Ph.D. Dissertation

A Study on Complexity-reduced Multi-user Signal Transmission in Massive MIMO Environments

대규모 다중 안테나 환경에서 낮은 복잡도의 다중 사용자 신호전송에 관한 연구

Geon-Woong Jung

August 2020

Department of Electrical Engineering and Computer Science
College of Engineering
Seoul National University

Abstract

Advanced wireless communication systems may employ massive multi-input multi-output (m-MIMO) techniques for performance improvement. A base station equipped with an m-MIMO configuration can serve a large number of users by means of beamforming. The m-MIMO channel becomes asymptotically orthogonal to each other as the number of antennas increases to infinity. In this case, we may optimally transmit signal by means of maximum ratio transmission (MRT) with affordable implementation complexity. However, the MRT may suffer from inter-user interference in practical m-MIMO environments mainly due to the presence of insufficient channel orthogonality. The use of zero-forcing beamforming can be a practical choice in m-MIMO environments since it can easily null out inter-user interference. However, it may require huge computational complexity for the generation of beam weight. Moreover, it may suffer from performance loss associated with the interference nulling, referred to transmission performance loss (TPL). The TPL may become serious when the number of users increases or the channel correlation increases in spatial domain.

In this dissertation, we consider complexity-reduced multi-user signal transmission in m-MIMO environments. We determine the beam weight to maximize the signal-to-leakage plus noise ratio (SLNR) instead of signal-to-interference plus noise ratio (SINR). We determine the beam direction assuming combined use of MRT and partial ZF that

partially nulls out interference. For further reduction of computational complexity, we determine the beam weight based on the approximated SLNR.

We consider complexity-reduced ZF beamforming that generates the beam weight in a group-wise manner. We partition users into a number of groups so that users in each group experience low TPL. We approximately estimate the TPL for further reduction of computational complexity. Finally, we determine the beam weight for each user group based on the approximated TPL.

Keywords: massive MIMO, ZF beamforming, complexity-reduced beamforming, user grouping.

Student number: 2013-20874

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Demand for mobile data traffic has been increasing explosively as personal mobile devices and data-intensive mobile applications become popular [1, 2]. Advanced wireless communication systems employ massive multi-input multi-output (m-MIMO) techniques to support the demand [3]–[5]. A base station (BS) equipped with an m-MIMO configuration can serve a large number of users by means of beamforming. The m-MIMO channel becomes asymptotically orthogonal to each other as the number of antennas increases to infinity [6]. In this case, we may optimally transmit signal by means of maximum ratio transmission (MRT) with affordable implementation complexity. However, the channel may not be orthogonal to each other in real operating environments, where the number of antennas is not sufficiently large. The MRT may suffer from inter-user interference due to the presence of insufficient channel orthogonality [7, 8].

The use of zero-forcing (ZF) beamforming can be a practical choice in m-MIMO environments since it can easily null out inter-user interference [9]–[13]. The ZF can null out the inter-user interference by projecting the desired channel onto the null space of interference channel as illustrated in Figure 1-1. However, the interference nulling may

require large computational complexity of an order of $O\left(KM^2 + K^3\right)$, where K is the number of users and M is the number of antennas. It can be seen from Figure 1-2 that the complexity for the generation of ZF beam weight increases exponentially proportional to M, which may not be affordable in m-MIMO operating environments.

A number of works considered the generation of beam weight with reduced computational complexity [14]–[28]. We may reduce the computational complexity by performing matrix inversion in an approximated manner. We may perform matrix inversion approximation (MIA) based on Neumann series by exploiting asymptotic characteristics of the largest and the smallest eigenvalue of Wishart matrix [14, 15] or exploiting channel hardening properties in m-MIMO environments [16]. We may improve the approximation accuracy using an eigen-based MIA technique with coefficients optimized by means of eigenvalue estimation and least-square fitting [17]. We may perform MIA by means of Kapteyn series expansion [18]. We may improve the approximation accuracy by optimizing polynomial coefficients for the MIA in a deterministic equivalent form [19]–[21]. We may reduce the computational complexity by using an iterative technique [22]–[28]. We may avoid the matrix inversion by using Gauss-Seidel method [22]. We can achieve performance similar to the ZF by using Jacobi method without matrix inversion [23]. We may further improve the approximation accuracy by using Jacobi and steepest descent method with affordable increase of complexity compared to the use of Jacobi method [24]. We can generate the beam weight using Newton iteration method by exploiting the property of diagonal dominance of an invertible matrix [25]. We may improve the approximation accuracy of Newton iteration method by using Tchebychev polynomials as an initial matrix for the iteration [26]. We

may reduce the computational complexity by using a relaxation parameter for the generation of beam weight in an iterative manner [27] or using a weighted coefficient to perform the MIA in an iterative manner [28].

These approximation approaches may generate the ZF beam weight with reduced computational complexity, but they may suffer from performance loss due to the approximation inaccuracy. Moreover, when applied to a large number of users, they may not effectively reduce the computational complexity since they may require the computational complexity in proportion to the number of users.

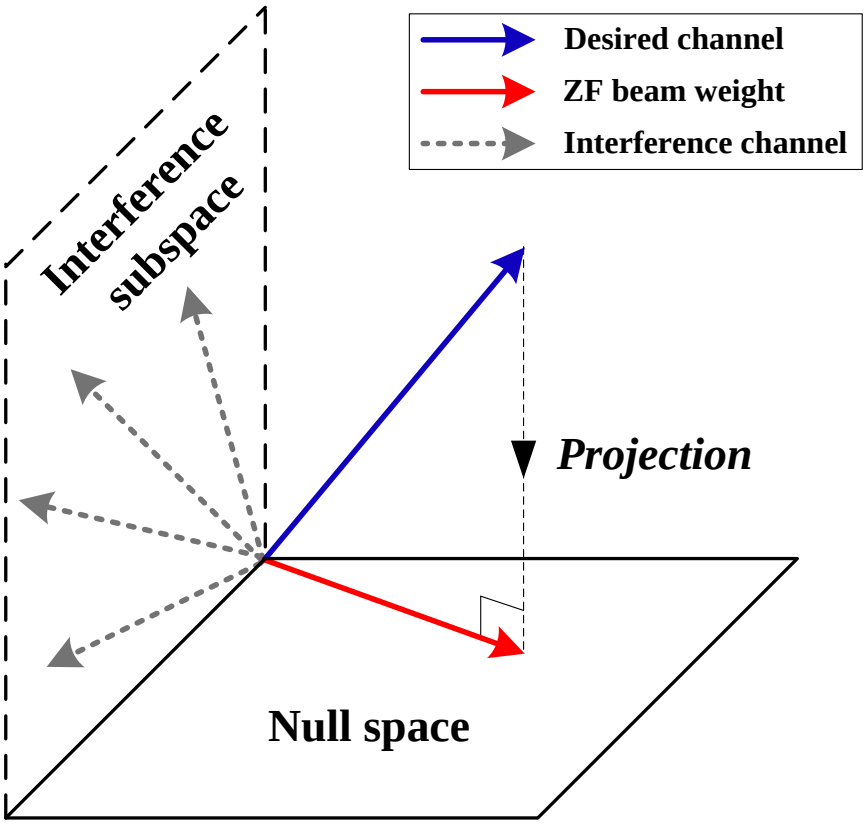


Figure 1-1. Interference nulling by ZF.

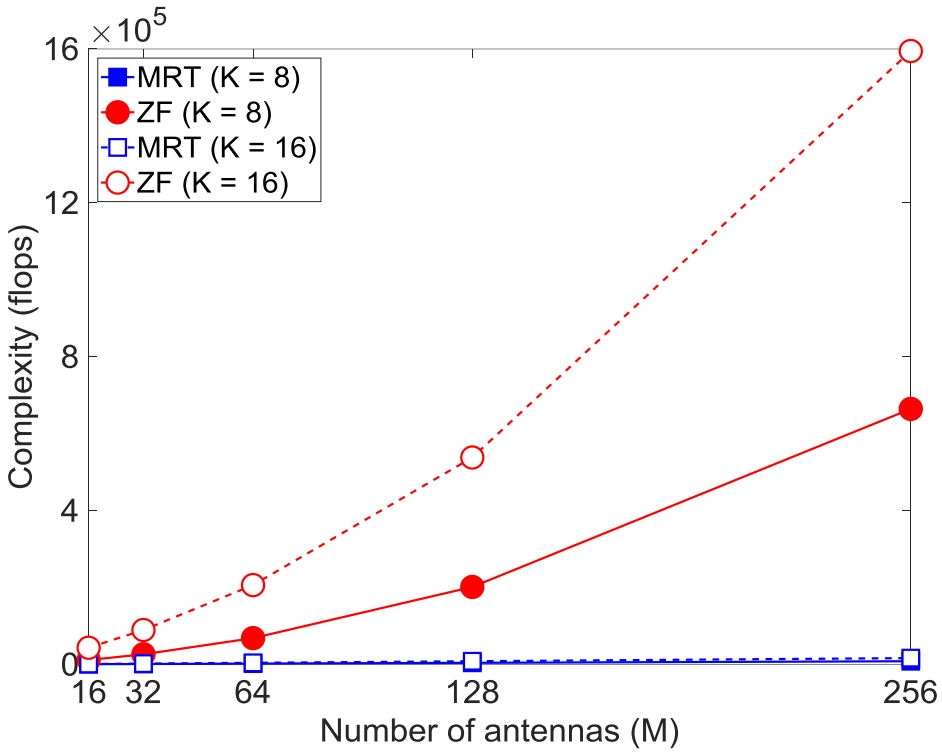


Figure 1-2. Computational complexity according to *M*.

We may generate the beam weight for ZF with reduced computational complexity by decreasing the dimension of nulling subspace [29, 30]. Moreover, the reduced dimension of nulling subspace may increase the spatial degree of freedom (DoF), making it possible to enhance the power of desired signal without noticeable increase of interference. An egoistic ZF technique can decrease the dimension of nulling subspace by only considering intra-cell interfering channel [29]. However, it cannot null out the inter-cell interference, suffering from performance loss. A cell-edge-aware (CEA)-ZF technique can decrease the dimension of nulling subspace while reducing the inter-cell interference by means of selective interference nulling to inter-cell users based on the path loss

between the BS and the inter-cell users [30]. However, it may not provide noticeable performance improvement since it decreases the dimension of nulling subspace without consideration of transmission environments other than the path loss.

In this dissertation, we consider multi-user beamforming with reduced complexity in m-MIMO environments. We may determine the beam weight to maximize the signal-to-interference plus noise ratio (SINR), which may require large computational complexity mainly due to the non-convexity property of the SINR [31]–[34]. Instead, we determine the beam weight to maximize the signal-to-leakage plus noise ratio (SLNR). The SLNR is defined by the ratio between the desired signal power to target user and the total interference power to other users (*i.e.*, the leakage power) plus noise. The Max SLNR technique may maximize the SLNR by invoking Rayleigh–Ritz quotient theorem [35, 36]. However, it may require computational complexity in cubic proportion to the number of antennas, which may not be affordable in m-MIMO environments. For further reduction of computational complexity, we consider the use of a parameterized beamforming (PB) technique that employs MRT and ZF beamforming, where we may use a single parameter to determine the beam weight [37, 38]. As the number of users increases, the ZF may suffer from performance loss due to the presence of insufficient spatial DoF. To alleviate this problem, we may employ a partial ZF (PZF) beamforming technique that selectively nulls out interference, making it possible to reduce the dimension of nulling subspace (*i.e.*, to exploit more spatial DoF). We may reduce the performance loss by increasing the spatial DoF, while reducing the computational complexity. With combined use of MRT and PZF, we design a multi-user parameterized beamforming (MUPB) scheme that maximizes the SLNR. For further reduction of computational complexity, we

approximately estimate the SNLR by exploiting channel characteristics in m-MIMO environments [39]–[41]. We analytically design the MUPB based on the approximated SNLR by applying the zero-gradient condition [42]. We analyze the effect of channel state information (CSI) inaccuracy on the transmission performance and the computational complexity [43, 44]. Finally, we verify the performance of MUPB by computer simulation.

The ZF may suffer from performance loss due to the interference nulling, referred to transmission performance loss (TPL). This is mainly because the beam direction of ZF may be different from that of desired channel. The ZF can effectively null out the interference when the beam weight lies in the null space of interference channel [9]. It can be seen from Figure 1-3 that the TPL may be serious unless the spatial DoF is sufficiently large when the number of users is large or the channel is spatially correlated [45]–[48]. It can also be seen that when the angular spread (AS) is small, users may seriously suffer from the TPL mainly due to the presence of channel correlation [49, 50].

We may alleviate the TPL problem by means of user grouping [51]–[60]. We may perform the user grouping to maximize the sum rate by means of exhaustive searching. However, it may not be practical mainly due to prohibitively large complexity. The use of a greedy grouping technique may be a practical choice [51, 52]. However, it may still require large computational complexity since it needs to generate the beam weight for each group.

A number of works considered complexity-reduced user grouping [53]–[60]. We may partition users into a number of groups to minimize the sum of channel correlation in each group [53] or to keep the channel correlation lower than a certain level [54]. We

may select a user with the largest channel gain and then allocate it to a group to minimize the maximum of channel correlation [55]. We may partition users into a number of groups by means of clustering [56]–[59]. We first partition users into a number of clusters by exploiting the orthogonality of spatial correlation matrix (SCM) [56]–[58] or the channel gain [59]. We may generate user groups by selecting users from each cluster based on an approximated sum rate [57], the largest eigenvalue of SCM [58] or a random manner [56, 59]. We may partition users into a number of groups based on the chordal distance by employing an agglomerative hierarchical clustering (AHC) method [60, 61]. However, these schemes do not use a grouping metric that may properly characterize the TPL.

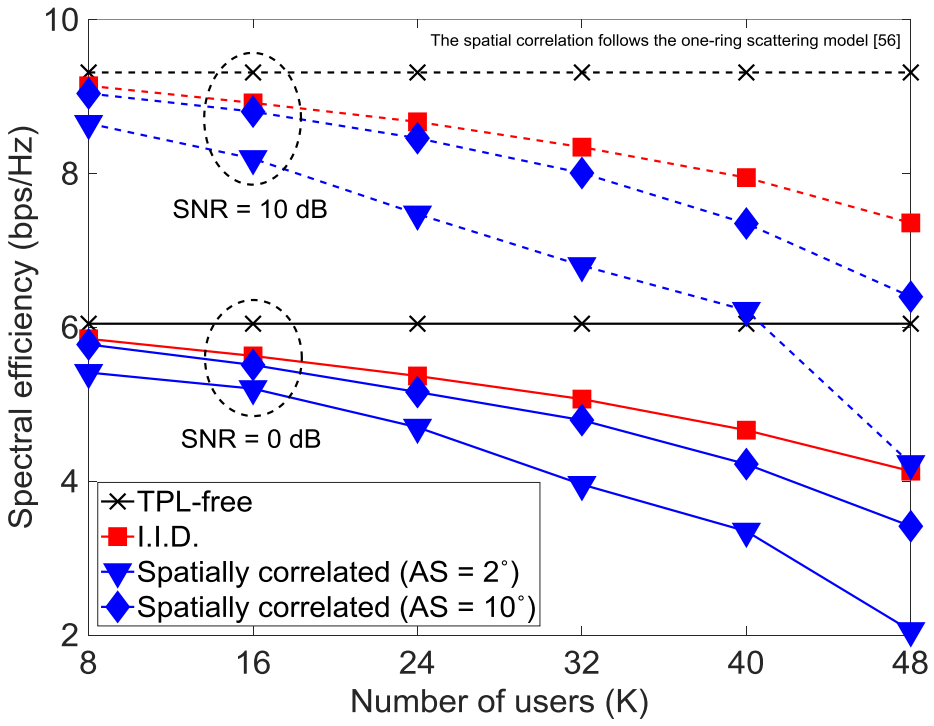


Figure 1-3. Spectral efficiency of ZF according to K when $M = 64$.

In this dissertation, we consider complexity-reduced ZF beamforming by means of user grouping. We consider the user grouping by using the TPL as the grouping metric. However, it may require a complexity of $O\left(M^3\right)$ for the estimation of TPL, which may not be affordable in m-MIMO environments. We approximately estimate the TPL in a deterministic equivalent form without consideration of beam weight. We first analyze the received signal in a deterministic equivalent form and then approximate the TPL by exploiting the property of m-MIMO channel. We partition users into a number of groups based on the approximated TPL so that users in each group experience TPL lower than a certain level. We generate the beam weight for each user group assuming that we allocate transmission resource to the user groups orthogonal to each other. When users are densely located in a small area, they may experience channel highly correlated to each other, suffering from the increase of TPL [56, 57]. We may reduce the TPL by increasing the number of groups, but we may have to allocate reduced transmission resource to each group (*i.e.*, reduction of transmission rate). We may alleviate this problem by means of re-grouping of users based on the approximated TPL. We analyze the effect of CSI inaccuracy on the transmission performance and the computational complexity. Finally, we evaluate the performance of the proposed scheme by computer simulation.

Following Introduction, Chapter 2 describes the system model in consideration. Chapter 3 describes the proposed multi-user beamforming scheme with reduced complexity. Chapter 4 describes the proposed ZF transmission based on user grouping. Finally, Chapter 5 summarizes concluding remarks and further research issues.

Notations: Boldface letter $\mathbf{X}$ and $\mathbf{x}$ denote a matrix and a column vector, respectively. $\mathbf{X}^T$, $\mathbf{X}^H$, and $\mathrm{Tr}\left(\mathbf{X}\right)$ denote the transpose, the conjugate transpose, and the trace of $\mathbf{X}$,

respectively. $\mathbf{I}_M$ and $\mathbf{0}_M$ denote an $(M \times M)$ identity matrix and an $(M \times M)$ zero matrix, respectively. $\|\mathbf{x}\|$ denotes the Euclidean norm of $\mathbf{x}$. $\mathbf{x} \sim CN(\mathbf{m}, \mathbf{C})$ refers to that $\mathbf{x}$ is a circularly-symmetric complex Gaussian random vector with mean $\mathbf{m}$ and covariance $\mathbf{C}$. $\mathbb{C}^{M \times 1}$ denotes all sets of M dimensional complex column vectors. The notation "$\xrightarrow[M \to \infty]{\text{a.s.}}$" refers to almost sure convergence.

Chapter 2

System model

We consider a cellular system comprising B BSs, where BS i serves K_i users and $1 \le i \le B$. We assume that each BS and users are equipped with M antennas and a single antenna, respectively, and that the BS has perfect channel state information (CSI) of all users. Let Ω, U, and K_{total} be the set of BSs, the set of total users, and the total number of users, respectively. Then, we may represent U as

$$U = \sum_{i \in \Omega} U_i \tag{2.1}$$

where U_i denotes the set of users of BS i and can be represented as

$$U_i = \left\{ k \mid \sum_{j=1}^{i-1} K_j < k \le \sum_{j=1}^{i-1} K_j + K_i \right\}. \tag{2.2}$$

We consider two transmission techniques; coordinated transmission and joint transmission [62, 63]. BS i can transmit signal to a user in U_i by means of coordinated transmission, while mitigating interference to intra-cell users and inter-cell users as well. We may represent the signal received by user k as

$$y_k = \sqrt{p_{i,k}\alpha_{i,k}}\mathbf{h}_{i,k}^H \mathbf{v}_{i,k} s_k + \sum_{l \in U_i \setminus \{k\}} \sqrt{p_{i,l}\alpha_{i,k}}\mathbf{h}_{i,k}^H \mathbf{v}_{i,l} s_l + \sum_{j \in \Omega \setminus \{i\}} \sum_{l \in U_j} \sqrt{p_{j,l}\alpha_{j,k}}\mathbf{h}_{j,k}^H \mathbf{v}_{j,l} s_l + n_k \quad (2.3)$$

where $p_{i,k}$ is the transmit power of BS i to user k, $\alpha_{i,k}$ is the path loss between BS i and user k, $\mathbf{h}_{i,k} \in \mathbb{C}^{M\times 1} \sim CN(\mathbf{0}_M, \mathbf{I}_M)$ is the channel between BS i and user k, $\mathbf{v}_{i,k} \in \mathbb{C}^{M\times 1}$ is the normalized beam weight for user k of BS i, s_k is the data for user k, and n_k is zero-mean complex circular-symmetric additive white Gaussian noise (AWGN). We define the signal-to-interference plus noise ratio (SINR) of user k by

$$\text{SINR}_k = \frac{S_k}{I_k + \sigma_n^2} \quad (2.4)$$

where S_k is the desired signal power of user k and can be represented as

$$S_k = p_{i,k}\alpha_{i,k}\left|\mathbf{h}_{i,k}^H \mathbf{v}_{i,k}\right|^2 \quad (2.5)$$

I_k is the interference power to user k and can be represented as

$$I_k = I_k^{\text{Intra}} + I_k^{\text{Inter}} \quad (2.6)$$

and σ_n^2 is the noise power. Here, I_k^{Intra} is the interference power to user k generated by the signal transmission to intra-cell users and can be represented as

$$I_k^{\text{Intra}} = \sum_{l \in U_i \setminus \{k\}} I_{l,k} \quad (2.7)$$

and I_k^{Inter} is the interference power to user k generated by the signal transmission to inter-cell users and can be represented as

$$I_k^{\text{Inter}} = \sum_{j \in \Omega \setminus \{i\}} \sum_{l \in U_j} I_{l,k} \quad (2.8)$$

where $I_{l,k}$ is the interference power to user *k* generated by the signal transmission to user *l* and can be represented as

$$I_{l,k} = \begin{cases} p_{i,l}\alpha_{i,k}\left|\mathbf{h}_{i,k}^H \mathbf{v}_{i,l}\right|^2, & \text{for } l \in U_i \setminus \{k\} \\ p_{j,l}\alpha_{j,k}\left|\mathbf{h}_{j,k}^H \mathbf{v}_{j,l}\right|^2, & \text{for } l \in U_j,\ j \in \Omega \setminus \{i\} \end{cases}. \tag{2.9}$$

Multiple BSs can jointly transmit signal to a user in *U* by means of joint transmission. We may represent the signal received by user *k* as

$$y_k = \mathbf{h}_k^H \mathbf{v}_k s_k + \sum_{l \in U \setminus \{k\}} \mathbf{h}_k^H \mathbf{v}_l s_l + n_k \tag{2.10}$$

where $\mathbf{h}_k = \left[\sqrt{\alpha_{1,k}}\mathbf{h}_{1,k}^T \cdots \sqrt{\alpha_{B,k}}\mathbf{h}_{B,k}^T\right]^T \in \mathbb{C}^{BM\times 1}$ is the channel vector from *B* BSs to user *k*, $\mathbf{v}_k = \left[\sqrt{p_{1,k}}\mathbf{v}_{1,k}^T \cdots \sqrt{p_{B,k}}\mathbf{v}_{B,k}^T\right]^T \in \mathbb{C}^{BM\times 1}$ is the beam weight for user *k*. We may represent the desired signal power of user *k* as

$$S_k = \left|\sum_{i\in\Omega} \sqrt{p_{i,k}\alpha_{i,k}}\mathbf{h}_{i,k}^H \mathbf{v}_{i,k}\right|^2. \tag{2.11}$$

We may represent the interference power to user *k* as

$$I_k = \sum_{l \in U \setminus \{k\}} I_{l,k} \tag{2.12}$$

where $I_{l,k}$ denotes the interference power to user *k* generated by the signal transmission to user *l* and can be represented as

$$I_{l,k} = \left|\sum_{i\in\Omega} \sqrt{p_{i,l}\alpha_{i,k}}\mathbf{h}_{i,k}^H \mathbf{v}_{i,l}\right|^2. \tag{2.13}$$

We may represent the corresponding achievable transmission rate of user k as

$$r_k = \log_2\left(1+\mathrm{SINR}_k\right). \tag{2.14}$$

We may represent the corresponding sum rate as

$$r_{\mathrm{sum}} = \sum_{i\in\Omega}\sum_{k\in U_i} r_k \ . \tag{2.15}$$

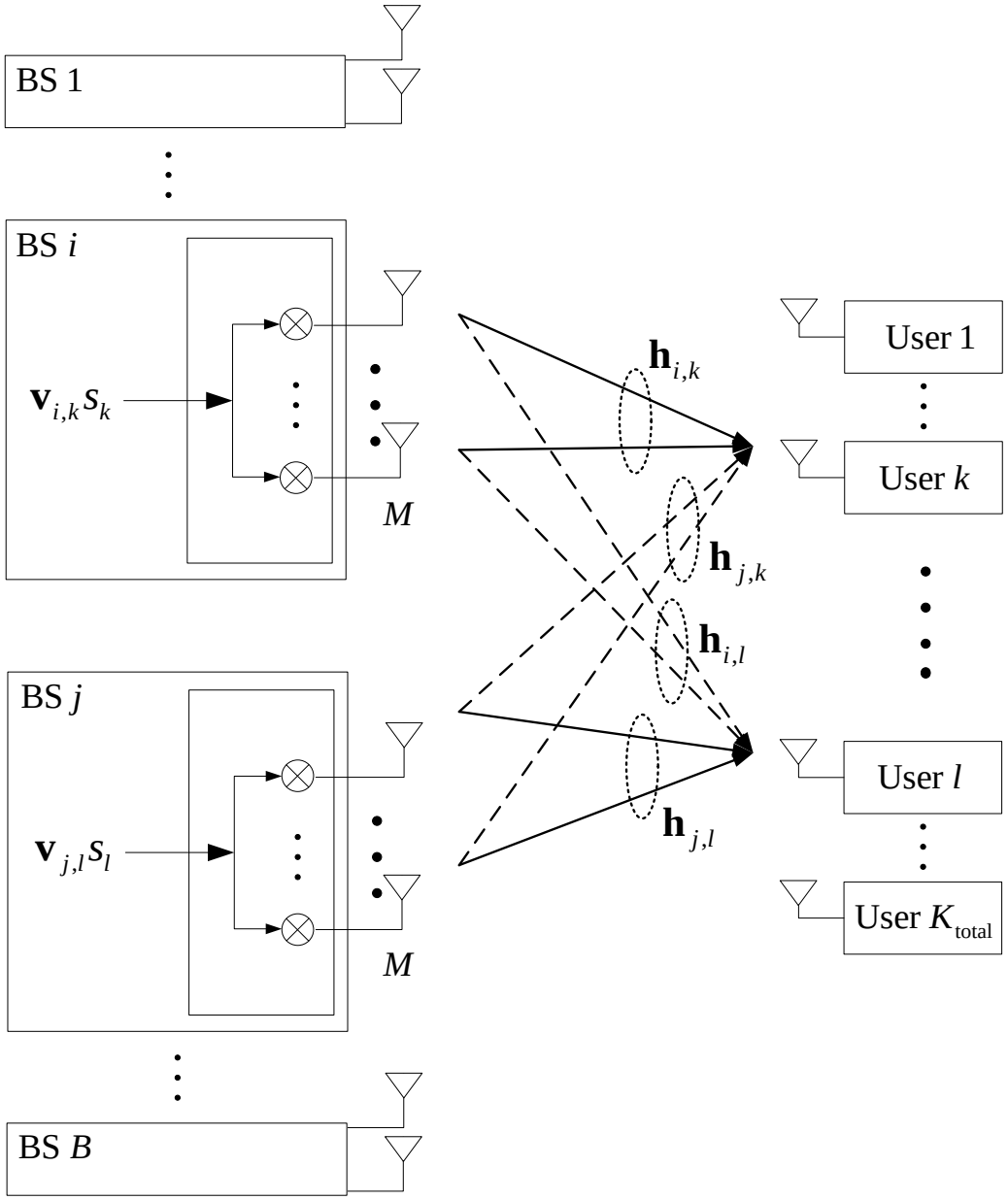


Figure 2-1. A cellular system with an m-MIMO configuration.

Chapter 3

Complexity-reduced multi-user signal transmission

In this chapter, we consider complexity-reduced multi-user transmission in m-MIMO environments. We determine the beam weight to maximize the SLNR instead of SINR. We consider the signal transmission by using MRT and partial ZF that partially nulls out interference. For further reduction of computational complexity, we approximately estimate the SLNR for the generation of beam weight.

3.1. Previous works

We briefly review previous works relevant to the design of transmission schemes in this dissertation.

3.1.1. Maximum ratio transmission (MRT)

The use of MRT may be a practical choice in consideration of low implementation complexity [6]. The MRT beam weight for user k of BS i can be determined by

$$\mathbf{v}_{i,k}^{\mathrm{MRT}} = \frac{\mathbf{h}_{i,k}}{\left\|\mathbf{h}_{i,k}\right\|}. \tag{3.1}$$

It can be shown that as M increases to infinity, the channel is asymptotically orthogonal to each other [6]

$$\frac{1}{M}\mathbf{h}_{i,k}^{H}\mathbf{h}_{i,l} \xrightarrow[M\to\infty]{\text{a.s.}} 0. \tag{3.2}$$

Then, it can be shown that

$$I_{l,k} \xrightarrow[M\to\infty]{\text{a.s.}} 0. \tag{3.3}$$

When M increases to infinity, the MRT can provide optimal transmission performance due to the disappearance of interference. However, the MRT may suffer from inter-user interference due to the presence of insufficient orthogonality in practical operating environments [7, 8].

3.1.2. Zero-forcing (ZF) beamforming

The use of ZF beamforming may be a practical choice in m-MIMO environments since it can easily null out inter-use interference [9]–[13]. The ZF beam weight for user k of BS i can be determined by

$$\mathbf{v}_{i,k}^{\mathrm{ZF}} = \frac{\mathbf{w}_{i,k}}{\left\|\mathbf{w}_{i,k}\right\|} \tag{3.4}$$

where $\mathbf{w}_{i,k}$ is the k-th column vector of $\mathbf{H}_i\mathbf{G}_i^{-1}$ and $\mathbf{G}_i = \mathbf{H}_i^{H}\mathbf{H}_i$. Here,

$$\mathbf{H}_i = \left[\mathbf{H}_{i,1}\cdots\mathbf{H}_{i,B}\right] \tag{3.5}$$

where $\mathbf{H}_{i,j}$ is the channel between BS *i* and users of BS *j*, and can be represented as

$$\mathbf{H}_{i,j} = \left[\mathbf{h}_{i,l_1} \cdots \mathbf{h}_{i,l_{K_j}} \right]. \tag{3.6}$$

Here, l_n denotes the *n*-th user in U_j.

