that minimizes $PAPR^{\text{MIMO}}$. To avoid solving such a quadratic equation so many times, here we introduce a low complexity method to find the appropriate step size δ.

Algorithm Smart Gradient-Project Method

1: Find the peak magnitude on the antenna $a_{\max}$ by

$$A = \left|\mathbf{W}^i\left[a_{\max}, n_{\max}\right]\right|. \tag{4.28}$$

2: For all the samples , compute the projection part of $\mathbf{W}^i_{\text{clip}}\left[a, n\right]$ along the phase of $\mathbf{W}^i\left[a, n\right]$:

$$\widetilde{\mathbf{W}}^{\text{proj}}_{\text{clip}}\left[a, n\right] = \frac{\text{Re}\left\{\mathbf{W}^i\left[a, n\right]\left(\widetilde{\mathbf{W}}^i_{\text{clip}}\left[a, n\right]\right)^*\right\}}{\left|\mathbf{W}^i\left[a, n\right]\right|}. \tag{4.29}$$

Algorithm Smart Gradient-Project Method (continued)

3: Compute the candidates of step size for only the samples where $\widetilde{\mathbf{W}}_{\text{clip}}^{\text{proj}}[a,n] > 0$ by solving

$$A + \delta[a,n]\,\widetilde{\mathbf{W}}_{\text{clip}}^{\text{proj}}[a_{\max}, n_{\max}] = \left|\mathbf{W}^{i}[a,n]\right| + \delta[a,n]\,\widetilde{\mathbf{W}}_{\text{clip}}^{\text{proj}}[a,n]\,. \tag{4.30}$$

4: Choose the $\delta[a,n]$ that minimizes $PAPR^{\text{MIMO}}$. If this $\delta[a,n]$ is negative, stop the entire PAPR reduction algorithm (the reason is left below).

There are two cases for explaining why this projected method works:

Case 1) $|\mathbf{W}_{\text{clip}}[a,n]| \neq 0$

This is the clipped portion of the sample which exceeds the threshold D. Although it will be distorted due to the ACE space constraints (the Step 4 of approximate gradient-project method). But it is still likely that $\angle\mathbf{W}_{\text{clip}}[a,n] \cong \angle\widetilde{\mathbf{W}}_{\text{clip}}[a,n]$. Based on (4.17)-(4.19), we know $\angle\mathbf{W}_{\text{clip}}[a,n] = -\angle\mathbf{W}[a,n]$, it implies that $\widetilde{\mathbf{W}}_{\text{clip}}^{\text{proj}}[a,n]$ holds the most energy of $\widetilde{\mathbf{W}}_{\text{clip}}[a,n]$. Hence, there is little loss between solving (4.26) and (4.30), but the latter has lower complexity. Since $\angle\widetilde{\mathbf{W}}_{\text{clip}}[a,n] \cong -\angle\mathbf{W}[a,n]$, if the $\delta[a,n]$ is negative, it may increases the peak power. That's why we stop the algorithm if this happens. In the Step 3 of the smart gradient-project method, we consider only the samples where $\widetilde{\mathbf{W}}_{\text{clip}}^{\text{proj}}[a,n] > 0$, since the $\widetilde{\mathbf{W}}_{\text{clip}}^{\text{proj}}[a,n] < 0$ situations originate from those samples with large amplitude, i.e., the $|\mathbf{W}_{\text{clip}}[a,n]| \neq 0$ samples. Reducing the peak sample to the level of such samples will result in little PAPR reduction.

Case 2) $|\mathbf{W}_{\text{clip}}[a,n]| = 0$

This is the sample which falls below the threshold D. Even with ACE distortion, it is likely that $\left|\widetilde{\mathbf{W}}_{\text{clip}}[a,n]\right| \cong 0$. Therefore, $\widetilde{\mathbf{W}}_{\text{clip}}[a,n]$ has little effect on the result of PAPR reduction, so projecting these samples is nothing matter.

4.3.4 Modifications of ACE

As mentioned above, it requires an expensive PA with wide power range to transmit a high PAPR signal, that is the reason why we are eager to reduce PAPR. In practical implementations, the signal before PA will be converted into analog type through a digital-to-analog (D/A) converter. Generally, the peak power increases after D/A conversion. Therefore, the performance of PAPR reduction will be reduced if we ignore the D/A effect.

To approximate the analog signal, we can oversample the digital signal by a factor of L. Assume there is a frequency domain vector $\mathbf{x} = [x_1, \ldots, x_M]^T$ with frequency-time relation

$$\mathbf{w} = \text{IDFT}(\mathbf{x}). \tag{4.31}$$

The L times oversampling can be implemented by inserting $M(L-1)$ zeros to $\mathbf{x}$ [23]:

$$\mathbf{w}^L = \text{IDFT}\left([x_1, \ldots, x_M, 0_{M+1}, \ldots 0_{ML}]^T\right). \tag{4.32}$$

The effect of oversampling will be shown in the simulation results of this chapter.

