Fall 1-21-2019

Channel feedback in FDD massive MIMO systems with multiple-antenna users

Mahmoud AlaaEldin Mahmoud
The American University in Cairo

AMERICAN UNIVERSITY IN CAIRO

MASTER OF SCIENCE THESIS

Channel Feedback in FDD Massive MIMO Systems with Multiple-Antenna Users

---

Author:
Mahmoud AlaaEldin

Supervisor:
Dr. Karim Seddik

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science

in Electronics and Communications Engineering

January 20, 2019

Declaration of Authorship

I, Mahmoud AlaaEldin, declare that this thesis titled, "Channel Feedback in FDD Massive MIMO Systems with Multiple-Antenna Users" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

_____

Date:

_____

AMERICAN UNIVERSITY IN CAIRO

Abstract

School of Sciences and Engineering

Electronics and Communications Engineering Department

Master of Science

Channel Feedback in FDD Massive MIMO Systems with Multiple-Antenna Users

by Mahmoud AlaaEldin

In this thesis, we consider the problem of Angle of Departure (AoD) based channel feedback in Frequency Division Duplex (FDD) massive Multiple-Input Multiple-Output (MIMO) systems with multiple antennas at the users. We consider the use of Zero-Forcing Block Diagonalization (BD) as the downlink precoding scheme. We consider two different cases; one in which the number of streams intended for a user equals the number of antennas at that user and the other case in which the number of streams is less than the number of user antennas. BD requires the feedback of the subspace spanned by the channel matrix at the user or a subspace of it in the case of having a smaller number of streams than the number of antennas at a specific user. Based on our channel model, we propose a channel feedback scheme that requires less feedback overhead compared to feeding back the whole channel matrix. Then, we quantify the rate gap between the rate of the system with perfect Channel State Information (CSI) at the massive MIMO Basestation (BS) and our proposed channel feedback scheme for a given number of feedback bits. Finally, we design feedback codebooks based on optimal subspace packing in the Grassmannian manifold. We show that our proposed codes achieve performance that is very close to the performance of the system with perfect CSI at the BS. We also propose a vector quantization scheme to quantize the channel matrix of the user when optimal power allocation across multiple streams is adopted. Simulation results show that the vector quantization scheme combined with power optimization across the streams outperforms the subspace quantization scheme at the low SNR regime. However, the situation is reversed at high SNR levels and subspace quantization with uniform power allocation becomes better.

Acknowledgements

I would like to thank my mother for always being supportive during the whole time I spent in AUC. She stood by me during all my master journey. I would give a special thanks to my thesis advisor, Dr. Karim Seddik. He is a model advisor that any student wishes to work with. He gave me technical as well as moral support and he always considered me as a younger brother ...

Contents

List of Figures

List of Abbreviations

AoD	**A**ngle **o**f **D**eparture
BD	**Z**ero-**F**orcing **B**lock **D**iagonalization
BS	**B**ase**s**tation
BW	**B**and**W**idth
CSI	**C**hannel **S**tate **I**nformation
FDD	**F**requency **D**ivision **D**uplex
i.i.d.	**I**ndependently and **I**dentically **D**istributed
LAA	**L**ens **A**ntenna **A**rray
MIMO	**M**ultiple-**I**nput **M**ultiple-**O**utput
MISO	**M**ultiple **I**nput **S**ingle **O**utput
MMSE	**M**inimum **M**ean **S**quare **E**rror
mmWave	**M**illimeter **W**ave
MSE	**M**ean **S**quare **E**rror
MU-MIMO	**M**ulti **U**ser **M**ultiple-**I**nput **M**ultiple-**O**utput
MUSIC	**MU**ltiple **SI**gnal **C**lassification
NTCQ	**N**oncoherent **T**rellis-**C**oded **Q**uantization
OFDM	**O**rthogonal **F**requency **D**ivision **M**ultiplexing
PCA	**P**rinciple **C**omponent **A**nalysis
RDSC	**R**educed-**D**imensional **S**ubspace **C**odebook
RVQ	**R**andom **V**ector **Q**uantization
SISO	**S**ingle **I**nput **S**ingle **O**utput
SNR	**S**ignal to **N**oise **R**atio
SVD	**S**ingular **V**alue **D**ecomposition
TDD	**T**ime **D**ivision **D**uplex
ULA	**U**niform **L**inear **A**rray
ZF	**Z**ero-**F**orcing

List of Symbols

λ_c	carrier wavelength
f_c	carrier frequency
$\mathbb{E}$	expected value
$\mathbb{C}$	set of complex numbers
$\mathcal{G}$	Grassmannian manifold
c	speed of light
d_l	antenna spacing
ϕ	angle of departure
$\mathbf{a}(\phi)$	spatial direction
P_k	number of resolvable paths at user k
$\mathbf{A}_k(\phi_1, \phi_2, \cdots, \phi_P)$	channel subspace of user k
$\mathbf{H}_k$	channel matrix of user k
$\mathbf{h}_k$	channel vector of user k
$\mathbf{G}_k$	path gains matrix of user k
$\mathbf{g}_k$	path gains vector of user k
G_r	Gram matrix
M	number of antennas at BS
K	number of users in the cell
N_k	number of receive antennas at user k
B	number of feedback bits
θ	principle angle between subspaces
$\|\|\|_F$	Frobenius norm
$\mathbf{F}_k$	precoding matrix for user k
$\mathbf{u}_k$	data streams vector intended to user k
$\mathbf{x}$	transmitted signal vector for the BS
$\mathbf{y}_k$	received signal vector at user k

$\mathbf{n}_k$	additive noise vector at user k
$\widetilde{\mathbf{H}}$	subspace spanned by the rows of $\mathbf{H}$
$\widehat{\mathbf{H}}$	quantized version of $\mathbf{H}$
$\mathcal{C}$	quantization codebook
$\mathbf{C}$	subspace quantization elements in $\mathcal{C}$
γ	total transmit power
σ	singular value element
Σ	singular value matrix
$\mathbf{U}$	left singular vectors matrix
$\mathbf{V}$	right singular vectors matrix
R	data rate in (bits/second/Hz)
ΔR	rate gap
m_k	number of streams
$\mathbf{I}$	identity matrix

Chapter 1

Introduction

1.1 General overview on massive MIMO

The field of wireless communications has been growing very fast through the last decades. Many improvements have been done to the field and a lot of ideas have been proposed too in order to make the wireless systems more reliable and provide higher data rates to every user. A lot of techniques have been proposed to increase the throughput like the multi antenna systems. It was proved in the literature that equipping the wireless systems with more than one antenna at both the transmitter and the receiver side can raise the capacity of the systems by orders of magnitude. Then, massive MIMO, which is a new technology in MIMO systems, has been recently proposed in the literature, and it is expected to be the future technology to be applied in the fifth generation of mobile technology (5G). Massive MIMO wireless communication systems are the systems that are equipped with a very large number of antennas (hundreds of antennas) at the base station side, and they have been shown to introduce massive improvements in both spectral efficiency and energy efficiency [1], [2], [3]. MIMO systems attracted the researchers through the last three decades because of their high achievable data rates and their ability to scale up the capacity of the wireless channels. These improvements in the achievable throughput led to more robust and fast data transfer in our daily life. Thanks to MIMO systems, new emerging technologies have appeared like WIFI, 4G and 5G. Initially, previous research in the literature focused on point to point MIMO systems where both the transmitter and the receiver have multiple antennas.

However, the focus now is towards more practical systems which are called multi-user MIMO (MU-MIMO) systems, where a base station with a huge number of antennas serves a number of single-antenna users at the same time. Massive MIMO is a special case of MIMO systems where we exploit providing the base station with a large number of antennas in order to serve multiple users at a time and hence increase the network throughput. Massive MIMO can improve the capacity and reliability of wireless systems. Massive MIMO has moved the complexity to the base station only and the user equipment's can be cheap single antenna devices. Although conventional MIMO systems, where the base station has only a few number of antennas, achieved significant improvement in the spectral efficiency, this enhancement is still modest and we are in need for more utilization of the resources.

The results from random matrix theory [4] prove that the number of users per cell does not depend on the size of it, and the energy per bit needed for transmission goes to zero as the number of antennas in a MIMO cell goes high to infinity. Also, we can use very simple signal processing algorithms, like matched filter precoding and detection, in order to achieve such benefits. As long as the number of antennas at the base station will be much larger than the number of users in a MU-Massive MIMO cell, there are many degrees of freedom that can be utilized to form the transmitted signals with a low cost from the hardware point of view or to cancel interference [5]. In order to realize such a system in practical use, there are algorithms and techniques for massive MIMO that are studied to reduce the computational complexity as much as possible.

Not only the spectral efficiency that can be improved using massive MIMO, but also the energy efficiency of such a system is way better than the corresponding single antenna systems. It is shown that each single antenna user in a massive MIMO system can reduce its transmitted power [6]

proportional to the number of antennas at the base station. This proportionality in scaling down the transmitted power can be achieved with perfect CSI. However, the reduction in power can be proportional to the square root of the number of antennas in the case of imperfect CSI. Although massive MIMO can reduce the power significantly in this way, it can achieve the same performance as the opposite single input single output (SISO) systems. This is a great result from the energy efficiency point of view, and it is very important for future wireless communications when the consumed energy is a big issue [7], [8].

Channel estimation is a crucial part in massive MIMO systems as all the linear transmit precoding schemes like minimum mean square error (MMSE) and zero-forcing (ZF) cannot be done without knowing the CSI as perfect as possible. The performance of massive MIMO systems based on various linear receivers has been studied from various perspectives [9], [10], [11] and [12]. Not only channel estimation but also the feedback of the estimated CSI to the base station again is very important to perform precoding and digital beam-forming on the transmitted signals. The challenges that face massive MIMO is that it has a very large number of antennas and hence the transmitted overhead (pilot signals) is also very large which will decrease the system throughput. The choice of the pilot signals used in channel estimation is a design issue and they must be chosen in a way that minimizes the overhead [13] and [14]. After estimation of the downlink channels at the user equipment, the feedback stage comes. This feedback data must be compressed in an efficient way that minimizes the feedback overhead without affecting the performance of the system or losing some information in the compression process.

1.2 Organization of the thesis

This thesis presents some work and contributions in massive MIMO Frequency Division Duplex (FDD) systems and how to efficiently feed back

the CSI of the cell users to the BS to perform precoding and beamforming. The thesis is organized as follows:

- Chapter (2) presents a literature review of the channel feedback problem in massive MIMO systems.
- Chapter (3) discusses the massive MIMO channel model as well as the proposed line packing based quantization of the massive MIMO channel in the case of single antenna users. We also compare the performance of line packing against random vector quantization in the chapter.
- In chapter (4), the channel feedback problem is discussed in the case of multiple antenna users. We also presents mathematical analysis to quantify the rate gap between the ideal and the quantized channel cases.
- Chapter (5) discusses a power optimization scheme that optimally allocates the total power over multiple data streams.
- In chapter (6), we provide the conclusions of this thesis and propose some directions for future research.

1.3 List of contributions of this thesis

- In chapter (3), we present an efficient channel feedback scheme from the users to the BS. The feedback codebook is designed based on a Grassmannian codebook designed by line packing on the Grassmannian manifold.
- In chapter (4), we consider the problem of channel feedback in FDD massive MIMO systems with multiple antennas at the users.
- We design efficient feedback codebooks based on optimal subspace packing on the Grassmannian manifold

- We consider two different cases; one in which the number of streams intended for a user equals the number of antennas at that user and the other case in which the number of streams is less than the number of user antennas.

- We propose a channel feedback scheme that requires less feedback overhead compared to feeding back the whole channel matrix as BD requires the feedback of the subspace spanned by the channel matrix at the user or a subspace of it in the case of having less number of streams than the number of antennas at a specific user.

- We quantify the rate gap between the rate of the system with perfect CSI at the massive MIMO BS and our proposed channel feedback scheme for a given number of feedback bits.

- In chapter (5), we propose an optimal power allocation scheme that allocates the total power over the multiple data streams. We propose a vector quantization scheme that is combined with power optimization and the simulations show that it is better than subspace quantization in the low SNR regime.

Chapter 2

Literature review

2.1 Summary of latest work on massive MIMO channel estimation and feedback

Generally, in FDD systems, the downlink channels are different from the uplink ones. This means that the base station transmits its data to the users in the call on channels with certain frequencies and receive data form the users on channels with other frequencies. The base station performs channel estimation for the uplink by making the users send different pilot signals. The pilot transmission for uplink channel estimation requires time which does not depend on the number of antennas at the base station. However, to acquire CSI for the downlink channels in FDD systems, we need two stages. First, the base station transmits the pilot signals to the users in the cell, then the users use these known pilots to estimate the downlink channels. After that, the users feedback these estimates, complete or partial, to the BS. The time needed for sending pilot signals to estimate the downlink channels is proportional to the number of antennas at the base station side. So, as massive MIMO size becomes larger, the transmitted overhead becomes larger too and hence the traditional methodology for channel estimation becomes infeasible.

The problem of channel estimation in massive MIMO FDD systems has been discussed in many recent papers. However, it still needs improvement and more research to bring it from theory to practice. For example,

[15] discussed a spatially common sparsity based adaptive channel estimation and feedback scheme for FDD systems based massive MIMO systems. The proposed scheme in this paper tried to adapt the training symbols as well as the pilot design to make reliable estimates of the downlink channel with significantly reduced overhead. They proposed a non-orthogonal pilot design scheme which is very different from the conventional orthogonal pilot design. Then they exploited the fact that the massive MIMO channels are sparse in the angle domain, so a compressive sensing based adaptive scheme for acquiring CSI was proposed. The overhead in this scheme only adaptively depends on the sparsity level of the channels. Then they proposed a joint estimation of the channels having multiple subcarriers. Then the paper suggested a closed-loop channel tracking algorithm exploiting the temporal correlation of massive MIMO systems. This scheme adaptively designs orthogonal pilot signals based on the previous channel estimations in order to enhance the performance of estimation. Finally, the paper discussed a distributive sparsity adaptive matching pursuit algorithm which makes use of the spatially common sparsity of the channels in massive MIMO to estimate a group of channels jointly at one time. Compared with the conventional techniques, this algorithm highly reduces the required overhead with the same computational complexity.

[16] proposed an angular domain pilot design and channel estimation for massive MIMO FDD systems. It proposed a pilot design scheme in the angular domain exploiting what is known as the sparsity of the channels in the angular domain. By exploiting this sparsity, we can dramatically reduce the required overhead in order to estimate all the channels in massive MIMO systems. The paper first proposed how to estimate the dominant angular set of the downlink channels from the uplink channels using the directional reciprocity of the FDD channels. It introduces an index calibration algorithm to get the dominant downlink angular set from those in uplink taking into account the difference in wavelengths between both sets in FDD systems. After estimating the dominant angular set, it proposes

two different schemes to estimate channels in that dominant set. One of the two schemes, is complete orthogonal pilot design, while the other one is partial orthogonal. After that, the paper discussed the feedback schemes corresponding to the two estimation schemes in order to send the estimated channels at each user back to the base station. Simulation results showed that the proposed angular domain pilot design has a good performance with low mean square error (MSE) while reducing the pilot signals overhead.

A non-coherent trellis-coded quantization (NTCQ) with encoding complexity that scales linearly with the number of BS antennas was proposed in [17]. The authors exploited the duality between source encoding in a Grassmannian manifold in order to find a vector in the codebook that maximizes beamforming gain) and non-coherent sequence detection (for maximum likelihood decoding with respect to to uncertainty in the channel gain). They obtained a low-complexity implementation of NTCQ encoding using Viterbi algorithm applied to standard trellis coded quantization. The authors exploited that non-coherent detection can be realized near-optimally using a bank of coherent detectors in order to reduce the complexity. Channel properties such as temporal and spatial correlations were exploited in order to develop advanced NTCQ schemes.

In multiuser massive MIMO systems, when the users are able to exchange the CSI, it is not clear whether they should feed back the channel or the precoder itself. [18] compared the precoder feedback scheme against the channel feedback scheme. The authors found that when there are sufficient number of bits to exchange CSI, the precoder feedback scheme is better than channel feedback scheme in the sense that it reduces the interference leakage to $1/(K-1)$ of the channel feedback scheme, where K is the number of users. Additionally, when the number of feedback bits increases, the interference leakage of the precoder feedback scheme reduces faster than the channel feedback scheme .