As illustrated in Figure 3-1, the ZF can suppress interference by projecting the desired channel vector onto the null space of interference channel. However, the generation of ZF beam weight may require a complexity of $O\left(K_{\text{total}} M^2 + K_{\text{total}}^3\right)$, which may not be affordable in m-MIMO environments. Moreover, the ZF may suffer from performance loss, referred to transmission performance loss (TPL), due to the interference nulling. The TPL may occur since the beam direction of ZF may be different from that of desired channel due to the projection operation. The TPL may increase as the number of users increases or the channel correlation increases [45]–[48].

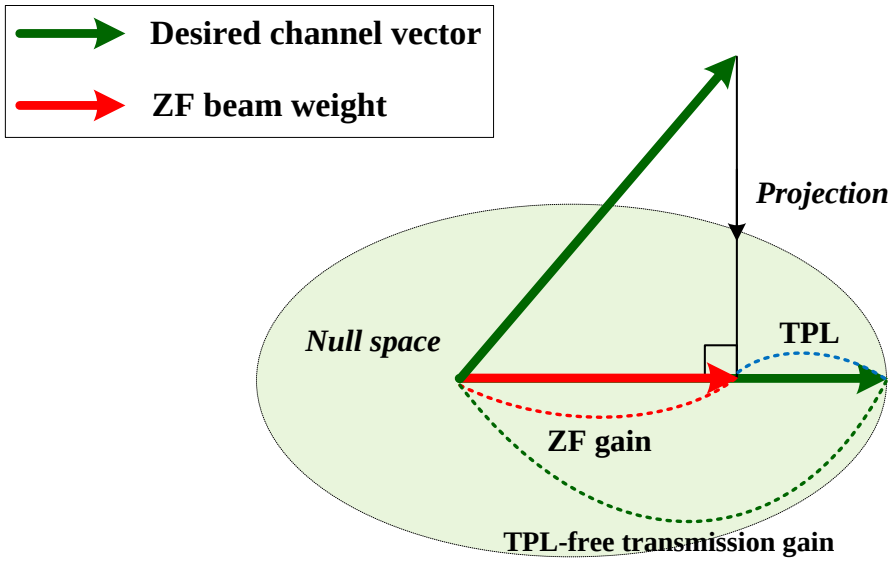


Figure 3-1. TPL due to interference nulling of ZF.

3.1.3. ZF based on Neumann series

We may reduce the computational complexity of ZF by exploiting Neumann series (NS) [14]. As the number of antennas and users increases, the eigenvalues of $\mathbf{G}_i$ converge to a fixed deterministic distribution, referred to Marchenko-Pastur distribution [64]. As M and K_{total} increase to infinity, the largest and the smallest eigenvalue of $\mathbf{G}_i$ converge to, respectively, [7]

$$\lambda_{i,\max}\left(\mathbf{G}_i\right)=\left(1+\frac{1}{\sqrt{\beta}}\right)^2 \tag{3.7}$$

$$\lambda_{i,\min}\left(\mathbf{G}_i\right)=\left(1-\frac{1}{\sqrt{\beta}}\right)^2 \tag{3.8}$$

where

$$\beta=\frac{M}{K_{\text{total}}}. \tag{3.9}$$

By scaling $\mathbf{G}_i$ by a factor of $\beta/\left(1+\beta\right)$, it can be shown that

$$\lambda_{i,\max}\left(\frac{\beta}{1+\beta}\mathbf{G}_i\right)=1+2\frac{\sqrt{\beta}}{1+\beta} \tag{3.10}$$

$$\lambda_{i,\max}\left(\frac{\beta}{1+\beta}\mathbf{G}_i\right)=1-2\frac{\sqrt{\beta}}{1+\beta}. \tag{3.11}$$

It can also be shown that

$$\lim_{\beta\to+\infty}\left(-2\frac{\sqrt{\beta}}{1+\beta},2\frac{\sqrt{\beta}}{1+\beta}\right)=(-0,0) \tag{3.12}$$

and

$$\lim_{n\to+\infty}\left(\mathbf{I}_{K_{\text{total}}}-\frac{1}{M+K_{\text{total}}}\mathbf{G}_i\right)^n \simeq \mathbf{0}_{K_{\text{total}}}. \tag{3.13}$$

When $\mathbf{G}_i$ satisfies the condition (3.13), the inverse of $\mathbf{G}_i$ can be approximated as [65]

$$\mathbf{G}_i^{-1}\approx\frac{\delta}{M+K_{\text{total}}}\sum_{n=0}^{N_{\text{iter}}}\left(\mathbf{I}_{K_{\text{total}}}-\frac{\delta}{M+K_{\text{total}}}\mathbf{G}_i\right)^n \tag{3.14}$$

where N_{iter} is the number of iterations, $\delta(<1)$ is an attenuation factor to be determined. Note that the equality holds when N_{iter} increases to infinity.

The accuracy of this approximation increases as N_{iter} increases. However, we may reduce the complexity for the matrix inversion by exploiting NS with a small N_{iter}. It may be desirable to consider the trade-off between the implementation complexity and the performance loss due to the inaccuracy of approximation.

3.1.4. ZF based on weighted two-stage

We may reduce the computational complexity of ZF by using a weighted two-stage method [28]. We may approximately inverse a matrix using a two-stage method with the use of a weighted coefficient in an iterative manner.

When we employ the ZF, we may represent the transmitted signal by BS *i* as

$$\mathbf{t}_i = \mathbf{G}_i^{-1}\mathbf{s}_i \tag{3.15}$$

where $\mathbf{s}_i = \left[s_1 \cdots s_{K_{\text{total}}} \right]^T$. We may approximate $\mathbf{G}_i^{-1}$ with a weighted coefficient μ in an iterative manner as

$$\hat{\mathbf{t}}_i^{(n+1)} = (1-\mu)\mathbf{t}_i^{(n+1)} + \mu\mathbf{t}_i^{(n+1/2)} \tag{3.16}$$

where

$$\mathbf{t}_i^{(n+1/2)} = -(\mathbf{D}_i + \mathbf{L}_i)^{-1}\mathbf{L}_i^H\mathbf{t}_i^{(n)} + (\mathbf{D}_i + \mathbf{L}_i)^{-1}\mathbf{s}_i \tag{3.17}$$

$$\mathbf{t}_i^{(n+1)} = -(\mathbf{D}_i + \mathbf{L}_i^H)^{-1}\mathbf{L}_i\mathbf{t}_i^{(n+1/2)} + (\mathbf{D}_i + \mathbf{L}_i^H)^{-1}\mathbf{s}_i . \tag{3.18}$$

Here, $\mathbf{L}_i$ and $\mathbf{D}_i$ respectively denote the strict lower triangular and diagonal matrix of $\mathbf{G}_i$, and μ is the weighted coefficient and can be represented as

$$\mu = \left(\frac{K_{\text{total}}}{M} \right)^2 . \tag{3.19}$$

We can generate the beam weight by (3.16) with reduced computational complexity. The accuracy of matrix inversion may increase as the number of iterations increases. It may be desirable to consider the trade-off between the computational complexity and the performance loss due to the approximation inaccuracy.

3.1.5. Egoistic ZF

We may reduce the computational complexity by reducing the dimension of nulling subspace. An egoistic ZF scheme only considers interference nulling to intra-cell users, as illustrated in Figure 3-2 [29]. Let the nulling user set be a set of users whose interference is nulled out. BS *i* can determine the nulling user set by

$$U_{\mathrm{E},i}^{\mathrm{o}} = U_i \,. \tag{3.20}$$

The beam weight for user *k* of BS *i* by the egoistic ZF can be determined by

$$\mathbf{v}_{i,k}^{\text{E-ZF}} = \frac{\mathbf{w}_{\mathrm{E},i,k}^{\mathrm{o}}}{\left\| \mathbf{w}_{\mathrm{E},i,k}^{\mathrm{o}} \right\|} \tag{3.21}$$

where $\mathbf{w}_{\mathrm{E},i,k}^{\mathrm{o}}$ is the *k*-th column vector of $\mathbf{H}_{\mathrm{E},i,k}^{\mathrm{o}} \left(\left(\mathbf{H}_{\mathrm{E},i,k}^{\mathrm{o}} \right)^H \mathbf{H}_{\mathrm{E},i,k}^{\mathrm{o}} \right)^{-1}$. Here,

$$\mathbf{H}_{\mathrm{E},i,k}^{\mathrm{o}} = \mathbf{H}_{i,i} \,. \tag{3.22}$$

The egoistic ZF may reduce the computational complexity by only taking care of interference nulling to intra-cell users. However, it may suffer from the interference from inter-cell BSs.

3.1.6. Cell-edge-aware zero-forcing (CEA-ZF)

The CEA-ZF can selectively null out the interference to inter-cell users based on the path loss between the BS and the inter-cell users, as illustrated in Figure 3-3 [30]. Let $U_{\mathrm{C},i}$ be the set of inter-cell users whose distance from BS *i* is shorter than other BSs except the serving BS. Then, BS *i* may determine the nulling user set by

$$U_{\mathrm{C},i}^{\mathrm{o}} = U_i \bigcup U_{\mathrm{C},i} . \tag{3.23}$$

The beam weight for user k of BS i by the CEA-ZF can be determined by

$$\mathbf{v}_{i,k}^{\text{CEA-ZF}} = \frac{\mathbf{w}_{\mathrm{C},i,k}^{\mathrm{o}}}{\left\| \mathbf{w}_{\mathrm{C},i,k}^{\mathrm{o}} \right\|} \tag{3.24}$$

where $\mathbf{w}_{\mathrm{C},i,k}^{\mathrm{o}}$ is the k-th column vector of $\mathbf{H}_{\mathrm{C},i,k}^{\mathrm{o}} \left(\left(\mathbf{H}_{\mathrm{C},i,k}^{\mathrm{o}} \right)^H \mathbf{H}_{\mathrm{C},i,k}^{\mathrm{o}} \right)^{-1}$. Here,

$$\mathbf{H}_{\mathrm{C},i,k}^{\mathrm{o}} = \left[\mathbf{H}_{i,i} \ \mathbf{H}_{\mathrm{C},i} \right] \tag{3.25}$$

where $\mathbf{H}_{\mathrm{C},i}$ is the channel between BS i and inter-cell users in $U_{\mathrm{C},i}$ and can be represented as

$$\mathbf{H}_{\mathrm{C},i} = \left[\mathbf{h}_{i,l_{\mathrm{C},i,1}} \cdots \mathbf{h}_{i,l_{\mathrm{C},i,K_{\mathrm{C},i}}} \right]. \tag{3.26}$$

Here, $l_{\mathrm{C},i,n}$ denotes the n-th user in $U_{\mathrm{C},i}$ and $K_{\mathrm{C},i}$ is the number of users in $U_{\mathrm{C},i}$.

The CEA-ZF may reduce the computational complexity by decreasing the nulling subspace. It may reduce the inter-cell interference by adaptively determining the nulling user set based on the path loss between BSs and inter-cell users. However, it may not provide noticeable performance improvement since it decreases the dimension of nulling subspace without consideration of transmission environments other than the path loss.

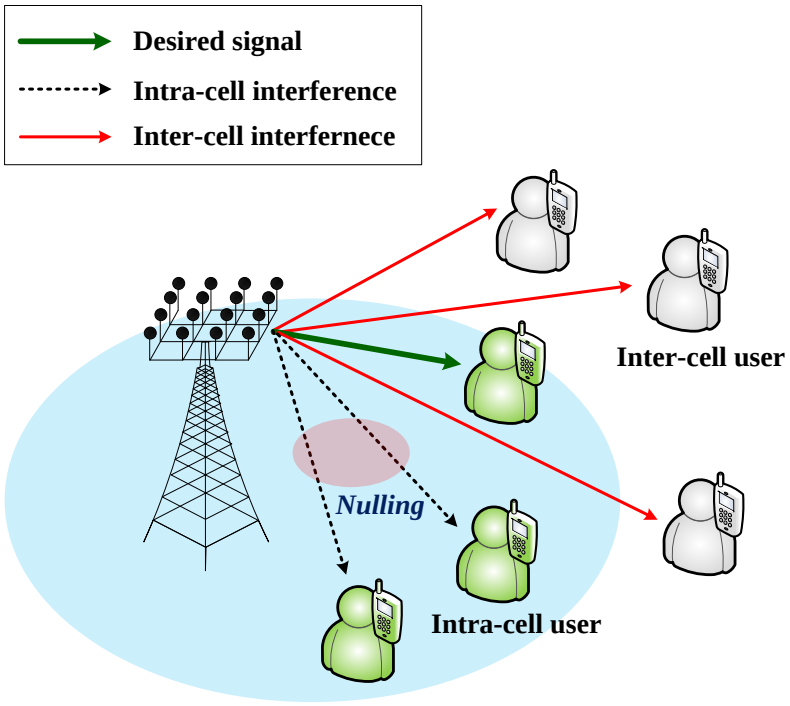


Figure 3-2. An egoistic ZF scheme.

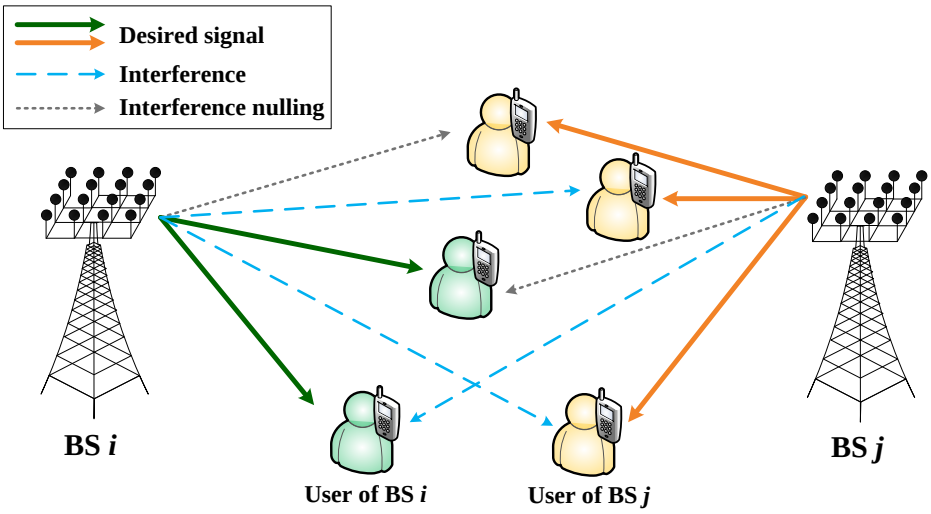


Figure 3-3. A CEA-ZF scheme.

3.2. Proposed scheme

3.2.1. Beam weight design

We may determine the beam weight to maximize the SINR [31]–[34]. However, it may require large computational complexity due to the non-convexity of the SINR. To alleviate the complexity problem, we determine the beam weight to maximize the SLNR. We may represent the SLNR of user *k* as [35, 36]

$$\mathrm{SLNR}_k = \frac{S_k}{L_k + \sigma_n^2} \tag{3.27}$$

where L_k is the leakage power to the other users by $\mathbf{v}_{i,k}$ and can be represented as

$$L_k = L_k^{\mathrm{Intra}} + L_k^{\mathrm{Inter}}. \tag{3.28}$$

Here, L_k^{Intra} is the leakage power to intra-cell users and can be represented as

$$L_k^{\mathrm{Intra}} = \sum_{l \in U_i \setminus \{k\}} L_{k,l} \tag{3.29}$$

and L_k^{Inter} is the leakage power to inter-cell users and can be represented as

$$L_k^{\mathrm{Inter}} = \sum_{j \in \Omega \setminus \{i\}} \sum_{l \in U_j} L_{k,l} \tag{3.30}$$

where $L_{k,l}$ is the leakage power to user *l* by $\mathbf{v}_{i,k}$ and can be represented as

$$L_{k,l} = \begin{cases} p_{i,k}\alpha_{i,l}\left|\mathbf{h}_{i,l}^H \mathbf{v}_{i,k}\right|^2, & \text{for } l \in U_i \setminus \{k\} \\ p_{i,k}\alpha_{j,l}\left|\mathbf{h}_{j,l}^H \mathbf{v}_{i,k}\right|^2, & \text{for } l \in U_j,\ j \in \Omega \setminus \{i\} \end{cases}. \tag{3.31}$$

It can be seen that the SLNR of user *k* is a function of $\mathbf{v}_{i,k}$. Thus, we can individually determine the beam weight for each user.

We determine the beam weight for user *k* of BS *i* to maximize the SLNR by

$$\mathbf{v}_{i,k}^{*} = \underset{\|\mathbf{v}_{i,k}\|^2=1}{\arg\max}\ \mathrm{SLNR}_k,\ k \in U_i,\ i \in \Omega. \tag{3.32}$$

We may determine the beam weight by using Rayleigh-Ritz quotient theorem as [35, 36]

$$\mathbf{v}_{i,k}^{*} = \frac{\left(\sum_{i\in\Omega}\mathbf{H}_i\mathbf{H}_i^H + \frac{\sigma_n^2}{p_{i,k}}\mathbf{I}_M\right)^{-1}\mathbf{h}_{i,k}}{\left\|\left(\sum_{i\in\Omega}\mathbf{H}_i\mathbf{H}_i^H + \frac{\sigma_n^2}{p_{i,k}}\mathbf{I}_M\right)^{-1}\mathbf{h}_{i,k}\right\|}. \tag{3.33}$$

It may require a computational complexity of $O\left(M^3 + K_{\text{total}}M^2\right)$ for the calculation of (3.33), which may not be affordable in m-MIMO environments.

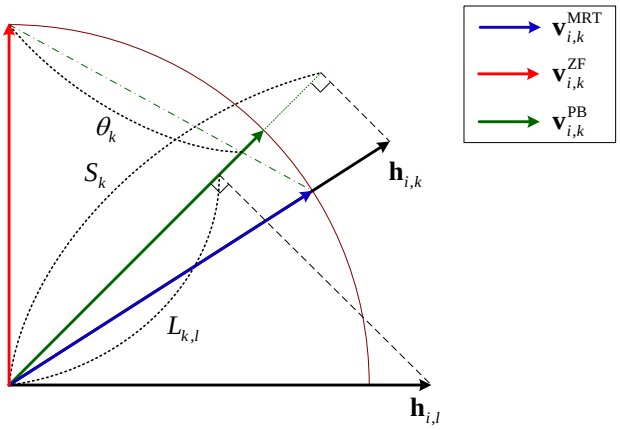
We consider a sub-optimum solution of (3.32) for the reduction of computational complexity. We consider the use of parameterized beamforming (PB) that determines the beam weight for user *k* of BS *i* by [37, 38]

$$\mathbf{v}_{i,k}^{\mathrm{PB}} = \beta_{i,k}\left[\theta_k \mathbf{v}_{i,k}^{\mathrm{MRT}} + (1-\theta_k)\mathbf{v}_{i,k}^{\mathrm{ZF}}\right] \tag{3.34}$$

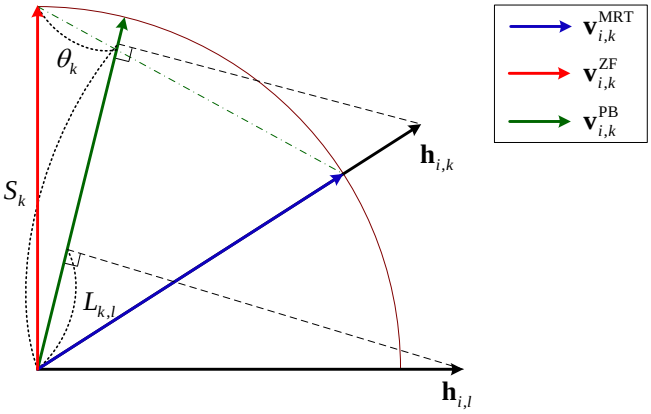
where $\theta_k\ (\in[0,1])$ is a scalar parameter to be determined and $\beta_{i,k}$ is a coefficient for normalization. We may determine the beam weight by optimizing θ_k as

$$\theta_k^{*} = \underset{\theta_k\in[0,1]}{\arg\max}\ \mathrm{SLNR}_k,\ k \in U_i,\ i \in \Omega. \tag{3.35}$$

Figure 3-4 illustrates the behavior of PB according to θ_k. It can be seen that as θ_k increases, S_k increases and $L_{k,l}$ may increase as well. On the other hand, as θ_k decreases, $L_{k,l}$ and S_k may decrease. Thus, it may be desirable to determine θ_k by taking consideration of the desired signal power and the leakage power as well.



(a) When θ_k is large.



(b) When θ_k is small.

Figure 3-4. Behavior of PB according to parameter θ_k.

As the number of users increases, the ZF may not provide desired signal power mainly due to the presence of insufficient spatial DoF [45]–[48]. As a result, the PB may not provide desired performance in multi-user environments. To alleviate this problem, we consider the use of ZF that selectively nulls out interference, referred to partial ZF (PZF). The PZF can reduce the dimension of nulling subspace, allowing to exploit higher spatial DoF. Moreover, it may be effective for the reduction of computational complexity for the generation of beam weight as the dimension of nulling subspace decreases.

BS i can partition user set U into two sets, U_i^{o} and $U_i^{\text{-o}}$, where U_i^{o} represents the set of users whose channel vector lies in the null space by the PZF and $U_i^{\text{-o}} = U - U_i^{\mathrm{o}}$. Let user n_i^{o} be the n-th user in U_i^{o}. When user k is the u-th user in U_i^{o}, the PZF beam weight for user k of BS i can be determined by

$$\mathbf{v}_{i,k}^{\mathrm{PZF}} = \frac{\mathbf{w}_{u_i^{\mathrm{o}}}^{\mathrm{o}}}{\left\| \mathbf{w}_{u_i^{\mathrm{o}}}^{\mathrm{o}} \right\|} \tag{3.36}$$

where $\mathbf{w}_{u_i^{\mathrm{o}}}^{\mathrm{o}}$ is the u-th column vector of $\mathbf{H}_i^{\mathrm{o}} \left(\left(\mathbf{H}_i^{\mathrm{o}} \right)^H \mathbf{H}_i^{\mathrm{o}} \right)^{-1}$. Here,

$$\mathbf{H}_i^{\mathrm{o}} = \left[\mathbf{h}_{i,1_i^{\mathrm{o}}} \cdots \mathbf{h}_{i,\left(K_i^{\mathrm{o}}\right)_i^{\mathrm{o}}} \right] \tag{3.37}$$

where K_i^{o} is the number of users in U_i^{o}. The PZF beam weight for user k in $U_i^{\text{-o}}$ can be determined by

$$\mathbf{v}_{i,k}^{\mathrm{PZF}} = \frac{\mathbf{w}_1^{\text{-o}}}{\left\| \mathbf{w}_1^{\text{-o}} \right\|} \tag{3.38}$$

where $\mathbf{w}_1^{\text{-o}}$ is the first column vector of $\mathbf{H}_{i,k}^{\text{-o}} \left(\left(\mathbf{H}_{i,k}^{\text{-o}} \right)^H \mathbf{H}_{i,k}^{\text{-o}} \right)^{-1}$. Here,

$$\mathbf{H}_{i,k}^{\text{-o}} = \left[\mathbf{h}_{i,k} \ \mathbf{H}_i^{\text{o}} \right]. \tag{3.39}$$

It can be conjectured that when K_i^{o} is small, the desired signal power may increase. However, users in $U_i^{\text{-o}}$ may suffer from interference since their channel may not lie in the null space. It may be desirable to determine $U_i^{\text{-o}}$ so that the interference-to-noise ratio (INR) is lower than a certain level. When user *k* is in the non-nulling user set of all BSs, user *k* may experience the INR represented as

$$\delta_k = \frac{\sum_{l \in U_i \setminus \{k\}} p_{i,l} \alpha_{i,k} \left| \mathbf{h}_{i,k}^H \mathbf{v}_{i,l}^{\text{PZF}} \right|^2 + \sum_{j \in \Omega \setminus \{i\}} \sum_{l \in U_j} p_{j,l} \alpha_{j,k} \left| \mathbf{h}_{j,k}^H \mathbf{v}_{j,l}^{\text{PZF}} \right|^2}{\sigma_n^2}. \tag{3.40}$$

We may estimate the INR by using the PZF beam weight. However, we may not know the PZF beam weight since the nulling user set is yet to be determined. As *M* goes to infinity, the channel may become deterministic [39]–[41], making it possible to estimate the INR without the PZF beam weight. We approximately estimate the INR assuming that the channel is deterministic. For ease of description, we introduce the following lemmas [39].

Lemma 3-1. Let $\mathbf{A} \in \mathbb{C}^{M \times M}$ and $\mathbf{x} \sim CN\left(\mathbf{0}_M, (1/M)\mathbf{I}_M\right)$. When the spectral norm of **A** is uniformly bounded and **x** is independent of **A**, it can be shown that [39]

$$\mathbf{x}^H \mathbf{A} \mathbf{x} - \frac{1}{M} \text{Tr}(\mathbf{A}) \xrightarrow[M \to \infty]{\text{a.s.}} 0. \tag{3.41}$$

Lemma 3-2. Let **A** be in Lemma 3-1, and **x** and $\mathbf{y} \sim CN\left(\mathbf{0}_M, (1/M)\mathbf{I}_M\right)$. When **x** and **y** are independent of **A**, it can be shown that [39]

$$\mathbf{x}^H \mathbf{A}\mathbf{y} \xrightarrow[M\to\infty]{\text{a.s.}} 0. \tag{3.42}$$

It can be shown from Lemma 3-1 that

$$\mathbf{h}_{j,k}^H \mathbf{v}_{j,l}^{\text{PZF}} \left(\mathbf{v}_{j,l}^{\text{PZF}}\right)^H \mathbf{h}_{j,k} - \text{Tr}\left(\mathbf{v}_{j,l}^{\text{PZF}} \left(\mathbf{v}_{j,l}^{\text{PZF}}\right)^H\right) \xrightarrow[M\to\infty]{\text{a.s.}} 0. \tag{3.43}$$

Note that

$$\mathbf{v}_{j,l}^{\text{PZF}} = \frac{\mathbf{P}_{j,l}^{\text{o}} \mathbf{h}_{j,l}}{\left\| \mathbf{P}_{j,l}^{\text{o}} \mathbf{h}_{j,l} \right\|} \tag{3.44}$$

where $\mathbf{P}_{j,l}^{\text{o}}$ is the projection matrix onto the null space by the PZF for user l of BS j [11]. It can be shown that

$$\begin{aligned} \text{Tr}\left(\mathbf{v}_{j,l}^{\text{PZF}} \left(\mathbf{v}_{j,l}^{\text{PZF}}\right)^H\right) &= \frac{\text{Tr}\left(\mathbf{P}_{j,l}^{\text{o}} \mathbf{h}_{j,l} \mathbf{h}_{j,l}^H \left(\mathbf{P}_{j,l}^{\text{o}}\right)^H\right)}{\left\| \mathbf{P}_{j,l}^{\text{o}} \mathbf{h}_{j,l} \right\|^2} \\ &= 1. \end{aligned} \tag{3.45}$$

It can also be shown that

$$\delta_k - \bar{\delta}_k \xrightarrow[M\to\infty]{\text{a.s.}} 0 \tag{3.46}$$

where $\bar{\delta}_k$ is a deterministic equivalent of δ_k and can be represented as

$$\bar{\delta}_k = \frac{\sum_{l \in U_i \setminus \{k\}} p_{i,l} \alpha_{i,k} + \sum_{j \in \Omega \setminus \{i\}} \sum_{l \in U_j} p_{j,l} \alpha_{j,k}}{\sigma_n^2}. \tag{3.47}$$

If $\bar{\delta}_k < \delta_{\text{th}}$, where δ_{th} is a threshold to be determined, user *k* can be included in the non-nulling user set of all BSs. Otherwise, the BS yielding the highest interference to user *k* excludes user *k* from its non-nulling user set. According to the change of non-nulling user set, we may update the deterministic INR of user *k*. We repeat this process until $\bar{\delta}_k < \delta_{\text{th}}$ or all BSs exclude user *k* from their non-nulling user set. Then, we can generate the PZF beam weight so that it generates interference to users in the non-nulling user set lower than a certain level. Figure 3-5 summarizes the process to determine a non-nulling user set.

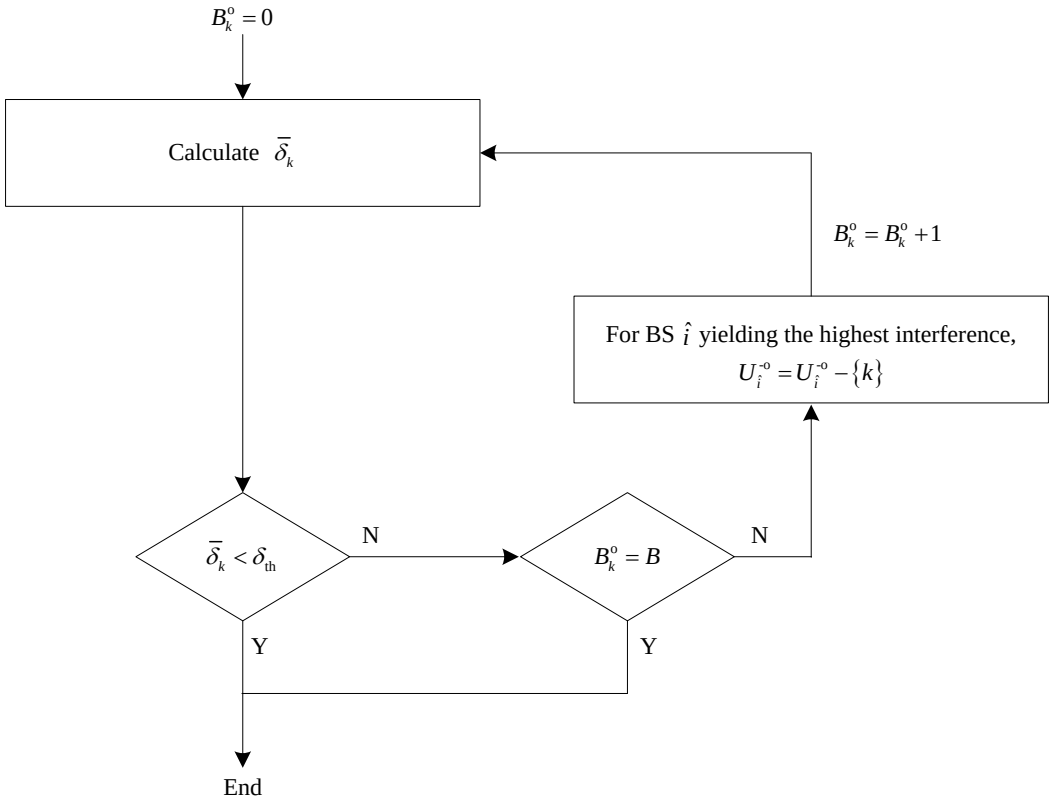


Figure 3-5. Generation of a non-nulling user set for user *k*.