One of the drawbacks of ACE is the increased average power. However, the power increment can be kept to an acceptable value if we set a target peak level D^{tar} to our algorithm. If the PA was able to transmit a signal with peak level D^{tar}, it is not necessary to reduce the peak far below D^{tar}. Hence if the δ chosen by the Step 3 of the smart gradient-project method satisfies

$$A - \left|\delta \widetilde{\mathbf{W}}^i_{\text{clip}}[a_{\max}, n_{\max}]\right| < D^{\text{tar}}, \tag{4.33}$$

we can meet the target level by replacing δ with

$$\delta = \frac{A - D^{\text{tar}}}{\left|\widetilde{\mathbf{W}}^i_{\text{clip}}[a_{\max}, n_{\max}]\right|}. \tag{4.34}$$

Although we make this modification, the peak level may not be D^{tar} exactly, since the peak location would be changed after every iteration of ACE. But our simulations show that it indeed reduces the increment of average power.

4.4 Simulation Results

Since the PAPR is not sensitive to the channel model, so for brevity, we consider only one BS which serves two users ($K = 2$) in the simulation of this chapter. The BS equips four antennas ($N_t = 4$) and each user has two antennas ($N_r = 2$). The total number of subcarrier is 256 ($M = 256$) with 15 KHz bandwidth each, unless otherwise stated. The channel is 17-taps Rayleigh distributed with uniform power delay profile. At first, the OFDM symbols on each antenna are generated by the power allocation algorithm introduced in section 4.2, the available constellations are QPSK and 16-QAM and the rate requirement is $r_k^{\text{tar}} = 3,\ \forall k$ with tolerable bit error rate $BER^{\text{tar}} = 10^{-2}$. Next, these signals are fed to the multiuser ACE algorithm. The clipped level D is 4.86 dB above the average power.

We first show the performance of power allocation. Figure 4.4 plots the transmit SNR versus different rate requirements. The transmit SNR is defined as

$$\frac{\sum_{k=1}^{K}\sum_{m=1}^{M}\sum_{l=1}^{N_r} p_{k,m,l}}{MN_0} = \frac{E_s}{MN_0}. \tag{4.35}$$

Since in a wireless environment, the singular value (channel quality) of the main spatial stream is usually much stronger than the others. So in the fixed allocation, only the main stream on each subcarrier is used to carry data. At $r_k^{\text{tar}} = 3$, the SNR gain is about 1.7 dB. This gain would be even larger if the fixed scheme use the secondary spatial stream.

Figure 4.5 and 4.6 shows the distribution of the data symbols with QPSK and 16-QAM after ACE. As expected, they lie in the constraint space and the symbols except the regular constellation points increase the average transmit power.

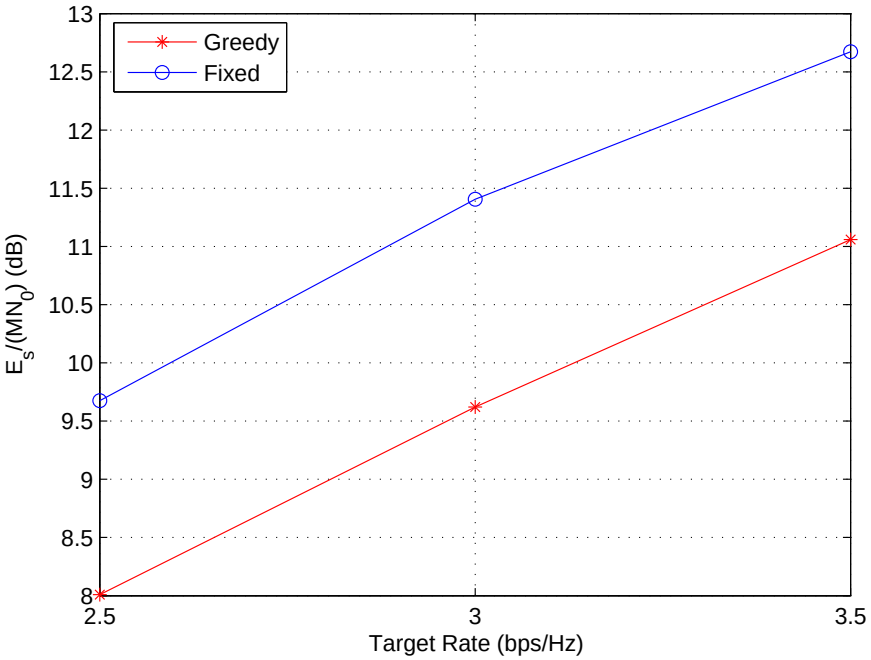

Figure 4.4: The SNR versus different rate requirements.