The problem of pilot beam pattern design for channel estimation in massive MIMO systems with large number of transmit antennas at the BS was proposed in [19]. The authors proposed a new algorithm for pilot beam pattern design for optimal channel estimation assuming that the channel is a stationary Gauss-Markov random process. They exploited Kalman filtering properties, the prediction error covariance matrices and the channel statistics such as spatial correlation in order to design the pilot beam pattern sequentially. The design generated an optimal sequential sequence of pilot beam patterns with low computational complexity for a given set of system parameters.

[20] discussed a distributed compressive channel estimation scheme for FDD massive MIMO systems was proposed and a feedback scheme to send the estimated channels to the base station was also proposed. The paper used the compressive sensing technique in order to reduce the amount of training overhead as well as the feedback in the estimation process in multi user massive MIMO systems. This technique makes use of the hidden sparsity structure which is an important feature in the user channel matrices in massive MIMO systems. The paper proposes a distributed compressive sampling scheme instead of applying the conventional one so that each user locally gets the compressed measurements while the CSI recovery is jointly done at the base station. The base station uses the joint orthogonal matching pursuit algorithm to jointly estimate the channels for each user with the aid of using the hidden joint sparsity in the users' channels. The paper introduced some insights of how to exploit the joint sparsity to improve the performance of the channel estimation process, simulation results showed that the estimated channels have good quality in terms of the mean square error.

[21] introduced a compressed channel feedback scheme for correlated massive MIMO channels is proposed. The channels are quantized based on compressive sensing technique in order to be fed back to the base station

with low overhead. This paper used a model which consider that the base station has a 2-dimentional uniform square array of antennas which has channels with high correlations. The paper used the zero-forcing precoder as a multiple access technique, so it must know all the channel matrix in order to form the precoders.

A new sparsifying basis that reflects the long-term characteristics of the channel was proposed in [22]. It does not need to change as long as the spatial correlation model does not change. They presented a new reconstruction algorithm for compressed sensing, and also proposed dimensionality reduction as a compression method. They proposed a new codebook for the compressed channel quantization assuming no other-cell interference in order to feed back the compressed CSI in practice. They provided simulation results to confirm that the proposed channel feedback schemes show better performance in point-to-point (single-user) and point-to-multi-point (multi-user) scenarios.

A limited feedback scheme for massive MIMO systems based on principle component analysis (PCA) was discussed by [23]. The proposed scheme makes a good use of the spatial correlations of massive MIMO channels. The mobile station compresses the spatially correlated channels into lower dimensional channel state information. The resultant compressed channel information is then fed back to the base station with reduced overhead as well as lower complexity for codebook search. Each mobile station uses a compression matrix in order to perform compression process, and this matrix is attained by applying the principle component analysis on the channel state information which is estimated over the long run at the mobile station. The numerical results show that the PCA based scheme can significantly reduce the feedback overhead without degrading the system performance dramatically.

In [24], a reduced-dimensional subspace codebook (RDSC) was proposed for Millimeter Wave (mmWave) massive MIMO systems relying on a lens antenna array (LAA). They first generated the large-dimensional vectors in the channel subspace, which is determined by the angles-of-departure of the dominant paths, specifically under the concept of angle coherence time. Then, based on the generated vectors in the channel subspace, they considered both the lens and the beam selector to create the RDSC. Finally, the equivalent channel is quantized using the proposed RDSC and fed back to the BS. The authors provided mathematical performance analysis of their proposed RDSC and showed that its feedback overhead is reduced proportional to the number of dominant paths per user which is relatively small.

2.2 AoD-adaptive subspace codebook for channel feedback in FDD massive MIMO systems

An adaptive subspace code book based on angles of departure for channel feedback was proposed by [25]. The scheme proposed a code book design which has a significantly reduced size which will result in much reduced feedback overhead. The paper utilized the idea of the angle coherence time which is much larger than the paths' gains coherence time. The path angles of departure vary much slower than the channel gains, consequently we can exploit this fact to reduce the frequency of the feedback information which is sent to the base station. The paper discussed an adaptive subspace codebook which is based on the angles of departure in order to quantize the massive MIMO channel vector. The paper used the random vector quantization as a proof of concept, but it does not consider a static design for the quantizer such that it minimizes the quantization error. The paper was built on a simple philosophy which is the channel vector always locates in the low dimensional subspace from the ambient full dimensional subspace, so you need not to feedback all entries of the

channel vector but all you need is to feedback the projections of this channel vector on the basis vectors of the subspace. This argument is very useful that it will reduce the amount of feedback bits dramatically because it compresses the channel information in a very efficient way. As the number of quantization vectors increase, the performance will reach the ideal case of perfect channel state information, but the number of feedback bits will also increase and hence increase the feedback overhead. The paper provided some performance analysis to prove that the number of feedback bits is linearly proportional to the number of basis vectors of the channel subspace which is also the number of resolvable paths each has a different angle of departure. Also, to maintain a constant gab between the ideal case and the proposed scheme with all signal to noise ratio values, the paper proved that we have to increase the number of quantization vectors for the codebook and hence this will result in increasing the number of feedback bits too. Fig. (2.1) shows the performance curve that exist in the original paper [25]. The figure shows the per-user rate of a massive MIMO cell with one BS having 128 transmit antennas serving 8 single antenna users. The graph shows that the proposed scheme gives a performance which is near optimal, but again the number of feedback bits increases as the signal to noise ratio (SNR) values increase. As SNR increases, the system additive noise becomes insignificant compared to the noise that comes from the channel quantizer. Consequently, more feedback bits are needed to quantize the channel vectors at the users in order to maintain a constant rate gap.

Now, we find that most of the papers cited above do not consider the design of the channel feedback scheme, or just assume random vector quantizers as a proof of concept. However, these quantizers do not result in the best performance. Furthermore, random vector quantizers do not provide a systematic was to design the feedback quantizer. Furthermore, most of the recent papers in MU-massive MIMO do not consider more than one antenna at each user in their system model. They assume that the BS has a very large number of antennas, but each user has only a single antenna.

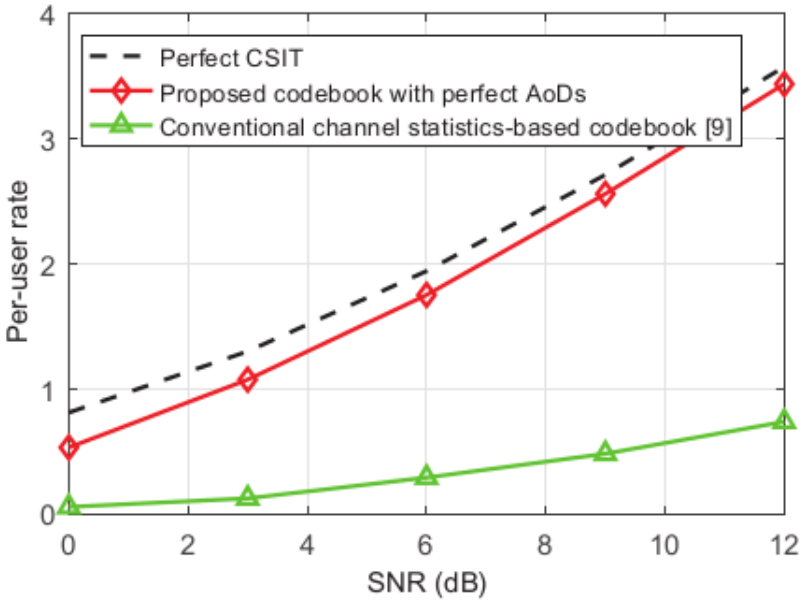

FIGURE 2.1: Comparison of the per-user rate between the ideal case of perfect CSIT and the practical cases of limited channel feedback [25]

However, it could be useful to consider multiple antennas at the user side because this will provide better estimation of the massive MIMO channels by exploiting the correlations between the channels of the users' antennas.

Chapter 3

Channel Feedback in Massive MIMO Systems

This chapter is divided into two sections. In the first section, we discuss the massive MIMO channel model in details. Then, the second section presents a systematic approach to design the channel feedback codebooks in multi-user massive MIMO systems with a single BS and multiple single antenna users. In this case, the codebook design is formulated as a line packing problem over the Grassmannian manifold.

3.1 The classical narrowband ray-based downlink channel model

In this section, the downlink channel model that is commonly assumed in massive MIMO systems is discussed. To be concrete, we focus on the uniform linear arrays in the case of MISO channels [26], where the transmit antennas are spaced on a straight line.

3.1.1 Line of sight MISO channel

The simplest MISO channel has only one line-of-sight from the transmit antenna array to a single receive antenna. Hence, there only exist free space without any reflectors or scatterers, and only a single path from each transmit antenna to the receive antenna. The antenna separation is $\Delta_t \lambda_c$, where

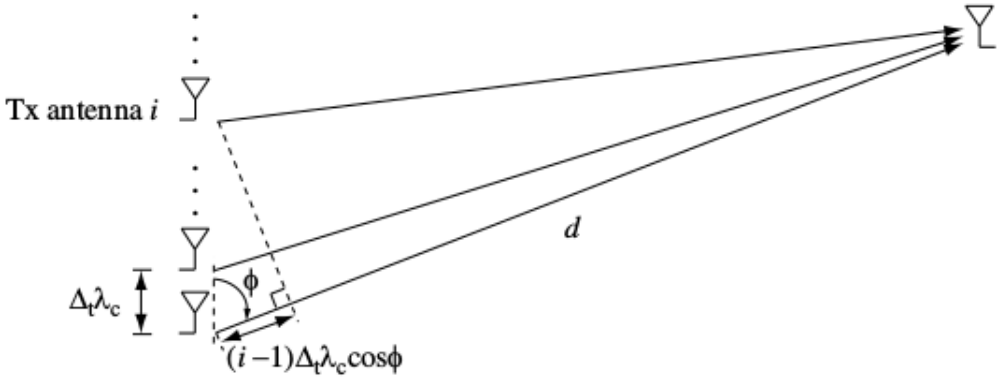

FIGURE 3.1: Line-of-sight channel with multiple transmit antennas and single receive antenna [26]

λ_c is the wavelength of the carrier and Δ_t is the normalized transmit antenna separation, normalized to the unit of λ_c. The transmit antenna array has a dimension which is much smaller than the distance between it and the receive antenna. The impulse response of the channel between the i^{th} transmit antenna and the receive antenna is given by

$$h_i(\tau) = a\,\delta(\tau - d_i/c), \quad i = 1, 2, \cdots, n_t, \tag{3.1}$$

where d_i is the distance between the i^{th} transmit antenna and the receive antenna, a is the attenuation of the channel and c is the speed of light. Let $d_i/c << 1/BW$ in order that the path delay be smaller than the data symbol time duration which results in flat fading, where BW is the bandwidth of the channel. Then the complex baseband channel coefficient is given as

$$h_i = a \exp\left(-\frac{j2\pi f_c d_i}{c}\right) = a \exp\left(-\frac{j2\pi d_i}{\lambda_c}\right), \tag{3.2}$$

where f_c is the frequency of the carrier. Thus, the MISO channel can be written as

$$y = \mathbf{h}^{\mathrm{T}}\mathbf{x} + n, \tag{3.3}$$

where $\mathbf{x} \in \mathbb{C}^{n_t \times 1}$ is transmitted signal vector from the antenna array and $n \sim \mathcal{CN}(0, N_0)$ is the additive noise. The channel coefficients vector $\mathbf{h}^{\mathrm{T}} = [h_1, h_2, \cdots, h_{n_t}]$ is called the signal direction or the spatial signature of the

transmit antenna array. The channel paths from each of the transmit antennas to the receive antenna are, to a first order, parallel as the distance between the transmit antenna array and the receive antenna is much larger than the size of the antenna array. Hence, we can express d_i as

$$d_i \approx d - (i-1)\Delta_t \lambda_c cos\phi, \quad i = 1, \cdots, n_t, \tag{3.4}$$

where d is the distance from the receive antenna to the first transmit antenna in the array and ϕ is the angle of departure of the line of sight from the transmit array to the receive antenna. The value $(i-1)\Delta_t \lambda_c cos\phi$ is the displacement of the i^{th} transmit antenna from transmit antenna 1. Hence, the spatial signature is given as

$$\mathbf{h} = a \exp\left(-\frac{j2\pi d}{\lambda_c}\right) \begin{bmatrix} 1 \\ \exp(j2\pi\Delta_t cos\phi) \\ \exp(j2\pi 2\Delta_t cos\phi) \\ \vdots \\ \exp(j2\pi(n_t-1)\Delta_t cos\phi) \end{bmatrix}. \tag{3.5}$$

Hence, the signals received at the receive antenna from each of the transmit antennas differ in phase by $2\pi\Delta_t cos\phi$. Finally, let's define the unit spatial signature in the directional cosine $cos\phi$ as [26]

$$\mathbf{e}_t(cos\phi) = \frac{1}{\sqrt{n_t}} \begin{bmatrix} 1 \\ \exp(-j2\pi\Delta_t cos\phi) \\ \exp(-j2\pi 2\Delta_t cos\phi) \\ \vdots \\ \exp(-j2\pi(n_t-1)\Delta_t cos\phi) \end{bmatrix}. \tag{3.6}$$

3.1.2 MISO channel model with reflectors and scatterings

In this part, the MISO channel model is discussed when there are reflectors that surround the BS and the receive antenna is located in a multi-path

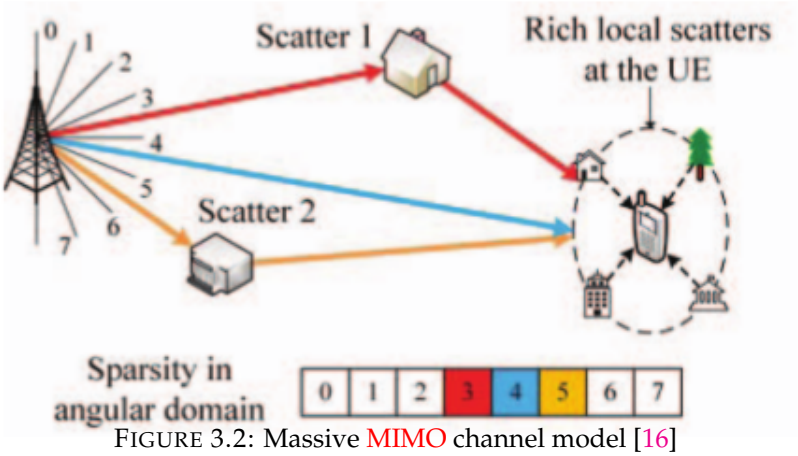

FIGURE 3.2: Massive MIMO channel model [16]

fading environment. The number of resolvable paths P seen by the BS depends on the scatters around the BS as in Fig. (3.2). Each one of these resolvable paths has a different Angle of Departure (AoD) from the BS. Then, the resolvable paths are each received at the receive antenna after passing through a multiplicative fading channel. Hence, each path is multiplied by a complex rayleigh fading coefficient $g_i \sim \mathcal{CN}(0,1)$. Therefore, the overall channel vector seen at the receive antenna is given as

$$\mathbf{h} = \sum_{i=1}^{P} g_i \mathbf{e}_i(cos\phi_i), \tag{3.7}$$

where $\mathbf{e}_i(cos\phi_i)$ is the spatial signature of the i^{th} resolvable path between the transmit antenna array and the receive antenna.

3.2 Line packing based channel feedback in massive MIMO systems with single antenna users

The massive MIMO channel model that we discussed in the previous section was used in [25]. It was proposed a random quantization scheme of the channel vectors. The random quantization scheme is not structured and cannot be used in real systems.