Assuming the use of MRT and PZF, referred to multi-user parameterized beamforming (MUPB), we may determine the beam weight for user *k* by

$$\mathbf{v}_{i,k}^{\mathrm{MUPB}} = \beta_{i,k}\left[\theta_k \mathbf{v}_{i,k}^{\mathrm{MRT}} + \left(1-\theta_k\right)\mathbf{v}_{i,k}^{\mathrm{PZF}}\right]. \tag{3.48}$$

We may represent the SINR of user *k* with the use of beam weight $\mathbf{v}_{i,k}^{\mathrm{MUPB}}$ as

$$\mathrm{SLNR}_k = \frac{S_k}{L_k^{\mathrm{o}} + \sigma_n^2} \tag{3.49}$$

where L_k^{o} is the leakage power to the other users in U_i^{o}. Note that θ_k only affects the leakage power to users in the nulling user set.

Since $\mathbf{P}_{i,k}^{\mathrm{o}}$ is a Hermitian and idempotent matrix (*i.e.*, $\left(\mathbf{P}_{i,k}^{\mathrm{o}}\right)^H = \mathbf{P}_{i,k}^{\mathrm{o}}$ and $\left(\mathbf{P}_{i,k}^{\mathrm{o}}\right)^2 = \mathbf{P}_{i,k}^{\mathrm{o}}$) [67], it can be shown that

$$\begin{aligned} \mathbf{h}_{i,k}^H \mathbf{v}_{i,k}^{\mathrm{MRT}} &= \frac{\mathbf{h}_{i,k}^H \mathbf{h}_{i,k}}{\left\|\mathbf{h}_{i,k}\right\|} \\ &= \left\|\mathbf{h}_{i,k}\right\| \geq 0 \end{aligned} \tag{3.50}$$

$$\begin{aligned} \mathbf{h}_{i,k}^H \mathbf{v}_{i,k}^{\mathrm{PZF}} &= \frac{\mathbf{h}_{i,k}^H \mathbf{P}_{i,k}^{\mathrm{o}} \mathbf{h}_{i,k}}{\left\|\mathbf{P}_{i,k}^{\mathrm{o}} \mathbf{h}_{i,k}\right\|} \\ &= \frac{\mathbf{h}_{i,k}^H \mathbf{P}_{i,k}^{\mathrm{o}} \mathbf{P}_{i,k}^{\mathrm{o}} \mathbf{h}_{i,k}}{\left\|\mathbf{P}_{i,k}^{\mathrm{o}} \mathbf{h}_{i,k}\right\|} \\ &= \left\|\mathbf{P}_{i,k}^{\mathrm{o}} \mathbf{h}_{i,k}\right\| \geq 0. \end{aligned} \tag{3.51}$$

It can also be shown that

$$
\begin{aligned}
S_k &= p_{i,k}\alpha_{i,k}\left|\mathbf{h}_{i,k}^H \mathbf{v}_{i,k}^{\mathrm{MUPB}}\right|^2 \\
&= p_{i,k}\alpha_{i,k}\beta_k^2\left|\theta_k \mathbf{h}_{i,k}^H \mathbf{v}_{i,k}^{\mathrm{MRT}} + (1-\theta_k)\mathbf{h}_{i,k}^H \mathbf{v}_{i,k}^{\mathrm{PZF}}\right|^2 \\
&\overset{(a)}{=} p_{i,k}\alpha_{i,k}\beta_k^2\left(\theta_k \mathbf{h}_{i,k}^H \mathbf{v}_{i,k}^{\mathrm{MRT}} + (1-\theta_k)\mathbf{h}_{i,k}^H \mathbf{v}_{i,k}^{\mathrm{PZF}}\right)^2 \\
&\overset{(b)}{=} \frac{p_{i,k}\alpha_{i,k}\left(\theta_k \mathbf{h}_{i,k}^H \mathbf{v}_{i,k}^{\mathrm{MRT}} + (1-\theta_k)\mathbf{h}_{i,k}^H \mathbf{v}_{i,k}^{\mathrm{PZF}}\right)^2}{\theta_k^2 + 2\theta_k(1-\theta_k)\left(\mathbf{v}_{i,k}^{\mathrm{MRT}}\right)^H \mathbf{v}_{i,k}^{\mathrm{PZF}} + (1-\theta_k)^2} \\
&= \frac{p_{i,k}\alpha_{i,k}\left(\sqrt{\Delta G_k}\,\theta_k + \sqrt{G_k^{\mathrm{PZF}}}\right)^2}{2(1-q_k)\theta_k(\theta_k - 1) + 1}
\end{aligned}
\tag{3.52}
$$

where (a) and (b) follow from the fact that $\mathbf{h}_{i,k}^H \mathbf{v}_{i,k}^{\mathrm{MRT}}$ and $\mathbf{h}_{i,k}^H \mathbf{v}_{i,k}^{\mathrm{PZF}}$ are non-negative. Here,

$$
\Delta G_k = \left(\sqrt{G_k^{\mathrm{MRT}}} - \sqrt{G_k^{\mathrm{PZF}}}\right)^2 \tag{3.53}
$$

$$
G_k^{\mathrm{MRT}} = \left|\mathbf{h}_{i,k}^H \mathbf{v}_{i,k}^{\mathrm{MRT}}\right|^2 \tag{3.54}
$$

$$
G_k^{\mathrm{PZF}} = \left|\mathbf{h}_{i,k}^H \mathbf{v}_{i,k}^{\mathrm{PZF}}\right|^2 \tag{3.55}
$$

$$
q_k = \left(\mathbf{v}_{i,k}^{\mathrm{MRT}}\right)^H \mathbf{v}_{i,k}^{\mathrm{PZF}}. \tag{3.56}
$$

Let $U_{i,k}^{\mathrm{o}}$ be the set of users whose interference is nulled out by $\mathbf{v}_{i,k}^{\mathrm{PZF}}$ and user $n_{i,k}^{\mathrm{o}}$ be the *n*-th user in $U_{i,k}^{\mathrm{o}}$. Then, $U_{i,k}^{\mathrm{o}}$ and L_k^{o} can be represented as, respectively,

$$
U_{i,k}^{\mathrm{o}} = \begin{cases} U_i^{\mathrm{o}} \setminus \{k\} & \text{for } k \in U_i^{\mathrm{o}} \\ U_i^{\mathrm{o}} & \text{for } k \notin U_i^{\mathrm{o}} \end{cases} \tag{3.57}
$$

$$L_k^{\mathrm{o}} = \sum_{n_{i,k}^{\mathrm{o}} \in U_{i,k}^{\mathrm{o}}} L_{k,n_{i,k}^{\mathrm{o}}} \tag{3.58}$$

where $L_{k,n_{i,k}^{\mathrm{o}}}$ is the leakage power to user $n_{i,k}^{\mathrm{o}}$. It can be shown that

$$\begin{aligned} L_{k,n_{i,k}^{\mathrm{o}}} &= p_{i,k}\alpha_{i,n_{i,k}^{\mathrm{o}}} \left| \mathbf{h}_{i,n_{i,k}^{\mathrm{o}}}^H \mathbf{v}_{i,k}^{\mathrm{MUPB}} \right|^2 \\ &= \frac{p_{i,k}\alpha_{i,n_{i,k}^{\mathrm{o}}} \left| \theta_k \mathbf{h}_{i,n_{i,k}^{\mathrm{o}}}^H \mathbf{v}_{i,k}^{\mathrm{MRT}} + (1-\theta_k)\mathbf{h}_{i,n_{i,k}^{\mathrm{o}}}^H \mathbf{v}_{i,k}^{\mathrm{PZF}} \right|^2}{2(1-q_k)\theta_k(\theta_k - 1) + 1} \\ &= \frac{L_{k,n_{i,k}^{\mathrm{o}}}^{\mathrm{MRT}} \theta_k^2}{2(1-q_k)\theta_k(\theta_k - 1) + 1} \end{aligned} \tag{3.59}$$

where the last equality comes from $\mathbf{h}_{i,n_{i,k}^{\mathrm{o}}}^H \mathbf{v}_k^{\mathrm{PZF}} = 0$. Here,

$$L_{k,n_{i,k}^{\mathrm{o}}}^{\mathrm{MRT}} = p_{i,k}\alpha_{i,n_{i,k}^{\mathrm{o}}} \left| \mathbf{h}_{i,n_{i,k}^{\mathrm{o}}}^H \mathbf{v}_{i,k}^{\mathrm{MRT}} \right|^2 . \tag{3.60}$$

It may still require large complexity to determine θ_k maximizing the SLNR. We consider approximate estimation of the SLNR for further reduction of computational complexity.

Lemma 3-3. A deterministic equivalent of MRT gain can be represented as [40]

$$G_k^{\mathrm{MRT}} - \bar{G}_k^{\mathrm{MRT}} \xrightarrow[M\to\infty]{\text{a.s.}} 0 \tag{3.61}$$

where $\bar{G}_k^{\mathrm{MRT}} = M$.

Lemma 3-4. A deterministic equivalent of PZF gain can be represented as

$$G_k^{\mathrm{PZF}} - \bar{G}_k^{\mathrm{PZF}} \xrightarrow[M\to\infty]{\text{a.s.}} 0 \tag{3.62}$$

where $\bar{G}_k^{\mathrm{PZF}} = M - K_{i,k}^{\mathrm{o}}$.

Proof. Refer to Appendix A.

Lemma 3-5. A deterministic equivalent of q_k can be approximated to 1.

Proof. Refer to Appendix B.

It can be shown from Lemma 3-5 that

$$S_k - \bar{S}_k \xrightarrow[M\to\infty]{\text{a.s.}} 0 \tag{3.63}$$

where $\bar{S}_k$ is a deterministic equivalent of S_k and can be approximated as

$$\bar{S}_k \approx p_{i,k}\alpha_{i,k}\left(\sqrt{\Delta\bar{G}_k}\theta_k + \sqrt{\bar{G}_k^{\mathrm{PZF}}}\right)^2. \tag{3.64}$$

Here,

$$\Delta\bar{G}_k = \left(\sqrt{\bar{G}_k^{\mathrm{MRT}}} - \sqrt{\bar{G}_k^{\mathrm{PZF}}}\right)^2. \tag{3.65}$$

It can be shown from Lemma 3-1 that

$$L_{k,n_{i,k}^{\mathrm{o}}}^{\mathrm{MRT}} - \bar{L}_{k,n_{i,k}^{\mathrm{o}}}^{\mathrm{MRT}} \xrightarrow[M\to\infty]{\text{a.s.}} 0 \tag{3.66}$$

where $\bar{L}_{k,n_{i,k}^{\mathrm{o}}}^{\mathrm{MRT}}$ is a deterministic equivalent of $L_{k,n_{i,k}^{\mathrm{o}}}^{\mathrm{MRT}}$ and can be represented as

$$\bar{L}_{k,n_{i,k}^{\mathrm{o}}}^{\mathrm{MRT}} = p_{i,k}\alpha_{i,n_{i,k}^{\mathrm{o}}}. \tag{3.67}$$

It can also be shown from Lemma 3-5 that

$$L_{k,n_{i,k}^{\mathrm{o}}} - \bar{L}_{k,n_{i,k}^{\mathrm{o}}} \xrightarrow[M\to\infty]{\text{a.s.}} 0 \tag{3.68}$$

where $\bar{L}_{k,n_{i,k}^{\mathrm{o}}}$ is a deterministic equivalent of $L_{k,n_{i,k}^{\mathrm{o}}}$ and can be approximated as

$$\bar{L}_{k,n_{i,k}^{\mathrm{o}}} \approx \bar{L}_{k,n_{i,k}^{\mathrm{o}}}^{\mathrm{MRT}} \theta_k^2 . \tag{3.69}$$

Then, it can easily be seen that

$$L_k^{\mathrm{o}} - \bar{L}_k^{\mathrm{o}} \xrightarrow[M\to\infty]{\text{a.s.}} 0 \tag{3.70}$$

where $\bar{L}_k^{\mathrm{o}}$ is a deterministic equivalent of L_k^{o} and can be approximated as

$$\bar{L}_k^{\mathrm{o}} \approx \bar{L}_{U_{i,k}^{\mathrm{o}}}^{\mathrm{MRT}} \theta_k^2 . \tag{3.71}$$

Here,

$$\bar{L}_{U_{i,k}^{\mathrm{o}}}^{\mathrm{MRT}} = \sum_{n_{i,k}^{\mathrm{o}} \in U_{i,k}^{\mathrm{o}}} \bar{L}_{k,n_{i,k}^{\mathrm{o}}}^{\mathrm{MRT}} . \tag{3.72}$$

It can also be seen that

$$\mathrm{SLNR}_k - \overline{\mathrm{SLNR}}_k \xrightarrow[M\to\infty]{\text{a.s.}} 0 \tag{3.73}$$

where $\overline{\mathrm{SLNR}}_k$ is a deterministic equivalent of SLNR_k and can be approximated as

$$\overline{\mathrm{SLNR}}_k \approx \frac{p_{i,k}\alpha_{i,k}\left(\sqrt{\Delta\bar{G}_k}\theta_k + \sqrt{\bar{G}_k^{\mathrm{PZF}}}\right)^2}{\bar{L}_{U_{i,k}^{\mathrm{o}}}^{\mathrm{MRT}}\theta_k^2 + \sigma_n^2} . \tag{3.74}$$

We can determine θ_k maximizing (3.74) by invoking the zero-gradient condition [42]. The gradient of $\overline{\mathrm{SLNR}}_k$ with respect to θ_k can be represented as

$$\nabla_{\theta_k} \overline{\mathrm{SLNR}}_k = \Phi_k \Gamma_k \tag{3.75}$$

where

$$\Phi_k = \frac{2 p_{i,k} \alpha_{i,k} \left(\sqrt{\Delta \bar{G}_k} \theta_k + \sqrt{\bar{G}_k^{\mathrm{PZF}}} \right)}{\left(\bar{L}_{U_{i,k}^{\mathrm{o}}}^{\mathrm{MRT}} \theta_k^2 + \sigma_n^2 \right)^2} \tag{3.76}$$

$$\Gamma_k = \sqrt{\Delta \bar{G}_k} \sigma_n^2 - \bar{L}_{U_{i,k}^{\mathrm{o}}}^{\mathrm{MRT}} \sqrt{\bar{G}_k^{\mathrm{PZF}}} \theta_k \,. \tag{3.77}$$

It can be seen from $\Phi_k > 0$ that Γ_k should be zero to satisfy the zero-gradient condition. The parameter θ_k can optimally be determined by

$$\theta_k^* = \frac{\sigma_n^2}{\bar{L}_{U_{i,k}^{\mathrm{o}}}^{\mathrm{MRT}}} \sqrt{\frac{\Delta \bar{G}_k}{\bar{G}_k^{\mathrm{PZF}}}}. \tag{3.78}$$

Note that θ_k^* is unique and globally optimal since $\overline{\mathrm{SLNR}}_k$ is a strict quasi-concave function of θ_k [42]. For the proof, refer to Appendix C. It can be shown that

$$\theta_k^* = \frac{\sigma_n^2}{\bar{L}_{U_{i,k}^{\mathrm{o}}}^{\mathrm{MRT}}} \frac{M \left(\sqrt{1 - \frac{K_{i,k}^{\mathrm{o}}}{M}} - 1 \right) + K_{i,k}^{\mathrm{o}}}{M - K_{i,k}^{\mathrm{o}}}. \tag{3.79}$$

Let $\bar{\alpha}_{i,k}^{\mathrm{o}}$ be the average path loss between BS i and users in $U_{i,k}^{\mathrm{o}}$. It can be shown that

$$\bar{\alpha}_{i,k}^{\mathrm{o}} = \frac{\sum_{n_{i,k}^{\mathrm{o}} \in U_{i,k}^{\mathrm{o}}} \alpha_{i,n_{i,k}^{\mathrm{o}}}}{K_{i,k}^{\mathrm{o}}} \tag{3.80}$$

$$\begin{aligned}\bar{L}_{U_{i,k}^{\mathrm{o}}}^{\mathrm{MRT}} &= \sum_{n_{i,k}^{\mathrm{o}} \in U_{i,k}^{\mathrm{o}}} p_{i,k} \alpha_{i,n_{i,k}^{\mathrm{o}}} \\ &= K_{i,k}^{\mathrm{o}} p_{i,k} \bar{\alpha}_{i,k}^{\mathrm{o}}.\end{aligned} \tag{3.81}$$

It can also be shown from the first-order Taylor approximation [42] that

$$\theta_k^* \approx \frac{\sigma_n^2}{2 p_{i,k} \bar{\alpha}_{i,k}^{\mathrm{o}} \eta_{i,k}^{\mathrm{o}}} \tag{3.82}$$

where $\eta_{i,k}^{\mathrm{o}}$ is the effective spatial DoF for user k of BS i. When $\eta_{i,k}^{\mathrm{o}}$ is large, the PZF gain is also large, implying that reducing performance loss by the PZF with a large θ_k^* may not be effective for the maximization of approximated SLNR. In this case, the use of a small θ_k may be effective for the maximization of approximated SLNR, while reducing the leakage power. On the other hand, as $\eta_{i,k}^{\mathrm{o}}$ decreases, θ_k^* increases, increasing the desired signal power. In this case, the use of a large θ_k may be effective for the maximization of approximated SLNR, while avoiding noticeable performance loss by the PZF.

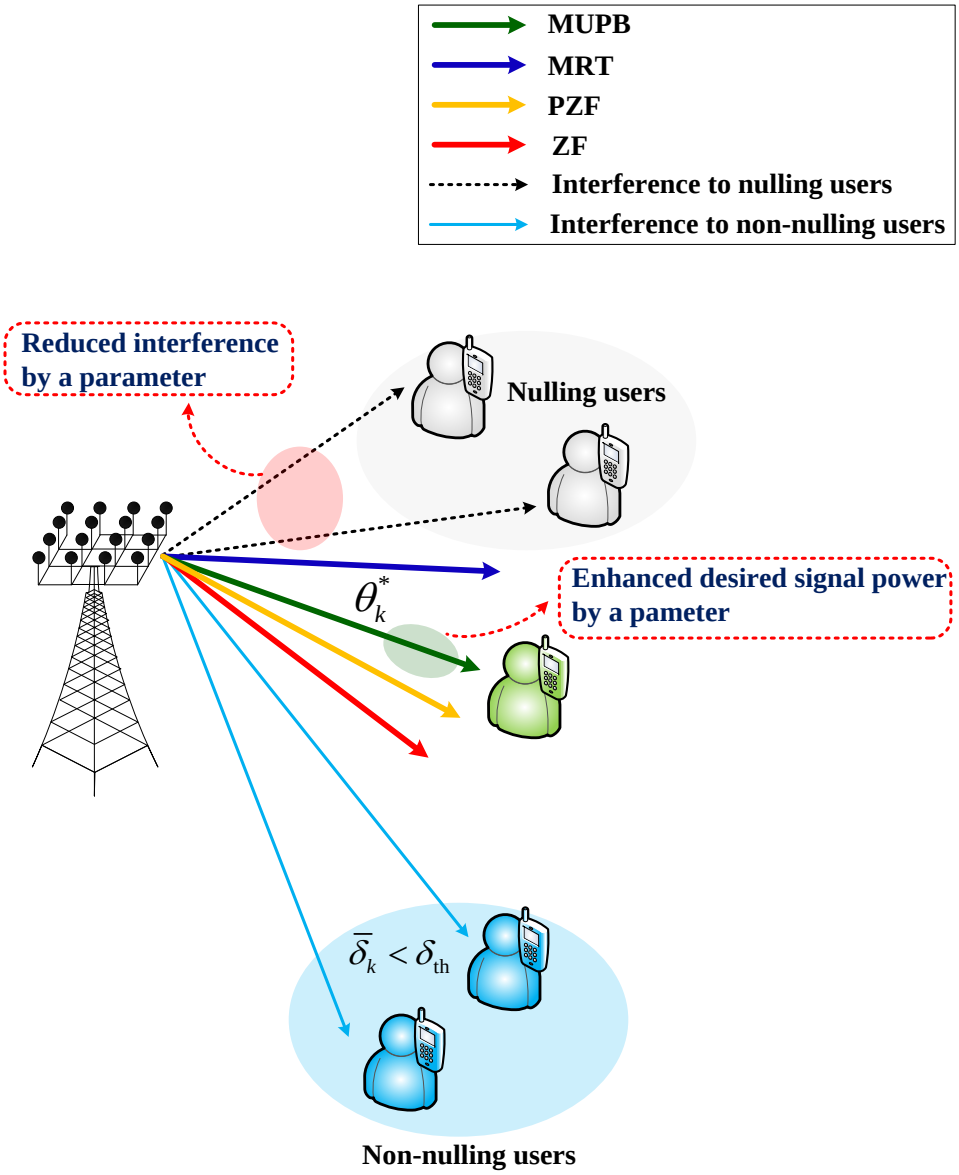


Figure 3-6. Proposed MUPB.

3.2.2. Effect of CSI inaccuracy

In this subsection, we analyze effect of CSI inaccuracy on the SLNR. In the presence of CSI inaccuracy, we may represent the true channel between BS *i* and user *k* as [41]

$$\hat{\mathbf{h}}_{i,k} = \sqrt{1-\tau_{i,k}^2}\,\mathbf{h}_{i,k} + \tau_{i,k}\mathbf{z}_{i,k} \tag{3.83}$$

where $\tau_{i,k}\left(\in[0,1]\right)$ denotes the channel inaccuracy for user k of BS i and $\mathbf{z}_{i,k} \in \mathbb{C}^{M\times 1} \sim CN\left(\mathbf{0}_M, \mathbf{I}_M\right)$. For ease of analysis, we assume that $\tau_{i,k} = \tau$ for all *i* and *k*.

In the presence of CSI inaccuracy, we may represent the desired signal power of user *k* as

$$\begin{aligned} S_{\tau,k} &= p_{i,k}\alpha_{i,k}\left|\hat{\mathbf{h}}_{i,k}^H \mathbf{v}_{i,k}^{\text{MUPB}}\right|^2 \\ &= p_{i,k}\alpha_{i,k}\left|\sqrt{1-\tau^2}\,\mathbf{h}_{i,k}^H \mathbf{v}_{i,k}^{\text{MUPB}} + \tau \mathbf{z}_{i,k}^H \mathbf{v}_{i,k}^{\text{MUPB}}\right|^2 \\ &= \left(1-\tau^2\right)S_k + p_{i,k}\alpha_{i,k}\left[2\tau\sqrt{1-\tau^2}\,\text{Re}\left(\varphi_{k,k}^{\text{MUPB}}\right) + \tau^2 Z_{k,k}^{\text{MUPB}}\right] \end{aligned} \tag{3.84}$$

where

$$\varphi_{k,k}^{\text{MUPB}} = \mathbf{h}_{i,k}^H \mathbf{v}_{i,k}^{\text{MUPB}} \left(\mathbf{v}_{i,k}^{\text{MUPB}}\right)^H \mathbf{z}_{i,k} \tag{3.85}$$

$$Z_{k,k}^{\text{MUPB}} = \left|\mathbf{z}_{i,k}^H \mathbf{v}_{i,k}^{\text{MUPB}}\right|^2 . \tag{3.86}$$

Since $\mathbf{h}_{i,k}$ and $\mathbf{z}_{i,k}$ are independent of each other, it can be shown from Lemma 3-2 that

$$\varphi_{k,k}^{\text{MUPB}} \xrightarrow[M\to\infty]{\text{a.s.}} 0 . \tag{3.87}$$

It can also be shown from Lemma 3-1 that

$$\mathbf{z}_{i,k}^{H}\mathbf{v}_{i,k}^{\mathrm{MUPB}}\left(\mathbf{v}_{i,k}^{\mathrm{MUPB}}\right)^{H}\mathbf{z}_{i,k}-\mathrm{Tr}\left(\mathbf{v}_{i,k}^{\mathrm{MUPB}}\left(\mathbf{v}_{i,k}^{\mathrm{MUPB}}\right)^{H}\right)\xrightarrow[M\to\infty]{\text{a.s.}}0\,. \tag{3.88}$$

Then, it can be shown that

$$Z_{k,k}^{\mathrm{MUPB}}-\bar{Z}_{k,k}^{\mathrm{MUPB}}\xrightarrow[M\to\infty]{\text{a.s.}}0 \tag{3.89}$$

where $\bar{Z}_{k,k}^{\mathrm{MUPB}}$ is a deterministic equivalent of $Z_{k,k}^{\mathrm{MUPB}}$ and can be represented as

$$\begin{aligned}\bar{Z}_{k,k}^{\mathrm{MUPB}}&=\mathrm{Tr}\left(\mathbf{v}_{i,k}^{\mathrm{MUPB}}\left(\mathbf{v}_{i,k}^{\mathrm{MUPB}}\right)^{H}\right)\\&=1.\end{aligned} \tag{3.90}$$

It can also be shown that

$$S_{\tau,k}-\bar{S}_{\tau,k}\xrightarrow[M\to\infty]{\text{a.s.}}0 \tag{3.91}$$

where $\bar{S}_{\tau,k}$ is a deterministic equivalent of $S_{\tau,k}$ and can be approximated as

$$\bar{S}_{\tau,k}\approx\bar{S}_{k}-e_{\tau,k}^{S}\,. \tag{3.92}$$

where $e_{\tau,k}^{S}$ denotes the performance loss in the presence of CSI inaccuracy and can be represented as

$$e_{\tau,k}^{S}=\tau^{2}\left(\bar{S}_{k}-p_{i,k}\alpha_{i,k}\right). \tag{3.93}$$

It can be seen that as τ increases, $e_{\tau,k}^{S}$ may increase. This implies that as τ increases, the performance loss increases.

In the presence of CSI inaccuracy, we may represent the leakage power to users in $U_{i,k}^{\mathrm{o}}$ as

$$L_{\tau,k}^{\mathrm{o}} = \sum_{n_{i,k}^{\mathrm{o}} \in U_{i,k}^{\mathrm{o}}} L_{\tau,k,n_{i,k}^{\mathrm{o}}} \ . \tag{3.94}$$

It can be shown that

$$\begin{aligned} L_{\tau,k,n_{i,k}^{\mathrm{o}}} &= p_{i,k}\alpha_{i,n_{i,k}^{\mathrm{o}}} \left| \hat{\mathbf{h}}_{i,n_{i,k}^{\mathrm{o}}}^{H} \mathbf{v}_{i,k}^{\mathrm{MUPB}} \right|^2 \\ &= p_{i,k}\alpha_{i,n_{i,k}^{\mathrm{o}}} \left| \sqrt{1-\tau^2}\mathbf{h}_{i,n_{i,k}^{\mathrm{o}}}^{H} \mathbf{v}_{i,k}^{\mathrm{MUPB}} + \tau \mathbf{z}_{i,n_{i,k}^{\mathrm{o}}}^{H} \mathbf{v}_{i,k}^{\mathrm{MUPB}} \right|^2 \\ &= \left(1-\tau^2\right) L_{k,n_{i,k}^{\mathrm{o}}} + p_{i,k}\alpha_{i,n_{i,k}^{\mathrm{o}}} \left[2\tau\sqrt{1-\tau^2}\,\mathrm{Re}\left(\varphi_{k,n_{i,k}^{\mathrm{o}}}^{\mathrm{MUPB}}\right) + \tau^2 Z_{k,n_{i,k}^{\mathrm{o}}}^{\mathrm{MUPB}} \right] \end{aligned} \tag{3.95}$$

where

$$\varphi_{k,n_{i,k}^{\mathrm{o}}}^{\mathrm{MUPB}} = \mathbf{h}_{i,n_{i,k}^{\mathrm{o}}}^{H} \mathbf{v}_{i,k}^{\mathrm{MUPB}} \left(\mathbf{v}_{i,k}^{\mathrm{MUPB}}\right)^{H} \mathbf{z}_{i,n_{i,k}^{\mathrm{o}}} \tag{3.96}$$

$$Z_{k,n_{i,k}^{\mathrm{o}}}^{\mathrm{MUPB}} = \left| \mathbf{z}_{i,n_{i,k}^{\mathrm{o}}}^{H} \mathbf{v}_{i,k}^{\mathrm{MUPB}} \right|^2 . \tag{3.97}$$

Since $\mathbf{h}_{i,k}$ and $\mathbf{z}_{i,n_{i,k}^{\mathrm{o}}}$ are independent of each other, it can be shown from Lemma 3-2 that

$$\varphi_{k,n_{i,k}^{\mathrm{o}}}^{\mathrm{MUPB}} \xrightarrow[M\to\infty]{\text{a.s.}} 0 \, . \tag{3.98}$$

It can also be shown that

$$
\begin{aligned}
Z_{k,n_{i,k}^{\mathrm{o}}}^{\mathrm{MUPB}} &= \left| \mathbf{z}_{i,n_{i,k}^{\mathrm{o}}}^{H} \mathbf{v}_{i,k}^{\mathrm{MUPB}} \right|^2 \\
&= \beta_{i,k}^2 \left| \theta_k \mathbf{z}_{i,n_{i,k}^{\mathrm{o}}}^{H} \mathbf{v}_{i,k}^{\mathrm{MRT}} + \left(1-\theta_k\right) \mathbf{z}_{i,n_{i,k}^{\mathrm{o}}}^{H} \mathbf{v}_{i,k}^{\mathrm{PZF}} \right|^2 \\
&= \beta_{i,k}^2 \left[\theta_k^2 Z_{k,n_{i,k}^{\mathrm{o}}}^{\mathrm{MRT}} + 2\theta_k \left(1-\theta_k\right) \mathrm{Re}\left(\mu_{k,n_{i,k}^{\mathrm{o}}} \right) + \left(1-\theta_k\right)^2 Z_{k,n_{i,k}^{\mathrm{o}}}^{\mathrm{PZF}} \right]
\end{aligned}
\tag{3.99}
$$

where

$$
\mu_{k,n_{i,k}^{\mathrm{o}}} = \mathbf{z}_{i,n_{i,k}^{\mathrm{o}}}^{H} \mathbf{v}_{i,k}^{\mathrm{MRT}} \left(\mathbf{v}_{i,k}^{\mathrm{PZF}} \right)^{H} \mathbf{z}_{i,n_{i,k}^{\mathrm{o}}} . \tag{3.100}
$$

Here, $Z_{k,n_{i,k}^{\mathrm{o}}}^{\mathrm{MRT}}$ and $Z_{k,n_{i,k}^{\mathrm{o}}}^{\mathrm{PZF}}$ respectively denote the performance loss by $\mathbf{v}_{i,k}^{\mathrm{MRT}}$ and $\mathbf{v}_{i,k}^{\mathrm{PZF}}$ in the presence of CSI inaccuracy, and can be represented as, respectively,

$$
Z_{k,n_{i,k}^{\mathrm{o}}}^{\mathrm{MRT}} = \left| \mathbf{z}_{i,n_{i,k}^{\mathrm{o}}}^{H} \mathbf{v}_{i,k}^{\mathrm{MRT}} \right|^2 \tag{3.101}
$$

$$
Z_{k,n_{i,k}^{\mathrm{o}}}^{\mathrm{PZF}} = \left| \mathbf{z}_{i,n_{i,k}^{\mathrm{o}}}^{H} \mathbf{v}_{i,k}^{\mathrm{PZF}} \right|^2 . \tag{3.102}
$$

It can be shown from Lemma 3-1 that

$$
\mathbf{z}_{i,n_{i,k}^{\mathrm{o}}}^{H} \mathbf{v}_{i,k}^{\mathrm{MRT}} \left(\mathbf{v}_{i,k}^{\mathrm{PZF}} \right)^{H} \mathbf{z}_{i,n_{i,k}^{\mathrm{o}}} - \mathrm{Tr}\left(\mathbf{v}_{i,k}^{\mathrm{MRT}} \left(\mathbf{v}_{i,k}^{\mathrm{PZF}} \right)^{H} \right) \xrightarrow[M\to\infty]{\text{a.s.}} 0 . \tag{3.103}
$$

Then, it can be shown from Lemma 3-5 that

$$
\mu_{k,n_{i,k}^{\mathrm{o}}} - \bar{\mu}_{k,n_{i,k}^{\mathrm{o}}} \xrightarrow[M\to\infty]{\text{a.s.}} 0 \tag{3.104}
$$

where $\bar{\mu}_{k,n_{i,k}^{\mathrm{o}}}$ is a deterministic equivalent of $\mu_{k,n_{i,k}^{\mathrm{o}}}$ and can be approximated as

$$
\bar{\mu}_{k,n_{i,k}^{\mathrm{o}}} \approx 1. \tag{3.105}
$$

It can be shown from Lemma 3-1 that

$$Z_{k,n_{i,k}^{o}}^{\mathrm{MRT}} - \bar{Z}_{k,n_{i,k}^{o}}^{\mathrm{MRT}} \xrightarrow[M\to\infty]{\text{a.s.}} 0 \tag{3.106}$$

where $\bar{Z}_{k,n_{i,k}^{o}}^{\mathrm{MRT}}$ is a deterministic equivalent of $Z_{k,n_{i,k}^{o}}^{\mathrm{MRT}}$ and can be approximated as

$$\bar{Z}_{k,n_{i,k}^{o}}^{\mathrm{MRT}} \approx 1. \tag{3.107}$$

Similarly, it can be shown from Lemma 3-1 that

$$Z_{k,n_{i,k}^{o}}^{\mathrm{PZF}} - \bar{Z}_{k,n_{i,k}^{o}}^{\mathrm{PZF}} \xrightarrow[M\to\infty]{\text{a.s.}} 0 \tag{3.108}$$

where $\bar{Z}_{k,n_{i,k}^{o}}^{\mathrm{PZF}}$ is a deterministic equivalent of $Z_{k,n_{i,k}^{o}}^{\mathrm{PZF}}$ and can approximately be represented as

$$\bar{Z}_{k,n_{i,k}^{o}}^{\mathrm{PZF}} \approx 1. \tag{3.109}$$

It can be seen that the PZF may not completely null out interference to users in the nulling user set in the presence of CSI inaccuracy, yielding performance loss.