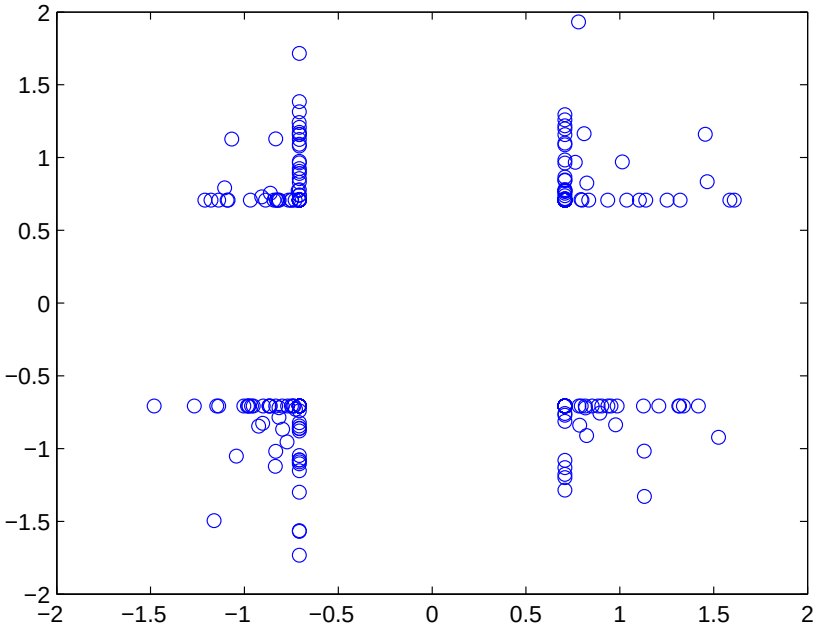

Figure 4.5: A snapshot of the data symbols with QPSK modulation after ACE.

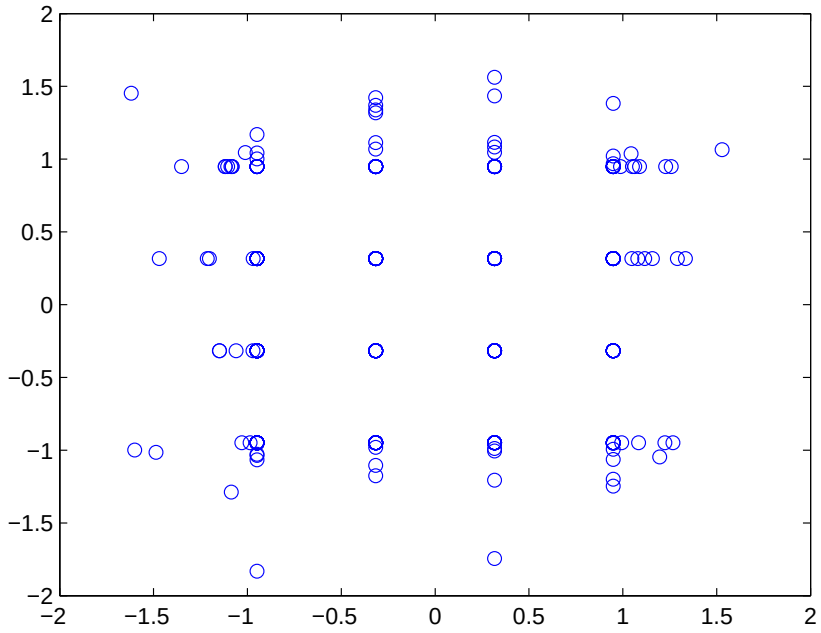

Figure 4.6: A snapshot of the data symbols with 16-QAM modulation after ACE.

Figure 4.7 plots the complementary cumulative distribution function (CCDF) of the PAPR after ACE. The upsampling factor L is set to two. The first iteration provides about 3.5 dB PAPR reduction gain at probability 10^{-3}, whereas the third provides negligible gains. Hence iteration more than two is meaningless.

The effect of different oversampling factor is shown in Figure 4.8. Since Figure 4.7 tells us that two iteration is sufficient, so we only show the performance up to iteration two. The scheme with $L = 4$ provides only additional 0.2 dB gain at probability 10^{-3}. However, the FFT size will become two times larger and therefore increases the system complexity.

In Figure 4.7, the power increments of the three interations are 1.95, 2.49, 2.53 dB. They can be kept low if we only reduce the PAPR to a certain level. Figure 4.9 plots the CCDF of the PAPR with different PAPR targets at iteration two when $L = 2$. The power increments with the PAPR targets 6.5 and 7.5 dB are reduced to 1.31 and 0.57 dB.