3.2.1 System model

It was assumed in [25] a mmWave massive MIMO broadcast (downlink) system with a single BS communicating with K single-antenna users. The

BS has M transmitting antennas while each user one single antenna. This model assumes that the number of transmitting antennas is much higher than the number of users (i.e., $M >> K$). This assumption is necessary in massive MIMO systems since it makes simple linear processing like zero forcing and MMSE precoders be optimal. They considered a narrowband ray-based downlink channel model for the downlink channel vectors $\mathbf{H}_k \in \mathbb{C}^{1\times M}$ at the k^{th} user [25]

$$\mathbf{H}_k = \mathbf{g}_k \mathbf{A}_k(\phi_{k,1}, \phi_{k,2}, \cdots, \phi_{k,P_k}). \tag{3.8}$$

The matrix $\mathbf{A}_k(\phi_{k,1}, \phi_{k,2}, \cdots, \phi_{k,P_k}) \in \mathbb{C}^{P_k\times M}$ is defined as:

$$\mathbf{A}_k(\phi_{k,1}, \phi_{k,2}, \cdots, \phi_{k,P_k}) = \begin{bmatrix} \mathbf{a}(\phi_{k,1}) \\ \mathbf{a}(\phi_{k,2}) \\ \vdots \\ \mathbf{a}(\phi_{k,P_k}) \end{bmatrix} \tag{3.9}$$

where P_k is the number of resolvable paths from the BS to the k^{th} user. The parameter $\phi_{k,i}(1 \leq i \leq P_k)$ represents the AoDs of the i^{th} path of the k^{th} user. Hence, $\mathbf{a}(\phi_{k,i}) \in \mathbb{C}^{1\times M}$ is the spatial signature of the i^{th} propagation path of the k^{th} user, and it can be written as

$$\mathbf{a}(\phi_{k,i}) = [1, e^{-j2\pi \frac{dl}{\lambda} \cos(\phi_{k,i})}, \cdots, e^{-j2\pi \frac{dl}{\lambda}(M-1)\cos(\phi_{k,i})}], \tag{3.10}$$

where λ is the signal wavelength and dl is the spacing between every two successive antennas at the BS. From Eq. (3.8), we can notice that the k^{th} user's channel vector for each antenna is a linear combination of its P_k steering vectors scaled by the complex gains of that path. $\mathbf{g}_k \in \mathbb{C}^{1\times P_k}$ is a vector whose elements contain the complex path gains at the k^{th} user (i.e., the entry $\mathbf{g}_k(i)$ represents the complex gain of the i^{th} path at user k). The complex path gains in $\mathbf{g}_k$ are assumed to be Independently and Identically Distributed (i.i.d.) circularly-symmetric complex Gaussian random variables with zero mean and unit variance.

Now, assuming that the spatial signatures $\mathbf{a}(\phi_{k,i})$ are known at the BS, only the complex path gains $\mathbf{g}_k$ are needed to be quantized and fed back to the BS. We extensively talk about the AoD-adaptive codebooks in chapter (4).

Now, we extend the work in [25] by structured codebooks to quantize the channels of the users and we show by simulations that the proposed codebooks give a better performance than the random codebooks. The quantization vectors ,that quantize $\mathbf{g}_k$, must be chosen in a way that they are maximally separated from each other, not random. We propose a systematic approach to quantize and feed back the channel vectors in multi-user massive MIMO systems with single antenna users. The codebook design is formulated as a line packing problem over the Grassmannian manifold.

3.2.2 Complex line/subspace quantization

In this sub-section, we present the random vector/subspace quantization as well as quantization based on packing of lines/subspaces over the Grassmannian manifold. Random subspace quantization are used to analyse the performance of systems by averaging over these random selections of codebooks. However, they cannot be used in practical systems due to their random nature. On the other hand, subspace packing over Grassmannian manifolds provides structured codebooks that can be used in practical systems, and its performance is always better than random quantization. Subspace packing quantization is always suitable for quantizing the channel vectors/matrices with complex Gaussian coefficients because the channel subspace, in that case, is isotropically distributed over the complex space. Subspace packing quantization is commonly used is regular MIMO and multi user MIMO systems.

Random Subspace Quantization codebooks

In general, the design of optimal quantization codebooks is a very hard problem, especially when the number of subspaces to be separated is large. Hence, the performance in such cases can be studied by averaging over

random codebooks [27]. It is easier to analyze the performance of random codes in this case, and this would provide us with some performance bounds for structured codes. In random subspace quantization, each one of the 2^B quantization subspaces is independently chosen from the isotropic distribution on the d-dimensional unit sphere. The set of all Q-dimensional subspaces in a d-dimensional space, where $Q < d$ represent a Grassmannian manifold, which is denoted by $\mathcal{G}_{d,Q}$. The 2^B random subspaces, that form the random quantization codebook, are uniformly distributed over $\mathcal{G}_{d,Q}$. A random subspace chosen uniformly over $\mathcal{G}_{d,Q}$ can be generated by generating an $Q \times d$ matrix whose elements are i.i.d. complex Gaussian. Then, an orthonormal basis for the row space of this matrix is calculated using QR decomposition.

Grassmannian Subspace Packing

The design of the quantized channel matrices is done using subspace packing in Grassmannian manifold. The packing problem tends to find 2^B subspaces in a higher dimensional space such that the minimum distance between two subspaces is maximized. There are many distance metrics that have been used for packing subspaces in the Grassmannian manifold. In this chapter, we adopt the chordal distance as our distance metric. The codebook design is done by solving the packing problem of 2^B Q-dimensional subspaces in a complex Euclidean space of dimensionality d. We follow the iterative algorithm stated in [28] in order to solve the subspace packing problem. The solution of this problem is usually simpler when the number of subspaces in the codebook 2^B is lower than d^2. In that case, the minimum distance between two subspaces in the codebook can reach the Rankin bound [28], which is the maximum attainable theoretical distance.

Complex line/subspace packings in Grassmannian manifolds We define first the complex Grassmannian manifold $\mathcal{G}(Q, \mathbb{C}^d)$ as the set of all Q-dimensional subspaces in the ambient d-dimensional space $\mathbb{C}^d$. We state our problem as we want to pack N subspaces in $\mathcal{G}(Q, \mathbb{C}^d)$, i.e. we want to

get N subspaces each has a dimension of Q from the ambient space $\mathbb{C}^d$ [28]. When these subspaces have a dimension of $Q = 1$, we call the Grassmannian manifold then as projective space and the problem reduces to packing lines that pass through the origin of a Euclidean space. In order to solve the packing problem, we solve an optimization problem which states that the minimum distance between two subspaces have to be maximized. By doing this, we will ensure that the resulting subspaces are well separated from each other. There are many interesting metrics in the Grassmannian manifolds that lead to different packing problems. We only consider metrics that are functions of what we call the principle angles between the two Q-dimensional subspaces. The metric that we use here to introduce our algorithm is called the chordal distance between two Q-dimensional subspaces, and it is given by the following relation [28],

$$dist_{chord}(S,T) = \sqrt{sin^2(\theta_1) + \cdots + sin^2(\theta_Q)} = \left[Q - \|S^H T\|_F^2\right]^{\frac{1}{2}}, \quad (3.11)$$

where $\theta_1, \cdots, \theta_Q$ are called the principle angles between the two Q-dimensional subspaces, $\|\|_F$ is the frobenius norm, and S and T are matrices that contains an orthonormal basis for each subspace on their columns. These principle angles can be efficiently determined by solving the singular value decomposition problem of the matrix $S^H T$ and obtaining the singular values of it. The produced singular values are the cosines of the desired principle angles that we want to determine. We define the Gram Matrix G_r as a configuration matrix which defines a collection of N subspaces each has a dimension of Q in an ambient space of dimension d i.e. it represents N Q-dimensional subspaces in $\mathcal{G}(Q, \mathbb{C}^d)$. In order to construct the G_r matrix, we define X as a matrix of dimension $d \times QN$ and it can be constructed like the following,

$$X = [X_1, X_2, \cdots, X_N] \quad (3.12)$$

where X_n is a nonunique $d \times Q$ matrix whose columns are an orthonormal basis for the n^{th} subspace out of the N subspaces that we have. So, back to the Gram matrix, we can define it as $G_r = X^H X$. By construction, the

rank of G_r always does not exceed d and it is positive semi-definite. The dimension of G_r is $QN \times QN$, so we can consider it as an $N \times N$ block matrix that consists of $Q \times Q$ blocks.

Packings with chordal distance Suppose that we want to get a packing of N subspaces in $\mathcal{G}(Q, \mathbb{C}^d)$ equipped with the chordal distance. Let X be a configuration matrix of the N subspaces which consists of N matrices, X_n, each has Q columns that represent an orthonormal basis for each subspace. We want to get X such that

$$\max_{m \neq n} \|X_m^H X_n\|_F \leq \mu \tag{3.13}$$

We can formulate this feasibility problem in terms of the Gram Matrix G_r. If we assume that there is a configuration matrix X that satisfy this feasibility condition, then its corresponding G_r matrix have the following properties [28]:

- G_r is Hermitian.
- Each diagonal block of G_r is an identity matrix.
- $\|G_{mn}\|_F \leq \mu$ for each $m \neq n$.
- G_r is positive semidefinite.
- G_r has a rank of d or less.
- G_r has trace QN.

The algorithm We notice from the above properties that the first three are structural properties as they control the entries of the G_r matrix directly, while the second three are spectral ones as they constrain the eigenvalues of the matrix. In order to solve our feasibility problem stated above, we have to find a matrix that satisfy both the constraint sets. This section addresses the algorithm of finding the configuration matrix X [28].

Input to the algorithm:

- $QN \times QN$ Hermitian matrix $G_r^{(0)}$
- Maximum number of iterations T_{max}

Output from the algorithm:

- A $QN \times QN$ matrix G_{out} that satisfy the spectral constraints and its diagonal blocks are the identity matrices

Algorithm 1: Subspace packing algorithm

1 Initialize $t = 0$
2 Determine a matrix $H^{(t)}$ that solves

$$\min_{H \in structuralconstraints} \|H - G_r^{(t)}\|_F$$

3 Determine a matrix $G_r^{(t+1)}$ that solves

$$\min_{G \in spectralconstraints} \|G_r - H^{(t)}\|_F$$

4 Increment t
5 If $t < T_{max}$ return to step 2
6 Define the block diagonal matrix $D = diag(G_r^{(T_{max})})$
7 Return the matrix $G_{out} = D^{\frac{-1}{2}} G_r^{(T_{max})} D^{\frac{-1}{2}}$

The algorithm is iterative and it is not guaranteed to converge in norm. Consequently, we have to stop the procedure after a fixed number of steps instead of looking at the behavior of the iterates. The output matrix G_{out} must share all the properties that were listed above except the third one that may be violated. This violation is because we usually do not know the minimum packing diameter that we wish to reach. It is like we are trying to reach an optimal solution that we do not know. Also, the output matrix always adheres to a factorization $G = X^H X$, where X is a $d \times QN$ configuration matrix that we want to calculate.

3.2.3 Simulation results

In this section, we compare the performance of the proposed line packing based quantization scheme with the random vector quantization which was proposed in [25]. The following figures show that line packing quantizer outperforms the random vector quantization. We simulated a single

massive MIMO cell with one BS and $K = 8$ single antenna users. The BS has $M = 128$ antennas forming a Uniform Linear Array (ULA). The ratio between the antenna spacing and the carrier wavelength is $\frac{dl}{\lambda} = 0.5$, and the number of resolvable paths at each user is $P = 3$. We only quantize the path gains vector of each user rather than quantizing the whole channel vector as the spatial direction $\mathbf{a}(\phi_{k,i})$ of user k is known at the BS. For Random Vector Quantization (RVQ) case, we pick $N = 2^B$ vectors at random on the complex P-dimensional unit sphere to form our random codebook. Then, the quantization of the path gain vector is performed using this codebook. After that, the performance of the system is averaged over the random codebooks as well as the random channel vectors. For line packing based quantization, we construct $N = 2^B$ vectors ($Q = 1$ dimensional subspaces) in a P-dimensional complex space using the line/subspace packing algorithm discussed in the previous sub-section. Once we construct the codebook, we fix it for all iterations while averaging only over the distribution of the channel vectors.

The following figures show the performance of linepacking against RVQ. It is noticed that as the number of quantization vectors $N = 2^B$ increase, the performance of RVQ approaches the performance of line packing based quantization. This is because that at large N, the random vectors in a random codebook may be distributed well in the space and they are well separated from each other since we have a large number of vectors elements.