It can be shown that

$$Z_{k,n_{i,k}^{o}}^{\mathrm{MUPB}} - \bar{Z}_{k,n_{i,k}^{o}}^{\mathrm{MUPB}} \xrightarrow[M\to\infty]{\text{a.s.}} 0 \tag{3.110}$$

where $\bar{Z}_{k,n_{i,k}^{o}}^{\mathrm{MUPB}}$ is a deterministic equivalent of $Z_{k,n_{i,k}^{o}}^{\mathrm{MUPB}}$ and can be approximated as

$$\bar{Z}_{k,n_{i,k}^{o}}^{\mathrm{MUPB}} \approx 1. \tag{3.111}$$

It can also be shown that

$$L_{\tau,k,n_{i,k}^{\mathrm{o}}} - \bar{L}_{\tau,k,n_{i,k}^{\mathrm{o}}} \xrightarrow[M\to\infty]{\text{a.s.}} 0 \tag{3.112}$$

where $\bar{L}_{\tau,k,n_{i,k}^{\mathrm{o}}}$ is a deterministic equivalent of $L_{\tau,k,n_{i,k}^{\mathrm{o}}}$ and can be approximated as

$$\bar{L}_{\tau,k,n_{i,k}^{\mathrm{o}}} \approx \left(1-\tau^2\right)\bar{L}_{k,n_{i,k}^{\mathrm{o}}} + \bar{L}_{k,n_{i,k}^{\mathrm{o}}}^{\mathrm{MRT}}\tau^2 . \tag{3.113}$$

It can also be shown that

$$L_k^{\mathrm{o}} - \bar{L}_{\tau,k}^{\mathrm{o}} \xrightarrow[M\to\infty]{\text{a.s.}} 0 \tag{3.114}$$

where $\bar{L}_{\tau,k}^{\mathrm{o}}$ is a deterministic equivalent of $L_{\tau,k}^{\mathrm{o}}$ and can be approximated as

$$\bar{L}_{\tau,k}^{\mathrm{o}} \approx \bar{L}_k^{\mathrm{o}} + e_{\tau,k}^{L} . \tag{3.115}$$

Here, $e_{\tau,k}^{L}$ denotes the undesired leakage power in the presence of CSI inaccuracy and can be represented as

$$e_{\tau,k}^{L} = \tau^2\left(1-\theta_k^2\right)\bar{L}_{U_{i,k}^{\mathrm{o}}}^{\mathrm{MRT}} . \tag{3.116}$$

It can be seen that as τ increases, $e_{\tau,k}^{L}$ may increase. This implies that as τ increases, the leakage power increases.

It can be shown that

$$\mathrm{SLNR}_{\tau,k} - \overline{\mathrm{SLNR}}_{\tau,k} \xrightarrow[M\to\infty]{\text{a.s.}} 0 \tag{3.117}$$

where $\overline{\mathrm{SLNR}}_{\tau,k}$ is a deterministic equivalent of $\mathrm{SLNR}_{\tau,k}$ and can be approximated as

$$\overline{\mathrm{SLNR}}_{\tau,k} \approx \frac{\bar{S}_k - e^S_{\tau,k}}{\bar{L}^{\mathrm{o}}_k + e^L_{\tau,k} + \sigma_n^2}. \tag{3.118}$$

Finally, it can also be shown that

$$\begin{aligned} \frac{\bar{S}_k - e^S_{\tau,k}}{\bar{L}^{\mathrm{o}}_k + e^L_{\tau,k} + \sigma_n^2} &= \frac{\bar{S}_k}{\bar{L}^{\mathrm{o}}_k + e^L_{\tau,k} + \sigma_n^2} - \frac{e^S_{\tau,k}}{\bar{L}^{\mathrm{o}}_k + e^L_{\tau,k} + \sigma_n^2} \\ &\le \overline{\mathrm{SLNR}}_k - e^{\mathrm{SLNR}}_{\tau,k} \end{aligned}. \tag{3.119}$$

where $e^{\mathrm{SLNR}}_{\tau,k}$ denotes the SLNR of user *k* in the presence of CSI inaccuracy and can be represented as

$$e^{\mathrm{SLNR}}_{\tau,k} = \frac{e^S_{\tau,k}}{\bar{L}^{\mathrm{o}}_k + e^L_{\tau,k} + \sigma_n^2}. \tag{3.120}$$

It can be seen that when τ is small, the proposed MUPB may effectively maximize the SLNR since $\overline{\mathrm{SLNR}}_k$ is not affected by τ. However, when τ is large, it may not maximize the SLNR due to the increase of $e^{\mathrm{SLNR}}_{\tau,k}$.

3.2.3. Computational complexity

We measure the computational complexity in terms of real-floating point operation (flop). We assume that a multiplication of two $(p \times q)$ and $(q \times r)$ complex matrices requires 8*pqr* flops, an inner product of two $(p \times 1)$ complex vectors requires 2*p* flops, a calculation of $(p \times q)$ complex Gram matrix requires pq^2 flops and an inversion of Hermitian matrix requires $(4/3)q^3$ flops [43, 44].

It can be shown that BS *i* requires $\psi_i^{\mathrm{MRT}} = 4K_iM$ flops to generate the beam weight for MRT, $\psi_{i,U_i^{\mathrm{o}}}^{\mathrm{PZF}} = K_i^{\mathrm{o}}M^2 + \left(8\left(K_i^{\mathrm{o}}\right)^2 + 4Q_i^{\mathrm{o}}\right)M + (4/3)\left(K_i^{\mathrm{o}}\right)^3$ flops and

$\psi_{i,U_i^{-o}}^{\mathrm{PZF}} = \left[\left(K_i^{o} + 1 \right) M^2 + \left(8\left(K_i^{o} + 1 \right)^2 + 4 \right) M + \left(4/3 \right) \left(K_i^{o} + 1 \right)^3 \right] Q_i^{-o}$ flops to generate the PZF beam weight for users in U_i^{o} and U_i^{-o}, respectively, and approximately $B\left(B-1 \right) K_{\mathrm{total}}$ flops to determine the nulling and non-nulling user sets, where Q_i^{o} and Q_i^{-o} are the number of users in U_i^{o} and U_i^{-o} served by BS *i*, respectively. Thus, BS *i* requires $\psi_i^{\mathrm{PZF}} = \psi_{i,U_i^{o}}^{\mathrm{PZF}} + \psi_{i,U_i^{-o}}^{\mathrm{PZF}}$ flops to generate the PZF beam weight. It may approximately require a computational complexity of $10K_{\mathrm{total}}M$ flops to generate the normalized beam weight with the use of MRT and PZF. Finally, the MUPB requires a computational complexity of $\psi^{\mathrm{MUPB}} = \sum_{i=1}^{B} \left(\psi_i^{\mathrm{MRT}} + \psi_i^{\mathrm{PZF}} \right) + \left[B\left(B-1 \right) + 10M \right] K_{\mathrm{total}}$ flops. The computational complexity of beamforming schemes is summarized in Table 3-1.

Table 3-1. Computational complexity for the generation of beam weight.

Scheme	**Computational complexity (flops)**
MRT	$4K_{\mathrm{total}}M$
ZF	$BK_{\mathrm{total}}M^2 + \left(8BK_{\mathrm{total}} + 4 \right) K_{\mathrm{total}}M + \frac{4}{3}BK_{\mathrm{total}}^3$
CEA-ZF	$\sum_{i=1}^{B} \left[\left(K_i + K_{\mathrm{C},i} \right) M^2 + \left(8\left(K_i + K_{\mathrm{C},i} \right)^2 + 4K_i \right) M + \frac{4}{3}\left(K_i + K_{\mathrm{C},i} \right)^3 \right]$
Max SLNR	$\frac{4}{3}K_{\mathrm{total}}M^3 + \left(\left(B+8 \right) K_{\mathrm{total}} + 2\left(B-1 \right) \right) M^2 + 6K_{\mathrm{total}}M$
PB	$BK_{\mathrm{total}}M^2 + \left(8BK_{\mathrm{total}} + 18 \right) K_{\mathrm{total}}M + \frac{4}{3}BK_{\mathrm{total}}^3$
MUPB	$\sum_{i=1}^{B} \left(\psi_i^{\mathrm{MRT}} + \psi_i^{\mathrm{PZF}} \right) + \left[B\left(B-1 \right) + 10M \right] K_{\mathrm{total}}$

3.3. Performance evaluation

We evaluate the performance of the proposed MUPB by computer simulation. We consider a three-cell cellular system as illustrated in Figure 3-7, where each cell equally serves K (*i.e.*, $K_i = K,\ \forall i$) users randomly distributed in cell boundary region. The other simulation parameters are summarized in Table 3-2 [68, 69].

Table 3-2. Simulation parameters.

Simulation parameters	**Value**
Number of BSs B	3
User distribution	Randomly distributed
Cell radius $R_{\max}$	300 m
Cell center radius R_{c}	200 m
Channel distribution	Independent and identically distributed (i.i.d.)
Path loss $\alpha_{i,k},\ \forall i$ and k	$148.1+37.6\log_{10}\left(d_{i,k}\right)$ dB, where $d_{i,k}$ is the distance between BS i and user k in km
Maximum transmit power P_{T}	30 dBm
Power allocation $p_{i,k},\ \forall i$ and k	P_{T} / K
Noise power σ_n^2	-92 dBm

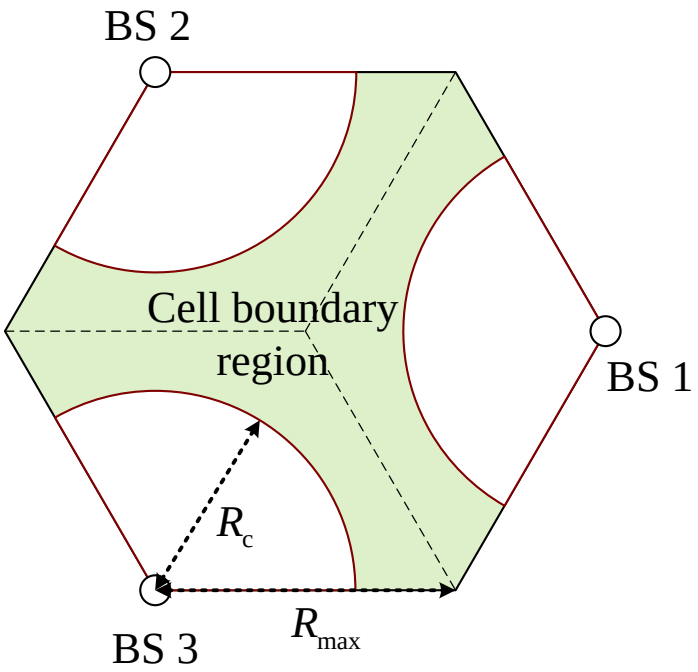
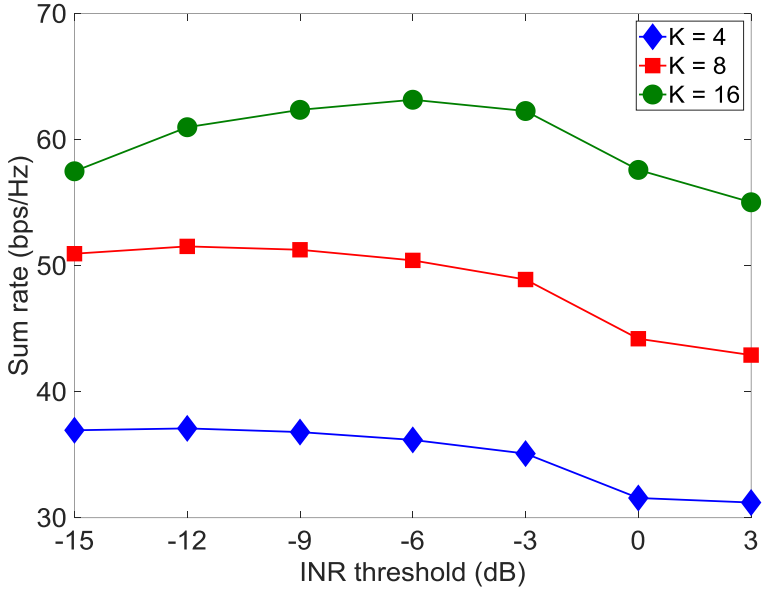
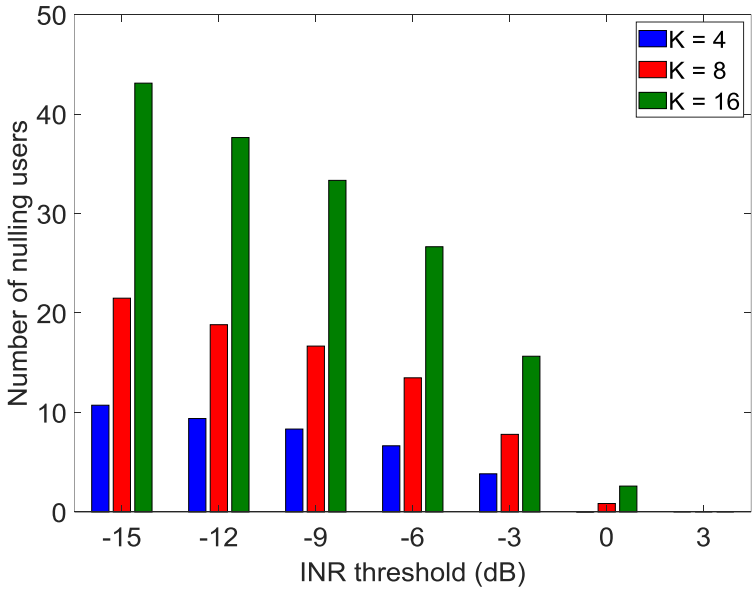


Figure 3-7. A three-cell cellular system.

Figure 3-8 depicts the sum rate and the number of users in the nulling user set of MUPB according to threshold δ_{th} when $M = 64$ and $K = 4, 8, 16$. It can be seen that as K increases, δ_{th} maximizing the sum rate and the number of users in the nulling user set increase. That is, when K is large, it may be desirable to reduce the number of users in the nulling user set for the reduction of performance loss by the PZF. Otherwise, the MUPB may suffer from performance loss mainly due to the presence of insufficient spatial DoF for the PZF. It can also be seen that when δ_{th} is large, the performance loss becomes serious even when K is large. This is mainly due to the presence of large interference which is generated by a small number of users in the nulling user set. It may be desirable to optimize the number of users in the nulling user set for the generation of MUPB beam weight.



(a) Sum rate.



(b) Number of users in the nulling user set.

Figure 3-8. Performance of MUPB according to δ_{th} when $M = 64$.

Figure 3-9 depicts the normalized mean square error (NMSE) of SLNR approximation according to M when $K = 4,\ 8,\ 16$. We define the NMSE by

$$\mathrm{NMSE} = \frac{1}{K_{\mathrm{total}}} \sum_{k=1}^{K_{\mathrm{total}}} \frac{\left(\mathrm{SLNR}_k^{*} - \mathrm{SLNR}_k^{\mathrm{MUPB}}\right)^2}{\left(\mathrm{SLNR}_k^{*}\right)^2} \tag{3.121}$$

where SLNR_k^{*} is the optimum SLNR of user k obtained by Rayleigh–Ritz quotient theorem, $\mathrm{SLNR}_k^{\mathrm{MUPB}}$ is the approximated SLNR of user k and threshold δ_{th} is determined to minimize the NMSE. It can be seen that the validity of the proposed approximation increases as M increases. This is mainly because the deterministic equivalent property, applied to the approximation of SLNR, becomes accurate as M increases. As a result, the proposed MUPB can provide performance slightly worse than the Max SLNR, while significantly reducing the computational complexity.

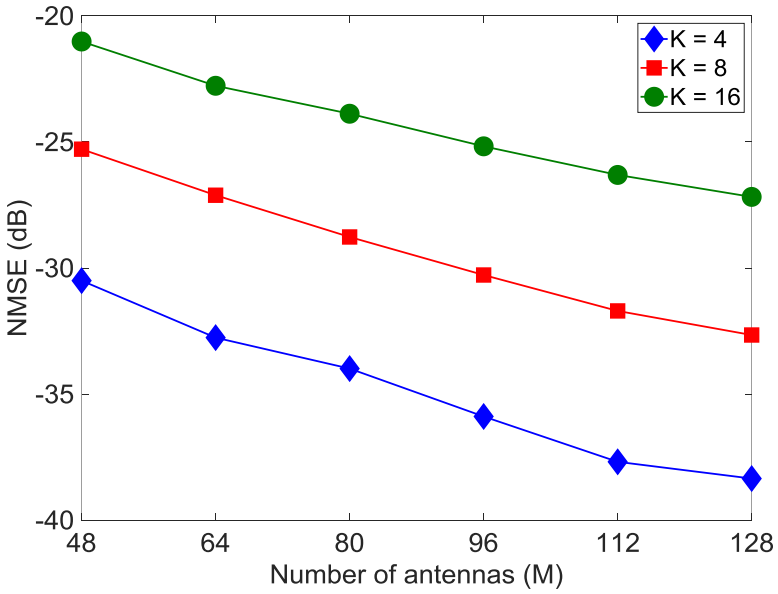


Figure 3-9. NMSE performance of MUPB according to M.

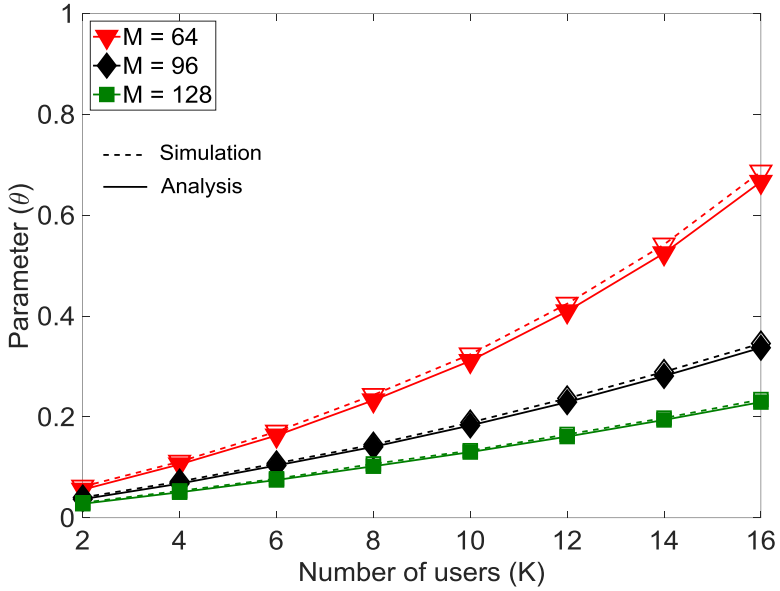
Figure 3-10 depicts θ according to K when $M = 64,\ 96,\ 128$ and $\delta_{\text{th}} = -12,\ -6$ dB. It can be seen that as K increases, θ also increases. When K is large, the PZF may experience performance loss mainly due to the presence of insufficient spatial DoF. In this case, we may need to increase θ to maximize the approximated SLNR. It can be seen that as M increases, θ decreases. When M is large, the approximated SLNR can be maximized by reducing the leakage power with the use of a small value of θ.

Figure 3-11 depicts the performance of the proposed MUPB in terms of the sum rate and the computational complexity according to K when $M = 64$ and $\delta_{\text{th}} = -6$ dB. For comparison, we also evaluate the performance of MRT [6], ZF [9], CEA-ZF [30], Max SNLR [35], and PB that maximizes the approximated SLNR by (3.78) with combined use of MRT and ZF. We also consider an upper bound of the sum rate achieved by the iterative algorithm based on the SINR in [32]. It can be seen that the MUPB provides performance slightly worse than the Max SLNR, while outperforming the other schemes. It can also be seen that the MRT and the ZF may be effective in certain operation environments, and that the CEA-ZF performs better than the ZF since it does not require full dimensional nulling subspace. However, since the CEA-ZF determines the beam weight without consideration of SINR or SLNR, it may not provide noticeable performance improvement. The PB and the MUPB perform better than the CEA-ZF since their beam weight is determined to maximize the approximated SLNR. However, as K increases, the performance gap between the PB and the MUPB increases since the PB may suffer from the presence of insufficient spatial DoF for ZF. It can also be seen that the MUPB provides performance quite close to the upper bound, and that it requires the

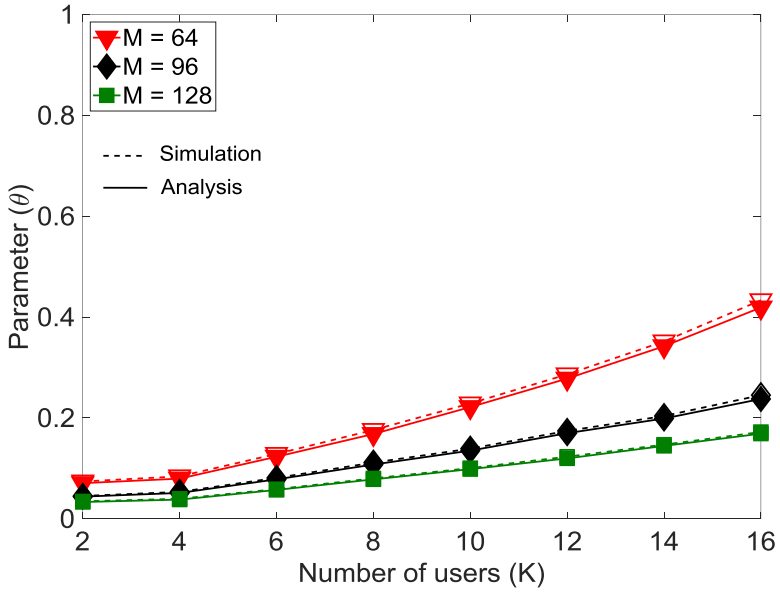
computational complexity much lower than the Max SLNR, the ZF and the PB. The MUPB outperforms the CEA-ZF, while requiring the computational complexity similar to that of CEA-ZF.

Figure 3-12 depicts the sum rate and the computational complexity according to M when $K=16$ and $\delta_{\text{th}}=-6$ dB. It can be seen that the MUPB provides performance improvement over the other schemes except the Max SLNR that may require unaffordable computational complexity.

Figure 3-13 depicts the sum rate according to the channel inaccuracy τ when $M=64$, $K=16$, and $\delta_{\text{th}}=-6$ dB. It can be seen that the MUPB can provide performance similar to the Max SLNR in the presence of CSI inaccuracy. On the other hand, as τ increases, the ZF experiences severe performance loss. In the presence of inaccurate CSI, the ZF may not properly null out the interference. It can also be seen that the MRT performs better than the ZF since the interference by the MRT is not affected by τ. It can also be seen that combined use of MRT and PZF makes the MUPB robust to the presence of CSI inaccuracy, compared to the ZF.

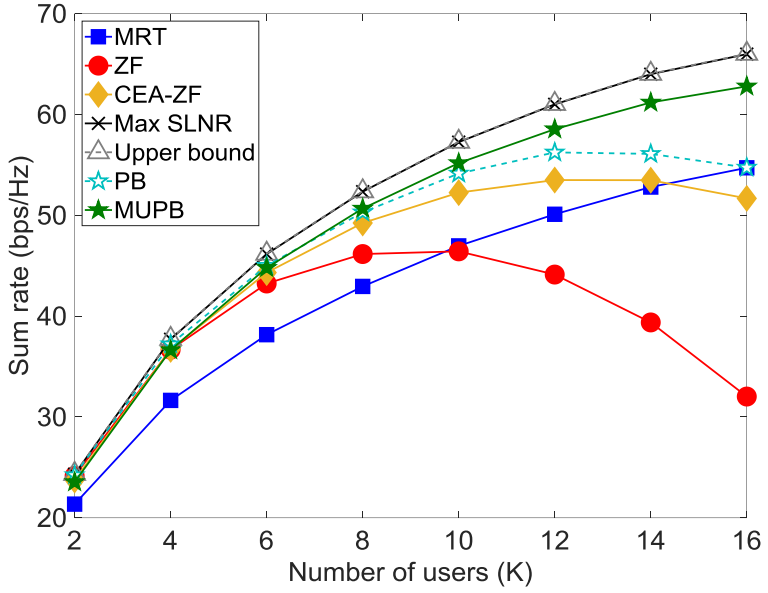


(a) $\delta_{\text{th}} = -12$ dB.

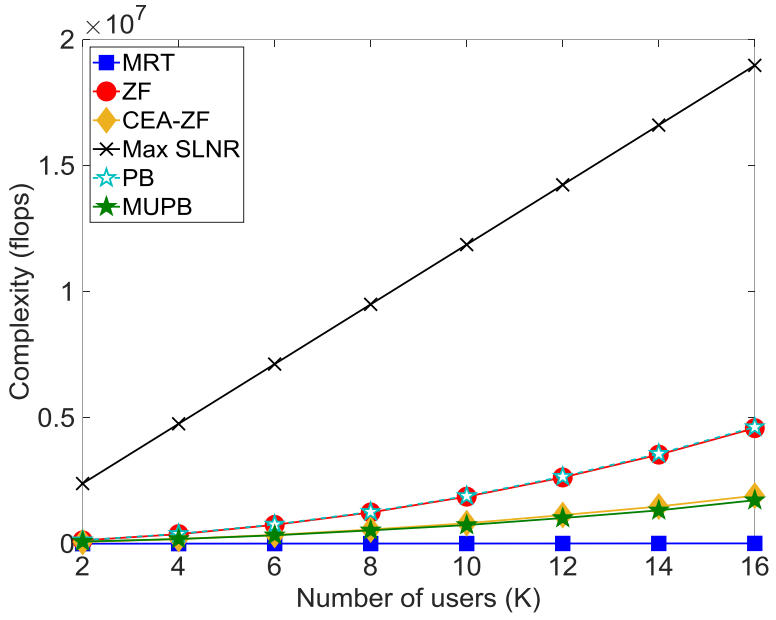


(b) $\delta_{\text{th}} = -6$ dB.

Figure 3-10. Parameter θ according to *K*.