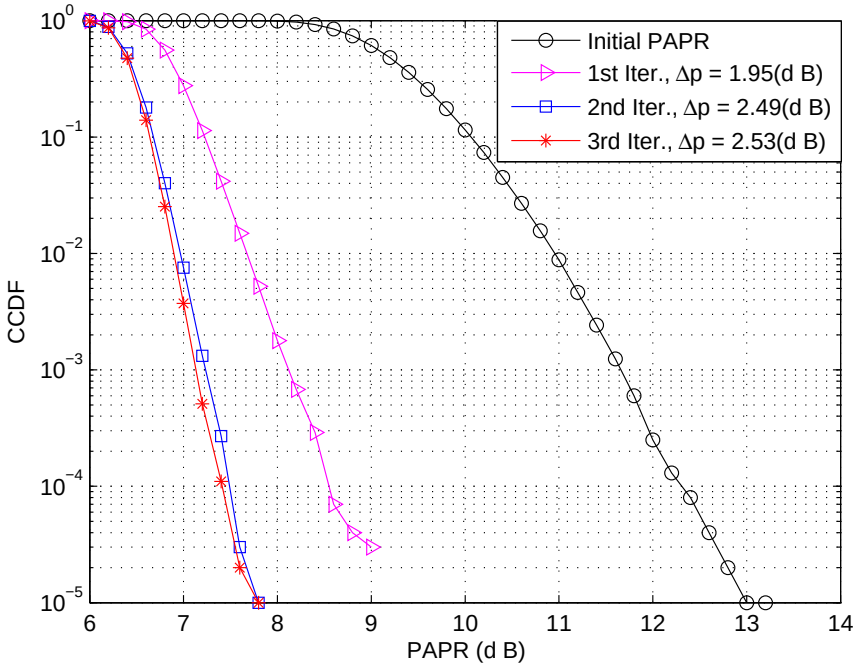

Figure 4.7: The CCDF of the PAPR after ACE.

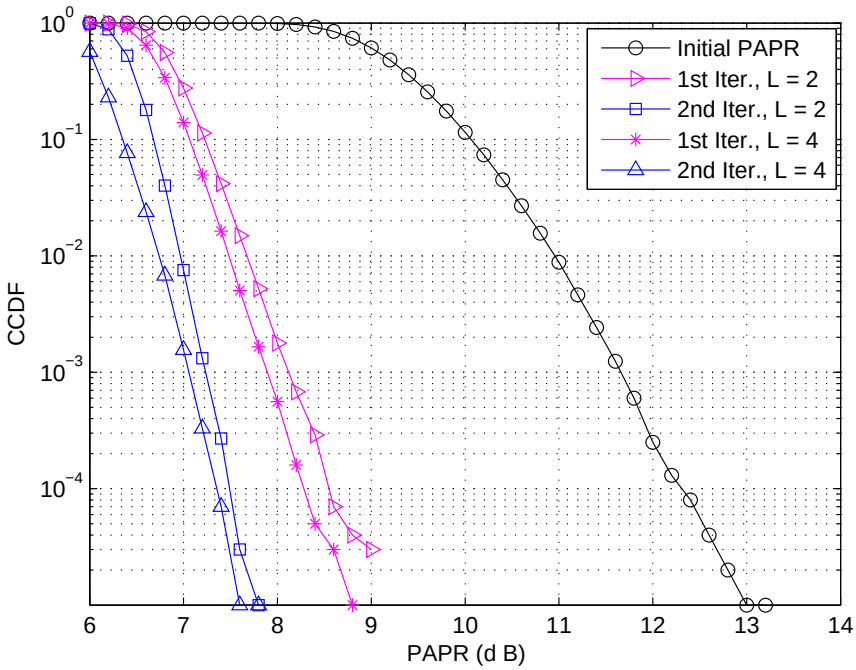

Figure 4.8: The CCDF of the PAPR after ACE with different oversampling factors.

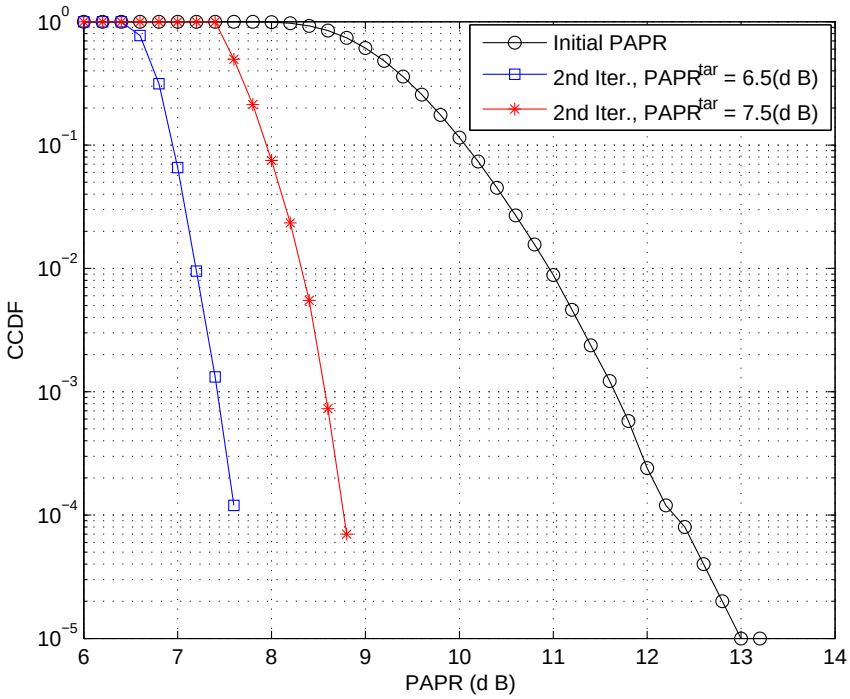

Figure 4.9: The CCDF of the PAPR after ACE with different PAPR targets.