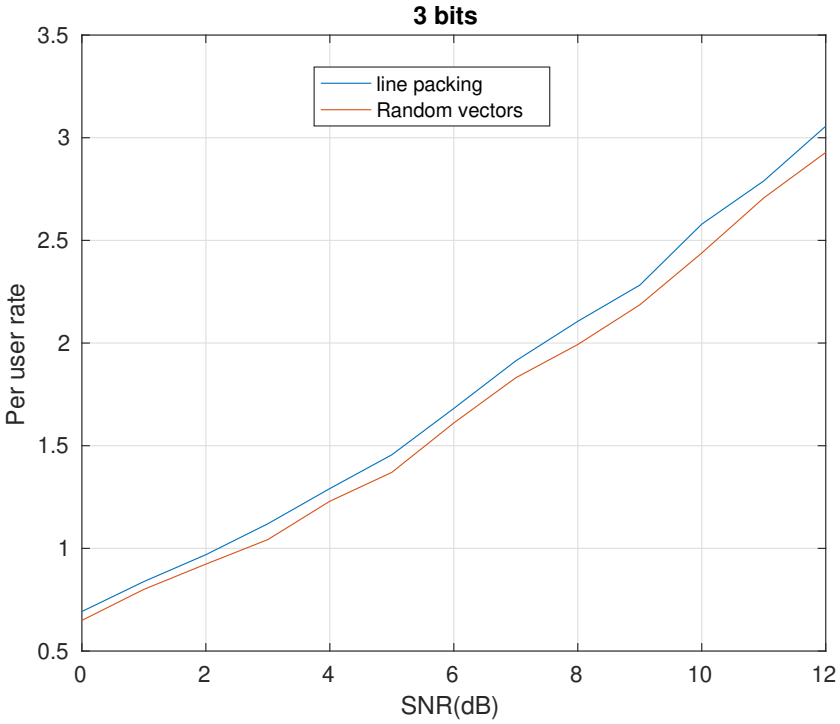

FIGURE 3.3: Line packing vs RVQ in case of 3 bits/user

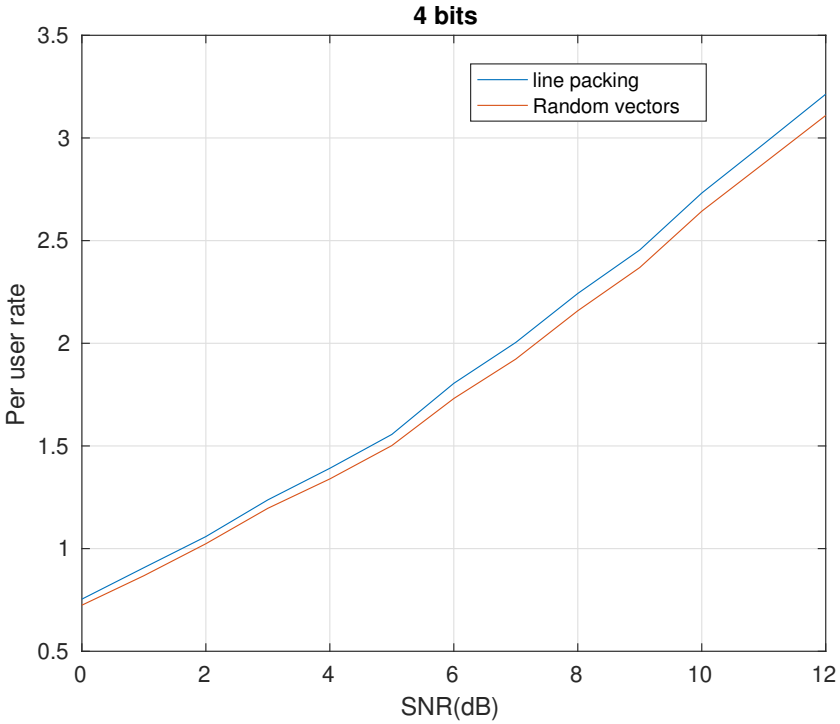

FIGURE 3.4: Line packing vs RVQ in case of 4 bits/user

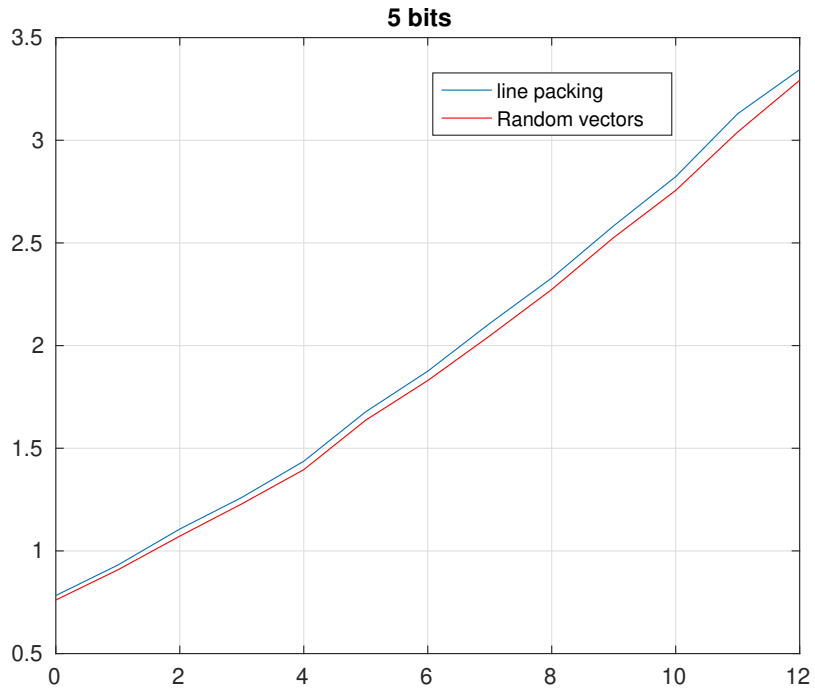

FIGURE 3.5: Line packing vs RVQ in case of 5 bits/user

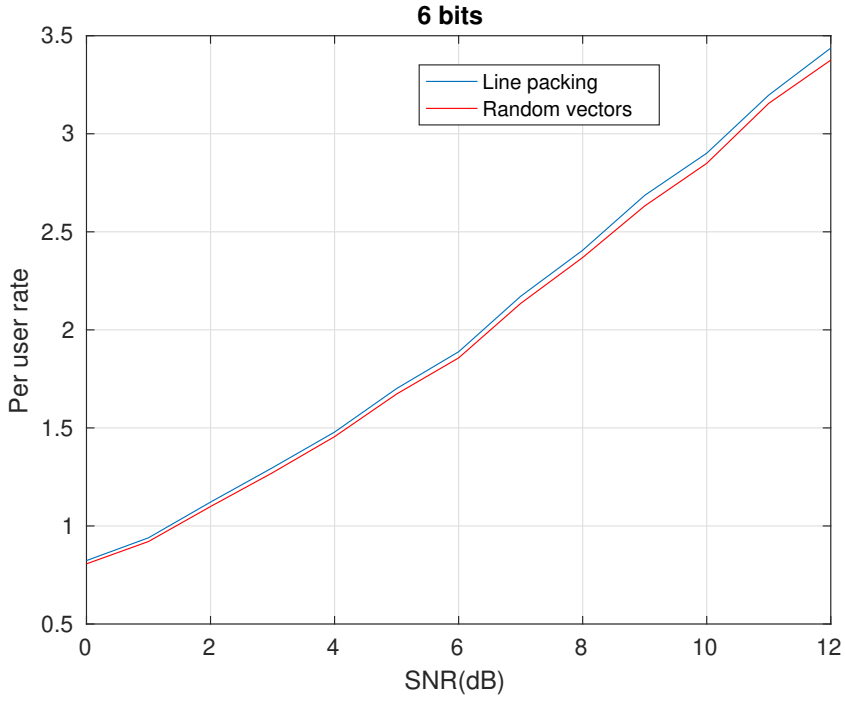

FIGURE 3.6: Line packing vs RVQ in case of 6 bits/user

Chapter 4

Channel Feedback in Massive MIMO Systems with Multi-Antenna Users

[29]

4.1 System Model

4.1.1 Downlink Massive MIMO Channel Model

In this chapter, we assume a mmWave massive MIMO broadcast (downlink) system with a single BS communicating with K multi-antenna users. We assume in this chapter the same system model discussed in 3.2.1, but the $k^{\text{th}}(\forall k \in \{1, 2, \cdots, K\})$ user has N_k receiving antennas. We consider a narrowband ray-based downlink channel model for the downlink channel matrices $\mathbf{H}_k \in \mathbb{C}^{N_k \times M}$ at the k^{th} user as in Chapter (3)

$$\mathbf{H}_k = \mathbf{G}_k \mathbf{A}_k(\phi_{k,1}, \phi_{k,2}, \cdots, \phi_{k,P_k}), \tag{4.1}$$

where $\mathbf{A}_k$ is the spatial direction, and $\mathbf{G}_k \in \mathbb{C}^{N_k \times P_k}$ is the complex path gains matrix.

During the angle coherence time of $\phi_{k,i}$, the channel vector of each antenna of user k is only distributed in a P_k-dimensional subspace, called as

the channel subspace in this thesis, of the full M-dimensional space. We assume throughout the thesis that the channel subspace $\mathbf{A}_k$, which is a function of the AoDs, is known at both user k and the BS. The AoDs can be estimated at the k^{th} user using the standard Multiple Signal Classification (MUSIC) algorithm [30], then they are fed back to the BS once after every angle coherence time. Consequently, the BS only needs to know the low dimensional path gains matrix $\mathbf{G}_k \in \mathbb{C}^{N_k \times P_k}$ in order to generate the actual channel matrix $\mathbf{H}_k$. In this thesis, we neglect the overhead coming from reporting the AoDs to the BS because it is very low compared to the overhead coming from reporting the path gains in $\mathbf{G}_k$.

The BS sends m_k streams to user k, where $m_k \leq N_k$. Let $\mathbf{u}_k \in \mathbb{C}^{m_k \times 1}$ contains the m_k data symbols to be transmitted simultaneously to the k^{th} user such that

$$\mathbf{u}_k = [u_{k,1} u_{k,2} \cdots u_{k,m_k}]^T. \tag{4.2}$$

Before transmitting the users' data symbols over the channel, the k^{th} user symbol vector is multiplied by the precoding matrix $\mathbf{F}_k \in \mathbb{C}^{M \times m_k}$. Thus, the overall transmitted vector $\mathbf{x} \in \mathbb{C}^{M \times 1}$, which contains all the data symbols intended for all users, is given by:

$$\mathbf{x} = \sum_{j=1}^{K} \mathbf{F}_j \mathbf{u}_j \tag{4.3}$$

and the received signal at the k^{th} user can be written as:

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x} + \mathbf{n}_k = \mathbf{H}_k \mathbf{F}_k \mathbf{u}_k + \mathbf{H}_k \sum_{\substack{j=1 \\ j \neq k}}^{K} \mathbf{F}_j \mathbf{u}_j + \mathbf{n}_k, \tag{4.4}$$

where $\mathbf{n}_k \in \mathbb{C}^{N_k \times 1}$ is the circularly symmetric complex Gaussian noise vector at the k^{th} user with a zero vector mean and identity covariance matrix.

The second term in Eq. (4.4) represents the summation of the interference, from the signals intended to all other users in the cell, at user k. The users' precoding matrices, $\mathbf{F}_k$'s, are unitary matrices (i.e., $\mathbf{F}_k^{\text{H}} \mathbf{F}_k = \mathbf{I}_{m_k}$), and in order to adhere to the power constraint, we have $E\left[\|\mathbf{u}_k\|^2\right] = \frac{\gamma}{K}, \forall k \in$

$\{1, 2, \cdots, K\}$, where γ is the total transmit power at the BS.

4.1.2 Partial CSI Feedback

The training overhead to perform channel estimation at the receiver side increases in massive MIMO systems as the number of transmit antennas at the BS increases [1]. However, there are many effective downlink channel estimation schemes that address this problem with a highly reduced amount of training overhead [15], [31] and [16]. Consequently, we assume in this thesis that each user knows its downlink channel matrix.

The channel matrix $\mathbf{H}_k$ of each user is required at the BS in order to perform precoding and power allocation. However, we assume in this chapter that the total power of each user is uniformly allocated across its multiple data streams. Hence, in order to perform BD, which will be discussed thoroughly in Sec. 4.2.1, it is only required to feed back the spatial direction of each user's effective channel. The spatial direction of the k^{th} user is defined as the subspace spanned by the rows of $\widetilde{\mathbf{H}}_k \in \mathbb{C}^{m_k \times M}$, where $\widetilde{\mathbf{H}}_k$ represents the subspace of the effective channel of user k. In case of $m_k = N_k$, the spatial direction of the k^{th} user is the subspace spanned by the rows of its channel matrix itself $\mathbf{H}_k \in \mathbb{C}^{N_k \times M}$. The quantization of the spatial direction $\widetilde{\mathbf{H}}_k$, say $\widehat{\mathbf{H}}_k \in \mathbb{C}^{m_k \times M}$, is chosen from the codebook $\mathcal{C}_k = \{\mathbf{C}_{k,1}, \mathbf{C}_{k,2}, \cdots, \mathbf{C}_{k,2^B}\}$, that consists of 2^B matrices in $\mathbb{C}^{m_k \times M}$, where B is the number of feedback bits for each user and the rows of $\mathbf{C}_{k,i}$ are orthonormal. The details of the beamforming matrix design as well as the codebook design are discussed in Sec. 4.2 and Sec. 4.3, respectively. The k^{th} user quantizes its spatial direction $\widetilde{\mathbf{H}}_k$ to a quantization subspace $\widehat{\mathbf{H}}_k = \mathbf{C}_{k,Z_k}$, where the index Z_k is calculated such that:

$$Z_k = \underset{i \in [1,2^B]}{\arg\min}\, d^2(\widetilde{\mathbf{H}}_k, \mathbf{C}_{k,i}), \tag{4.5}$$

where $d(\mathbf{H}_k, \mathbf{C}_{k,i})$ is the distance metric between the two matrices $\mathbf{H}_k$ and $\mathbf{C}_{k,i}$. In this chapter, we adopt the chordal distance discussed in chapter (3) as our distance metric, which is given by Equ. (3.11). Note that we do not

feed back any channel magnitude information to the BS because we need to reduce the feedback overhead amount. Hence, we assume uniform power allocation across the data streams in this chapter.

4.2 Design of BD based beamforming matrices

In this section, we present the details of the block diagonalization (BD) precoding scheme. Then, we analyze the per-user data rates of the BD scheme.

4.2.1 Design of Users' Beamforming Matrices

In this chapter, we consider BD as our linear BS precoding technique. BD is a zero-forcing technique which completely nulls the interference at each user due to the signals transmitted to all other users. Thus, BD can be thought of as a generalization of channel inversion in cases of multiple antennas per user. Following the BD algorithm, each $\mathbf{F}_k$ is chosen under the constraint of having $\mathbf{H}_j\mathbf{F}_k = \mathbf{0}$, $\forall j \neq k$. This leads to obtaining an orthonormal basis for the null space of the matrix formed by stacking all $\{\mathbf{H}_j\}_{j\neq k}$ matrices. This procedure nulls the interference terms in Eq. (4.4) at each user. BD is different from the conventional Zero-Forcing (ZF) precoding, where every complex data symbol to be transmitted to the n^{th} antenna (among the N_k antennas) of the k^{th} user is precoded by a vector which is orthogonal to all the rows of $\mathbf{H}_j, j \neq k$, and is orthogonal to all rows of $\mathbf{H}_k$ except the n^{th} one. In other words, conventional ZF forces every transmitted data symbol to be received by only one antenna at the intended user. This results in more restrictions in designing the BS precoders and results in a degraded performance if compared to BD based precoders design.

However, in practice, we cannot achieve zero interference as the BS does not have perfect knowledge of $\{\mathbf{H}_k\}_{k=1}^{K}$. In the case of limited feedback, BS has access to a quantized version of the subspace spanned by the rows of each $\mathbf{H}_k$, namely $\widehat{\mathbf{H}}_k$. We follow the strategy in [32], where the BS treats the quantized subspaces $\widehat{\mathbf{H}}_1, \widehat{\mathbf{H}}_2, \cdots, \widehat{\mathbf{H}}_K$ as the true channel subspaces while

performing the BD procedure. In that case, we denote the generated precoding matrices as $\widehat{\mathbf{F}}_1, \widehat{\mathbf{F}}_2, \cdots, \widehat{\mathbf{F}}_K$ in order to distinguish them from those selected with perfect channel knowledge at the BS.

We assume in this chapter that the number of antennas of the k^{th} user, N_k, is smaller than the number of resolvable paths P_k, (i.e., $N_k < P_k$). Thus, all antennas of user k are independent from each other since they experience P_k independent paths with independent path gains (i.e., entries of $\mathbf{G}_k$ are independent). We consider two different cases when designing the precoding matrices $\widehat{\mathbf{F}}_k,\ \forall k \in \{1, 2, \cdots, K\}$ as follows.

Case I: $N_k = m_k$

In this case, it is assumed that the number of antennas of the k^{th} user, N_k, is equal to the number of complex data symbols m_k to be simultaneously transmitted to it. Define $\mathbf{W}_k$ as

$$\mathbf{W}_k = \left[\widehat{\mathbf{H}}_1^T \cdots \widehat{\mathbf{H}}_{k-1}^T \widehat{\mathbf{H}}_{k+1}^T \cdots \widehat{\mathbf{H}}_K^T\right]^T, \tag{4.6}$$

where $\widehat{\mathbf{H}}_k,\ k \in \{1, 2, \cdots, K\}$, is the quantized feedback version of the original spatial direction $\widetilde{\mathbf{H}}_k$ of the k^{th} user. The zero-interference constraint forces the precoding matrix $\widehat{\mathbf{F}}_k$ of the k^{th} user to lie in the null space of $\mathbf{W}_k$. The channel subspace of the k^{th} user $\mathbf{A}_k$ only depends on the AoDs of the user which are assumed to be independent from one user to another. Thus, we can conclude that the spatial directions of different users $\widehat{\mathbf{H}}_k$ are linearly independent from each other. Consequently, the rank of $\mathbf{W}_k$ of the k^{th} user is $\widetilde{L}_k = \text{rank}(\mathbf{W}_k) = N_R - N_k$, where N_R is the aggregate number of receive antennas (i.e., $N_R = \sum_{k=1}^{K} N_k$). Define the Singular Value Decomposition (SVD) of $\mathbf{W}_k$ as

$$\mathbf{W}_k = \mathbf{U}_k \mathbf{\Sigma_k} \left[\mathbf{V_k^{(1)}}\ \ \mathbf{V_k^{(0)}}\right]^{\text{H}}, \tag{4.7}$$

where $\mathbf{V}_k^{(1)}$ holds the first $\widetilde{L}_k$ right singular vectors, while $\mathbf{V}_k^{(0)}$ have the remaining $(M - \widetilde{L}_k)$ right singular vectors. Hence, $\mathbf{V}_k^{(0)}$ forms an orthonormal basis for the null space of $\mathbf{W}_k$, and therefore, its columns are candidates for the columns of the k^{th} user precoding matrix, $\widehat{\mathbf{F}}_k$.

The effective channel of the k^{th} user is the product $\widehat{\mathbf{H}}_k \mathbf{V}_k^{(0)}$. Due to nulling the interference of other users, this is now equivalent to the single-user MIMO capacity maximization problem and the best precoder is, thus, the right singular vectors of that effective channel [33]. Let $\bar{L}_k$ be the rank of the product $\widehat{\mathbf{H}}_k \mathbf{V}_k^{(0)}$ and it is upper bounded by $\bar{L}_k \leq min\{L_k, \widetilde{L}_k\}$, where L_k is the rank of $\widehat{\mathbf{H}}_k$. Thus, the SVD of the effective channel of the k^{th} user is given by:

$$\widehat{\mathbf{H}}_k \mathbf{V}_k^{(0)} = \mathbf{Q}_k \begin{bmatrix} \mathbf{\Lambda_k} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{R}_k^{(1)} & \mathbf{R}_k^{(0)} \end{bmatrix}^{\text{H}}, \tag{4.8}$$

where $\mathbf{\Lambda_k}$ is $\bar{L}_k \times \bar{L}_k$ and the columns of $\mathbf{R}_k^{(1)}$ are the first $\bar{L}_k$ singular vectors. Finally, the product $\mathbf{V}_k^{(0)} \mathbf{R}_k^{(1)}$ forms an orthonormal basis of dimension $\bar{L}_k$, and it represents the precoding matrix that maximizes the capacity of the k^{th} user while achieving zero interference.

$$\widehat{\mathbf{F}}_k = \mathbf{V}_k^{(0)} \mathbf{R}_k^{(1)}. \tag{4.9}$$

Case II: $N_k > m_k$

In this case, it is assumed that the number of antennas of the k^{th} user, N_k, is larger than the number of complex data symbols, m_k, to be simultaneously transmitted to that user. Adding more antennas at each receiver enhances the diversity gain at each user. In addition, having more antennas at the users than the number of data streams means that we only feed back a smaller subspace of the right singular vectors of the channel matrix $\mathbf{H}_k \in \mathbb{C}^{N_k \times M}$ of user k. This, in turn, enhances the capacity of the system. Let the SVD of the channel matrix $\mathbf{H}_k$ of the k^{th} user be:

$$\mathbf{H}_k = \mathbf{U}_k \Sigma_k \mathbf{V}_k^{\text{H}}, \tag{4.10}$$

where $\mathbf{U}_k \in \mathbb{C}^{N_k \times N_k}$ and $\mathbf{V}_k \in \mathbb{C}^{M \times M}$ are unitary matrices, and $\Sigma_k \in \mathbb{C}^{N_k \times M}$ is a rectangular matrix that has the singular values on its diagonal. Let $\mathbf{V}_{k,m_k}$ be a matrix that contains the first m_k columns of $\mathbf{V}_k$. From Eq.