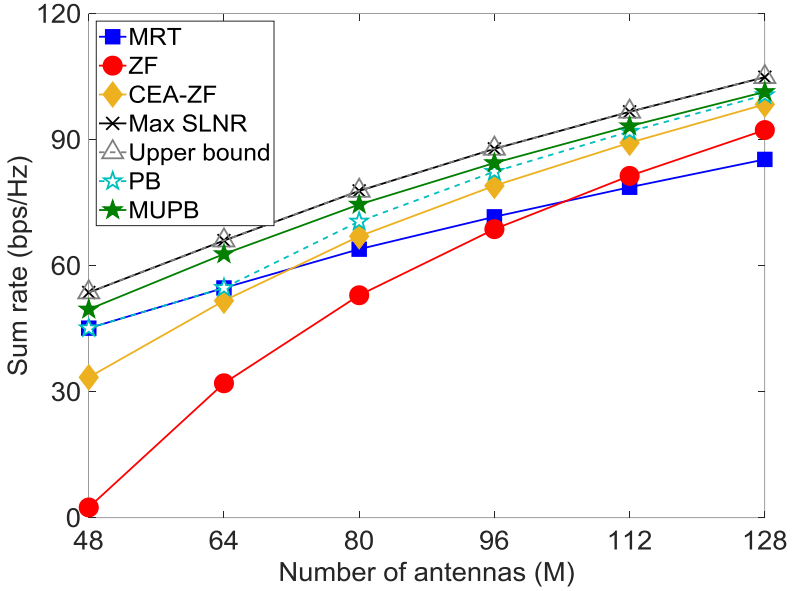


(a) Sum rate.

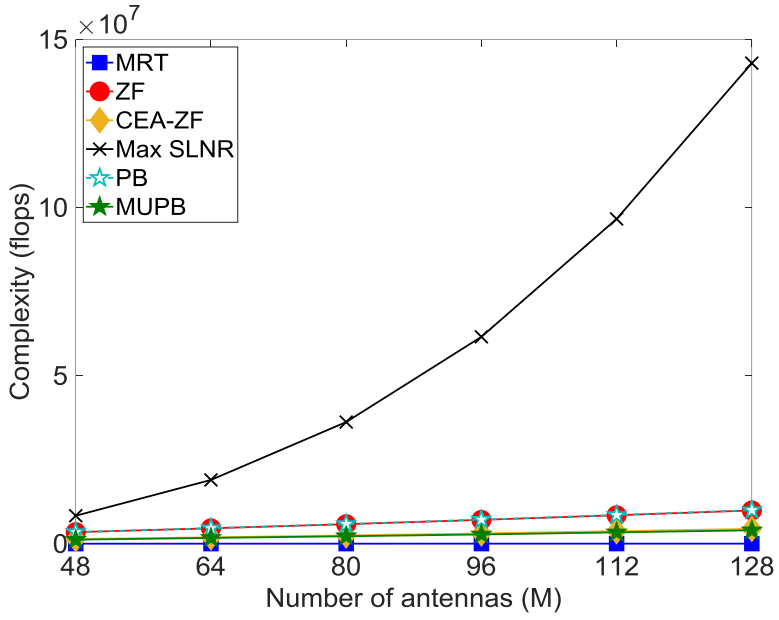


(b) Computational complexity.

Figure 3-11. Performance according to K when $M = 64$ and $\delta_{\text{th}} = -6$ dB.



(a) Sum rate.



(b) Computational complexity.

Figure 3-12. Performance according to M when $K=16$ and $\delta_{\text{th}}=-6$ dB.

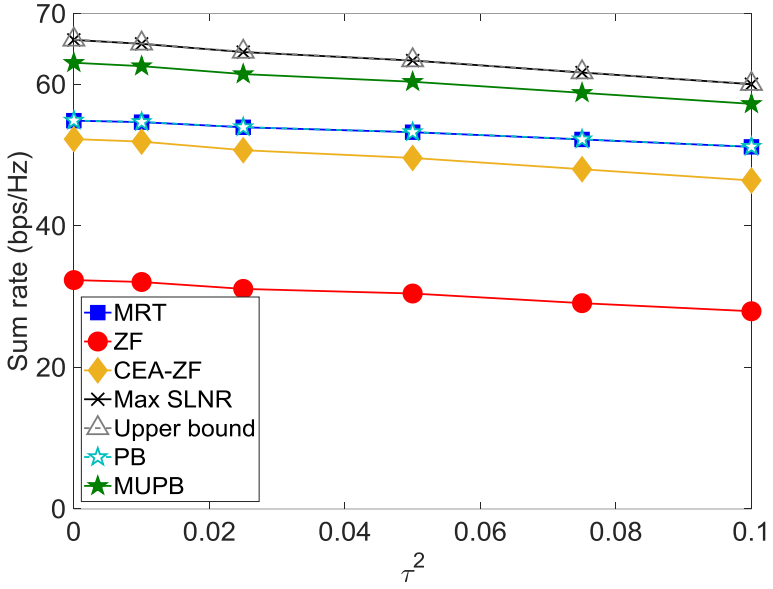


Figure 3-13. Sum rate according to τ.

Chapter 4

User grouping-based ZF transmission

In this chapter, we consider complexity-reduced ZF beamforming that generates the beam weight in a group-wise manner when the number of users is large. We partition users into a number of groups so that users in each group experience TPL lower than a certain level. We approximately estimate the TPL of each user for further reduction of computational complexity. Finally, we determine the beam weight for each user group based on the approximated TPL.

4.1. Spatially correlated channel

In the presence of spatial channel correlation, we may represent the channel $\mathbf{h}_{i,k}$ as [56]

$$\mathbf{h}_{i,k} = \mathbf{R}_{i,k}^{1/2}\tilde{\mathbf{h}}_{i,k} \tag{4.1}$$

where $\mathbf{R}_{i,k}$ is the spatial correlation matrix (SCM) for user k of BS i and $\tilde{\mathbf{h}}_{i,k} \in \mathbb{C}^{M\times 1} \sim CN(\mathbf{0}_M, \mathbf{I}_M)$. Here, [56]

$$\mathbf{R}_{i,k} = \frac{1}{2\Delta_{i,k}} \int_{\vartheta_{i,k}-\Delta_{i,k}}^{\vartheta_{i,k}+\Delta_{i,k}} \mathbf{a}_{i,k}(\vartheta)\mathbf{a}_{i,k}^{H}(\vartheta) d\vartheta \tag{4.2}$$

where $\vartheta_{i,k}$ is the angle of departure (AoD) for user *k* of BS *i*, $\Delta_{i,k}$ is the angular spread (AS) for user *k* of BS *i*, and $\mathbf{a}_{i,k}(\vartheta) \in \mathbb{C}^{M\times 1}$ is the array response vector for user *k* of BS *i* corresponding to ϑ, represented as

$$\mathbf{a}_{i,k}(\vartheta) = \left[1, \cdots, e^{-j2\pi D \sin\vartheta(M-1)}\right]^{T}. \tag{4.3}$$

Here, *D* is the normalized antenna spacing, defined by $D = d / \lambda_c$, where *d* is the antenna spacing and λ_c is the signal wavelength.

We may represent $\mathbf{R}_{i,k}$ as

$$\mathbf{R}_{i,k} = \mathbf{U}_{i,k}\mathbf{\Lambda}_{i,k}\mathbf{U}_{i,k}^{H} \tag{4.4}$$

where $\mathbf{U}_{i,k}$ is an $(M \times b_{i,k})$ matrix comprising orthonormal eigenvectors corresponding to $b_{i,k}$ non-zero eigenvalues of $\mathbf{R}_{i,k}$ and $\mathbf{\Lambda}_{i,k}$ is an $(b_{i,k} \times b_{i,k})$ diagonal matrix comprising non-zero eigenvalues of $\mathbf{R}_{i,k}$. We may represent $\mathbf{U}_{i,k}$ and $\mathbf{\Lambda}_{i,k}$ as, respectively,

$$\mathbf{U}_{i,k} = \left[\mathbf{u}_{i,k,1}, \cdots, \mathbf{u}_{i,k,b_{i,k}}\right] \tag{4.5}$$

$$\mathbf{\Lambda}_{i,k} = \mathrm{diag}\left(\lambda_{i,k,1}, \cdots, \lambda_{i,k,b_{i,k}}\right) \tag{4.6}$$

where $\lambda_{i,k,m}(\geq 0)$ is the *m*-th largest eigenvalue of $\mathbf{R}_{i,k}$ and $\mathbf{u}_{i,k,m}$ is the eigenvector corresponding to $\lambda_{i,k,m}$.

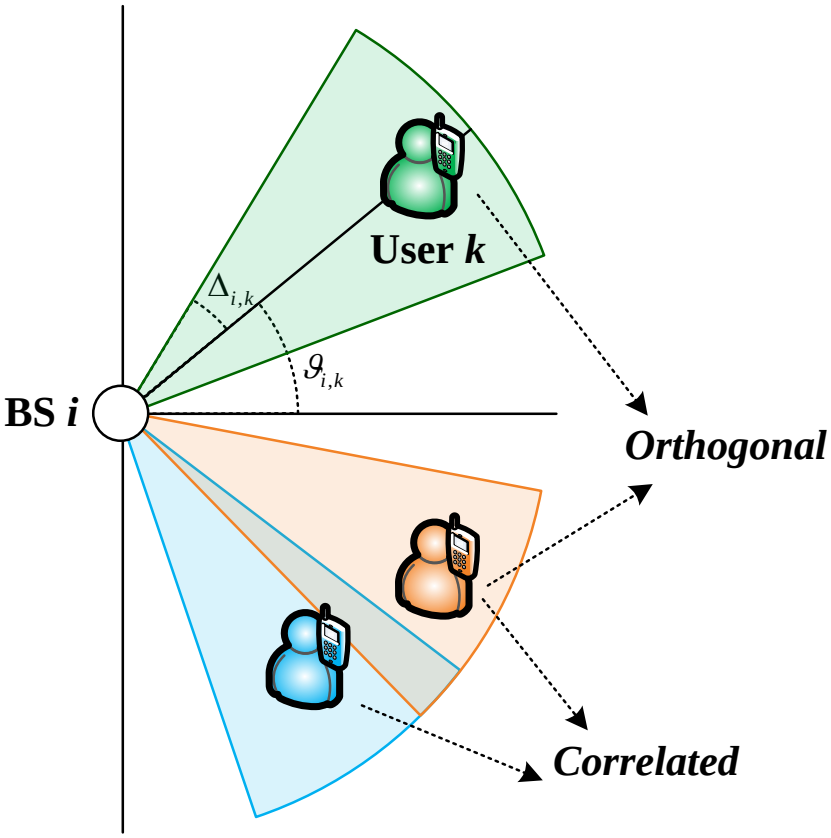


Figure 4-1. Presence of spatial channel correlation.

4.2. Previous works

We briefly review previous works relevant to the user-grouping. Let U, U_g, U_r, and K_g be the set of total users, the set of users selected to group g, the set of remaining users, and the number of users in group g (*i.e.*, $K_g = |U_g|$), respectively. We may represent U_r as

$$U_{\mathrm{r}} = U - \bigcup_{g=1}^{G} U_g \tag{4.7}$$

where G is the number of groups.

4.2.1. Greedy user grouping

We may perform user grouping by sequentially selecting users to maximize the sum rate, referred to greedy user grouping [51, 52]. We may select an initial user for group g by

$$k^* = \arg\max_{k \in U_{\mathrm{r}}} \left\| \mathbf{h}_k \right\|^2 . \tag{4.8}$$

After selecting user k^* as the initial user to group g, we may update the user sets as

$$U_g = U_g \cup \left\{ k^* \right\} \tag{4.9}$$

$$U_{\mathrm{r}} = U_{\mathrm{r}} - \left\{ k^* \right\} . \tag{4.10}$$

We may put a new user l^* to group g to maximize the sum rate by

$$l^* = \arg\max_{l \in U_{\mathrm{r}}} r_{\mathrm{sum}} \left(U_g \cup \{ l \} \right) \tag{4.11}$$

where $r_{\mathrm{sum}} \left(U_g \cup \{ l \} \right)$ is the sum rate of group g when user l is selected to group g and can be represented as

$$r_{\mathrm{sum}} \left(U_g \cup \{ l \} \right) = \sum_{g_k \in U_g} r_{g_k} + r_l . \tag{4.12}$$

Here, g_k denotes the *k*-th user in U_g. After selecting user l^*, we update U_g and U_{r} according to (4.9) and (4.10). We terminate the user grouping for group *g* if the sum rate of group *g* does not increase after the selection of a new user l^* to group *g*. We terminate the user selection for group *g* if

$$r_{\mathrm{sum}}\left(U_g \cup \left\{l^*\right\}\right) < r_{\mathrm{sum}}\left(U_g\right). \tag{4.13}$$

We perform the user selection for the next group $g+1$. We continue this process until all users are grouped (*i.e.*, $U_{\mathrm{r}} = \varnothing$).

The greedy grouping scheme may require huge computational complexity for the generation of beam weight for the user selection, which may not be affordable in m-MIMO operating environments.

4.2.2. Aggregated channel correlation-based user grouping

We may use the aggregation of channel correlation as the grouping metric [53]. We define the aggregated channel correlation of user *l* for group *g* by the sum of channel correlation between user *l* and users in group *g*, represented as

$$\rho_{\mathrm{sum},l,g} = \sum_{g_k \in U_g} \rho_{l,g_k} \tag{4.14}$$

where ρ_{l,g_k} is the channel correlation between user *l* and user g_k, represented as

$$\rho_{l,g_k} = \frac{\left|\mathbf{h}_l^H \mathbf{h}_{g_k}\right|}{\left\|\mathbf{h}_l\right\|\left\|\mathbf{h}_{g_k}\right\|}. \tag{4.15}$$

We may partition users into a number of groups based on the aggregated channel correlation.

We may partition users experiencing high channel correlation into different groups. If the channel correlation between users is higher than a threshold α, we may partition them into different groups. We continue this process until the presence of no user in U_{r} experiencing channel correlation higher than α. We sequentially allocate the rest of users to groups in consideration of aggregated channel correlation. We may randomly allocate a new user l^* in U_{r} to group g^* to minimize the aggregated channel correlation by

$$g^* = \arg\min_{g=1,\cdots,G} \rho_{\mathrm{sum},l^*,g} . \tag{4.16}$$

We continue this process until all users are grouped.

We may reduce the computational complexity for the user grouping based on the aggregated channel correlation without the generation of beam weight. However, the aggregated channel correlation may not properly characterize the TPL.

4.2.3. ε-orthogonality-based user grouping

We may perform the user grouping by exploiting ε-orthogonal condition that can be represented as

$$\rho_{k,l} < \varepsilon \tag{4.17}$$

where $\varepsilon\left(\in[0,1]\right)$ is a threshold to be determined. Based on the ε-orthogonality, we may partition users into a number of groups.

Let G_0 be the initial number of groups. We select an initial user for each group. We may select the initial user k^* for group g $(g=1,\cdots,G_0)$ based on the channel gain according to (4.8). After the initial user selection, we partition the rest of users into a number of groups. We may perform the grouping by a two-step process; user selection and group allocation. We select a new user l^* having the largest channel gain among the rest of users. For allocation of user l^* to a group, we calculate the maximum channel correlation of user l^* in each group. We may represent the maximum channel correlation of user l^* in group g as

$$\rho_{\max,l^*,g} = \max_{g_k \in U_g} \rho_{l^*,g_k} . \tag{4.18}$$

We may allocate user l^* to group g^* that minimizes the maximum channel correlation as

$$g^* = \arg\min_{1 \le g \le G_0} \rho_{\max,l^*,g} . \tag{4.19}$$

If the ε-orthogonal condition for group g^* is not satisfied, we increase the number of groups by one (*i.e.*, $G_0 \leftarrow G_0 + 1$) and allocate user l^* to the new group G_0. We continue this process until all users are grouped.

We may reduce the computational complexity by using the ε-orthogonality. However, it may not be effective to minimize performance loss by the interference nulling since the ε-orthogonality may not provide the reduction of TPL.

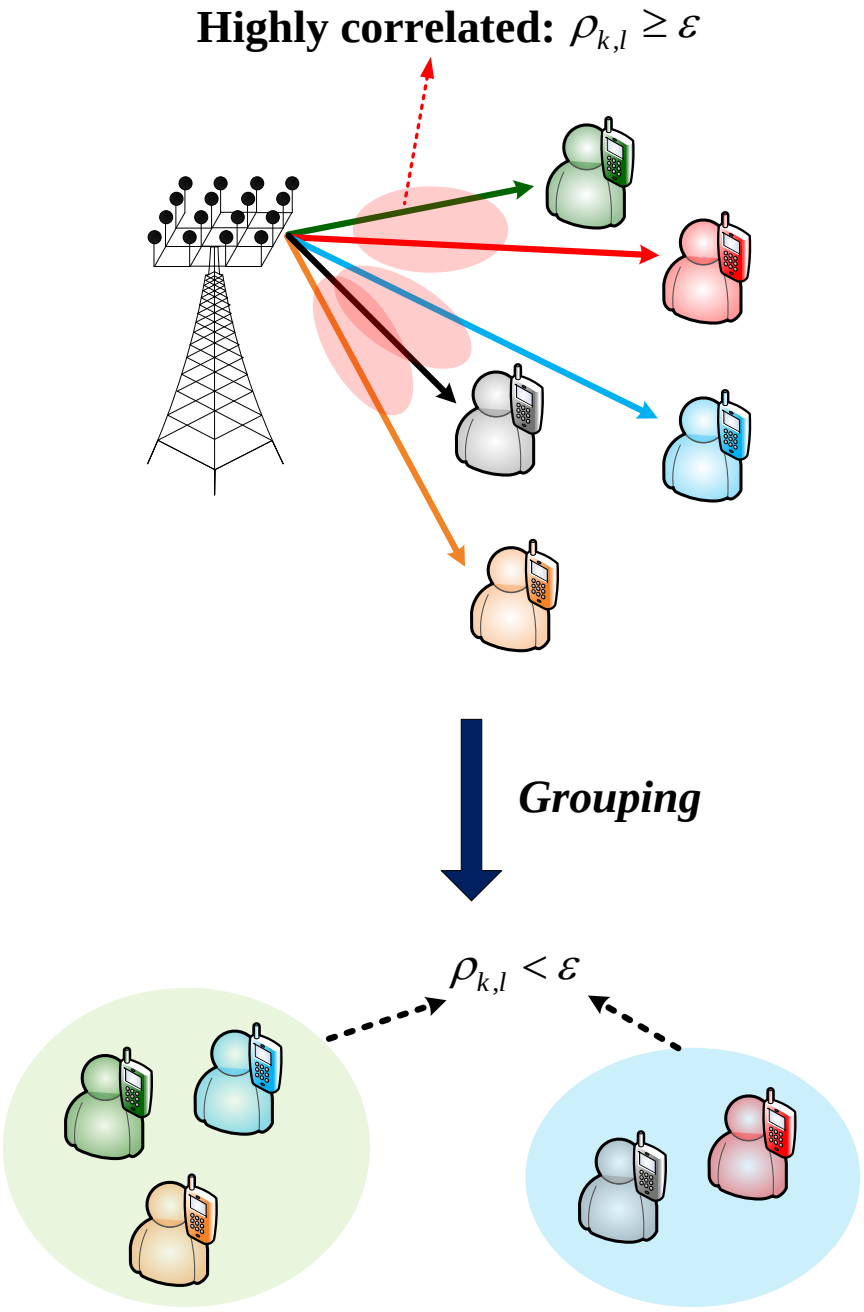


Figure 4-2. An ε-orthogonality-based user grouping scheme.

4.2.4. Chordal distance-based user grouping

We may consider the chordal distance as the grouping metric [60]. We define the chordal distance between user group *g* and *t* by

$$C(g,t)=\frac{\sum_{g_k\in U_g}\sum_{g_l\in U_g\setminus\{g_k\}} d(g_k,g_l)+\sum_{t_k\in U_t}\sum_{t_l\in U_t\setminus\{t_k\}} d(t_k,t_l)+2\sum_{g_k\in U_g}\sum_{t_k\in U_t} d(g_k,t_k)}{(K_g+K_t)(K_g+K_t-1)} \tag{4.20}$$

where $d(g_k,g_l)$ is the chordal distance between user g_k and user g_l, and can be represented as

$$d(g_k,g_l)=\sqrt{1-\frac{\left|\mathbf{h}_{g_k}^H\mathbf{h}_{g_l}\right|^2}{\left\|\mathbf{h}_{g_k}\right\|^2\left\|\mathbf{h}_{g_l}\right\|^2}}\,. \tag{4.21}$$

We may partition users into a number of groups by employing an agglomerative hierarchical clustering (AHC) method [61]. As illustrated in Figure 4-3, we allocate a single user to each group. We calculate the chordal distance between two groups by (4.20). Then, we merge a pair of groups with the maximum chordal distance into a single group so as to maximize the chordal distance of users in the group. In this manner, we may sequentially reduce the number of groups until the number of groups equals to a predetermined number *G*.

We may reduce the computational complexity by using the chordal distance without consideration of beam weight. However, the chordal distance may not properly characterize the TPL. It may be desirable to use a grouping metric related to the TPL, while requiring reduced-complexity for the implementation.

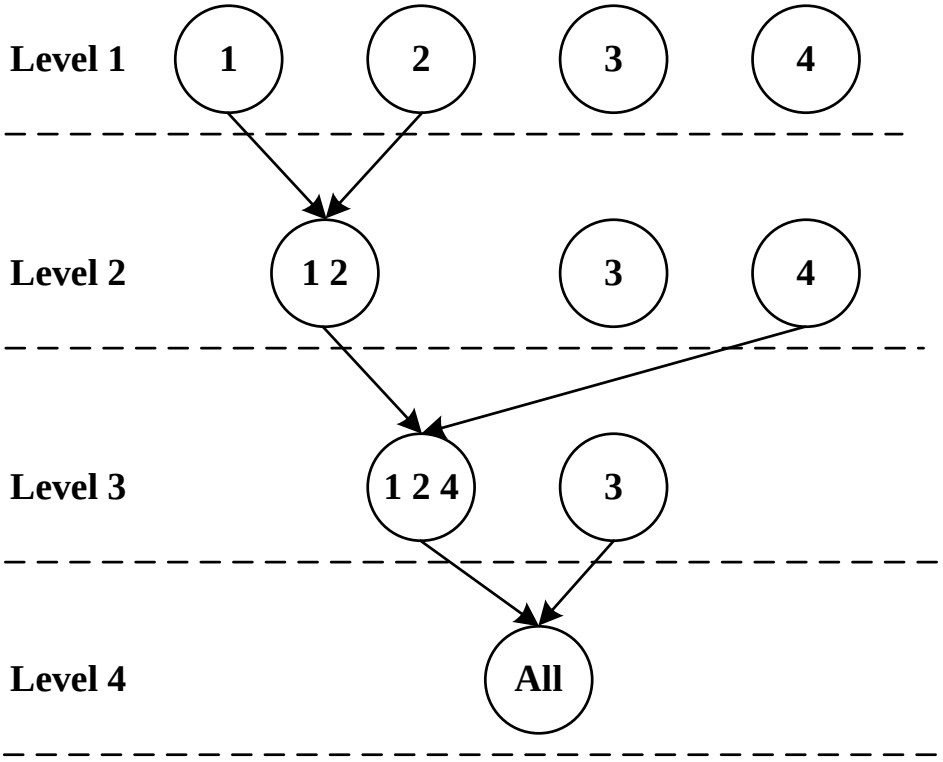


Figure 4-3. An example of user grouping based on AHC.

4.3. Proposed scheme

4.3.1. User grouping

Consider the case that multiple BSs jointly transmit signal by means of ZF beamforming. We may determine the beam weight for user *k* of BS *i* by [9]

$$\mathbf{v}_{i,k}^{\mathrm{ZF}} = \frac{\mathbf{w}_{i,k}}{\left\|\mathbf{w}_{i,k}\right\|} \tag{4.22}$$

where $\mathbf{w}_{i,k}$ is the *k*-th column of $\mathbf{H}_i\mathbf{G}_i^{-1}$ and $\mathbf{G}_i = \mathbf{H}_i^H\mathbf{H}_i$. Here,

$$\mathbf{H}_i = \left[\mathbf{h}_{i,1} \cdots \mathbf{h}_{i,K_{\text{total}}} \right]. \tag{4.23}$$

The ZF may suffer from performance loss due to the interference nulling (*i.e.*, the TPL). It may be desirable to partition users into different groups so that users in each group experience low TPL as illustrated in Figure 4-4. However, it may need very high computational complexity to estimate the TPL in m-MIMO environments since it requires the beam weight. For the reduction of computational complexity, we consider the estimation of TPL without the use of beam weight. To this end, we analyze the received signal in a deterministic equivalent form and then approximate the TPL by exploiting the property of m-MIMO channel. We partition users into a number of groups based on the approximated TPL. Finally, we generate the beam weight for each user group assuming that we allocate the transmission resource to user groups in an orthogonal manner.

The received signal power of user k can be represented as

$$\begin{aligned} S_k &= \left| \sum_{i \in \Omega} \sqrt{p_{i,k} \alpha_{i,k}} \mathbf{h}_{i,k}^H \mathbf{v}_{i,k}^{\text{ZF}} \right|^2 \\ &= \left(\sum_{i \in \Omega} \sqrt{p_{i,k} \alpha_{i,k}} \left| \mathbf{h}_{i,k}^H \mathbf{v}_{i,k}^{\text{ZF}} \right| \right)^2 \\ &= \sum_{i \in \Omega} S_{i,k} + \sum_{i \in \Omega} \sqrt{S_{i,k}} \sum_{j \in \Omega \setminus \{i\}} \sqrt{S_{j,k}} \end{aligned} \tag{4.24}$$

where

$$S_{i,k} = p_{i,k} \alpha_{i,k} \left| \mathbf{h}_{i,k}^H \mathbf{v}_{i,k}^{\text{ZF}} \right|^2 \tag{4.25}$$

and the second equality follows from the fact that $\mathbf{h}_{i,k}^H \mathbf{v}_{i,k}^{\text{ZF}}$ is non-negative.

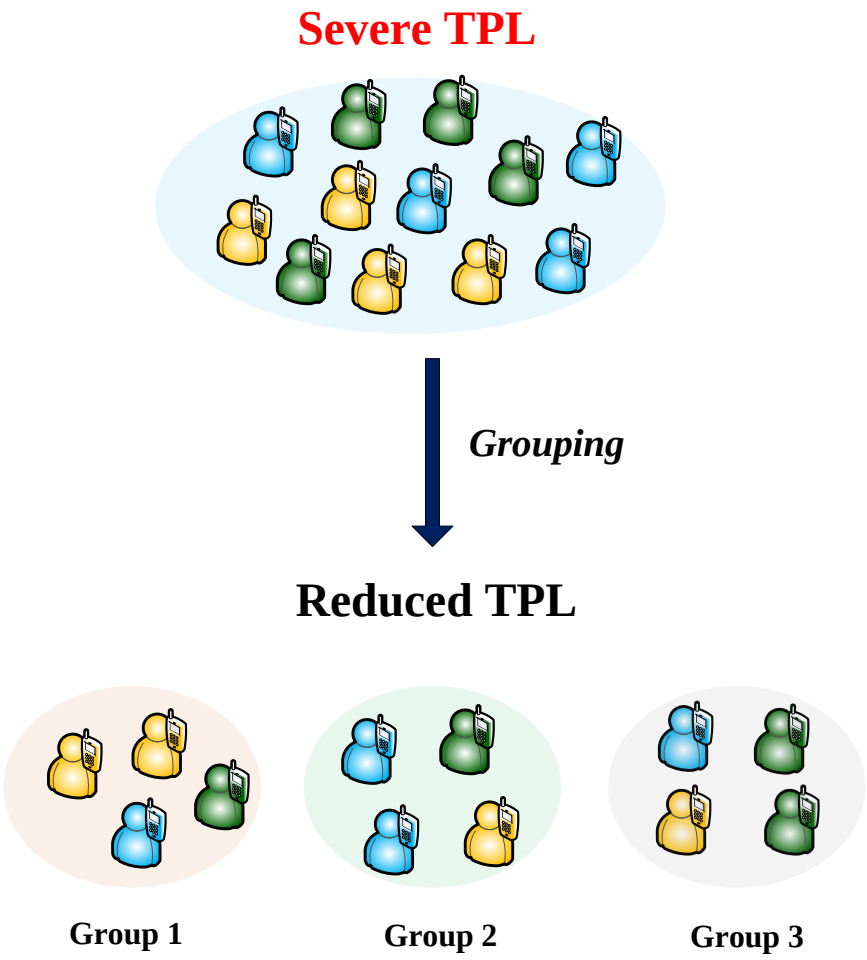


Figure 4-4. An example of user grouping to minimize the TPL.