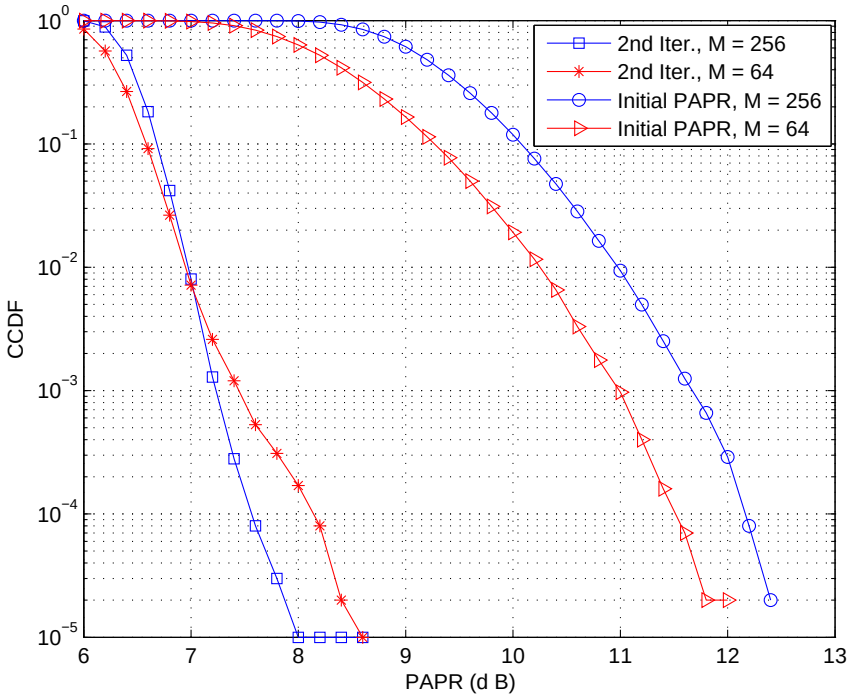

Figure 4.10: The CCDF of the PAPR after ACE with different subcarrier sizes.

Finally, we show the effect of different subcarier sizes. Figure 4.10 plots the CCDF of the PAPR with subcarrier size 64 and 256. The upsampling factor is two. Larger subcarrier size results in larger initial PAPR, however, it provides more space to design ACE signals, so the PAPR reduction performance will be better after ACE.

CHAPTER 5

CONCLUSION

In this thesis, we tackle three main problems including BS clustering, resource allocation and PAPR reduction for downlink LTE-A cellular system. In chapter2, the BS clustering problem is considered. CoMP is a promising way of suppressing inter-BS interference, but it requires the CSI and/or user data sharing between BSs. In order to reduce the overhead, only a limited number of BSs form a CoMP cluster. This raises the problem of which BSs should cooperate in a cluster to maximize the system performance. A straightforward way is static clustering, but it inherits the drawback that users in the edge of the static cluster still suffer from severe interference. In this thesis, we propose a dynamic clustering method based on modifying the method proposed in [10]. The method utilizes the instantaneous CSI and forms CoMP clusters adaptively, hence there is no fixed cluster edge in this scheme and no user will always prone to interference. Nevertheless, this dynamic method requires a central controller which runs the clustering algorithm. We make some modifications to [10], which can be listed below:

1. We replace the simple zero-forcing and MMSE precoder with a more powerful BD precoder.

2. In [10], it assumed the users always need CoMP regardless of their signal strength. But CoMP provides only marginal gain to the users with high SINR. Hence, we let each user measure their pilot SINR and feedback to the controller for making BS clustering decisions.

3. We provide three utility functions for the dynamic algorithm and make it more flexible.

In chapter 3, we consider the problem of resource allocation. Equal power allocation is applied in chapter 2, however, the performance can be further improved by adaptively allocating power to frequency and space domain. Based on [12], we introduce an optimal power allocation method that minimizes the total transmit power while satisfying BER and user rate constraints. Besides, we add practical per-antenna power constraints to the problem formulation. The main idea is to use the Lagrange dual decomposition to transform the original non-convex problem into convex form. In addition, an adaptive method is proposed to find the optimal solution. Simulation results show that the adaptive allocation provides large performance gain over the fixed allocation, and the transmit power on each antenna seldom exceeds the maximum power limit, as compared with the result of [12]. This improves the overall system power efficiency.

In chapter 4, we address the PAPR reduction problem. Although the average power can be reduced via power allocation, multicarrier systems still suffer from large PAPR. We adopt the concept called ACE [15] to minimize PAPR. The downside of this method is the increased transmit power, however, it can be kept low by achieving acceptable PAPR level. To get better performance, we oversample the digital signal to approximate analog signal. Simulation results shows that only two times oversampling and two iterations is sufficient to provide great PAPR reduction. Our contributions to this topic are

1. We propose a low complexity power allocation algorithm to minimize the average transmit power while satisfying user rate and constellation constraints. It also helps us to generate MIMO-OFDM signals.