(4.10), we can notice that each row of $\mathbf{H}_k$ is a linear combination of the complex conjugate of the first N_k columns of $\mathbf{V}_k$. Thus, the subspace spanned by the first N_k columns of $\mathbf{V}_k$ is equivalent to the subspace spanned by the complex conjugate of the rows of the channel matrix $\mathbf{H}_k \in \mathbb{C}^{N_k \times M}$. Consequently, we can conclude that the subspace spanned by the first N_k columns of $\mathbf{V}_k$ always lies in the subspace spanned by the rows of $\mathbf{A}_k^* \in \mathbb{C}^{P_k \times M}$. This is important since $\mathbf{A}_k$ is assumed to be already known at the BS. Then, we can use a low dimensional codebook, to be designed in Sec. 4.3, in order to quantize $\mathbf{V}_{k,m_k}$. It was proved in [34] that the columns of $\mathbf{V}_{k,m_k}$ are isotropically distributed on the subspace they lie in. Hence, a Grassmannian packing based codebook, to be presented in Sec. 3.2.2, can be used to quantize $\mathbf{V}_{k,m_k}$. Let the quantized version of $\mathbf{V}_{k,m_k}$ be $\widehat{\mathbf{V}}_{k,m_k} \in \mathbb{C}^{M \times m_k}$, and it is chosen from the codebook $\mathcal{C}$ according to Eq. (4.5).

Now, following the conventional BD procedure, let $\mathbf{S}_k \in \mathbb{C}^{M \times (M - \sum_{i=1, i \neq k}^{K} m_i)}$ represent the orthonormal basis of the null space of $\mathbf{W}_k$, where

$$\mathbf{W}_k = \left[\widehat{\mathbf{V}}_{1,m_1} \cdots \widehat{\mathbf{V}}_{k-1,m_{k-1}} \widehat{\mathbf{V}}_{k+1,m_{k+1}} \cdots \widehat{\mathbf{V}}_{K,m_K} \right]^{\mathrm{H}}. \tag{4.11}$$

The effective channel of the k^{th} user will be the product $\widehat{\mathbf{V}}_{k,m_k}^{\mathrm{H}} \mathbf{S}_k$. The SVD of this product is given by:

$$\widehat{\mathbf{V}}_{k,m_k}^{\mathrm{H}} \mathbf{S}_k = \mathbf{Q}_k \mathbf{\Lambda}_\mathbf{k} \left[\mathbf{R}_\mathbf{k}^{(\mathbf{1})} \;\; \mathbf{R}_\mathbf{k}^{(\mathbf{0})} \right]^{\mathrm{H}}, \tag{4.12}$$

where $\mathbf{R}_k^{(1)}$ represents the first m_k right singular vectors. Finally, the product $\mathbf{S}_k \mathbf{R}_k^{(1)}$ form an orthonormal basis of dimension m_k, and it represents the precoding matrix $\widehat{\mathbf{F}}_k \in \mathbb{C}^{M \times m_k}$ that maximizes the capacity of the k^{th} user while achieving zero interference. The precoding matrix $\widehat{\mathbf{F}}_k$ is given by

$$\widehat{\mathbf{F}}_k = \mathbf{S}_k \mathbf{R}_k^{(1)}. \tag{4.13}$$

Hence, the received vector $\mathbf{y}_k \in \mathbb{C}^{N_k \times 1}$ at user k becomes

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{F}_k \mathbf{u}_k + \sum_{j=1, j\neq k}^{K} \mathbf{H}_k \widehat{\mathbf{F}}_j \mathbf{u}_j + \mathbf{n}_k. \tag{4.14}$$

The received vector $\mathbf{y}_k$, in Eq. (4.14), is finally left multiplied by $\mathbf{U}_{k,m_k}^{\mathrm{H}}$, where $\mathbf{U}_{k,m_k} \in \mathbb{C}^{N_k \times m_k}$ is the matrix that contains the first m_k columns of the matrix $\mathbf{U}_k$ given in Eq. (4.10).

4.2.2 The Per-User Rate

The BS can perform downlink precoding on the data vectors $\mathbf{u}_k \in \mathbb{C}^{m_k \times 1}$ intended for each user based on the fed back quantized spatial directions represented by $\widehat{\mathbf{H}}_k$. As described above, we consider the BD based linear precoding at the BS to obtain the beamforming matrices for each user $\mathbf{F}_k$. The BD strategy involves linear precoding that eliminates the interference at each user due to all other users as discussed in Sec. 4.2.1. Hence, the second term in Eq. (4.4), which represents the interference at the k^{th} user due to all other users, is canceled in the case of perfect CSI at the BS (i.e., $\widehat{\mathbf{H}}_k \equiv \widetilde{\mathbf{H}}_k$). Then, the per-user ergodic rate for case I is given by [35]:

$$R_{\text{CSIT,I}}(\gamma) = \mathbb{E} \log_2 \left| \mathbf{I}_{m_k} + \frac{\gamma}{K m_k} \mathbf{H}_k \mathbf{F}_k \mathbf{F}_k^{\mathrm{H}} \mathbf{H}_k^{\mathrm{H}} \right|. \tag{4.15}$$

For case II, the total effective channel after left multiplying Eq. (4.14) by $\mathbf{U}_{k,m_k}^{\mathrm{H}}$ is $\mathbf{U}_{k,m_k}^{\mathrm{H}} \mathbf{H}_k \mathbf{F}_k$. Hence the per-user ergodic rate for case II is given by:

$$R_{\text{CSIT,II}}(\gamma) = \mathbb{E} \log_2 \left| \mathbf{I}_{m_k} + \frac{\gamma}{K m_k} \mathbf{U}_{k,m_k}^{\mathrm{H}} \mathbf{H}_k \mathbf{F}_k \mathbf{F}_k^{\mathrm{H}} \mathbf{H}_k^{\mathrm{H}} \mathbf{U}_{k,m_k} \right|, \tag{4.16}$$

where k is the user index, and a uniform power allocation policy is adopted. The expectation is evaluated over the distribution of the channel matrix, $\mathbf{H}_k$.

In the case of limited feedback of B bits for each user, the interference at the k^{th} user due to all other users cannot be completely eliminated because the quantized spatial direction spanned by the rows of $\widehat{\mathbf{H}}_k$ is not exactly the same as the original spatial direction spanned by the rows of $\widetilde{\mathbf{H}}_k$. As a

result, this quantization leads to residual interference power, and the per-user rate for case I is given by [32]:

$$R_{\text{QUANT,I}}(\gamma) = \mathbb{E}\log_2\left|\mathbf{I}_{m_k} + \frac{\gamma}{Km_k}\sum_{j=1}^{K}\mathbf{H}_k\widehat{\mathbf{F}}_j\widehat{\mathbf{F}}_j^{\text{H}}\mathbf{H}_k^{\text{H}}\right| - \mathbb{E}\log_2\left|\mathbf{I}_{m_k} + \frac{\gamma}{Km_k}\sum_{\substack{j=1\\ j\neq k}}^{K}\mathbf{H}_k\widehat{\mathbf{F}}_j\widehat{\mathbf{F}}_j^{\text{H}}\mathbf{H}_k^{\text{H}}\right|. \tag{4.17}$$

Similarly, the per-user rate for case II due to quantization is given by:

$$R_{\text{QUANT,II}}(\gamma) = \mathbb{E}\log_2\left|\mathbf{I}_{m_k} + \frac{\gamma}{Km_k}\sum_{j=1}^{K}\mathbf{U}_{k,m_k}^{\text{H}}\mathbf{H}_k\widehat{\mathbf{F}}_j\widehat{\mathbf{F}}_j^{\text{H}}\mathbf{H}_k^{\text{H}}\mathbf{U}_{k,m_k}\right| - \mathbb{E}\log_2\left|\mathbf{I}_{m_k} + \frac{\gamma}{Km_k}\sum_{\substack{j=1\\ j\neq k}}^{K}\mathbf{U}_{k,m_k}^{\text{H}}\mathbf{H}_k\widehat{\mathbf{F}}_j\widehat{\mathbf{F}}_j^{\text{H}}\mathbf{H}_k^{\text{H}}\mathbf{U}_{k,m_k}\right|, \tag{4.18}$$

where k is the user index, and the expectation is evaluated over the distribution of the channel matrices, $\mathbf{H}_k$ $\forall k \in \{1, 2, \cdots, K\}$, and the corresponding quantized precoding matrices, $\widehat{\mathbf{F}}_j$. The term $\mathbf{H}_k\widehat{\mathbf{F}}_k\widehat{\mathbf{F}}_k^{\text{H}}\mathbf{H}_k^{\text{H}}$ represents the useful signal intended for user k and, $\sum_{j=1,j\neq k}^{K}\mathbf{H}_k\widehat{\mathbf{F}}_j\widehat{\mathbf{F}}_j^{\text{H}}\mathbf{H}_k^{\text{H}}$ represents the muti-user interference at user k.

4.3 AoD-adaptive Subspace Codebook

The path angles of departure of the k^{th} user, $\phi_{k,i}$s, defined in Eq. (4.1) depend on the obstacles that surround the BS. These obstacles are expected to change their physical positions in a much longer time than the channel coherence time. On the other hand, for the path gains represented by $\mathbf{G}_k$ in Eq. (4.1), one resolvable path is formed by a set of scatters around user k, which consists of a number of unresolvable paths. Hence, path gains, $\mathbf{G}_k$'s, are expected to change much faster than path AoDs, $\phi_{k,i}$'s [25], [36]. However, the size of $\mathbf{G}_k$ is very low compared to the original channel matrix $\mathbf{H}_k$ and here comes the reuction in feedback overhead. During the angle coherence time, the spatial direction of the k^{th} user $\widetilde{\mathbf{H}}_k$ is isotropically distributed in the channel subspace, which is spanned by the rows of $\mathbf{A}_k(\phi_{k,1}, \phi_{k,2}, \cdots, \phi_{k,P_k})$.

As shown in Eq. (4.1), each row of the channel matrix $\mathbf{H}_k$ is composed of P_k paths, where $\mathbf{A}_k(\phi_{k,1}, \phi_{k,2}, \cdots, \phi_{k,P_k})$ is completely determined by the path AoDs. The reason for the uniform distribution of $\widetilde{\mathbf{H}}_k$ in the row space of $\mathbf{A}_k$ is that the rows of $\mathbf{A}_k$ (steering vectors) are asymptotically orthogonal to each other (i.e., $\mathbf{A}_k\mathbf{A}_k^{\mathrm{H}} \approx M\mathbf{I}_{P_k}$) [25]. Additionally, the path gains in $\mathbf{G}_k$ are modeled as i.i.d. circularly symmetric complex Gaussian random variables with zero mean and unit variance, which causes the user's spatial direction to be uniformly distributed in its channel subspace during the angle coherence time.

Due to limited scattering of mmWave, the number of paths P_k is much smaller than the number of transmit antennas M at the BS [37]. Therefore, the row space of $\mathbf{A}_k$ is only a subspace of the full M-dimensional space. Thus, assuming that the BS knows the AoDs, we can only quantize and feed back the path gains matrix $\mathbf{G}_k \in \mathbb{C}^{N_k \times P_k}$. Then, the quantization subspace $\mathbf{C}_{k,i}$ of the proposed AoD-adaptive subspace codebook $\mathcal{C} = \{\mathbf{C}_{k,1}, \mathbf{C}_{k,2}, \cdots, \mathbf{C}_{k,2^B}\}$ is formed by:

$$\mathbf{C}_{k,i} = \frac{1}{\sqrt{M}}\mathbf{X}_i\mathbf{A}_k, \tag{4.19}$$

where $\mathbf{X}_i \in \mathbb{C}^{m_k \times P_k}$ is a matrix whose rows are orthonormal, and its row space is isotropically distributed over the complex P_k-dimensional space.

4.3.1 Subspace Quantization codebooks

We use random subspace quantization as well as subspace packing over the Grassmannian manifold to quantize the path gains matrix $\mathbf{G}_k$. The random quantization is used to analyse the system by averaging over all the generated codebooks, however we use subspace packing quantization in order to design structured codebooks to be used in practical systems. For the random subspace quantization in our problem, a number of 2^B subspaces, each having a dimension of m_k, are picked at random in a P_k-dimensional Euclidean space. The set of all m_k-dimensional subspaces in a P_k-dimensional space represent a Grassmannian manifold, which is denoted by $\mathcal{G}_{P_k,m_k}$. The

2^B random subspaces, that form the random quantization codebook, are uniformly distributed over $\mathcal{G}_{P_k,m_k}$.

To obtain structured codebooks that can be used in practical systems, the design of the quantized path gain matrices $\mathbf{X}_i$ is done using subspace packing in Grassmannian manifold. The packing problem tends to find 2^B subspaces in a higher dimensional space such that the minimum distance between two subspaces is maximized. There are many distance metrics that have been used for packing subspaces in the Grassmannian manifold. In this chapter, we adopt the chordal distance, defined in **??**, as our distance metric. The codebook design is done by solving the packing problem of 2^B m_k-dimensional subspaces in a complex Euclidean space of dimensionality P_k. We follow the iterative algorithm discussed in Sec. (3.2.2) in order to solve the subspace packing problem. The solution of this problem is usually simpler when the number of subspaces in the codebook 2^B is lower than P_k^2. In that case, the minimum distance between two subspaces in the codebook can reach the Rankin bound [28], which is the maximum attainable theoretical distance.

4.4 Throughput analysis

In this section, we calculate the rate gap between the ideal rate and the rate using a random subspace quantization scheme. We study the rate gap for case I assuming that all users have the same number of receive antennas (i.e., $N_k = m_k = N$) and same number of resolvable paths (i.e., $P_k = P$). We derive an expression for the required number of feedback bits to achieve some constant rate gap, where we prove that the number of bits scales linearly with the transmit power γ_{dB} in dB.

4.4.1 Rate Gap

The per-user rate of the ideal case of case I is given by Eq. (4.15), and the per-user rate of the practical case of case I is given by Eq. (5.8). Following

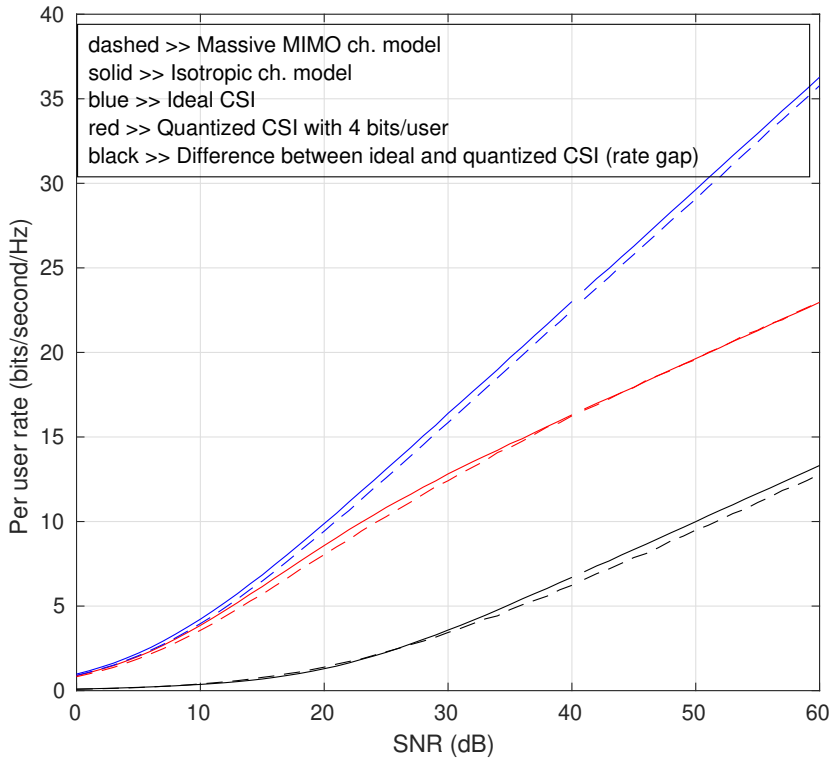

FIGURE 4.1: Simulation of the rate gap for both isotropically distributed channel and actual massive MIMO channel model using 4 bits/user with $N = 2, K = 8, M = 128$ and $P = 3$

Theorem 1 of [32], which gives an upper bound for the rate gap in Multi-User MIMO systems, we derive an expression for the per-user rate gap due to limited feedback in our massive MIMO system model. We make an approximation that the subspace spanned by the rows of the channel matrix of the k^{th} user $\mathbf{H}_k$ is isotropically distributed in the M-dimensional complex space. The channel subspace $\mathbf{A}_k$ has a certain structure based on the AoD's which make it hard to follow its exact distribution over the complex space. The following figures show that this approximation is valid as the rate gap of the isotropic distribution case is a bit larger than the rate gap of the actual distribution case of the massive MIMO channel model. Hence this approximation can be taken as an upper bound for the actual rate gap.