It can be shown that [10]

$$S_{i,k} = G_{i,k} - T_{i,k} \tag{4.26}$$

where $G_{i,k}$ is the TPL-free transmission gain of user k of BS i and can be represented as

$$G_{i,k} = p_{i,k}\alpha_{i,k}\left\|\mathbf{h}_{i,k}\right\|^2 \tag{4.27}$$

and $T_{i,k}$ is the TPL of user k due to the interference nulling of BS i and can be represented as

$$T_{i,k} = \sum_{l \neq k}^{K_{\text{total}}} T_{i,k,l} \ . \tag{4.28}$$

Here, $T_{i,k,l}$ is the TPL of user k due to the interference nulling for user l of BS i and can be represented as

$$T_{i,k,l} = p_{i,k}\alpha_{i,k}\left(-1\right)^{k+l+1}\mathbf{h}_{i,k}^{H}\mathbf{h}_{i,l}\Phi_{i,k,l} \tag{4.29}$$

where

$$\Phi_{i,k,l} = \frac{\det\left(\mathbf{G}_i^{\langle k,l\rangle}\right)}{\det\left(\mathbf{G}_i^{\langle k,k\rangle}\right)}. \tag{4.30}$$

We may represent $G_{i,k}$ and $T_{i,k}$ in a deterministic equivalent form. It can be shown from Lemma 3-1 that

$$G_{i,k} - \bar{G}_{i,k} \xrightarrow[M\to\infty]{\text{a.s.}} 0 \tag{4.31}$$

where $\bar{G}_{i,k}$ is a deterministic equivalent of $G_{i,k}$ and can be represented as [40]

$$\bar{G}_{i,k} = p_{i,k}\alpha_{i,k}M \ . \tag{4.32}$$

It can be shown that [70]

$$\det\left(\mathbf{G}_i^{\langle k,l\rangle}\right) = \sum_{m=1}^{K_{\text{total}}-1}\left(-1\right)^{m+q_{k,l}}\mathbf{h}_{i,a}^{H}\mathbf{h}_{i,k}\det\left(\left[\mathbf{G}_i^{\langle k,l\rangle}\right]^{(m,q_{k,l})}\right) \tag{4.33}$$

$$\det\left(\mathbf{G}_i^{\langle k,k\rangle}\right)=\sum_{m=1}^{K_{\text{total}}-1}(-1)^{m+q_{l,k}}\,\mathbf{h}_{i,a}^{H}\mathbf{h}_{i,l}\det\left(\left[\mathbf{G}_i^{\langle k,k\rangle}\right]^{(m,q_{l,k})}\right) \tag{4.34}$$

where

$$q_{k,l}=\begin{cases}k & \text{for } k<l\\ k-1 & \text{for } k>l\end{cases} \tag{4.35}$$

$$a=\begin{cases}m & \text{for } m<k\\ m+1 & \text{for } m>k\end{cases}. \tag{4.36}$$

Then, it can be shown that

$$\Phi_{i,k,l}=\frac{(-1)^{z_{k,l}}\sum_{m=1}^{K_{\text{total}}-1}w_{i,k,l}^{(m)}\mathbf{h}_{i,a}^{H}\mathbf{h}_{i,k}}{\sum_{m=1}^{K_{\text{total}}-1}w_{i,k,l}^{(m)}\mathbf{h}_{i,a}^{H}\mathbf{h}_{i,l}} \tag{4.37}$$

where

$$z_{k,l}=\begin{cases}k-l+1 & \text{for } k<l\\ k-l-1 & \text{for } k>l\end{cases} \tag{4.38}$$

$$w_{i,k,l}^{(m)}=(-1)^{m}\det\left(\left[\mathbf{G}_i^{\langle k,k\rangle}\right]^{(m,q_{l,k})}\right). \tag{4.39}$$

It can also be shown that

$$T_{i,k,l}=p_{i,k}\alpha_{i,k}(-1)^{k+l+z_{k,l}+1}\left(\frac{\sum_{m=1}^{K_{\text{total}}-1}w_{i,k,l}^{(m)}\mathbf{h}_{i,k}^{H}\mathbf{h}_{i,l}\mathbf{h}_{i,a}^{H}\mathbf{h}_{i,k}}{\sum_{m=1}^{K_{\text{total}}-1}w_{i,k,l}^{(m)}\mathbf{h}_{i,a}^{H}\mathbf{h}_{i,l}}\right). \tag{4.40}$$

Since $(-1)^{k+l+z_{k,l}+1}=1$ irrespective of k and l, it can be shown that

$$
\begin{aligned}
T_{i,k,l} &= p_{i,k}\alpha_{i,k}\left(\frac{\sum_{m=1}^{K_{\text{total}}-1} w_{i,k,l}^{(m)} N_{i,k,l}^{(m)}}{\sum_{m=1}^{K_{\text{total}}-1} w_{i,k,l}^{(m)} D_{i,k,l}^{(m)}}\right) \\
&= p_{i,k}\alpha_{i,k}\left(\frac{w_{i,k,l}^{(q_{l,k})} N_{i,k,l}^{(q_{l,k})} + \sum_{m\neq q_{l,k}}^{K_{\text{total}}-1} w_{i,k,l}^{(m)} N_{i,k,l}^{(m)}}{w_{i,k,l}^{(q_{l,k})} D_{i,k,l}^{(q_{l,k})} + \sum_{m\neq q_{l,k}}^{K_{\text{total}}-1} w_{i,k,l}^{(m)} D_{i,k,l}^{(m)}}\right)
\end{aligned}
\tag{4.41}
$$

where

$$
N_{i,k,l}^{(m)} = \mathbf{h}_{i,k}^{H}\mathbf{h}_{i,l}\mathbf{h}_{i,a}^{H}\mathbf{h}_{i,k} \tag{4.42}
$$

$$
D_{i,k,l}^{(m)} = \mathbf{h}_{i,a}^{H}\mathbf{h}_{i,l}\,. \tag{4.43}
$$

When $m = q_{l,k}$, it can be shown that

$$
\begin{aligned}
N_{i,k,l}^{(q_{l,k})} &= \mathbf{h}_{i,k}^{H}\mathbf{h}_{i,l}\mathbf{h}_{i,l}^{H}\mathbf{h}_{i,k} \\
&= \tilde{\mathbf{h}}_{i,k}^{H}\mathbf{R}_{i,k}^{1/2}\mathbf{h}_{i,l}\mathbf{h}_{i,l}^{H}\mathbf{R}_{i,k}^{1/2}\tilde{\mathbf{h}}_{i,k}\,.
\end{aligned}
\tag{4.44}
$$

It can be shown from Lemma 3-1 that

$$
\tilde{\mathbf{h}}_{i,k}^{H}\mathbf{R}_{i,k}^{1/2}\mathbf{h}_{i,l}\mathbf{h}_{i,l}^{H}\mathbf{R}_{i,k}^{1/2}\tilde{\mathbf{h}}_{i,k} - \mathrm{Tr}\left(\mathbf{R}_{i,k}^{1/2}\mathbf{h}_{i,l}\mathbf{h}_{i,l}^{H}\mathbf{R}_{i,k}^{1/2}\right) \xrightarrow[M\to\infty]{\text{a.s.}} 0\,. \tag{4.45}
$$

Since $\mathrm{Tr}\left(\mathbf{R}_{i,k}^{1/2}\mathbf{h}_{i,l}\mathbf{h}_{i,l}^{H}\mathbf{R}_{i,k}^{1/2}\right) = \tilde{\mathbf{h}}_{i,l}^{H}\mathbf{R}_{i,l}^{1/2}\mathbf{R}_{i,k}\mathbf{R}_{i,l}^{1/2}\tilde{\mathbf{h}}_{i,l}$, it can be shown from Lemma 3-1 that

$$
\mathrm{Tr}\left(\mathbf{R}_{i,k}^{1/2}\mathbf{h}_{i,l}\mathbf{h}_{i,l}^{H}\mathbf{R}_{i,k}^{1/2}\right) - \mathrm{Tr}\left(\mathbf{R}_{i,k}\mathbf{R}_{i,l}\right) \xrightarrow[M\to\infty]{\text{a.s.}} 0\,. \tag{4.46}
$$

Then, it can be shown that

$$
N_{i,k,l}^{(q_{l,k})} - \bar{N}_{i,k,l}^{(q_{l,k})} \xrightarrow[M\to\infty]{\text{a.s.}} 0 \tag{4.47}
$$

where $\bar{N}_{i,k,l}^{(q_{l,k})}$ is a deterministic equivalent of $N_{i,k,l}^{(q_{l,k})}$ and can be represented as

$$\bar{N}_{i,k,l}^{(q_{l,k})} = \mathrm{Tr}\left(\mathbf{R}_{i,k}\mathbf{R}_{i,l}\right). \tag{4.48}$$

Since $D_{i,k,l}^{(q_{l,k})} = \tilde{\mathbf{h}}_{i,l}^{H}\mathbf{R}_{i,l}^{1/2}\mathbf{R}_{i,l}^{1/2}\tilde{\mathbf{h}}_{i,l}$, it can also be shown from Lemma 3-1 that

$$D_{i,k,l}^{(q_{l,k})} - \bar{D}_{i,k,l}^{(q_{l,k})} \xrightarrow[M\to\infty]{\text{a.s.}} 0 \tag{4.49}$$

where $\bar{D}_{i,k,l}^{(q_{l,k})}$ is a deterministic equivalent of $D_{i,k,l}^{(q_{l,k})}$ and can be represented as

$$\bar{D}_{i,k,l}^{(q_{l,k})} = M\ . \tag{4.50}$$

When $m \neq q_{l,k}$, it can be shown that

$$\begin{aligned} N_{i,k,l}^{(m)} &= \mathbf{h}_{i,k}^{H}\mathbf{h}_{i,l}\mathbf{h}_{i,a}^{H}\mathbf{h}_{i,k} \\ &= \tilde{\mathbf{h}}_{i,a}^{H}\mathbf{R}_{i,a}^{1/2}\mathbf{h}_{i,k}\mathbf{h}_{i,k}^{H}\mathbf{R}_{i,l}^{1/2}\tilde{\mathbf{h}}_{i,l}. \end{aligned} \tag{4.51}$$

Since $\tilde{\mathbf{h}}_{i,a}$ and $\tilde{\mathbf{h}}_{i,l}$ are independent of each other, it can be shown from Lemma 3-2 that

$$\tilde{\mathbf{h}}_{i,a}^{H}\mathbf{R}_{i,a}^{1/2}\mathbf{h}_{i,k}\mathbf{h}_{i,k}^{H}\mathbf{R}_{i,l}^{1/2}\tilde{\mathbf{h}}_{i,l} \xrightarrow[M\to\infty]{\text{a.s.}} 0\ . \tag{4.52}$$

Then, it can be shown that

$$N_{i,k,l}^{(m)} - \bar{N}_{i,k,l}^{(m)} \xrightarrow[M\to\infty]{\text{a.s.}} 0 \tag{4.53}$$

where $\bar{N}_{i,k,l}^{(m)}$ is a deterministic equivalent of $N_{i,k,l}^{(m)}$ and can be represented as

$$\bar{N}_{i,k,l}^{(m)} = 0\ . \tag{4.54}$$

Similarly, since $D_{i,k,l}^{(m)} = \tilde{\mathbf{h}}_{i,k}^{H} \mathbf{R}_{i,k}^{1/2} \mathbf{R}_{i,l}^{1/2} \tilde{\mathbf{h}}_{i,l}$, it can be shown from Lemma 3-2 that

$$D_{i,k,l}^{(m)} - \bar{D}_{i,k,l}^{(m)} \xrightarrow[M \to \infty]{\text{a.s.}} 0 \tag{4.55}$$

where $\bar{D}_{i,k,l}^{(m)}$ is a deterministic equivalent of $D_{i,k,l}^{(m)}$ and can be represented as

$$\bar{D}_{i,k,l}^{(m)} = 0 . \tag{4.56}$$

It can also be shown that

$$T_{i,k,l} - \bar{T}_{i,k,l} \xrightarrow[M \to \infty]{\text{a.s.}} 0 \tag{4.57}$$

where $\bar{T}_{i,k,l}$ is a deterministic equivalent of $T_{i,k,l}$ and can be represented as

$$\bar{T}_{i,k,l} = \frac{p_{i,k} \alpha_{i,k} \mathrm{Tr}\left(\mathbf{R}_{i,k} \mathbf{R}_{i,l}\right)}{M} . \tag{4.58}$$

It can also be shown that

$$T_{i,k} - \bar{T}_{i,k} \xrightarrow[M \to \infty]{\text{a.s.}} 0 \tag{4.59}$$

where $\bar{T}_{i,k}$ is a deterministic equivalent of $T_{i,k}$ and can be represented as

$$\begin{aligned} \bar{T}_{i,k} &= \sum_{l \neq k}^{K_{\text{total}}} \bar{T}_{i,k,l} \\ &= p_{i,k} \alpha_{i,k} \sum_{l \neq k}^{K_{\text{total}}} \frac{\mathrm{Tr}\left(\mathbf{R}_{i,k} \mathbf{R}_{i,l}\right)}{M} . \end{aligned} \tag{4.60}$$

It can be shown that a deterministic equivalent of S_k can be represented as

$$\bar{S}_k = \sum_{i \in \Omega} \bar{S}_{i,k} + \sum_{i \in \Omega} \sqrt{\bar{S}_{i,k}} \left(\sum_{j \in \Omega \setminus \{i\}} \sqrt{\bar{S}_{j,k}} \right) \tag{4.61}$$

where $\bar{S}_{i,k}$ is a deterministic equivalent of $S_{i,k}$ and can be represented as

$$\bar{S}_{i,k} = p_{i,k}\alpha_{i,k}M - \bar{T}_{i,k}\,. \tag{4.62}$$

The estimation of $\bar{T}_{i,k}$ may require a complexity of $O\left(K_{\text{total}}M^3\right)$, which may not be affordable in m-MIMO environments. For further reduction of computational complexity, we consider the following lemmas.

Lemma 4-1. As $M \to \infty$, the eigenvectors of $\mathbf{R}_{i,k}$ can approximately be represented as [71]

$$Q_{i,k} \approx \left\{\mathbf{f}_m \mid m \in B_{i,k}\right\}. \tag{4.63}$$

Here, $\mathbf{f}_m$ is the *m*-th orthonormal basis vector in $Q = \left\{\mathbf{f}_m \mid m = 1,\cdots,M\right\}$ and can be represented as

$$\mathbf{f}_m = \frac{1}{\sqrt{M}}\left[1,\cdots,e^{-j2\pi\sin(\omega_m)(M-1)}\right]^T \tag{4.64}$$

where

$$\omega_m = \sin^{-1}\left(\frac{2(m-1)-M}{2M}\right) \tag{4.65}$$

and $B_{i,k}$ is the set of indices of basis vectors for user *k* of BS *i* and can be represented as

$$B_{i,k} = \left\{m_{\min,i,k},\cdots,m_{\max,i,k}\right\}. \tag{4.66}$$

Here,

$$m_{\min,i,k} = \left\lceil MD\sin\left(\vartheta_{i,k} - \Delta_{i,k}\right) + \frac{M}{2} + 1\right\rceil \tag{4.67}$$

$$m_{\max,i,k} = \left\lfloor MD\sin\left(\vartheta_{i,k} + \Delta_{i,k}\right) + \frac{M}{2} + 1 \right\rfloor. \tag{4.68}$$

Lemma 4-2. As $M \to \infty$, the eigenvalues of $\mathbf{R}_{i,k}$ can approximately be represented as [71]

$$\lambda_{i,k,m} \approx \frac{M}{b_{i,k}},\ m = 1, \cdots, b_{i,k} \tag{4.69}$$

where $b_{i,k}$ is the size of $B_{i,k}$ and can be approximated as

$$b_{i,k} \approx 2DM\cos\left(\vartheta_{i,k}\right)\sin\left(\Delta_{i,k}\right) + 1. \tag{4.70}$$

It can be shown that

$$\begin{aligned} \mathrm{Tr}\left(\mathbf{R}_{i,k}\mathbf{R}_{i,l}\right) &= \mathrm{Tr}\left(\mathbf{U}_{i,k}\mathbf{\Lambda}_{i,k}\mathbf{U}_{i,k}^{H}\mathbf{U}_{i,l}\mathbf{\Lambda}_{i,l}\mathbf{U}_{i,l}^{H}\right) \\ &= \mathrm{Tr}\left(\left(\mathbf{\Lambda}_{i,k}\mathbf{U}_{i,k}^{H}\mathbf{U}_{i,l}\right)\left(\mathbf{\Lambda}_{i,l}\mathbf{U}_{i,l}^{H}\mathbf{U}_{i,k}\right)\right) \\ &= \sum_{p=1}^{b_{i,k}} \lambda_{i,k,p}\left(\sum_{q=1}^{b_{i,l}} \rho_{i,k,p}^{i,l,q}\lambda_{i,l,q}\right) \end{aligned} \tag{4.71}$$

where

$$\rho_{i,k,p}^{i,l,q} = \left|\mathbf{u}_{i,k,p}^{H}\mathbf{u}_{i,l,q}\right|^{2}. \tag{4.72}$$

It can be shown from Lemma 4-1 and 4-2 that

$$\begin{aligned} \sum_{p=1}^{b_{i,k}} \lambda_{i,k,p}\left(\sum_{q=1}^{b_{i,l}} \rho_{i,k,p}^{i,l,q}\lambda_{i,l,q}\right) &\approx \frac{M^{2}}{b_{i,k}b_{i,l}} \sum_{m_{i,k}\in B_{i,k}}\left(\sum_{m_{i,l}\in B_{i,l}} \left|\mathbf{f}_{m_{i,k}}^{H}\mathbf{f}_{m_{i,l}}\right|^{2}\right) \\ &= c_{i,k,l}M^{2}. \end{aligned} \tag{4.73}$$

Here, $c_{i,k,l}$ is the performance degradation term associated with the channel correlation between user *k* and user *l*, and can be represented as

$$c_{i,k,l} = \frac{N\left(B_{i,k}, B_{i,l}\right)}{b_{i,k} b_{i,l}} \tag{4.74}$$

where $N\left(B_{i,k}, B_{i,l}\right)$ is the number of common elements between $B_{i,k}$ and $B_{i,l}$, and can be represented as

$$N\left(B_{i,k}, B_{i,l}\right) = \left|B_{i,k} \cap B_{i,l}\right|. \tag{4.75}$$

Using $\mathrm{Tr}\left(\mathbf{R}_{i,k}\mathbf{R}_{i,l}\right) \approx c_{i,k,l} M^2$, $\bar{T}_{i,k,l}$ can be approximated as

$$\begin{aligned} \bar{T}_{i,k,l} &\approx \tilde{T}_{i,k,l} \\ &= p_{i,k}\alpha_{i,k}c_{i,k,l}M. \end{aligned} \tag{4.76}$$

Then, it can be shown that

$$\begin{aligned} \bar{T}_{i,k} &\approx \tilde{T}_{i,k} \\ &= \sum_{l \neq k}^{K_{\text{total}}} \tilde{T}_{i,k,l}. \end{aligned} \tag{4.77}$$

It can also be shown that

$$\bar{S}_k \approx \left(A_{k,1} + A_{k,2}\right)M \tag{4.78}$$

where

$$A_{k,1} = \sum_{i \in \Omega} p_{i,k} \alpha_{i,k} - \frac{1}{M} \sum_{i \in \Omega} \tilde{T}_{i,k} \tag{4.79}$$

$$A_{k,2} = \sum_{i \in \Omega} \sqrt{p_{i,k} \alpha_{i,k} - \frac{\tilde{T}_{i,k}}{M}} \left(\sum_{j \in \Omega \setminus \{i\}} \sqrt{p_{j,k} \alpha_{j,k} - \frac{\tilde{T}_{j,k}}{M}} \right). \tag{4.80}$$

It can be shown from the first-order Taylor approximation [42] that

$$\sqrt{p_{i,k} \alpha_{i,k} - \frac{\tilde{T}_{i,k}}{M}} \approx \sqrt{p_{i,k} \alpha_{i,k}} - \frac{\tilde{T}_{i,k}}{2M \sqrt{p_{i,k} \alpha_{i,k}}}. \tag{4.81}$$

It can also be shown that

$$\begin{aligned} A_{k,2} &\approx \sum_{i \in \Omega} \left(\sqrt{p_{i,k} \alpha_{i,k}} - \frac{\tilde{T}_{i,k}}{2M \sqrt{p_{i,k} \alpha_{i,k}}} \right) \left[\sum_{j \in \Omega \setminus \{i\}} \left(\sqrt{p_{j,k} \alpha_{j,k}} - \frac{\tilde{T}_{j,k}}{2M \sqrt{p_{j,k} \alpha_{j,k}}} \right) \right] \\ &= A_{k,2,1} + A_{k,2,2} + A_{k,2,3} + A_{k,2,4} \end{aligned} \tag{4.82}$$

where

$$A_{k,2,1} = \sum_{i \in \Omega} \sqrt{p_{i,k} \alpha_{i,k}} \left(\sum_{j \in \Omega \setminus \{i\}} \sqrt{p_{j,k} \alpha_{j,k}} \right) \tag{4.83}$$

$$A_{k,2,2} = - \frac{\sum_{i \in \Omega} \frac{\tilde{T}_{i,k}}{\sqrt{p_{i,k} \alpha_{i,k}}} \left(\sum_{j \in \Omega \setminus \{i\}} \sqrt{p_{j,k} \alpha_{j,k}} \right)}{2M} \tag{4.84}$$

$$A_{k,2,3} = - \frac{\sum_{i \in \Omega} \sqrt{p_{i,k} \alpha_{i,k}} \left(\sum_{j \in \Omega \setminus \{i\}} \frac{\tilde{T}_{j,k}}{\sqrt{p_{j,k} \alpha_{j,k}}} \right)}{2M} \tag{4.85}$$

$$A_{k,2,4} = \frac{\sum_{i\in\Omega} \frac{\tilde{T}_{i,k}}{\sqrt{p_{i,k}\alpha_{i,k}}} \left(\sum_{j\in\Omega\setminus\{i\}} \frac{\tilde{T}_{j,k}}{\sqrt{p_{j,k}\alpha_{j,k}}} \right)}{4M^2}. \tag{4.86}$$

Since $A_{k,2,4} \approx 0$ for large M, it can be shown that

$$\bar{S}_k \approx \tilde{G}_k - \tilde{T}_k \tag{4.87}$$

where $\tilde{G}_k$ is the approximated TPL-free transmission gain of user k and $\tilde{T}_k$ is the approximated TPL of user k, and can be represented as, respectively,

$$\tilde{G}_k = \left[\sum_{i\in\Omega} p_{i,k}\alpha_{i,k} + \sum_{i\in\Omega} \sqrt{p_{i,k}\alpha_{i,k}} \left(\sum_{j\in\Omega\setminus\{i\}} \sqrt{p_{j,k}\alpha_{j,k}} \right) \right] M \tag{4.88}$$

$$\tilde{T}_k = \sum_{i\in\Omega} \tilde{T}_{i,k} + \sum_{i\in\Omega} \frac{\tilde{T}_{i,k}}{\sqrt{p_{i,k}\alpha_{i,k}}} \left(\sum_{j\in\Omega\setminus\{i\}} \sqrt{p_{j,k}\alpha_{j,k}} \right). \tag{4.89}$$

Based on the approximated TPL, we partition users into a number of groups in what follows. The proposed user grouping algorithm is summarized in Figure 4-5.

Step 1: For $g = 1$, we initialize the user sets by

$$U_g = \varnothing \tag{4.90}$$

$$U_{\mathrm{r}} = U \tag{4.91}$$

$$\hat{U}_g = U \tag{4.92}$$

where $\hat{U}_g$ is the set of users who can be selected to group g.

Step 2: We select a new user k^* that yields the lowest TPL to users in $\hat{U}_g$ by

$$k^* = \arg\min_{k \in \hat{U}_g} \sum_{l \in \hat{U}_g \setminus \{k\}} \tilde{T}_{l,k} \tag{4.93}$$

where $\tilde{T}_{l,k}$ is the approximated TPL of user *l* due to the interference nulling for user *k* and can be represented as

$$\tilde{T}_{l,k} = \sum_{i \in \Omega} \tilde{T}_{i,l,k} + \sum_{i \in \Omega} \frac{\tilde{T}_{i,l,k}}{\sqrt{p_{i,l}\alpha_{i,l}}} \left(\sum_{j \in \Omega \setminus \{i\}} \sqrt{p_{j,l}\alpha_{j,l}} \right). \tag{4.94}$$

After the user selection, we update user sets by

$$U_g = U_g \cup \left\{ k^* \right\} \tag{4.95}$$

$$U_{\mathrm{r}} = U_{\mathrm{r}} - \left\{ k^* \right\} \tag{4.96}$$

$$\hat{U}_g = \hat{U}_g - \left\{ k^* \right\}. \tag{4.97}$$

Step 3: To keep the TPL lower than a certain level, we update $\hat{U}_g$ for the selection of a new user by

$$\hat{U}_g = \left\{ k \,|\, \hat{U}_{g,1} \cap \hat{U}_{g,2},\ k \in \hat{U}_g \right\} \tag{4.98}$$

where $\hat{U}_{g,1}$ and $\hat{U}_{g,2}$ are defined by, respectively,

$$\hat{U}_{g,1} = \left\{ k \,|\, \tilde{T}_{g_l}\left(U_g\right) + \tilde{T}_{g_l,k} < \delta_{\mathrm{th},g_l},\ l \in U_g,\ k \in \hat{U}_g \right\} \tag{4.99}$$

$$\hat{U}_{g,2} = \left\{ k \,|\, \tilde{T}_k\left(U_g\right) < \delta_{\text{th},k},\ k \in \hat{U}_g \right\}. \tag{4.100}$$

Here, $\delta_{\text{th},k} = c\tilde{G}_k$ where $c\left(\in[0,1]\right)$ is a parameter for user k to be determined and $\tilde{T}_k\left(U_g\right)$ is the approximated TPL of user k due to the interference nulling for users in U_g. We may represent $\tilde{T}_k\left(U_g\right)$ as

$$\tilde{T}_k\left(U_g\right) = \sum_{g_l \in U_g} \tilde{T}_{k,g_l}\ . \tag{4.101}$$

By selecting a new user for group g among users in $\hat{U}_g$, we may keep the TPL of users in U_g lower than a certain level. If $\hat{U}_g = \varnothing$, we finish the user selection process for user group g and go to **Step 2** with $g \leftarrow g+1$. Otherwise, we go to **Step 4**.

Step 4: Let $\tilde{T}_g\left(k\right)$ be the total TPL of users in U_g due to the interference nulling for user k in $\hat{U}_g$. Then, we may represent $\tilde{T}_g\left(k\right)$ as

$$\tilde{T}_g\left(k\right) = \sum_{g_l \in U_g} \tilde{T}_{g_l,k}\ . \tag{4.102}$$

For user k in $\hat{U}_g$, we may represent the total TPL of group g as

$$\tilde{T}\left(U_g, k\right) = \tilde{T}_g\left(k\right) + \tilde{T}_k\left(U_g\right). \tag{4.103}$$

We select a new user k^* to minimize the total TPL for group g by

$$k^* = \arg\min_{k \in \hat{U}_g} \tilde{T}\left(U_g, k\right) \tag{4.104}$$

and update U_g, U_{r}, and $\hat{U}_g$ by (4.95), (4.96), and (4.97), respectively. If $U_{\text{r}} = \varnothing$, we finish the user selection process. Otherwise, we go to **Step 5**.

Step 5: If $K_g = M$, we go to **Step 2** for the selection of a new user in group g+1. Otherwise, we go to **Step 3**.

After the user grouping, we generate the beam weight for each user group with the use of transmission resource orthogonal to each other. We may represent the total achievable sum rate by means of proposed user grouping as [53]

$$r_{\text{sum}} = \sum_{g=1}^{G} \frac{K_g}{K} \sum_{k=1}^{K_g} \log_2 \left(1 + \text{SINR}_{g_k}\right). \tag{4.105}$$

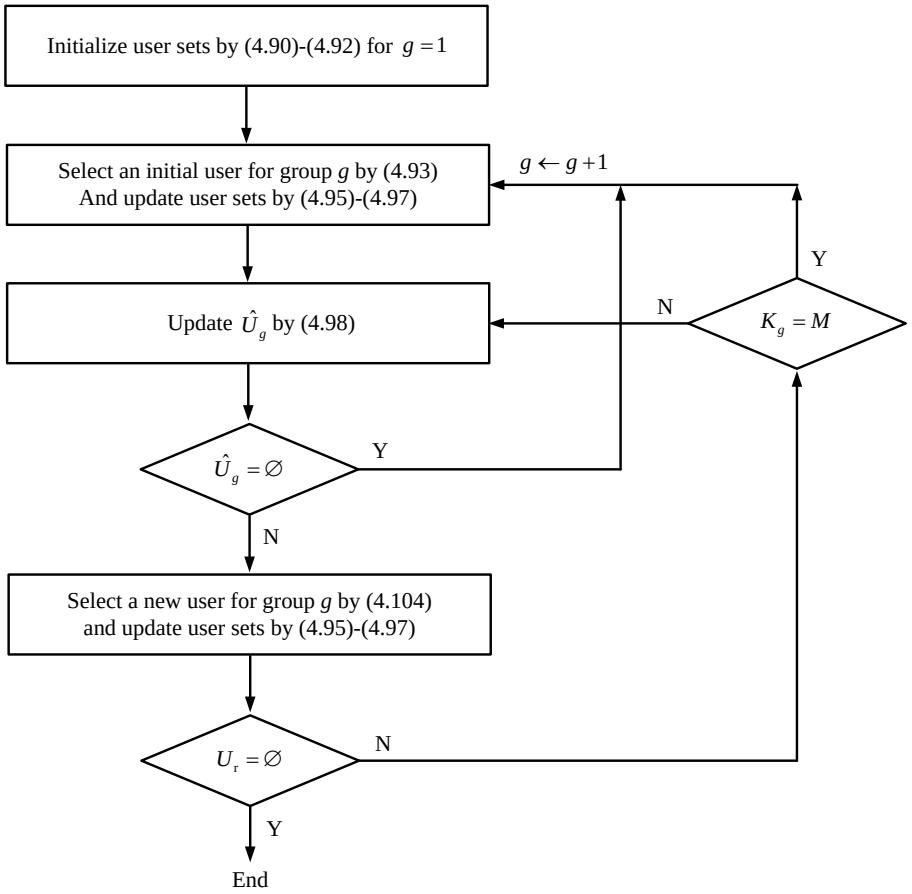


Figure 4-5. Proposed user grouping scheme.

When a large number of users are densely located in a small area, they may experience channel highly correlated to each other, suffering from large TPL [56, 57]. We may reduce the TPL by increasing the number of groups, but users may suffer from performance loss due to the use of reduced transmission resource allocated to each group. We may alleviate this problem by means of user re-grouping based on the approximated TPL.

When a group has a small number of users, it may not be possible to fully utilize the transmission resource to maximize the sum rate. Thus, when a group has users less than δM , where $0<\delta\le 1$, we may perform user re-grouping. The set of users required for the re-grouping can be represented as

$$U_{\mathrm{a}}=\left\{g_k\in U_g \mid K_g<\delta M,\ g=1,\cdots,G\right\}. \tag{4.106}$$

Let Ψ be the set of user groups having the number of users larger than a threshold. We re-allocate a user in U_{a} to another group in Ψ to minimize the total TPL by

$$\left(k^*,g^*\right)=\underset{k\in U_{\mathrm{a}},g\in\Psi}{\arg\min}\,\tilde{T}\left(U_g,k\right). \tag{4.107}$$

Then, we update the user set by $U_{\mathrm{a}}=U_{\mathrm{a}}-\left\{k^*\right\}$. We repeat this process until $U_{\mathrm{a}}=\varnothing$.

4.3.2. Effect of CSI inaccuracy

In this subsection, we analyze effect of CSI inaccuracy on the TPL. In the presence of CSI inaccuracy in spatially correlated channel environments, we may represent the channel between BS *i* and user *k* as [41]

$$\begin{aligned}\hat{\mathbf{h}}_{i,k} &= \sqrt{1-\tau_{i,k}^2}\,\mathbf{h}_{i,k} + \tau_{i,k}\mathbf{z}_{i,k} \\ &= \mathbf{R}_{i,k}^{1/2}\left(\sqrt{1-\tau_{i,k}^2}\,\tilde{\mathbf{h}}_{i,k} + \tau_{i,k}\tilde{\mathbf{z}}_{i,k}\right)\end{aligned} \tag{4.108}$$

where $\tilde{\mathbf{z}}_{i,k} \in \mathbb{C}^{M\times 1} \sim CN\left(\mathbf{0}_M, \mathbf{I}_M\right)$. For ease of analysis, we assume that $\tau_{i,k} = \tau$ for all *i* and *k*.