2. We modify the SISO ACE method in [15] to make it compatible to MU-MIMO systems.

APPENDIX A

DERIVATIONS FOR CHAPTER 3

A.1 Derivation of Competitive Water Filling Solution

The Lagrangian (3.13) has become the new object function that to be minimized. We add the constraint $p_{k,m,l} \geq 0$ here so that the problem can be formulated as

$$\begin{aligned} \underset{p_{k,m,l}}{\text{minimize}} \quad & \left(p_{k,m,l} - \mu_k r_{k,m,l} + \sum_{a=1}^{N_T} \kappa_a |\mathbf{F}_{k,m}[a,l]|^2 p_{k,m,l} \right) \\ \text{subject to} \quad & p_{k,m,l} \geq 0. \end{aligned} \tag{A.1}$$

We introduce a new Lagrange multiplier ν for the new power constraint, and write the Lagrangian of (A.1) as

$$\overline{\overline{\mathcal{L}}} = p_{k,m,l} - \mu_k r_{k,m,l} + \sum_{a=1}^{N_T} \kappa_a |\mathbf{F}_{k,m}[a,l]|^2 p_{k,m,l} - \nu p_{k,m,l}. \tag{A.2}$$

Then we obtain the *Karush-Kuhn-Tucker* (KKT) conditions:

$$p_{k,m,l} \geq 0, \tag{A.3}$$

$$\nu \geq 0, \tag{A.4}$$

$$\nu p_{k,m,l} = 0, \tag{A.5}$$

$$\frac{\partial \overline{\overline{\mathcal{L}}}}{\partial p_{k,m,l}} = 0. \tag{A.6}$$

We do the partial differentiation of (A.6) and get

$$1 + \sum_{a=1}^{N_T} \kappa_a |\mathbf{F}_{k,m}[a,l]|^2 - \frac{\mu_k}{\ln(2)} \frac{1}{\frac{\tau N_0}{s_{k,m,l}^2} + p_{k,m,l}} = \nu \geq 0, \tag{A.7}$$

where the inequality is based on (A.4). Replace the ν in (A.5) by the one in (A.7), we get

$$p_{k,m,l}\left(1+\sum_{a=1}^{N_T}\kappa_a|\mathbf{F}_{k,m}[a,l]|^2-\frac{\mu_k}{\ln(2)}\frac{1}{\frac{\tau N_0}{s_{k,m,l}^2}+p_{k,m,l}}\right)=0. \tag{A.8}$$

Case 1) $1+\sum_{a=1}^{N_T}\kappa_a|\mathbf{F}_{k,m}[a,l]|^2<\frac{\mu_k}{\ln(2)}\frac{1}{\frac{\tau N_0}{s_{k,m,l}^2}}$

In this case, the inequality in (A.7) only holds if $p_{k,m,l}>0$. Therefore, the second term in (A.8) must be zero, which yields

$$1+\sum_{a=1}^{N_T}\kappa_a|\mathbf{F}_{k,m}[a,l]|^2=\frac{\mu_k}{\ln(2)}\frac{1}{\frac{\tau N_0}{s_{k,m,l}^2}+p_{k,m,l}}. \tag{A.9}$$

Case 2) $1+\sum_{a=1}^{N_T}\kappa_a|\mathbf{F}_{k,m}[a,l]|^2\geq\frac{\mu_k}{\ln(2)}\frac{1}{\frac{\tau N_0}{s_{k,m,l}^2}}$

If $p_{k,m,l}>0$, it implies

$$1+\sum_{a=1}^{N_T}\kappa_a|\mathbf{F}_{k,m}[a,l]|^2\geq\frac{\mu_k}{\ln(2)}\frac{1}{\frac{\tau N_0}{s_{k,m,l}^2}}>\frac{\mu_k}{\ln(2)}\frac{1}{\frac{\tau N_0}{s_{k,m,l}^2}+p_{k,m,l}}, \tag{A.10}$$

which contradicts (A.7). Hence $p_{k,m,l}$ must be zero.

Combine the results of Case 1 and Case 2, we get

$$p_{k,m,l}=\begin{cases}\frac{\mu_k}{\ln(2)1+\sum_{a=1}^{N_T}\kappa_a\left|\mathbf{F}_{k,m}[a,l]\right|^2}-\frac{\tau N_0}{s_{k,m,l}^2} & , \frac{\mu_k}{\ln(2)\left(1+\sum_{a=1}^{N_T}\kappa_a\left|\mathbf{F}_{k,m}[a,l]\right|^2\right)}>\frac{\tau N_0}{s_{k,m,l}^2}\\ 0 & , \frac{\mu_k}{\ln(2)\left(1+\sum_{a=1}^{N_T}\kappa_a\left|\mathbf{F}_{k,m}[a,l]\right|^2\right)}\leq\frac{\tau N_0}{s_{k,m,l}^2}\end{cases} \tag{A.11}$$

which is equivalent to (3.14). This result is a transformation of the traditional water-filling solution. In [24], it has been shown that if we write the water-filling in a general form:

$$p=\max\left\{\kappa-\lambda^{-1},0\right\}, \tag{A.12}$$

where κ is the water-level and λ^{-1} is the riverbed. The capacity will be

$$c=\log_2\left(1+p\lambda\right). \tag{A.13}$$

Substitute (A.12) into (A.13), we obtain

$$\begin{aligned} c &= \log_2\left(1+\max\left\{\kappa-\lambda^{-1},0\right\}\lambda\right) \\ &= \log_2\left(1+\max\left\{\kappa\lambda-1,0\right\}\right) \\ &= \log_2\left(\max\left\{\kappa\lambda,1\right\}\right), \end{aligned} \tag{A.14}$$

where

$$\kappa = \frac{\mu_k}{\ln(2)\left(1+\sum_{a=1}^{N_T}\kappa_a\left|\mathbf{F}_{k,m}[a,l]\right|^2\right)} \tag{A.15}$$

and

$$\lambda = \frac{s_{k,m,l}^2}{\tau N_0} \tag{A.16}$$

in our case. This verifies (3.15).