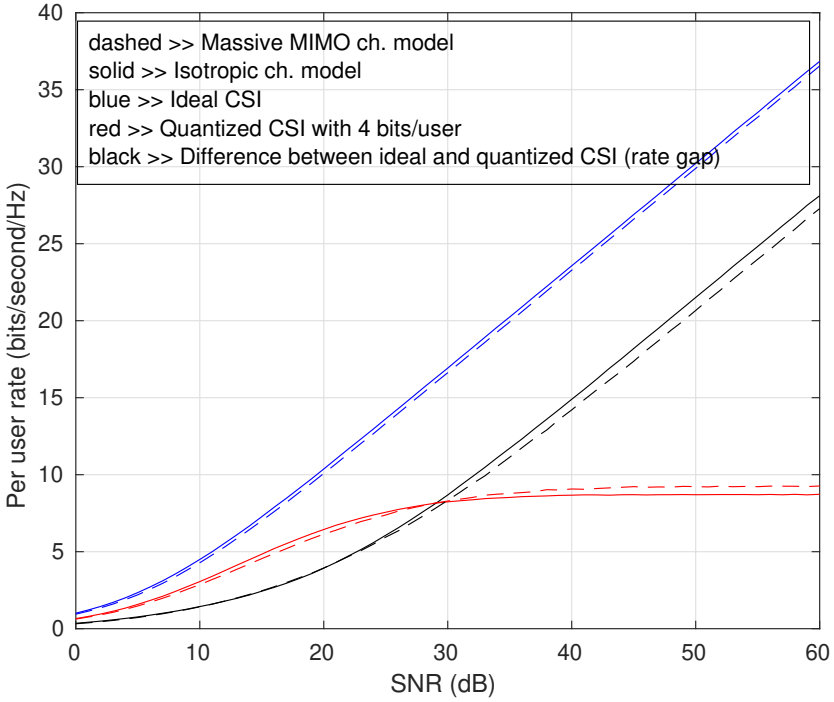

FIGURE 4.2: Simulation of the rate gap for both isotropically distributed channel and actual massive MIMO channel model using 4 bits/user with $N = 2, K = 8, M = 128$ and $P = 5$

The per-user rate gap $\Delta R(\gamma)$ is bounded as follows:

$$\Delta R(\gamma) = R_{\text{CSIT,I}}(\gamma) - R_{\text{QUANT,I}}(\gamma) \tag{4.20}$$

$$\begin{aligned} &\overset{(a)}{\leq} \mathbb{E}\log_2\left|\mathbf{I}_N + \frac{\gamma}{KN}\mathbf{H}_k\mathbf{F}_k\mathbf{F}_k^{\mathrm{H}}\mathbf{H}_k^{\mathrm{H}}\right| - \\ &\quad \mathbb{E}\log_2\left|\mathbf{I}_N + \frac{\gamma}{KN}\mathbf{H}_k\widehat{\mathbf{F}}_k\widehat{\mathbf{F}}_k^{\mathrm{H}}\mathbf{H}_k^{\mathrm{H}}\right| + \\ &\quad \mathbb{E}\log_2\left|\mathbf{I}_N + \frac{\gamma}{KN}\sum_{\substack{j=1\\ j\neq k}}^{K}\mathbf{H}_k\widehat{\mathbf{F}}_j\widehat{\mathbf{F}}_j^{\mathrm{H}}\mathbf{H}_k^{\mathrm{H}}\right| \end{aligned} \tag{4.21}$$

$$\overset{(b)}{=} \mathbb{E}\log_2\left|\mathbf{I}_N + \frac{\gamma}{KN}\sum_{\substack{j=1\\ j\neq k}}^{K}\mathbf{H}_k\widehat{\mathbf{F}}_j\widehat{\mathbf{F}}_j^{\mathrm{H}}\mathbf{H}_k^{\mathrm{H}}\right| \tag{4.22}$$

$$\overset{(c)}{=} \mathbb{E}\log_2\left|\mathbf{I}_N + \frac{\gamma}{KN}\widetilde{\mathbf{H}}_k\left(\sum_{j\neq k}\widehat{\mathbf{F}}_j\widehat{\mathbf{F}}_j^{\mathrm{H}}\right)\widetilde{\mathbf{H}}_k^{\mathrm{H}}\beta_k\right| \tag{4.23}$$

$$\overset{(d)}{\leq} \log_2\left|\mathbf{I}_N + \frac{\gamma MP}{K}(K-1)\mathbb{E}\left[\widetilde{\mathbf{H}}_k\widehat{\mathbf{F}}_j\widehat{\mathbf{F}}_j^{\mathrm{H}}\widetilde{\mathbf{H}}_k^{\mathrm{H}}\right]\right| \tag{4.24}$$

Here, (a) follows by neglecting the positive semi-definite interference terms in the quantity:

$$\mathbb{E}\log_2\left|\mathbf{I}_N+\frac{\gamma}{KN}\sum_{j=1}^{K}\mathbf{H}_k\widehat{\mathbf{F}}_j\widehat{\mathbf{F}}_j^{\mathrm{H}}\mathbf{H}_k^{\mathrm{H}}\right|. \tag{4.25}$$

Following the BD procedure, both $\mathbf{F}_k$ and $\widehat{\mathbf{F}}_k$ are isotropically distributed according to our approximation, and they are chosen independent of $\mathbf{H}_k$. This means that the first two terms in Equ. (4.21) are the same and hence gives (b). By writing $\mathbf{H}_k^H\mathbf{H}_k = \widetilde{\mathbf{H}}_k^H\beta_k\widetilde{\mathbf{H}}_k$, where the rows of $\widetilde{\mathbf{H}}_k \in \mathbb{C}^{N\times M}$ forms an orthonormal basis for the subspace spanned by the rows of $\mathbf{H}_k$, and β_k are the N non-zero and unordered eigenvalues of $\mathbf{H}_k^H\mathbf{H}_k$. Step (c) follows using the fact that $|\mathbf{I}+\mathbf{AB}|=|\mathbf{I}+\mathbf{BA}|$ for matrices $\mathbf{A}$ and $\mathbf{B}$. Finally, (d) follows from Jensen's inequality due to the concavity of log, noting that $\mathbb{E}[\beta_k] = MPN\mathbf{I}_N$.

The value $\mathbb{E}\left[\widetilde{\mathbf{H}}_k\widehat{\mathbf{F}}_j\widehat{\mathbf{F}}_j^{\mathrm{H}}\widetilde{\mathbf{H}}_k^{\mathrm{H}}\right]$ was evaluated in [32] as the following equation

$$\mathbb{E}\left[\widetilde{\mathbf{H}}_k\widehat{\mathbf{F}}_j\widehat{\mathbf{F}}_j^{\mathrm{H}}\widetilde{\mathbf{H}}_k^{\mathrm{H}}\right] = \frac{D}{KP-N}\,, \tag{4.26}$$

where D is the average subspace quantization error which is given by:

$$D = \mathbb{E}\left[d^2(\widetilde{H}_k,\widehat{H}_k)\right], \tag{4.27}$$

and $d(\widetilde{H}_k,\widehat{H}_k)$ is the chordal distance defined in Eq. (**??**).

Hence, the rate gap can be upper bounded using the following equation

$$\Delta R(\gamma) \le N\log_2\left(1+\frac{\gamma(K-1)MP}{K(KP-N)}D\right). \tag{4.28}$$

4.4.2 Quantization Error

In this subsection, we calculate the quantization error, D, of the spatial direction of user k when the AoD-adaptive subspace codebook is used. We have $\mathbf{C}_{k,Z_k} = \frac{1}{\sqrt{M}}\mathbf{X}_{Z_k}\mathbf{A}_k$ and $\widetilde{\mathbf{H}}_k = \frac{1}{\sqrt{M}}\widetilde{\mathbf{G}}_k\mathbf{A}_k$; then, the quantization error

is given by

$$D = \mathbb{E}\left[N - \left\|\widetilde{\mathbf{H}}_k \mathbf{C}_{k,Z_k}^{\mathrm{H}}\right\|_{\mathrm{F}}^2\right] = \mathbb{E}\left[N - \left\|\frac{1}{M}\widetilde{\mathbf{G}}_k \mathbf{A}_k \mathbf{A}_k^{\mathrm{H}} \mathbf{X}_{Z_k}^{\mathrm{H}}\right\|_{\mathrm{F}}^2\right] \tag{4.29}$$

$$\overset{(e)}{\approx} \mathbb{E}\left[N - \left\|\widetilde{\mathbf{G}}_k \mathbf{X}_{Z_k}^{\mathrm{H}}\right\|_{\mathrm{F}}^2\right] \tag{4.30}$$

where $\widetilde{\mathbf{G}}_k \in \mathbb{C}^{N\times P}$ is a matrix whose rows are orthonormal and its row space represents the subspace spanned by the rows of $\widetilde{\mathbf{G}}_k$. Step (e) is true due to $\mathbf{A}_k \mathbf{A}_k^{\mathrm{H}} \approx M\mathbf{I}_P$. Both $\widetilde{\mathbf{G}}_k$ and $\mathbf{X}_{Z_k}$ are isotropically distributed subspaces on the P-dimensional space. Then, we can bound the quantization error as [32]:

$$D \leq \bar{D} = \frac{\Gamma(\frac{1}{T})}{T}(C_{PN})^{-\frac{1}{T}} 2^{-\frac{B}{T}}, \tag{4.31}$$

where $T = N(P-N)$ and $C_{PN} = \frac{1}{T!}\prod_{i=1}^{N}\frac{(P-i)!}{(N-i)!}$.

4.4.3 Feedback Bits

Now, we discuss the required number of feedback bits B that results in a constant rate gap. After bounding the quantization error by $\bar{D}$, the rate loss can be bounded as:

$$\Delta R(\gamma) \leq \log_2\left(1 + \frac{\gamma(K-1)MP}{K(KP-N)}\bar{D}\right). \tag{4.32}$$

Let the rate gap be such that $\Delta R(\gamma) \leq \log_2(b)$ bps/Hz, and substituting for $\bar{D}$ from Eq. (4.31), then the number of feedback bits that guarantees this rate loss is given by:

$$\begin{aligned} B = 3.3\, T\log_{10}(\gamma) - T\log_2\left[(b^{\frac{1}{N}} - 1)\frac{K(KP-N)}{(K-1)MP}\right] \\ + T\log_2\left(\frac{\Gamma(\frac{1}{T})}{T}\right) - \log_2(C_{PN}), \end{aligned} \tag{4.33}$$

where B scales linearly with the transmit power γ_{dB} in dB.

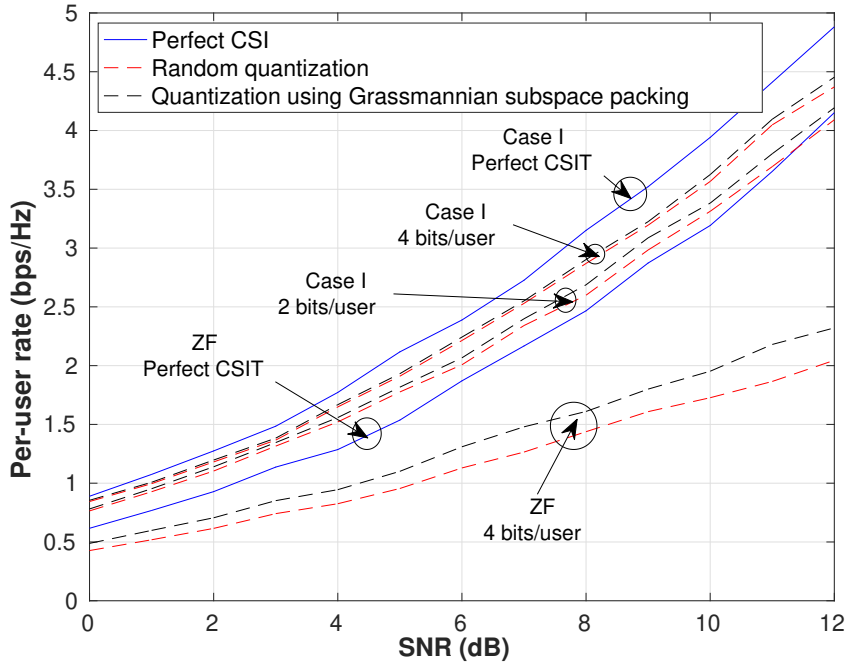

FIGURE 4.3: BD vs conventional ZF: case I with $N_k = m_k = 2$

4.5 Numerical Results

In this section, the performance of the proposed feedback system and codebook design is examined and verified. The system parameters are set as follows. The number of antennas at the BS is $M = 128$, the number of users in the system is $K = 8$, the number of antennas at each user is $N_k = N = 2$, the number of data streams transmitted simultaneously to each user is $m_k = 2$ and the number of resolvable paths is $P = 3$. The path AoDs of the users are independent and uniformly distributed in $\left[-\frac{1}{2}\pi, \frac{1}{2}\pi\right]$.

Fig. (4.3) compares the performance of BD and the conventional ZF scheme for case I with $N_k = m_k = 2$. Fig. (4.3) also compares the performance of the ideal case, where perfect CSI is assumed available at the BS, and the limited feedback case where quantized CSI is fed back to the BS with $B = 2$ and 4 per user. Note that in the case of conventional ZF scheme, the channel vector of each antenna at the k^{th} user is separately quantized and fed back to the BS; therefore, the feedback bits for each user are divided among its receive antennas in this case. This is because in the case of conventional ZF, any user antenna is used to receive a single stream

and all other streams must be nulled (even other streams intended for the same user), which is not the case for BD. In Fig. (4.3), we plot the per-user rate using the AoD-adaptive codebook with both random subspace quantization and using Grassmannian subspace packing based coodbook. From this figure, we can easily see the performance gains of the BD approach as compared to the conventional ZF approach. In addition, it can be noticed that Grassmannian codes are always better (or slightly better) than random codes. Note that Grassmannian codes are more structured, which deem them suitable for practical implementation, while random codes are impractical. Finally, it is clear that increasing the number of feedback bits enhances the system performance, and we can get arbitrary close to the performance of the ideal system with perfect CSI at the BS.

Fig. (4.4) compares the performance of BD with ideal and quantized CSI against the ideal and quantized CSI of the conventional ZF scheme for case II with $N_k = 3, m_k = 2$ with $B = 2$ and 4 per user. The same observations mentioned above while commenting on the results of Fig. (4.3) apply in this case as well. Moreover, it is noticeable that case II has higher per-user rate than case I for the same number of user streams. This is due to the fact that in case II we assume more receiving antennas at each user than the number of streams, which introduces diversity gain at the users.

In Fig. (4.5), we present numerical results for the practical per-user rate when using random quantization codebook. The required number of feedback bits is scaled as per Eq. (4.33) in order to guarantee a maximum rate gap of $\log_2(b)$, where we show the results for $b = 2$ and 4. We notice in Fig. (4.5) that the rate gap between the ideal (perfect CSI at the BS) and practical cases does not increase as the SNR increases; this is due to scaling the number of feedback bits B with the transmitted power γ_{dB} as explained above. It is clear that the rate gap at any SNR does not exceed the maximum value of $\log_2(b)$, which validates the expression in (4.33).