In the presence of CSI inaccuracy, we may represent the desired signal power of user *k* of BS *i* as

$$\begin{aligned}S_{\tau,i,k} &= p_{i,k}\alpha_{i,k}\left|\hat{\mathbf{h}}_{i,k}^H \mathbf{v}_{i,k}^{\mathrm{ZF}}\right|^2 \\ &= p_{i,k}\alpha_{i,k}\left|\sqrt{1-\tau^2}\,\mathbf{h}_{i,k}^H \mathbf{v}_{i,k}^{\mathrm{ZF}} + \tau\mathbf{z}_{i,k}^H \mathbf{v}_{i,k}^{\mathrm{ZF}}\right|^2 \\ &= \left(1-\tau^2\right)S_{i,k} + p_{i,k}\alpha_{i,k}\left[2\tau\sqrt{1-\tau^2}\,\mathrm{Re}\left(\varphi_{i,k,k}^{\mathrm{ZF}}\right) + \tau^2 Z_{i,k,k}^{\mathrm{ZF}}\right]\end{aligned} \tag{4.109}$$

where

$$\varphi_{i,k,k}^{\mathrm{ZF}} = \mathbf{h}_{i,k}^H \mathbf{v}_{i,k}^{\mathrm{ZF}}\left(\mathbf{v}_{i,k}^{\mathrm{ZF}}\right)^H \mathbf{z}_{i,k} \tag{4.110}$$

$$Z_{i,k,k}^{\mathrm{ZF}} = \left|\mathbf{z}_{i,k}^H \mathbf{v}_{i,k}^{\mathrm{ZF}}\right|^2. \tag{4.111}$$

It can be shown that

$$\begin{aligned}\varphi_{i,k,k}^{\mathrm{ZF}} &= \mathbf{h}_{i,k}^H \mathbf{v}_{i,k}^{\mathrm{ZF}}\left(\mathbf{v}_{i,k}^{\mathrm{ZF}}\right)^H \mathbf{z}_{i,k} \\ &= \tilde{\mathbf{h}}_{i,k}^H \mathbf{R}_{i,k}^{1/2} \mathbf{v}_{i,k}^{\mathrm{ZF}}\left(\mathbf{v}_{i,k}^{\mathrm{ZF}}\right)^H \mathbf{R}_{i,k}^{1/2}\tilde{\mathbf{z}}_{i,k}.\end{aligned} \tag{4.112}$$

Since $\tilde{\mathbf{h}}_{i,k}$ and $\tilde{\mathbf{z}}_{i,k}$ are independent of each other, it can be shown from Lemma 3-2 that

$$\varphi_{i,k,k}^{\mathrm{ZF}} \xrightarrow[M\to\infty]{\text{a.s.}} 0 . \tag{4.113}$$

Then, it can be shown that

$$\begin{aligned} Z_{i,k,k}^{\mathrm{ZF}} &= \left|\mathbf{z}_{i,k}^{H}\mathbf{v}_{i,k}^{\mathrm{ZF}}\right|^2 \\ &= \tilde{\mathbf{z}}_{i,k}^{H}\mathbf{R}_{i,k}^{1/2}\mathbf{v}_{i,k}^{\mathrm{ZF}}\left(\mathbf{v}_{i,k}^{\mathrm{ZF}}\right)^{H}\mathbf{R}_{i,k}^{1/2}\tilde{\mathbf{z}}_{i,k}. \end{aligned} \tag{4.114}$$

It can be shown from Lemma 3-1 that

$$\tilde{\mathbf{z}}_{i,k}^{H}\mathbf{R}_{i,k}^{1/2}\mathbf{v}_{i,k}^{\mathrm{ZF}}\left(\mathbf{v}_{i,k}^{\mathrm{ZF}}\right)^{H}\mathbf{R}_{i,k}^{1/2}\tilde{\mathbf{z}}_{i,k} - \mathrm{Tr}\left(\mathbf{R}_{i,k}^{1/2}\mathbf{v}_{i,k}^{\mathrm{ZF}}\left(\mathbf{v}_{i,k}^{\mathrm{ZF}}\right)^{H}\mathbf{R}_{i,k}^{1/2}\right)\xrightarrow[M\to\infty]{\text{a.s.}} 0 . \tag{4.115}$$

It can also be shown that

$$\begin{aligned} \mathrm{Tr}\left(\mathbf{R}_{i,k}^{1/2}\mathbf{v}_{i,k}^{\mathrm{ZF}}\left(\mathbf{v}_{i,k}^{\mathrm{ZF}}\right)^{H}\mathbf{R}_{i,k}^{1/2}\right) &= \left(\mathbf{v}_{i,k}^{\mathrm{ZF}}\right)^{H}\mathbf{R}_{i,k}\mathbf{v}_{i,k}^{\mathrm{ZF}} \\ &= \frac{\tilde{\mathbf{h}}_{i,k}^{H}\mathbf{R}_{i,k}^{1/2}\mathbf{P}_{i,k}\mathbf{R}_{i,k}\mathbf{P}_{i,k}\mathbf{R}_{i,k}^{1/2}\tilde{\mathbf{h}}_{i,k}}{\left\|\mathbf{P}_{i,k}\mathbf{h}_{i,k}\right\|^2}. \end{aligned} \tag{4.116}$$

It can be shown from Lemma 3-1 that

$$\tilde{\mathbf{h}}_{i,k}^{H}\mathbf{R}_{i,k}^{1/2}\mathbf{P}_{i,k}\mathbf{R}_{i,k}\mathbf{P}_{i,k}\mathbf{R}_{i,k}^{1/2}\tilde{\mathbf{h}}_{i,k} - \mathrm{Tr}\left(\mathbf{R}_{i,k}^{1/2}\mathbf{P}_{i,k}\mathbf{R}_{i,k}\mathbf{P}_{i,k}\mathbf{R}_{i,k}^{1/2}\right)\xrightarrow[M\to\infty]{\text{a.s.}} 0 . \tag{4.117}$$

Then, it can be shown that

$$Z_{i,k,k}^{\mathrm{ZF}} - \bar{Z}_{i,k,k}^{\mathrm{ZF}} \xrightarrow[M\to\infty]{\text{a.s.}} 0 \tag{4.118}$$

where $\bar{Z}_{i,k,k}^{\mathrm{ZF}}$ is a deterministic equivalent of $Z_{i,k,k}^{\mathrm{ZF}}$ and can be approximated as

$$\bar{Z}_{i,k,k}^{\mathrm{ZF}} \approx \frac{p_{i,k}\alpha_{i,k}\sum_{m=1}^{b_{i,k}}\lambda_{i,k,m}^{2}}{p_{i,k}\alpha_{i,k}M-\tilde{T}_{i,k}}. \tag{4.119}$$

It can also be shown that

$$S_{\tau,i,k}-\bar{S}_{\tau,i,k}\xrightarrow[M\to\infty]{\text{a.s.}}0 \tag{4.120}$$

where $\bar{S}_{\tau,i,k}$ is a deterministic equivalent of $S_{\tau,i,k}$ and can be approximated as

$$\bar{S}_{\tau,i,k}\approx\tilde{G}_{\tau,i,k}-\tilde{T}_{\tau,i,k}. \tag{4.121}$$

Here,

$$\tilde{G}_{\tau,i,k}=p_{i,k}\alpha_{i,k}M+\tau^{2}\tilde{T}_{i,k}+\tau^{2}\left[\frac{\left(p_{i,k}\alpha_{i,k}\right)^{2}\sum_{m=1}^{b_{i,k}}\lambda_{i,k,m}^{2}}{p_{i,k}\alpha_{i,k}M-\tilde{T}_{i,k}}\right] \tag{4.122}$$

$$\tilde{T}_{\tau,i,k}=\tilde{T}_{i,k}+\tau^{2}p_{i,k}\alpha_{i,k}M. \tag{4.123}$$

Finally, it can be shown that

$$\tilde{T}_{\tau,k}=\tilde{T}_{k}+e_{\tau,k}^{\mathrm{TPL}} \tag{4.124}$$

where $e_{\tau,k}^{\mathrm{TPL}}$ is the TPL of user k in the presence of CSI inaccuracy and can be represented as

$$e_{\tau,k}^{\mathrm{TPL}}=\tau^{2}\left[\sum_{i\in\Omega}p_{i,k}\alpha_{i,k}+\sum_{i\in\Omega}\sqrt{p_{i,k}\alpha_{i,k}}\left(\sum_{j\in\Omega\setminus\{i\}}\sqrt{p_{j,k}\alpha_{j,k}}\right)\right]M. \tag{4.125}$$

It can be seen that when τ is small, the proposed scheme may be effective for the reduction of TPL since $\tilde{T}_k$ is not affected by τ. However, when τ is large, it may not be effective mainly due to the increase of $e_{\tau,k}^{\mathrm{TPL}}$.

4.3.3. Computational complexity

We measure the computational complexity in terms of flop [43, 44]. It can be shown that the estimation of TPL requires $5BK_{\mathrm{total}}M\log_2 M$ flops to approximate the SCM by means of FFT [56] and $\left(BK_{\mathrm{total}}\left(K_{\mathrm{total}}-1\right)M\right)/2$ flops to calculate $\tilde{T}_{i,k,l}$ for $\forall i,\ k,\ l\neq k$. The generation of ZF beam weights for group g requires a computational complexity of $\psi_{K_g}^{\mathrm{ZF}}=BK_g\left[M^2+\left(8K_g+4\right)M+\left(4/3\right)K_g^2\right]$ flops. Finally, the proposed scheme requires a computational complexity of $5BK_{\mathrm{total}}M\log_2 M+BK_{\mathrm{total}}\left(K_{\mathrm{total}}-1\right)M/2+\sum_{g=1}^{G}\psi_{K_g}^{\mathrm{ZF}}$ flops. The computational complexity of various grouping schemes is summarized in Table 4-1.

Table 4-1. Computational complexity of grouping schemes.

Scheme	Complexity
Greedy	$BK_{\mathrm{total}}\left(K_{\mathrm{total}}+1\right)M+\sum_{g=1}^{G-1}\sum_{n=1}^{K_g}\psi_{n+1}^{\mathrm{ZF}}+\sum_{n=1}^{K_G-1}\psi_{n+1}^{\mathrm{ZF}}$
Chordal distance	$BK_{\mathrm{total}}\left(K_{\mathrm{total}}+1\right)M+\sum_{g=1}^{G}\psi_{K_g}^{\mathrm{ZF}}$
Random	$\sum_{g=1}^{G}\psi_{K_g}^{\mathrm{ZF}}$
Proposed	$5BK_{\mathrm{total}}M\log_2 M+\frac{BK_{\mathrm{total}}\left(K_{\mathrm{total}}-1\right)M}{2}+\sum_{g=1}^{G}\psi_{K_g}^{\mathrm{ZF}}$

4.4. Performance evaluation

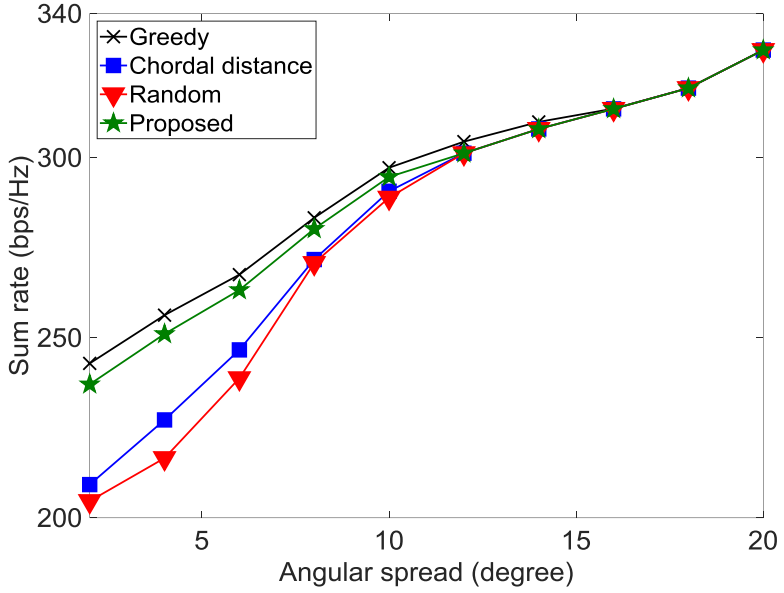
We evaluate the performance of the proposed scheme by computer simulation. We consider a three-cell cellular system as illustrated in Figure 3-7, where K users in each cell are randomly distributed in cell boundary region (*i.e.*, $K_i = K, \forall i$). The other simulation parameters are summarized in Table 4-2 [68, 69], [73]. For comparison, we consider the performance of greedy grouping scheme [52], chordal distance-based grouping scheme [60], and random grouping scheme. The greedy grouping scheme partitions users into a number of groups by sequentially selecting users to maximize the sum rate in consideration of beam weight [52]. The chordal distance-based grouping scheme partitions users into a number of groups to maximize the chordal distance in each group by employing the AHC method [61]. Random grouping scheme randomly selects users with consideration of the number of users in each group determined by the proposed scheme. For fair comparison, the design parameters of each scheme are optimally determined.

Table 4-2. Simulation parameters.

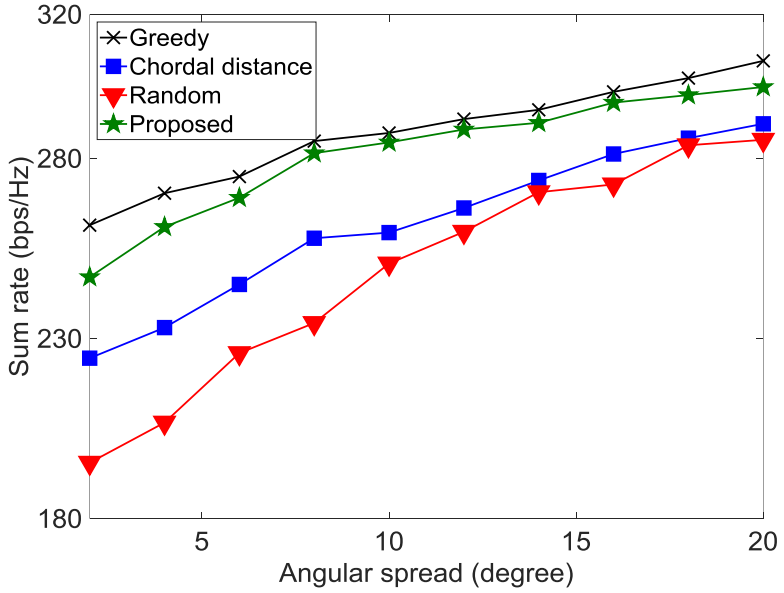
Simulation parameters	**Value**
Number of BSs B	3
Number of users in each cell K	16, 20, 24, 30
User distribution	Randomly distributed
Cell radius $R_{\max}$	300 m

Cell center radius R_c	200 m
Normalized antenna spacing D	0.5
AoD $\Delta_{i,k}$, $\forall i$ and k	$\left[-60^\circ, 60^\circ\right]$
Path loss $\alpha_{i,k}$, $\forall i$ and k	$148.1+37.6\log_{10}\left(d_{i,k}\right)$ dB, where $d_{i,k}$ is the distance between BS i and user k in km
Edge SNR	0 dB
Noise power σ_n^2	-92 dBm

Figure 4-6 depicts the sum rate according to the AS when $K=16,\ 24$. It can be seen that the proposed scheme provides almost the same performance as the greedy grouping scheme while outperforming the other schemes. This is mainly because the proposed scheme performs the user grouping based on approximated TPL. It can also be seen that the chordal distance-based grouping scheme may not provide noticeable performance improvement compared to the random grouping scheme since the chordal distance may not properly characterize the TPL. It can be seen from Figure 4-6 (a) that as the AS increases, the grouping gain becomes marginal. This is mainly because the channel correlation decreases as the AS increases [49, 50]. On the other hand, it can be seen from Figure 4-6 (b) that the grouping is effective even when the AS is large. This is mainly because when the number of users is large, users may suffer from the TPL due to the presence of high channel correlation. In this case, it may be desirable to perform the user grouping so that users in each group experience low TPL.



(a) $K = 16$.



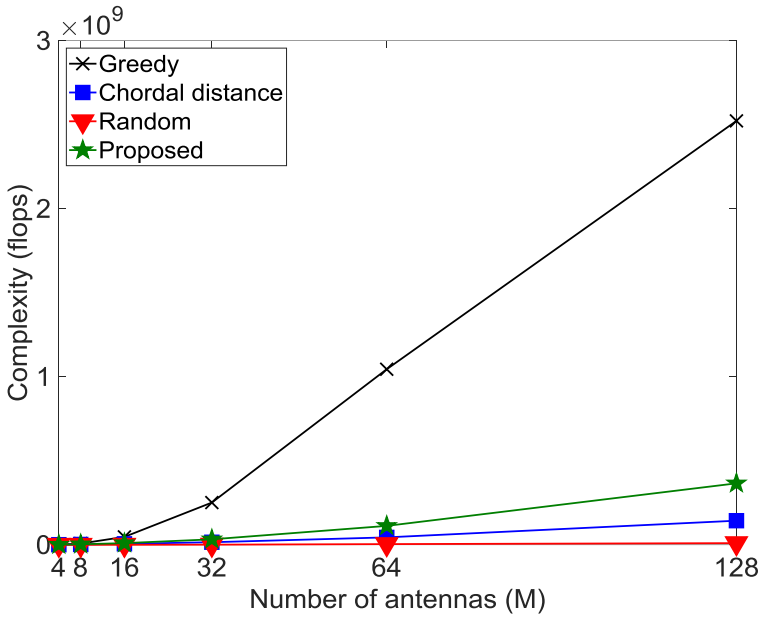
(b) $K = 24$.

Figure 4-6. Sum rate according to AS.

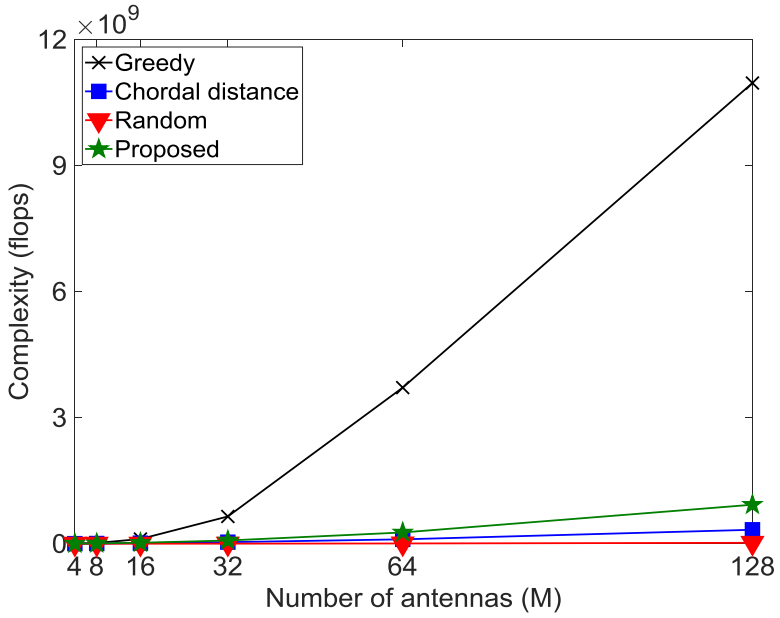
Figure 4-7 depicts the computational complexity according to the number of antennas M when $K = 16,\ 24$, and the AS is randomly distributed in 5° to 15°. It can be seen that the proposed scheme requires computational complexity much lower than the greedy grouping scheme since it performs the user grouping without consideration of beam weight. The proposed scheme outperforms the chordal distance-based grouping scheme with complexity slightly higher than that.

Figure 4-8 depicts the number of groups and the sum rate when users are located in a small cluster of a radius of $R_{\mathrm{C}} = 20\ \mathrm{m}$, $K = 30$, and the AS is randomly distributed in 5° to 15°. We assume that the cluster is randomly located in each cell and n percent of users in each cell are in the cluster, and the threshold for the re-grouping is optimally determined. It can be seen from Figure 4-8 (a) that as n increases, the proposed scheme increases the number of groups to reduce the TPL, while reducing the amount of transmission resource to each group. It can be seen from Figure 4-8 (b) that as n increases, the proposed scheme suffers from performance loss, showing the effectiveness of user re-grouping.

Figure 4-9 depicts the sum rate according to the channel inaccuracy τ when $K = 20$ and the AS is randomly distributed in 5° to 15°. It can be seen that the proposed scheme works better than the greedy grouping scheme in the presence of CSI inaccuracy. The greedy grouping scheme may not accurately calculate the transmission rate for the user grouping in the presence of CSI inaccuracy. The proposed scheme estimates the TPL based on spatial channel correlation, which is robust to the presence of CSI inaccuracy.

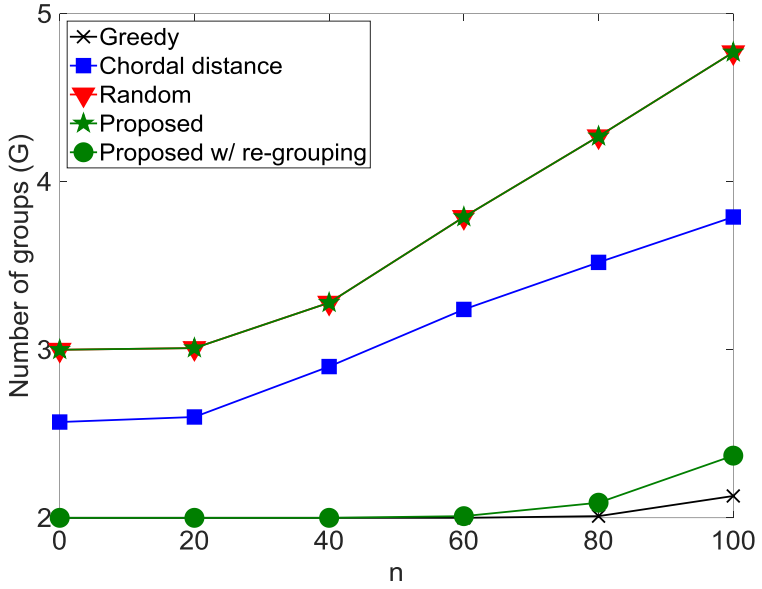


(a) $K = 16$

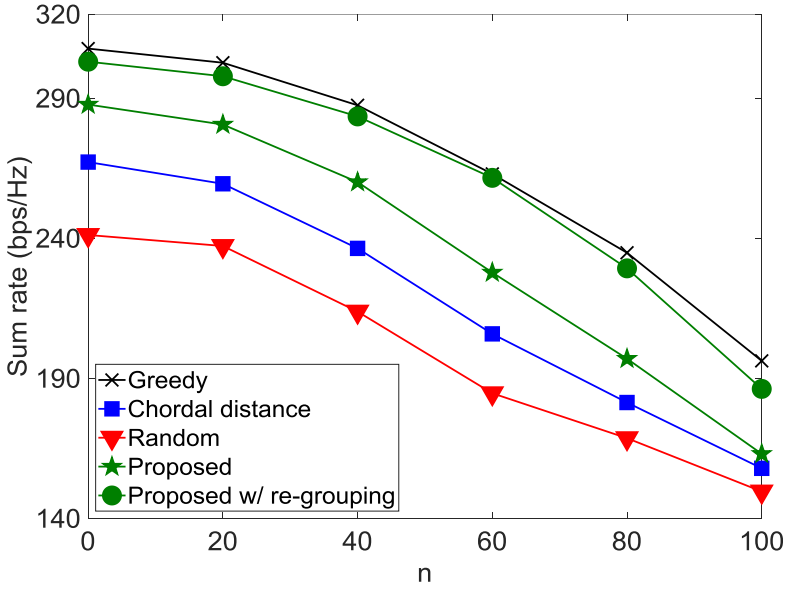


(b) $K = 24$.

Figure 4-7. Computational complexity according to *M*.



(a) Number of groups.



(b) Sum rate.

Figure 4-8. Performance according to *n*.

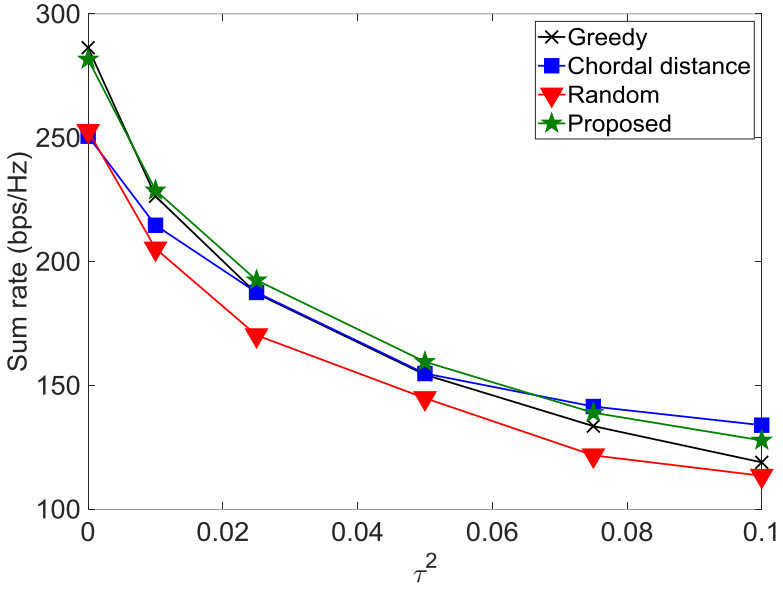


Figure 4-9. Sum rate according to τ.

Chapter 5

Conclusions and further research issues

In this dissertation, we have considered complexity-reduced multi-user signal transmission in m-MIMO environments. We have considered the use of MRT and PZF for signal transmission. We have used the SLNR instead of SINR for the generation of beam weight to reduce the computational complexity. For further reduction of computational complexity, we have approximated the SLNR. The simulation results show the effectiveness of the proposed scheme.

We have also considered complexity-reduced ZF beamforming by means of user-grouping so that users in each group experience low TPL. We have further reduced the complexity for the user grouping by approximating the TPL. The simulation results show the effectiveness of the proposed scheme especially in the presence of high channel correlation, with significantly reduced computational complexity.

There are some interesting issues that may require further research, including

- **Complexity-reduced multicast transmission**: Multicast transmission is a new technology that can support increasing wireless traffic demand [74]. It may be desirable to mitigate inter-group interference with affordable implementation complexity in multicast transmission environments. We may optimally generate multicast beam weight, which is an NP-hard problem [75]. We may sub-optimally determine the beam weight by means of semidefinite relaxation (SDR) [76] or successive convex approximation (SCA) [77]. However, these schemes may still require huge computational complexity, which may not be affordable in m-MIMO operating environments. The ZF beamforming can be applied to multicast transmission in m-MIMO environments [78], but it may require huge computational complexity. Thus, it may be of great concern to develop a complexity-reduced multicast transmission scheme.

- **Scalable cell-free m-MIMO:** It is of great interest to improve transmission performance in cell boundary region. We may consider the use of cell-free m-MIMO, where a central processing unit controls a massive number of access points (APs) [79]. However, it is of great concern how to realize a scalable cell-free m-MIMO scheme in real operating environments [80].

- **Spatial correlation-based V2I transmission:** Intelligent transport systems may require signal transmission of high data rate and low latency [81]. We may apply m-MIMO techniques to vehicle-to-infrastructure (V2I) communications [82]. It is of great concern to provide desired requirements in the presence of high mobility [83]. We may exploit spatial channel correlation to alleviate the

channel aging problem [84, 85]. However, it may require very high implementation complexity in V2I transmission environments.