A.2 Derivation of Supergradient

To match the dual function in our power minimization problem, we rewrite the definition of supergradient (3.17) by

$$d\left(\widetilde{\boldsymbol{\mu}},\widetilde{\boldsymbol{\kappa}}\right) \le d\left(\boldsymbol{\mu},\boldsymbol{\kappa}\right) + \boldsymbol{\chi}_1^T\left(\widetilde{\boldsymbol{\mu}}-\boldsymbol{\mu}\right) + \boldsymbol{\chi}_2^T\left(\widetilde{\boldsymbol{\kappa}}-\boldsymbol{\kappa}\right) \tag{A.17}$$

Based on the definition of the dual function (3.9), we know

$$\begin{aligned} & d\left(\widetilde{\boldsymbol{\mu}},\widetilde{\boldsymbol{\kappa}}\right) \\ &= \min_{\mathbf{r}} \mathcal{L}\left(\mathbf{r},\widetilde{\boldsymbol{\mu}},\widetilde{\boldsymbol{\kappa}}\right) \\ &= \min_{\mathbf{r}_m}\left\{\sum_{m=1}^{M} f_m\left(\mathbf{r}_m\right) + \widetilde{\boldsymbol{\mu}}^T\left(M\mathbf{r}^{\text{tar}} - \sum_{m=1}^{M}\mathbf{r}_m\right) + \widetilde{\boldsymbol{\kappa}}^T\left(\sum_{m=1}^{M} g_m\left(\mathbf{r}_m\right) - \mathbf{p}^{\text{con}}\right)\right\} \\ &\le \sum_{m=1}^{M} f_m\left(\mathbf{r}_m^*\right) + \widetilde{\boldsymbol{\mu}}^T\left(M\mathbf{r}^{\text{tar}} - \sum_{m=1}^{M}\mathbf{r}_m^*\right) + \widetilde{\boldsymbol{\kappa}}^T\left(\sum_{m=1}^{M} g_m\left(\mathbf{r}_m^*\right) - \mathbf{p}^{\text{con}}\right). \end{aligned} \tag{A.18}$$

We have the inequality because $\mathbf{r}^* = \min_{\mathbf{r}} \mathcal{L}(\mathbf{r}, \boldsymbol{\mu}, \boldsymbol{\kappa})$, which is not actually (but still possible) the optimal solution that minimize $\mathcal{L}(\mathbf{r}, \widetilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{\kappa}})$. After inserting some tricky terms, (A.18) becomes

$$
\begin{aligned}
& d(\widetilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{\kappa}}) \\
\leq & \sum_{m=1}^{M} f_m(\mathbf{r}_m^*) + \boldsymbol{\mu}^T \left(M\mathbf{r}^{\text{tar}} - \sum_{m=1}^{M} \mathbf{r}_m^* \right) + (\widetilde{\boldsymbol{\mu}} - \boldsymbol{\mu})^T \left(M\mathbf{r}^{\text{tar}} - \sum_{m=1}^{M} \mathbf{r}_m^* \right) \\
& + \boldsymbol{\kappa}^T \left(\sum_{m=1}^{M} g_m(\mathbf{r}_m^*) - \mathbf{p}^{\text{con}} \right) + (\widetilde{\boldsymbol{\kappa}} - \boldsymbol{\kappa})^T \left(\sum_{m=1}^{M} g_m(\mathbf{r}_m^*) - \mathbf{p}^{\text{con}} \right) \\
= & d(\boldsymbol{\mu}, \boldsymbol{\kappa}) + (\widetilde{\boldsymbol{\mu}} - \boldsymbol{\mu})^T \left(M\mathbf{r}^{\text{tar}} - \sum_{m=1}^{M} \mathbf{r}_m^* \right) + (\widetilde{\boldsymbol{\kappa}} - \boldsymbol{\kappa})^T \left(\sum_{m=1}^{M} g_m(\mathbf{r}_m^*) - \mathbf{p}^{\text{con}} \right) \\
= & d(\boldsymbol{\mu}, \boldsymbol{\kappa}) + \left(M\mathbf{r}^{\text{tar}} - \sum_{m=1}^{M} \mathbf{r}_m^* \right)^T (\widetilde{\boldsymbol{\mu}} - \boldsymbol{\mu}) + \left(\sum_{m=1}^{M} g_m(\mathbf{r}_m^*) - \mathbf{p}^{\text{con}} \right)^T (\widetilde{\boldsymbol{\kappa}} - \boldsymbol{\kappa}),
\end{aligned}
\tag{A.19}
$$

which satisfies the definition of the supergradient (A.17) and hence verifies (3.18) and (3.19).