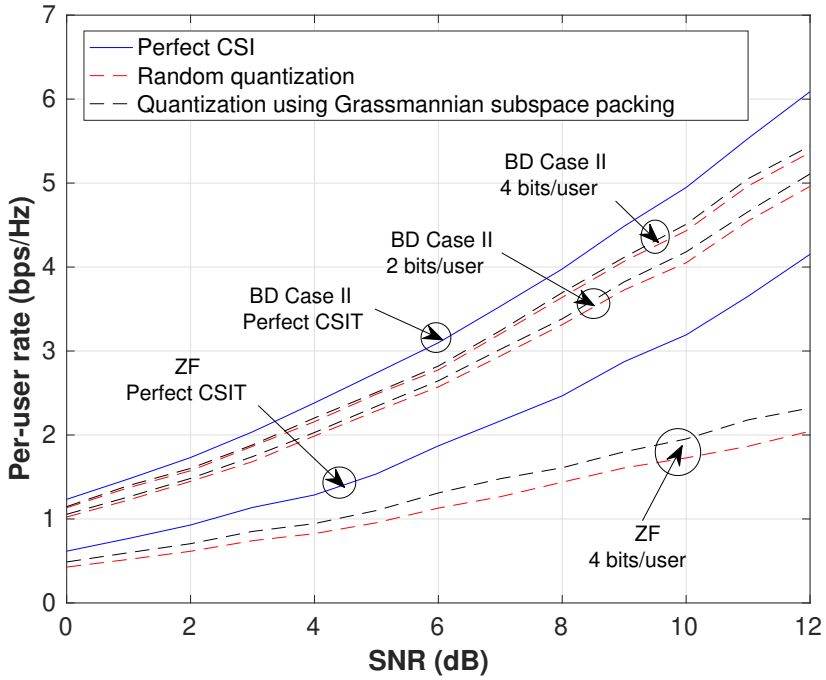

FIGURE 4.4: BD vs conventional ZF: case II with $N_k = 3$, $m_k = 2$

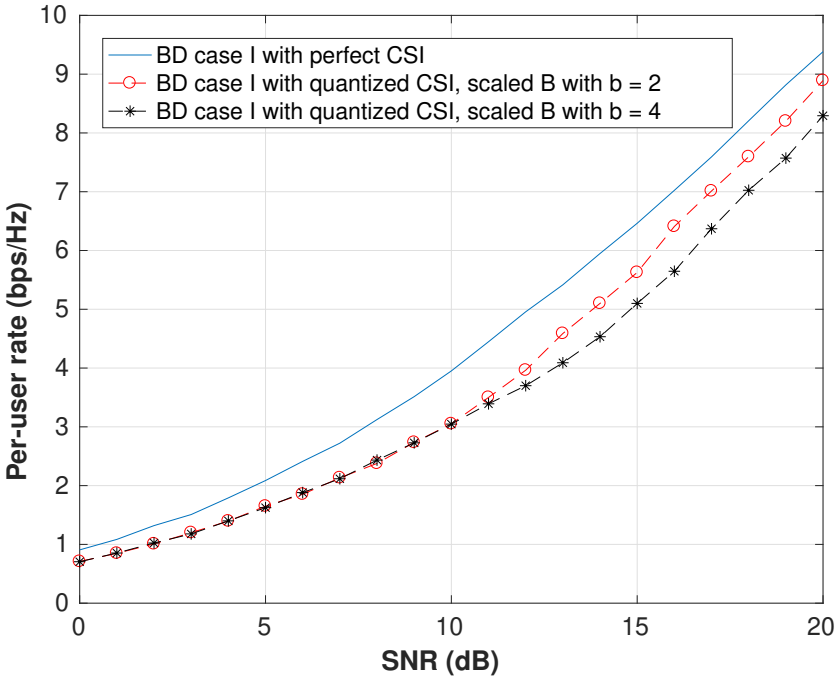

FIGURE 4.5: Ideal vs quantized CSI for case I with B as in (4.33)

Chapter 5

Water Filling Based Channel Quantization and Feedback

In this chapter, we consider optimal power allocation accross the multiple data streams that are sent to the users. Compared to chapter (4), in which we only feedback the subspace information, we now need to quantize and feedback the channel vector of each antenna separately. We first present a vector quantization scheme in order to quantize the channel matrices. Then, we discuss optimal power optimization scheme that maximizes the system performance specially in the low SNR regime.

5.1 Water filling based channel quantization and feedback

In this section, we discuss the quantization and feedback of the channel matrix $\mathbf{H}_k$ of the k^{th} user when optimal power allocation accross the multiple data streams is applied. In the case of uniform power allocation, only the subspace spanned by the columns of $\mathbf{H}_k$ i.e. (the spatial direction of user k) is needed at the BS. However, to allow the BS to perform the algorithm, extra information about the channel matrix of each user $\mathbf{H}_k$ is needed. The waterfilling algorithm needs to know the direction of each channel vector of user k, i.e. the direction of the rows of $\mathbf{H}_k$, and it needs also to know the magnitude information of each direction, i.e. the frobenius norm of each

row of $\mathbf{H}_k$. Consequently, we cannot use the subspace quantization codebook that we used in the previous sections, but instead we shall quantize and feedback each of the N_k channel vectors separately in order to maintain the direction information in each of them. As will become later, there will be a fundamental tradeoff when we compare the water-filling approach we are presenting in this chapter to the subspace based channel quantization discussed in chapter (4). At low SNRs, water-filling approach gives a better performance than subspace quantization. However, at high SNRs, subspace quantization approach outperforms the other approach. This can be attributed to the fact that at low SNRs, the performance is limited by the additive noise in the system. However, at high SNRs, the system performance is limited by the quantization noise.

5.1.1 Proposed channel quantizer for water filling

Now, the proposed quantization technique of the channel matrix $\mathbf{H}_k$ of the k^{th} user is discussed. The total number of bits B allocated for quantizing $\mathbf{H}_k$ are equally distributed among the rows of $\mathbf{H}_k$. As discussed in Sec. 4.3, we only need to feedback the path gains matrix $\mathbf{G}_k$ because the channel subspace of each user $\mathbf{A}_k$ is assumed to be known at the BS. Hence, each row of the path gains matrix $\mathbf{G}_k \in \mathbb{C}^{N_k \times P_k}$ is quantized separately using vector quantization. The quantization of the i^{th} row, $\mathbf{g}_{k,i} \in \mathbb{C}^{1\times P_k}$, of the path gains matrix $\mathbf{G}_k$ is chosen from the codebook $\mathcal{C}_{k,i} = \{\mathbf{c}_{k,i,1}, \mathbf{c}_{k,i,2}, \cdots, \mathbf{c}_{k,i,2^{B^*}}\}$, where $\mathbf{c}_{k,i,j} \in \mathbb{C}^{1\times P_k}$ is the j^{th} quantization vector in the codebook and B^* is the number of bits used to quantize each row, $\mathbf{g}_{k,i}$, of $\mathbf{G}_k$. The k^{th} user quantizes its $\mathbf{G}_k$ to N_k quantization vectors that form the quantized path gains matrix $\widehat{\mathbf{G}}_k$, where $\widehat{\mathbf{G}}_k$ is defined as:

$$\widehat{\mathbf{G}}_k = \left[\hat{\mathbf{g}}_{k,1}^T, \hat{\mathbf{g}}_{k,2}^T, \cdots, \hat{\mathbf{g}}_{k,N_k}^T\right]^T, \tag{5.1}$$

and $\hat{\mathbf{g}}_{k,i} = \mathbf{c}_{k,i,Z_{k,i}}$ is chosen from the codebook $\mathcal{C}_{k,i}$, where the index $Z_{k,i}$ is calculated such that:

$$Z_{k,i} = \underset{j \in \left[1, 2^{B^*}\right]}{\arg\min}\, 1 - \left| \mathbf{g}_{k,i} \mathbf{c}_{k,i,j}^{\mathrm{H}} \right|^2 . \tag{5.2}$$

Additionally, the squared frobenius norm of the i^{th} row of $\mathbf{G}_k$, $\|\mathbf{g}_{k,i}\|_F^2$, is quantized and fed back to the BS. The magnitude information of the channel vectors does contribute in the power allocation solution because it affects the calculation of the singular values of the effective channel, $\mathbf{H}_k\mathbf{F}_k$, of user k, where $\mathbf{F}_k$ is the precoding matrix. We show next, in Sec. 5.1.2, the procedures of the power allocation solution and how it is computed using the singular values of $\mathbf{H}_k\mathbf{F}_k$. The squared frobenius norm $\|\mathbf{g}_{k,i}\|_F^2$ has an Erlang distriution with parameters $(P_k, 1)$ because the entries of $\mathbf{g}_{k,i} \in \mathbb{C}^{1\times P_k}$ are modeled as complex Gaussian random variables with zero mean and unit variance. Hence, we can quantize $\|\mathbf{g}_{k,i}\|_F^2$ using Lloyd-Max scalar quantizer [38], [39] based on Erlang distribution and number of quantization levels equals to $2^{B_{norm}}$, where B_{norm} is the number of bits allocated for quantization of $\|\mathbf{g}_{k,i}\|_F^2$.

The proposed channel quantizer in this section is suboptimal when we do not perform water filling to do power allocation among the data streams intended to the users. The subspace quantization proposed in Sec. 4.3 is better than this quantizer when the BD with uniform power allocation is used. The reason for this is that the BD procedure does not require the direction information of each row vector of $\mathbf{H}_k$ but only the spatial direction of it as discussed in Sec. 4.1.2 i.e. (the subspace spanned by the rows of $\mathbf{H}_k$). Hence, the subspace quantization perform better in this case because it puts all the bits in quantizing the spatial direction of the channel matrix $\widetilde{\mathbf{H}}_k$. However, we show next that the proposed vector quantizer in this section performs better than subspace quantization when combined with optimal power allocation accross the data streams at low-to-mid SNR regime.

Fig. (5.1) compares the performance of the two different quantizers

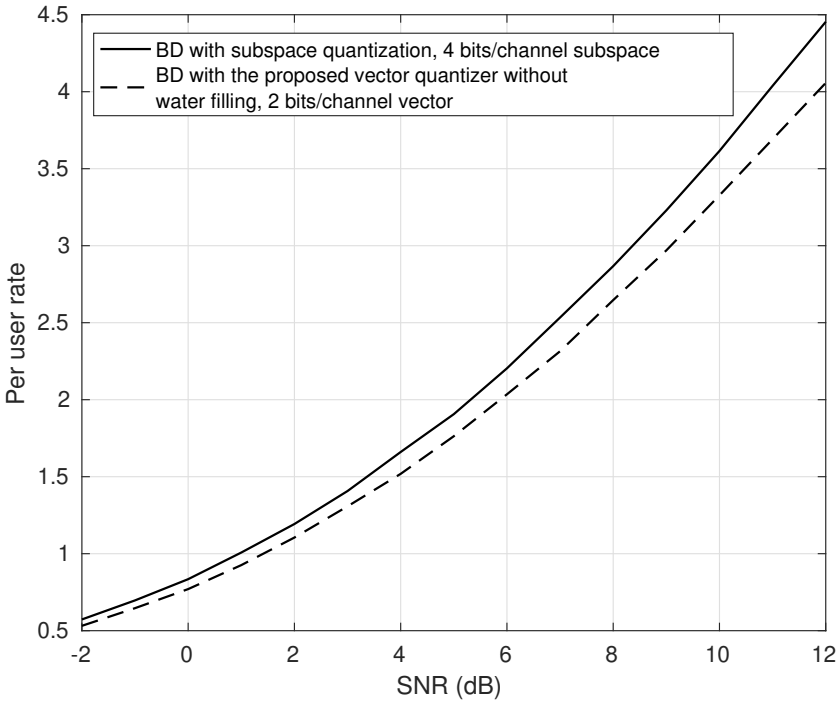

FIGURE 5.1: Performance comparison between BD with subspace quantization vs the proposed vector quantization when no water filling is used: with $N_k = 2$, $P = 3$, $K = 8$, and $M = 128$

when uniform power allocation accross the data streams is adopted. The figure shows that subspace quantization clearly outperforms the proposed vector quantization in all SNR ranges. This shows that subspace quantization of $\mathbf{H}_k$ is always better than vector quantization when using BD with uniform power allocation strategy.

5.1.2 Power optimization algorithm

Due to limitations of the zero-forcing techniques in general when the system noise is high, BD based precoding is sub-optimal in the low SNR regime. Consequently, in this section, we introduce a power optimization technique aiming to maximize the total sum rate of the system. Power optimization accross the data streams is discussed under both perfect and limited CSI feedback cases.

Power optimization assuming ideal CSI at the BS

In this part, the power optimization algorithm based on ideal CSI at the BS is discussed. We are interested in finding the power allocation diagonal matrices $\mathbf{\Delta_k}, \forall \mathbf{k} \in \{\mathbf{1}, \mathbf{2}, \cdots, \mathbf{K}\}$ that maximize the sum rate of the whole system. Hence, the precoding matrix at the BS for user k becomes $\mathbf{F}_k \mathbf{\Delta_k^{1/2}}$. When considering the BD based linear precoding at the BS, the inter-user interference is totally cancelled and the sum rate of the system assuming perfect CSI knowledge at the BS becomes:

$$R_{\text{tot}} = \mathbb{E}\left\{\sum_{k=1}^{K} \log_2 \left|\mathbf{I}_{N_k} + \mathbf{H}_k \mathbf{F}_k \mathbf{\Delta_k} \mathbf{F_k^H} \mathbf{H_k^H}\right|\right\}, \tag{5.3}$$

where $\mathbf{F}_k$ is the precoding matrix at user k and the diagonal elements of $\mathbf{\Delta_k}$ scale the power transmitted into each of the columns of $\mathbf{F}_k$. Because of the nature of the BD structure, the BS sees every user as a point-to-point MIMO link. Therefore, the sum information rate of the system can be calculated as:

$$R_{\text{tot}} = \mathbb{E}\left\{\sum_{k=1}^{K} \log_2 \left|\mathbf{I}_{N_k} + \mathbf{\Lambda_k^2} \mathbf{\Delta_k}\right|\right\}, \tag{5.4}$$

where $\mathbf{\Lambda_k}$ is a diagonal matrix whose elements $\sigma_{k,i}$ are the singular values of the effective channel, $\mathbf{H}_k \mathbf{F}_k$, of user k and $\mathbf{\Delta_k}$ is a diagonal matrix whose elements $\delta_{k,i}$ are the power values transmitted into each of the N_k data streams of user k. Now, the power allocation problem that maximizes the sum-rate can be rewritten as:

$$\begin{aligned} \max_{\delta_{k,i}} \quad & \sum_{k=1}^{K} \sum_{i=1}^{N_k} \log_2 \left(1 + \sigma_{k,i}^2 \delta_{k,i}\right) \\ \text{s.t.} \quad & \sum_{k=1}^{K} \sum_{i=1}^{N_k} \delta_{k,i} \leq \gamma, \end{aligned}$$

where γ is the total transmitted power at the BS. The above problem is clearly a convex optimization problem and it can be solved using the standard solutions. The solution of this problem is well known and it has a closed form expression [33] when solved using the Lagrange multiplier

method. Let a Lagrange multiplier α to be used with the power constraint at the BS, hence the corresponding unconstrained maximization problem is given by:

$$\mathcal{L}(\alpha, \delta_{k,i}) = \sum_{k=1}^{K} \sum_{i=1}^{N_k} \log_2 \left(1 + \sigma_{k,i}^2 \delta_{k,i}\right) - \alpha \left(\sum_{k=1}^{K} \sum_{i=1}^{N_k} \delta_{k,i} - \gamma \right).$$

Due to the KKT conditions, the partial derevatives of $\mathcal{L}$ w.r.t. $\delta_{k,i}$ are

$$\frac{\partial \mathcal{L}}{\partial \delta_{k,i}} = \begin{cases} = 0, & \text{when } \delta_{k,i} > 0 \\ \leq 0, & \text{when } \delta_{k,i} = 0 \end{cases} \tag{5.5}$$

Then, the power allocation solution for the i^{th} data stream of the k^{th} user is derived as:

$$\delta_{k,i}^{*} = \left(\frac{1}{\alpha} - \frac{1}{\sigma_{k,i}^2} \right)^{+}. \tag{5.6}$$

The value of α is determined such that the total power constraint at the BS is satisfied. Hence, α is the solution of the following equation:

$$\sum_{k=1}^{K} \sum_{i=1}^{N_k} \left(\frac{1}{\alpha} - \frac{1}{\sigma_{k,i}^2} \right)^{+} = \gamma. \tag{5.7}$$

Power optimization algorithm considering vector quantization of CSI at the BS

Now, the power optimization algorithm when considering limited feedback of CSI at the BS is discussed. We shall use the proposed vector quantizer as discussed in Sec. 5.1.1. The path gains matrix $\mathbf{G}_k$ is first quantized to $\widehat{\mathbf{G}}_k$ at the k^{th} user then fed back to the BS. Then, the BS uses $\widehat{\mathbf{G}}_k$ to form $\widehat{\mathbf{H}}_k$ and hence going through the BD procedure to generate the precoding matrices $\widehat{\mathbf{F}}_k$ as discussed in Sec. 4.2.1. Then, power optimization accross the data streams is adopted at the BS. Algorithm (2) concludes these steps in an easy way.