Appendix

A. Proof of Lemma 3-4

Let user *k* be the *u*-th user in U_i^{o}. G_k^{PZF} can be represented as [10]

$$G_k^{\mathrm{PZF}} = G_k^{\mathrm{MRT}} - D_k^{\mathrm{PZF}} \tag{A.1}$$

where

$$D_k^{\mathrm{PZF}} = \sum_{n \neq u}^{K_i^{\mathrm{o}}} D_{u,n}^{\mathrm{PZF}} . \tag{A.2}$$

Here, $D_{u,n}^{\mathrm{PZF}}$ can be represented as

$$D_{u,n}^{\mathrm{PZF}} = (-1)^{u+n+1} \mathbf{h}_{i,k}^H \mathbf{h}_{i,n_i^{\mathrm{o}}} \Theta_{u,n} \tag{A.3}$$

where

$$\Theta_{u,n} = \frac{\det\left(\mathbf{R}^{\langle u,n \rangle}\right)}{\det\left(\mathbf{R}^{\langle u,u \rangle}\right)} . \tag{A.4}$$

Here, $\mathbf{R} = \left(\mathbf{H}_i^{\mathrm{o}}\right)^H \mathbf{H}_i^{\mathrm{o}}$ and $\mathbf{R}^{\langle u,n \rangle}$ denotes $\mathbf{R}$ without the *u*-th row and the *n*-th column. It can be shown that [70]

$$\Theta_{u,n} = \frac{\sum_{p=1}^{K_{i,k}^{o}} (-1)^{p+l_{u,n}} \mathbf{h}_{i,b_i^o}^H \mathbf{h}_{i,k} \det\left(\left[\mathbf{R}^{\langle u,n \rangle} \right]^{\langle p,l_{u,n} \rangle} \right)}{\sum_{p=1}^{K_{i,k}^{o}} (-1)^{p+l_{n,u}} \mathbf{h}_{i,b_i^o}^H \mathbf{h}_{i,n_i^o} \det\left(\left[\mathbf{R}^{\langle u,u \rangle} \right]^{\langle p,l_{n,u} \rangle} \right)} = \frac{(-1)^{z_{u,n}} \sum_{p=1}^{K_{i,k}^{o}} w_{u,n}^{(p)} \mathbf{h}_{i,b_i^o}^H \mathbf{h}_{i,k}}{\sum_{p=1}^{K^{o}} w_{u,n}^{(p)} \mathbf{h}_{i,b_i^o}^H \mathbf{h}_{i,n_i^o}} \tag{A.5}$$

where

$$l_{u,n} = \begin{cases} u & \text{for } u < n \\ u-1 & \text{for } u > n \end{cases} \tag{A.6}$$

$$z_{u,n} = \begin{cases} u-n+1 & \text{for } u < n \\ u-n-1 & \text{for } u > n \end{cases} \tag{A.7}$$

$$b = \begin{cases} p & \text{for } p < u \\ p+1 & \text{for } p > u \end{cases} \tag{A.8}$$

$$w_{u,n}^{(p)} = (-1)^p \det\left(\left[\mathbf{R}^{\langle u,u \rangle} \right]^{\langle p,l_{n,u} \rangle} \right). \tag{A.9}$$

Since $(-1)^{u+n+1}(-1)^{z_{u,n}} = 1$ irrespective of u and n, it can also be shown that

$$D_{u,n}^{\text{PZF}} = \frac{\sum_{p=1}^{K_{i,k}^{o}} f_{u,n}^{(p)}}{\sum_{p=1}^{K_{i,k}^{o}} g_{u,n}^{(p)}} = \frac{f_{u,n}^{(l_{n,u})} + \sum_{p \neq l_{n,u}}^{K_{i,k}^{o}} f_{u,n}^{(p)}}{g_{u,n}^{(l_{n,u})} + \sum_{p \neq l_{n,u}}^{K_{i,k}^{o}} g_{u,n}^{(p)}} \tag{A.10}$$

where

$$f_{u,n}^{(p)} = w_{u,n}^{(p)} \mathbf{h}_{i,b_i^o}^H \mathbf{h}_{i,k} \mathbf{h}_{i,k}^H \mathbf{h}_{i,n_i^o} \tag{A.11}$$

$$g_{u,n}^{(p)} = w_{u,n}^{(p)} \mathbf{h}_{i,b_i^o}^H \mathbf{h}_{i,n_i^o} . \tag{A.12}$$

For $b \neq n$, it can be seen from Lemma 3-2 that

$$\mathbf{h}_{i,b_i^o}^H \mathbf{h}_{i,k} \mathbf{h}_{i,k}^H \mathbf{h}_{i,n_i^o} \xrightarrow[M\to\infty]{\text{a.s.}} 0 \tag{A.13}$$

$$\mathbf{h}_{i,b_i^o}^H \mathbf{h}_{i,n_i^o} \xrightarrow[M\to\infty]{\text{a.s.}} 0 . \tag{A.14}$$

Then, it can be shown that

$$\sum_{p\neq l_{n,u}}^{K_{i,k}^o} f_{u,n}^{(p)} \xrightarrow[M\to\infty]{\text{a.s.}} 0 \tag{A.15}$$

$$\sum_{p\neq l_{n,u}}^{K_{i,k}^o} g_{u,n}^{(p)} \xrightarrow[M\to\infty]{\text{a.s.}} 0 . \tag{A.16}$$

For $b = n$, it can be shown from Lemma 3-1 that

$$\mathbf{h}_{i,k}^H \mathbf{h}_{i,n_i^o} \mathbf{h}_{i,n_i^o}^H \mathbf{h}_{i,k} - \mathrm{Tr}\left(\mathbf{h}_{i,n_i^o} \mathbf{h}_{i,n_i^o}^H\right) \xrightarrow[M\to\infty]{\text{a.s.}} 0 . \tag{A.17}$$

Thus, it can be shown that

$$\frac{f_{u,n_i^{\mathrm{o}}}^{(l_{n,u})}}{g_{u,n_i^{\mathrm{o}}}^{(l_{n,u})}} - 1 \xrightarrow[M\to\infty]{\text{a.s.}} 0 . \tag{A.18}$$

It can be shown that

$$D_k^{\mathrm{PZF}} - K_{i,k}^{\mathrm{o}} \xrightarrow[M\to\infty]{\text{a.s.}} 0 . \tag{A.19}$$

Finally, it can be shown that

$$G_k^{\mathrm{PZF}} - \bar{G}_k^{\mathrm{PZF}} \xrightarrow[M\to\infty]{\text{a.s.}} 0 \tag{A.20}$$

where $\bar{G}_k^{\mathrm{PZF}} = M - K_{i,k}^{\mathrm{o}}$.

B. Proof of Lemma 3-5

It can be shown from Lemma 3-1 that

$$\mathbf{h}_{i,k}^H \mathbf{P}_{i,k}^{\mathrm{o}} \mathbf{h}_{i,k} - \mathrm{Tr}\left(\mathbf{P}_{i,k}^{\mathrm{o}}\right) \xrightarrow[M\to\infty]{\text{a.s.}} 0 . \tag{B.1}$$

It can be seen from Lemma 3-3 and 3-4 that

$$q_k - \bar{q}_k \xrightarrow[M\to\infty]{\text{a.s.}} 0 \tag{B.2}$$

where $\bar{q}_k$ is a deterministic equivalent of q_k and can be represented as

$$\bar{q}_k = \frac{\mathrm{Tr}\left(\mathbf{P}_{i,k}^{\mathrm{o}}\right)}{\sqrt{M}\sqrt{M - K_{i,k}^{\mathrm{o}}}} . \tag{B.3}$$

Since $\mathrm{Tr}\left(\mathbf{P}_{i,k}^{\mathrm{o}}\right) = M - K_{i,k}^{\mathrm{o}}$, it can be shown that

$$\bar{q}_k = \frac{\mathrm{Tr}\left(\mathbf{P}_{i,k}^{\mathrm{o}}\right)}{\sqrt{M}\sqrt{M - K_{i,k}^{\mathrm{o}}}} = \frac{M - K_{i,k}^{\mathrm{o}}}{M\sqrt{1 - \frac{K_{i,k}^{\mathrm{o}}}{M}}}. \tag{B.4}$$

Thus, it can be seen that $\lim_{M\to\infty} \bar{q}_k \approx 1$.

C. Proof of strict quasi-concavity of $\overline{\mathrm{SLNR}}_k$

We prove that $\overline{\mathrm{SLNR}}_k$ is a strict quasi-concave function of θ_k by using the following Theorem [42].

Theorem. A continuous function $f : \mathbf{R} \to \mathbf{R}$ is strictly quasi-concave if and only if there is a point $c \in \mathrm{dom}\, f$ such that *f* is an increasing function for $x < c$ ($x \in \mathrm{dom}\, f$), and *f* is a decreasing function for $x > c$ ($x \in \mathrm{dom}\, f$). The point *c* can be chosen by any point that globally maximizes *f*.

Let c_k be a global maximizer of $\overline{\mathrm{SLNR}}_k$. Without loss of generality, we can assume that c_k can be set to

$$c_k = \frac{\sigma_n^2}{\bar{L}_{U_k^{\mathrm{o}}}^{\mathrm{MRT}}} \sqrt{\frac{\Delta\bar{G}_k}{\bar{G}_k^{\mathrm{PZF}}}} \tag{C.1}$$

since the global optimum occurs at which the gradient vanishes [42]. For $\theta_k < c_k$, θ_k can be set to $\theta_k^* - \varepsilon$, where ε is an arbitrary positive real number satisfying $\theta_k^* - \varepsilon \in [0,1]$. For $\theta_k < c_k$, Φ_k in (3.76) and Γ_k in (3.77) can be represented as, respectively,

$$\Phi_k = \frac{2p_{i,k}\alpha_{i,k}}{\left(\bar{L}_{U_{i,k}^o}^{\mathrm{MRT}}\theta_k^2 + \sigma_n^2\right)^2}\left[\sqrt{\bar{G}_k^{\mathrm{PZF}}}\left(\frac{\sigma_n^2\Delta\bar{G}_k}{\bar{L}_{U_{i,k}^o}^{\mathrm{MRT}}\bar{G}_k^{\mathrm{PZF}}}+1\right)-\sqrt{\Delta\bar{G}_k}\varepsilon\right] \tag{C.2}$$

$$\Gamma_k = \bar{L}_{U_{i,k}^o}^{\mathrm{MRT}}\sqrt{\bar{G}_k^{\mathrm{PZF}}}\theta_k\varepsilon\,. \tag{C.3}$$

It can be shown that

$$\nabla_{\theta_k}\overline{\mathrm{SLNR}}_k = \varepsilon\frac{N_k}{D_k} \tag{C.4}$$

where

$$N_k = 2p_{i,k}\alpha_{i,k}\bar{L}_{U_{i,k}^o}^{\mathrm{MRT}}\sqrt{\bar{G}_k^{\mathrm{PZF}}}\left(\sqrt{\Delta\bar{G}_k}\theta_k+\sqrt{\bar{G}_k^{\mathrm{PZF}}}\right) \tag{C.5}$$

$$D_k = \left(\bar{L}_{U_{i,k}^o}^{\mathrm{MRT}}\theta_k^2+\sigma_n^2\right)^2. \tag{C.6}$$

It can be seen that $N_k > 0$ since $\theta_k \geq 0$ and $D_k > 0$. Then, it can be shown for $\theta_k < c_k$ that $\overline{\mathrm{SLNR}}_k$ is a strictly increasing function since $\nabla_{\theta_k}\overline{\mathrm{SLNR}}_k > 0$.

For $\theta_k > c_k$, θ_k can be set to $\theta_k^* + \varepsilon$, where ε is an arbitrary positive real number satisfying $\theta_k^* + \varepsilon \in [0,1]$. It can be shown for $\theta_k > c_k$ that

$$\nabla_{\theta_k}\overline{\mathrm{SLNR}}_k = -\varepsilon\frac{N_k}{D_k}\,. \tag{C.7}$$

It can be seen for $\theta_k > c_k$ that $\overline{\mathrm{SLNR}}_k$ is a strictly decreasing function since $\nabla_{\theta_k}\overline{\mathrm{SLNR}}_k < 0$. Finally, it can be shown from Theorem that $\overline{\mathrm{SLNR}}_k$ is a strict quasi-concave function of θ_k.

References

[1] Z. Zhang *et al*., "6G wireless networks: Vision, requirements, architecture, and key technologies," *IEEE Veh. Technol. Mag*., vol. 14, no. 3, pp. 28–41, Sept. 2019.

[2] S. Chen, Y.-C. Liang, S. Sun, S. Kang, W. Cheng, and M. Peng, "Vision, requirements, and technology trend of 6G: How to tackle the challenges of system coverage, capacity, user data-rate and movement speed," *IEEE Wireless Commun*., vol. 27, no. 2, pp. 218–228, Apr. 2020.

[3] P. Yang, Y. Xiao, M. Xiao, and S. Li, "6G wireless communications: Vision and potential techniques," *IEEE Netw*., vol. 33, no. 4, pp. 70–75, July 2019.

[4] K. Maruta and F. Falcone, "Massive MIMO systems: Present and future," *Electronics*, vol. 9, no. 385, pp. 1–4, Feb. 2020.

[5] H. Viswanathan and P. Mogensen "Communications in the 6G era," *IEEE Access*, vol. 8, pp. 57063–57074, Mar. 2020.

[6] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun*., vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[7] F. Rusek *et al*., "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag*., vol. 30, no. 1, pp. 40–60, Jan. 2013.

[8] L. Lu *et al*., "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process*., vol. 8, no. 5, pp. 742–758, Oct. 2014.

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

[10] K. Wong and Z. Pan, "Array gain and diversity order of multiuser MISO antenna systems," *Int. J. Wireless Inf. Net*., vol. 15, no. 2, pp 82–89, June 2008.

[11] S.-H. Moon, C. Lee, S.-R. Lee, and I. Lee, "Joint user scheduling and adaptive intercell interference cancelation for MISO downlink cellular systems," *IEEE Trans. Veh. Technol.*, vol. 62, no. 1, pp. 172–181, Jan. 2013.

[12] K. Hosseini, W. Yu, and R. S. Adve, "Large-scale MIMO versus network MIMO for multicell interference mitigation," *IEEE J. Sel. Topics Signal Process*., vol. 8, no. 5, pp. 930–941, Oct. 2014.

[13] G. Geraci, A. Garcia-Rodriguez, D. López-Pérez, A. Bonfante, L. G. Giordano, and H. Claussen, "Operating massive MIMO in unlicensed bands for enhanced coexistence and spatial reuse," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1282–1293, June 2017.

[14] H. Prabhu, J. Rodrigues, O. Edfors, and F. Rusek, "Approximative matrix inverse computations for very-large MIMO and applications to linear pre-coding systems," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, pp. 2710–2715, Apr. 2013.

[15] Q. Deng, X. Liang, X. Wang, M. Huang, C. Dong, and Y. Zhang, "Fast converging iterative precoding for massive MIMO systems: An accelerated weighted Neumann series-steepest descent approach," *IEEE Access*, vol. 8, pp. 50244–50255, Mar. 2020.

[16] M. Wu, B. Yin, A. Vosoughi, C. Studer, J. R. Cavallaro, and C. Dick, "Approximate matrix inversion for high-throughput data detection in the large-scale MIMO uplink," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, pp. 2155–2158, May 2013.

[17] K. K. C. Lee and C. E. Chen, "An eigen-based approach for enhancing matrix inversion approximation in massive MIMO systems," *IEEE Trans. Veh. Tech.*, vol. 66, no. 6, pp. 5483–5487, June 2017.

[18] Y. Man, Z. Li, F. Yan, S. Xing, and L. Shen, "Massive MIMO pre-coding algorithm based on truncated Kapteyn series expansion," in *Proc. IEEE Int. Conf. Commun. Syst. (ICCS)*, pp. 1–5, Dec. 2016.

[19] A. Kammoun, A. Müller, E. Björnson, and M. Debbah, "Linear precoding based on polynomial expansion: Large-scale multi-cell MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 861–875, Oct. 2014.

[20] T. Cheng, Y. He, L. Shi, Y. Wu, Y. Huang, and Y. Sui, "A novel adaptive hybrid truncation precoding strategy in massive MIMO," *IEEE Commun. Lett*., vol. 22, no. 11, pp. 2298–2301, Nov. 2018.

[21] A. Benzin, G. Caire, Y. Shadmi, and A. M. Tulino, "Low-complexity truncated polynomial expansion DL precoders and UL receivers for massive MIMO in correlated channels," *IEEE Trans. Wireless Commun*., vol. 18, no. 2, pp. 1069–1084, Feb. 2019.

[22] X. Gao, L. Dai, J. Zhang, and S. Han, and C.-L. I, "Capacity-approaching linear precoding with low-complexity for large-scale MIMO systems," in *Proc. IEEE Int. Commun. Conf. (ICC)*, pp. 1577–1582, June 2015.

[23] J. Minango and C. de Almeida, “A low-complexity linear precoding algorithm based on Jacobi method for massive MIMO systems,” in *Proc. IEEE Veh. Tech. Conf. (VTC-Spring)*, pp. 1–5, June 2018.

[24] X. Qin, Z. Yan, and G. He, “A near-optimal detection scheme based on joint steepest descent and Jacobi method for uplink massive MIMO systems,” *IEEE Commun. Lett*., vol. 20, no. 2, pp. 276–279, Feb. 2016.

[25] C. Tang, C. Liu, L. Yuan, and Z. Xing, “High precision low complexity matrix inversion based on Newton iteration for data detection in the massive MIMO,” *IEEE Commun. Lett*., vol. 20, no. 3, pp. 490–493, Mar. 2016.

[26] J. Minango and C. de Almeida, “Low complexity zero forcing detector based on Newton-Schultz iterative algorithm for massive MIMO systems,” *IEEE Trans. Veh. Tech.*, vol. 67, no. 12, pp. 11759–11766, Dec. 2018.

[27] T. Xie, L. Dai, X. Gao, X. Dai, and Y. Zhao, “Low-complexity SSOR-based precoding for massive MIMO systems,” *IEEE Commun. Lett*., vol. 20, no. 4, pp. 744–747, Apr. 2016.

[28] Y. Liu, J. Liu, Q. Wu, Y. Zhang, and M. Jin, “A near-optimal iterative linear precoding with low complexity for massive MIMO systems,” *IEEE Commun. Lett*., vol. 23, no. 6, pp. 1105–1108, June 2019.

[29] C. Lee, S.-H. Moon, S.-R. Lee, and I. Lee “Adaptive beamforming selection methods for inter-cell interference cancellation in multicell multiuser systems,” in *Proc. IEEE Int. Commun. Conf. (ICC)*, pp. 5684–5688, June 2013.

[30] H. H. Yang, G. Geraci, T. Q. S. Quek, and J. G. Andrews, “Cell-edge-aware precoding for downlink massive MIMO cellular networks,” *IEEE Trans. Signal Process*., vol. 65, no. 13, pp. 3344–3358, July 2017.

[31] L. Venturino, N. Prasad, and X. Wang, “Coordinated linear beamforming in downlink multi-cell wireless networks,” *IEEE Trans. Wireless Commun*., vol. 9, no. 4, pp. 1451–1461, Apr. 2010.

[32] Q. Shi, M. Razaviyayn, Z. Luo, and C. He, “An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel,” *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, Sept. 2011.

[33] J. Escudero Garzás, M. Hong, A. Garcia, and A. García-Armada, “Interference pricing mechanism for downlink multicell coordinated beamforming,” *IEEE Trans. Commun*., vol. 62, no. 6, pp. 1871–1883, June 2014.

[34] B. Dai and W. Yu, “Sparse beamforming and user-centric clustering for downlink cloud radio access network,” *IEEE Access*, vol. 2, pp. 1326–1339, Oct. 2014.

[35] M. Sadek, A. Tarighat, and A. H. Sayed, “A leakage-based precoding scheme for downlink multi-user MIMO channels,” *IEEE Trans. Wireless Commun*., vol. 6, no. 5, pp. 1711–1721, May 2007.

[36] I. Boukhedimi, A. Kammoun, and M.-S. Alouini, “Coordinated SLNR based precoding in large-scale heterogeneous networks,” *IEEE J. Sel. Topics Signal Process*., vol. 11, no. 3, pp. 534–548, Apr. 2017.

[37] E. A. Jorswieck, E. G. Larsson, and D. Danev, “Complete characterization of the Pareto boundary for the MISO interference channel,” *IEEE Trans. Signal Process*., vol. 56, no. 10, pp. 5292–5296, Aug. 2008.

[38] R. Mochaourab and E. Jorswieck, “Optimal beamforming in interference networks with perfect local channel information,” *IEEE Trans. Signal Process*., vol. 59, no. 3, pp. 1128–1141, Mar. 2011.

[39] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications*, Cambridge University Press, 2011.

[40] H. Ngo, E. Larsson, and T. Marzetta, “Energy and spectral efficiency of very large multiuser MIMO systems,” *IEEE Trans. Commun*., vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[41] S. Wagner, R. Couillet, M. Debbah, and D. T. M. Slock, “Large system analysis of linear precoding in correlated MISO broadcast channels under limited feedback,” *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4509–4537, July 2012.

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

[43] R. Hunger, “Floating point operations in matrix-vector calculus,” Technische Universität München, München, Germany, Oct. 2005.

[44] S. He, Y. Huang, L. Yang, and B. Ottersten, “Coordinated multicell multiuser precoding for maximizing weighted sum energy efficiency,” *IEEE Trans. Signal Process*., vol. 62, no. 3, pp. 741–751, Feb. 2014.

[45] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, “Massive MIMO performance evaluation based on measured propagation data,” *IEEE Trans. Wireless Commun.*, vol. 14, no. 7, pp. 3899–3911, July 2015.

[46] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, “Linear pre-coding performance in measured very-large MIMO channels,” in *Proc. IEEE Veh. Tech. Conf. (VTC-Fall)*, pp. 1–5, Sept. 2011.

[47] X. Gao, F. Tufvesson, O. Edfors, and F. Rusek, “Measured propagation characteristics for very-large MIMO at 2.6 GHz,” in *Proc. Asilomar Conf. Signals, Syst. Comput.*, pp. 295–299, Nov. 2012.

[48] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, “Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, capacity limits,” *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 7112–7139, Nov. 2014.

[49] Z. Jiang, A. F. Molisch, G. Caire, and Z. Niu, “Achievable rates of FDD massive MIMO systems with spatial channel correlation,” *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2868–2882, May 2015.

[50] Y. Jang, T. Kim, K. Min, M. Jung, and S. Choi, “Statistical beamforming based on effective channel gain for spatially correlated massive MIMO systems,” *IEEE Commun. Lett.*, vol. 22, no. 1, pp. 197–200, Jan. 2018.

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

[52] F. Riera-Palou and G. Femenias, "Cluster-based cooperative MIMO-OFDMA cellular networks: Scheduling and resource allocation," *IEEE Trans. Veh. Technol*., vol. 67, no. 2, pp. 1202–1216, Feb. 2018.

[53] M. Alkhaled, E. Alsusa, and W. Pramudito, "Adaptive user grouping algorithm for the downlink massive MIMO systems," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, pp. 1–6, Apr. 2016.

[54] H. Yang, "User scheduling in massive MIMO," in *Proc. IEEE SPAWC*, pp. 1–5, June 2018.

[55] K. U. Storek and A. Knopp, "Fair user grouping for multibeam satellites with MU-MIMO precoding," in *Proc. IEEE Global Commun. Conf (GLOBECOM)*, pp. 1–7, Dec. 2017.

[56] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing-The large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.

[57] J. Nam, A. Adhikary, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing: Opportunistic beamforming, user grouping and simplified downlink scheduling," *IEEE J. Sel. Topics Signal Process*., vol. 8, no. 5, pp. 876–890, Oct. 2014.

[58] X. Li, S. Jin, X. Gao, and R. W. Heath, "Three-dimensional beamforming for large-scale FD-MIMO systems exploiting statistical channel state information," *IEEE Trans. Veh. Technol*., vol. 65, no. 11, pp. 8992–9005, Nov. 2016.

[59] J. Duan, X. Lagrange, and F. Guilloud, “Analysis of different user grouping algorithms in a C-RAN downlink system,” in *Proc. IEEE Int. Commun. Conf. (ICC)*, pp. 1-6, May 2017.

[60] B. Hu, C. Hua, C. Chen, X. Ma, and X. Guan, “MUBFP: Multi-user beamforming and partitioning for sum capacity maximization in MIMO systems,” *IEEE Trans. Veh. Technol.*, vol. 66, no. 1, pp. 233–245, Jan. 2017.

[61] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*, Cambridge University Press, 2008.

[62] D. Lee *et al*., “Coordinated multipoint transmission and reception in LTE-Advanced: Deployment scenarios and operational challenges,” *IEEE Commun. Mag.*, vol. 50, no. 2, pp. 148–155, Feb. 2012.

[63] G. Nigam, P. Minero and M. Haenggi, “Coordinated multipoint joint transmission in heterogeneous networks,” *IEEE Trans. Commun*., vol. 62, no. 11, pp. 4134–4146, Nov. 2014.

[64] F. Götze and A. Tikhomirov, “Rate of convergence in probability to the Marchenko-Pastur law,” *Bernoulli*, vol. 10, no. 4, pp. 503–548, 2004.

[65] G. W. Stewart, *Matrix Algorithms: Volume 1, Basic Decompositions*, Society for Industrial Mathematics, 1998.

[66] X. Gao, L. Dai, Y. Ma, and Z. Wang, “Low-complexity near-optimal signal detection for uplink large-scale MIMO systems,” *Electron. Lett*., vol. 50, no. 18, pp. 1326–1328, Aug. 2014.

[67] R. Horn and C. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.

[68] S. Parkvall *et al*., "LTE-advanced–Evolving LTE towards IMT-advanced," in *Proc. IEEE Veh. Tech. Conf. (VTC)*, pp. 1–5, Sept. 2008.

[69] C. Pan, H. Ren, M. Elkashlan, A. Nallanathan, and L. Hanzo, "The noncoherent ultra-dense C-RAN is capable of outperforming its coherent counterpart at a limited fronthaul capacity," *IEEE J. Sel. Areas Commun*., vol. 36, no. 11, pp. 2549–2560, Nov. 2018.

[70] H. E. Rose, *Linear Algebra: A Pure Mathematical Approach*, Springer, 2002.

[71] C. Zhang *et al*., "Sum-rate analysis for massive MIMO downlink with joint statistical beamforming and user scheduling," *IEEE Trans. Wireless Commun*., vol. 16, no. 4, pp. 2181–2194, Apr. 2017.

[72] C.-S. Park, "Guaranteed-stable sliding DFT algorithm with minimal computational requirements," *IEEE Trans. Signal Process*., vol. 65, no. 20, pp. 5281–5288, Oct. 2017.

[73] A. Papadogiannis, H. J. Bang, D. Gesbert, and E. hardouin, "Efficient selective feedback design for multicell cooperative networks," *IEEE Trans. Veh. Technol*., vol. 60, no. 1, pp. 196–205, Jan. 2011.

[74] Z. Xiang, M. Tao, and X. Wang, "Massive MIMO multicasting in noncooperative cellular networks," *IEEE J. Sel. Areas Commun*., vol. 32, no. 6, pp. 1180–1193, June 2014.

[75] N. D. Sidiropoulos, T. Davidson, and Z. Q. Luo, "Transmit beamforming for physical layer multicasting," *IEEE Trans. Signal Process*., vol. 54, no. 6, pp. 2239–2251, June 2006.

[76] E. Karipidis, N. D. Sidiropoulos, and Z.-Q. Luo, “Quality of service and max-min fair transmit beamforming to multiple cochannel multicast groups,” *IEEE Trans. Signal Process*., vol. 56, no. 3, pp. 1268–1279, Mar. 2008.

[77] O. Mehanna, K. Huang, B. Gopalakrishnan, A. Konar, and N. D. Sidiropoulos, “Feasible point pursuit and successive approximation of non-convex QCQPs,” *IEEE Signal Process. Lett*., vol. 22, no. 7, pp. 804–808, July 2015.

[78] M. Sadeghi, E. Björnson, E. G. Larsson, C. Yuen, and T. L. Marzetta, “Max–min fair transmit precoding for multi-group multicasting in massive MIMO,” *IEEE Trans. Wireless Commun*., vol. 17, no. 2, pp. 1358–1373, Feb. 2018.

[79] J. Zhang, S. Chen, Y. Lin, J. Zheng, B. Ai, and L. Hanzo, “Cell-free massive MIMO: A new next-generation paradigm,” *IEEE Access*, vol. 7, pp. 99878–99888, Aug. 2019.

[80] E. Björnson and L. Sanguinetti, “Scalable cell-free massive MIMO systems,” *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4247–4261, July 2020.

[81] S. Chen *et al*., “Vehicle-to-everything (V2X) services supported by LTE-based systems and 5G,” *IEEE Commun. Stand. Mag*., vol. 1, no. 2, pp. 70–76, July 2017.

[82] A. Pfadler, C. Ballesteros, J. Romeu and L. Jofre, “Hybrid massive MIMO for urban V2I: Sub-6 GHz vs mmWave performance assessment,” *IEEE Trans. Veh. Technol*., vol. 69, no. 5, pp. 4652–4662, May 2020.

[83] K. T. Truong and R. W. Heath, Jr., “Effects of channel aging in massive MIMO systems,” *J. Commun. Netw.*, vol. 15, no. 4, pp. 338–351, Aug. 2013.

[84] L. You, J. Wang, W. Wang, and X. Gao, “Secure multicast transmission for massive MIMO with statistical channel state information,” *IEEE Signal Process. Lett.*, vol. 26, no. 6, pp. 803–807, June 2019.

[85] K. Li, L. You, J. Wang and X. Gao, “Physical layer multicasting in massive MIMO systems with statistical CSIT,” *IEEE Trans. Veh. Technol.*, vol. 69, no. 2, pp. 1651–1665, Feb. 2020.

Korean Abstract

차세대 무선 통신 시스템에서 성능 향상을 위해 대규모 다중 안테나 (massive MIMO) 기술들을 사용할 수 있다. 대규모 안테나를 가진 기지국은 많은 수의 사용자들을 빔포밍 (beamforming)으로 서비스해줄 수 있다. 안테나 수가 무한히 증가함에 따라서 채널은 점근적으로 서로 직교 (orthogonal)하게 된다. 이러한 경우, 낮은 실장 복잡도를 가지는 최대 비 전송 (maximum ratio transmission)을 사용함으로써 신호전송을 최적화할 수 있다. 하지만, 현실적인 대규모 다중 안테나 환경에서는 채널 직교성이 충분하지 못하기 때문에 최대 비 전송은 간섭에 의한 성능 저하를 겪을 수 있다. 제로-포싱 (zero-forcing) 빔포밍은 간섭을 쉽게 제거할 수 있기 때문에 대규모 다중 안테나 환경에서 현실적인 선택이 될 수 있다. 하지만, 제로-포싱은 빔 가중치 (beam weight) 생성으로 인해 높은 계산 복잡도를 요구할 수 있다. 뿐만 아니라, 제로-포싱은 간섭 제거에 대한 대가로 심각한 성능 저하 (즉, transmission performance loss; TPL)를 겪을 수 있다. TPL 은 사용자 수가 많거나 채널의 공간 상관도가 클 때 더 심각해질 수 있다.

본 논문에서 대규모 다중 안테나 환경에서 낮은 복잡도의 다중 사용자 신호전송을 고려한다. 제안 기법은 신호-대-간섭 및 잡음 비 (signal-to-interference plus noise ratio) 대신 신호-대-유출 및 잡음 비 (signal-to-leakage plus noise ratio)를 최대화하는 빔 가중치를 결정한다. 제안 기법은

최대 비 전송과 간섭을 선택적으로 제거하는 부분 제로-포싱 (partial zero-forcing)의 사용을 기반으로 빔 방향을 결정한다. 계산 복잡도를 더 감소시키기 위해서, 제안 기법은 근사화된 신호-대-유출 및 잡음비를 사용하여 빔 가중치를 결정한다.

본 논문에서 그룹 기반으로 빔 가중치를 생성하는 낮은 복잡도의 제로-포싱 빔포밍 전송을 고려한다. 제안 기법은 사용자들이 낮은 TPL 을 갖도록 사용자들을 다수의 그룹으로 분리시킨다. 계산 복잡도를 더 감소시키기 위해서, 제안 기법은 TPL 을 근사적으로 추정한다. 마지막으로, 제안 기법은 근사화된 TPL 을 기반으로 형성된 각 사용자 그룹에 대하여 빔 가중치를 결정한다.

주요어: 대규모 다중 안테나, 제로-포싱 빔포밍, 낮은 복잡도의 빔포밍, 사용자 그룹핑.

학번: 2013-20874