REFERENCES

[1] S. Catreux, P. F. Driessen, and L. J. Greenstein, "Simulation results for an interference-limited multiple-input multiple-output cellular system," *IEEE Commun. Lett.*, vol. 4, no. 11, pp. 334–336, Nov. 2000.

[2] D. Gesbert, S. Hanly, H. Huang, S. Shamai, O. Simeone, and W. Yu, "Multicell MIMO cooperative networks: A new look at interference," *IEEE J. Select. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.

[3] M. K. Karakayali, G. J. Foschini, and R. A. Valenzuela, "Network coordination for spectrally efficient communications in cellular systems," *IEEE Wireless Commun.*, vol. 13, no. 4, pp. 56–61, Aug. 2006.

[4] H. Zhang and H. Dai, "Cochannel interference mitigation and cooperative processing in downlink multicell multiuser MIMO networks," *EURASIP J. Wireless Commun.*, pp. 222–235, Dec. 2004.

[5] J. Zhang, R. Chen, J. G. Andrews, A. Ghosh, and R. W. Heath Jr., "Networked MIMO with clustered linear precoding," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 1910–1921, Apr. 2009.

[6] P. Marsch and G. Fettweis, "A framework for optimizing the downlink performance of distributed antenna systems under a constrained backhaul," in *Proc. European Wireless Conf. (EW)*, Paris, France, Apr. 2007.

[7] U. Jang, H. Son, J. Park, and S. Lee, "CoMP-CSB for ICI nulling with user selection," *IEEE Trans. Wireless Commun.*, vol. 10, no. 9, pp. 2982–2993, Sep. 2011.

[8] D. Gesbert, S. G. Kiani, A. Gjendemsj, and G. E. Ien, "Adaptation, coordination, and distributed resource allocation in interference-limited wireless networks," in *Proc. IEEE*, vol. 95, no. 12, pp. 2393–2409, Dec. 2007.

[9] P. Marsch and G. Fettweis, "Static Clustering for cooperative multi-point (CoMP) in mobile communications," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kyoto, Japan, pp. 1–6, Jun. 2011.

[10] A. Papadogiannis, D. Gesbert, and E. Hardouin, "A dynamic clustering approach in wireless networks with multi-cell cooperative processing." in *Proc. IEEE Int. Conf. Commun. (ICC)*, Beijing, China, pp. 4033–4037, May 2008.

[11] K. Seong, M. Mohseni, and J. M. Cioffi, "Optimal resource allocation for OFDMA downlink systems," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Seattle, USA, pp. 1394–1398, Jul. 2006.

[12] W. W. L. Ho and Y. C. Liang, "Optimal resource allocation for multiuser MIMO-OFDM systems with user rate constraints," *IEEE Trans. Veh. Technol.*, vol. 58, no. 3, pp. 1190–1203, Mar. 2009.

[13] L. Choi and R.D. Murch, "A transmit preprocessing technique for multiuser MIMO systems using a decomposition approach," *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 773–786, Jul. 2003.

[14] B. S. Krongold and D. L. Jones, "An active-set approach for OFDM PAR reduction via tone reservation," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 495–509, Feb. 2004.

[15] B. S. Krongold and D. L. Jones, "PAR reduction in OFDM via active constellation extension," *IEEE Trans. Broadcast.*, vol. 49, no. 3, pp. 258–268, Sep. 2003.

[16] B. S. Krongold, G. R. Woo, and D. L. Jones, "Fast active constellation extension for MIMO-OFDM PAR reduction," in *Proc. Asilomar Conf. Signals, Syst., Comupt.*, California, USA, pp. 1476–1479, Oct. 2005.

[17] A. J. Goldsmith and S. -G. Chua. "Variable-rate variable-power MQAM for fading channels," *IEEE Trans. Commun.*, vol. 45, no. 10, pp. 1218–1230, Oct. 1997.

[18] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge Univ. Press, 2004.

[19] W. Yu and R. Lui, "Dual methods for nonconvex spectrum optimization of multicarrier systems," *IEEE Trans. Commun.*, vol. 54, no. 7, pp. 1310–1322, Jul. 2006.

[20] R. Freund, "15.084J/6.252J Nonlinear Programming," *MIT OpenCourseWare*, Spring 2004.

[21] S. Sesia, I. Toufik, and M. Baker, *LTE: The UMTS Long Term Evolution*. John Wiley and Sons, 2009.

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

[23] S. Litsyn, *Peak Power Control in Multicarrier Communications*. Cambridge Univ. Press, 2007.

[24] F. R. Farrokhi, G. J. Foschini, A. Lozano, and R. A. Valenzuela, "Link-optimal space-time processing with multiple transmit and receive antennas," *IEEE Commun. Lett.*, vol. 5, no. 3, pp. 85–87, Mar. 2001.