Algorithm 2: Power optimization across the data streams

1 User k quantizes $\mathbf{G}_k$ into $\widehat{\mathbf{G}}_k$ based on vector quantization and send it to the BS

2 The BS calculates $\widehat{\mathbf{H}}_k = \widehat{\mathbf{G}}_k \mathbf{A}_k$

3 The BS performs the BD procedures and generates the precoding matrices $\widehat{\mathbf{F}}_k$

4 The BS computes the singular values, $\mathbf{\Lambda_k}$, of the effective channel, $\widehat{\mathbf{H}}_k \widehat{\mathbf{F}}_k$, of user k

5 The BS calculates the value α that satisfies this equation

$$\sum_{k=1}^{K} \sum_{i=1}^{N_k} \left(\frac{1}{\alpha} - \frac{1}{\hat{\sigma}_{k,i}^2} \right)^+ = \gamma$$

6 The BS calculates the optimal power scaling values, $\hat{\delta}_{k,i}$, according to

$$\hat{\delta}_{k,i}^* = \left(\frac{1}{\alpha} - \frac{1}{\hat{\sigma}_{k,i}^2} \right)^+$$

7 The final scaled precoding matrix of user k becomes $\widehat{\mathbf{F}}_k \mathbf{\Delta_k^{1/2}}$

Due to limited feedback of CSI at the BS, the interference on user k due to signals of other users is not totally canceled. Because of this residual interference, we cannot express the information rate of user k as in Equ. 5.4. Hence, the per-user rate considering limited feedback of CSI and power optimization across the data streams is given by:

$$\begin{aligned} R_{\text{limited},k}(\mathbf{\Delta_k}) &= \mathbb{E} \log_2 \left| \mathbf{I}_{N_k} + \sum_{j=1}^{K} \mathbf{H}_k \widehat{\mathbf{F}}_j \mathbf{\Delta_j} \widehat{\mathbf{F}}_\mathbf{j}^{\mathrm{H}} \mathbf{H}_\mathbf{k}^{\mathrm{H}} \right| - \\ & \mathbb{E} \log_2 \left| \mathbf{I}_{N_k} + \sum_{\substack{j=1 \\ j \neq k}}^{K} \mathbf{H}_k \widehat{\mathbf{F}}_j \mathbf{\Delta_j} \widehat{\mathbf{F}}_\mathbf{j}^{\mathrm{H}} \mathbf{H}_\mathbf{k}^{\mathrm{H}} \right|. \end{aligned} \tag{5.8}$$

5.1.3 Numerical results

In this part, the performance of the derived power allocation scheme based on the proposed vector quantizer is examined and verified. Also, we compare the power allocation scheme based on vector quantization with the subspace quantization technique discussed in Sec. 4.3. The system parameters are set as follows. The number of antennas at the BS is $M = 128$, the number of users in the system is $K = 8$, the number of antennas at each

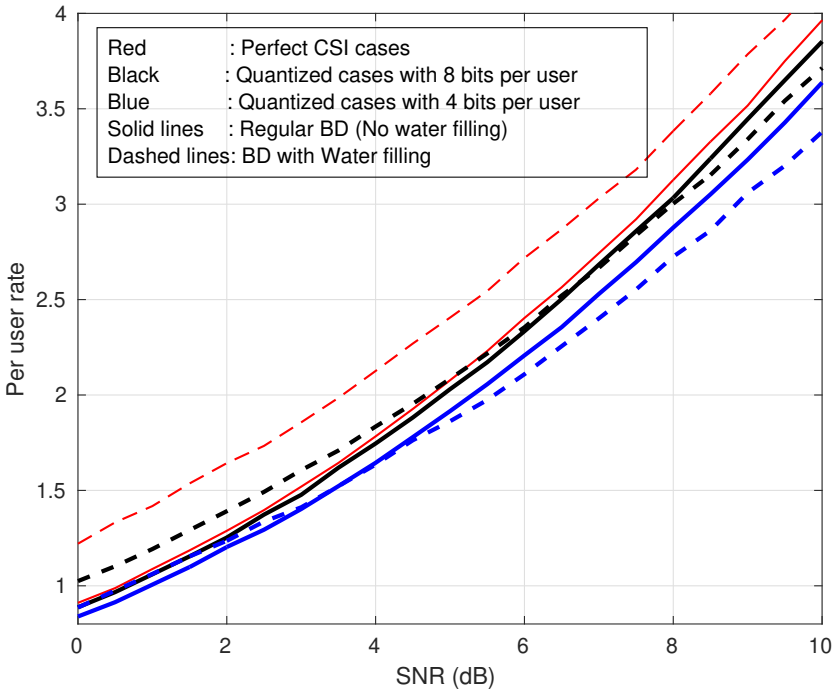

FIGURE 5.2: Performance of BD without power optimization vs BD with power optimization for both ideal and quantization cases with $B = 4, 8$ bits/user

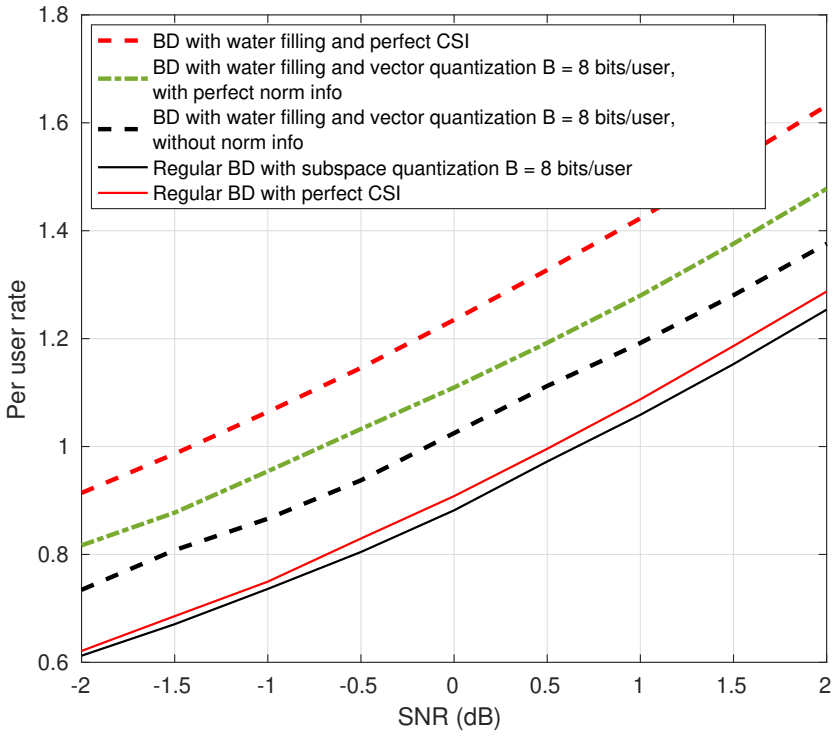

FIGURE 5.3: Performance of BD based suspace quantization vs BD based vector quantization with power optimization at low SNR range with 8 bits/user

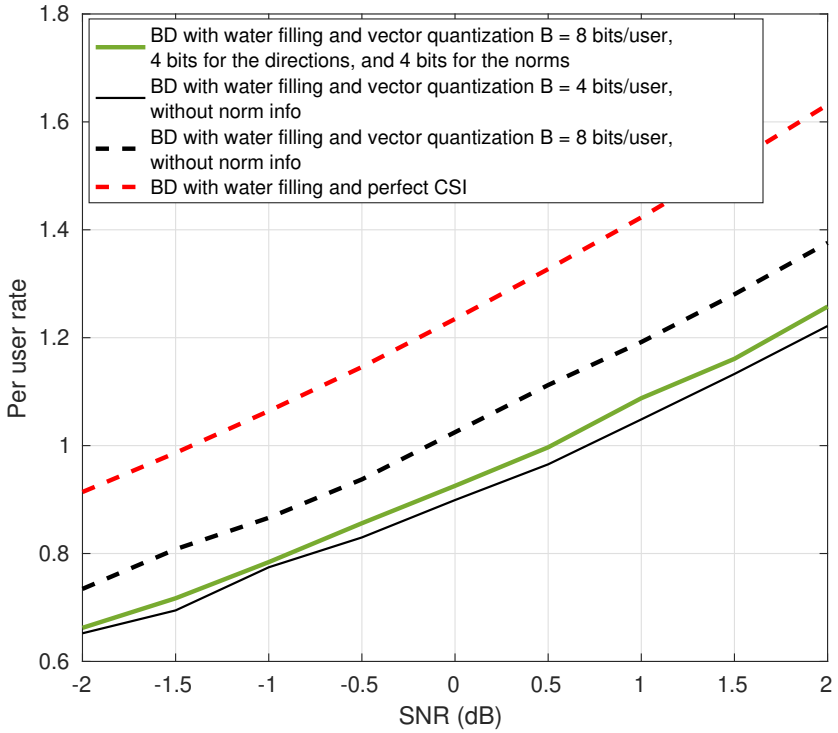

FIGURE 5.4: Performance comparison of BD based vector quantization with water filling when only quantizing the direction information vs quantizing both directions and norms

user is equal to the number of data streams transmitted simultaneously to each user which is $N_k = N = 2$, and the number of resolvable paths is $P = 3$.

In Fig. (5.2), the performance of the BD scheme without power optimization is compared against the performance of BD with power optimization for both ideal and quantization cases. We use a total of $B = 8$ and 4 bits per user whether for subspace quantization or for vector quantization. The graph clearly shows that ideal BD with power optimization (water filling) outperforms the regular ideal BD for all SNR ranges. However, the gap between both schemes gets decreased in higher SNR regions. The reason of this fact is that water filling resists the system noise. Hence, at high SNR values, the system noise nearly becomes insignificant and the regular BD approaches BD with water filling.

Fig. (5.2) also shows that there is a trade-off between both schemes in

the case of limited feedback (quantization). In the low SNR regime, waterfilling with the proposed vector quantization scheme is better than the regular BD with subspace quantization. The reason of this trade-off is that we have two sources of noise in that case, the system noise and the quantization noise. At low SNR, the system noise is more dominant than the quantization noise, hence using water filling along with vector quantization is more useful as it is more immune to the system noise. However, at high SNR, the noise resulted from the quantization of $\mathbf{H}_k$ is more dominant. Hence, no gain is added from using the water filling solution, and subspace quantization, which is the better quantizer as in Fig. (5.1), gives a better performance.

As long as BD based power optimization scheme is useful at the low SNR regime, Fig. (5.3) shows a comparison of both schemes in the range from SNR $= -2dB$ to SNR $= 2dB$ with $B = 8$ bits/user. The graph shows that BD based water filling using the proposed vector quantization clearly outperforms the regular BD based subspace quantization at these low SNR levels. There is nearly a gain of 1dB in SNR when using water filling with vector quantization. This gain is moderately good at these low SNR levels. Fig. (5.3) also shows the effect of having the ideal frobenius norms of the channel vectors $\mathbf{g}_{k,i}$ on the water filling solution at the BS. It is clear that these norm information does enhance the BD based vector quantization with water filling because it affects the singular values calculation of the effective channels of the users as discussed before. However, the performance of the water filling scheme is still better than the regular BD based subspace quantization if all the bits are put in quantizing only the directions of $\mathbf{g}_{k,i}$ and no information is fed back about the norms of $\mathbf{g}_{k,i}$.

Finally, we show in Fig. (5.4) that there is no any gain of quantizing the frobenius norms of $\mathbf{g}_{k,i}$ when we have a limited number of feedback bits per user, but we shall put all the bits in quantizing only the directions. As previously discussed in Sec. 5.1.1, we used the Lloyd-Max algorithm to quantize $\|\mathbf{g}_{k,i}\|_F^2$ which has an Erlang distribution. The performance of BD based vactor quantization with water filling is better when we allocate the

total $B = 8$ bits for quantizing the directions of $\mathbf{g}_{k,i}$ than allocating 4 bits for the directions and another 4 bits for the squared norm values.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

In this work, we have considered the problem of channel feedback in FDD massive MIMO systems with single and multiple antenna users. We have considered the line packing problem on the Grassmannian manifold to quantize the channel vectors in the case of single antenna users. We have considered the use of BD at the massive basestation in the case of multi antenna users. Based on the nature of our channel model, we have devised a channel feedback scheme to reduce the required feedback bits. We showed that the joint feedback of the channel vectors of the receive antennas based on subspace packing reduces the feedback overhead dramatically compared to the ZF scheme on each receive antenna separately. The gain in this case comes from two sources, the first one is using BD instead of ZF. The second source of gain is using subspace quantization rather than vector quantization. Then, we have quantified the rate loss due to the use of channel feedback (compared to the case with perfect CSI at the BS). Also, we have proposed a systematic approach to design the channel feedback codebooks in which the codebook design is formulated as a subspace packing problem over the Grassmannian manifold. These structured codebook have a better performance than RVQ as well as they can be used in practical systems. Finally, we proposed a power allocation scheme to distribute the total power at the BS over the multiple data streams intended to the users. We also combined power optimization with a proposed vector quantization scheme of the channels and this led to a better system performance in the

low SNR regime. A fundamental trade-off appears when we compared vector quantization scheme against subspace quantization. Simulation results prove that at high SNR, subspace quantization is better than vector quantization, however at low SNR, the vector quantization scheme outperforms subspace quantization scheme. The reason behind this is that at low SNR, the additive noise is very high compared to quantization noise. However, at high SNR, the additive system noise becomes insignificant compared to the quantization noise resulted from quantizing the channel vectors.

6.2 Future directions

For future work, we need to study the mathematical analysis of the trade off between the waterfilling scheme and subspace quantization at low SNRs. Also, we need to quantify the effect of quantizing the frobenius norms of the channel vectors on the water-filling algorithm and how much this contributes to the rate loss of the system users.

Bibliography

[1] L. Lu et al. “An Overview of Massive MIMO: Benefits and Challenges”. In: *IEEE Journal of Selected Topics in Signal Processing* 8.5 (2014), pp. 742–758. ISSN: 1932-4553. DOI: 10.1109/JSTSP.2014.2317671.

[2] E. G. Larsson et al. “Massive MIMO for next generation wireless systems”. In: *IEEE Communications Magazine* 52.2 (2014), pp. 186–195. ISSN: 0163-6804. DOI: 10.1109/MCOM.2014.6736761.

[3] T. L. Marzetta. “Massive MIMO: An Introduction”. In: *Bell Labs Technical Journal* 20 (2015), pp. 11–22. ISSN: 1538-7305. DOI: 10.15325/BLTJ.2015.2407793.

[4] T. L. Marzetta. “Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas”. In: *IEEE Transactions on Wireless Communications* 9.11 (2010), pp. 3590–3600. ISSN: 1536-1276. DOI: 10.1109/TWC.2010.092810.091092.

[5] F. Rusek et al. “Scaling Up MIMO: Opportunities and Challenges with Very Large Arrays”. In: *IEEE Signal Processing Magazine* 30.1 (2013), pp. 40–60. ISSN: 1053-5888. DOI: 10.1109/MSP.2011.2178495.

[6] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta. “Energy and Spectral Efficiency of Very Large Multiuser MIMO Systems”. In: *IEEE Transactions on Communications* 61.4 (2013), pp. 1436–1449. ISSN: 0090-6778. DOI: 10.1109/TCOMM.2013.020413.110848.

[7] G. Y. Li et al. “Energy-efficient wireless communications: tutorial, survey, and open issues”. In: *IEEE Wireless Communications* 18.6 (2011), pp. 28–35. ISSN: 1536-1284. DOI: 10.1109/MWC.2011.6108331.

[8] C. Xiong et al. “Energy- and Spectral-Efficiency Tradeoff in Downlink OFDMA Networks”. In: *IEEE Transactions on Wireless Communications*

10.11 (2011), pp. 3874–3886. ISSN: 1536-1276. DOI: 10.1109/TWC.2011.091411.110249.

[9] J. Hoydis, S. ten Brink, and M. Debbah. "Massive MIMO in the UL/DL of Cellular Networks: How Many Antennas Do We Need?" In: *IEEE Journal on Selected Areas in Communications* 31.2 (2013), pp. 160–171. ISSN: 0733-8716. DOI: 10.1109/JSAC.2013.130205.

[10] M. Matthaiou et al. "Sum Rate Analysis of ZF Receivers in Distributed MIMO Systems". In: *IEEE Journal on Selected Areas in Communications* 31.2 (2013), pp. 180–191. ISSN: 0733-8716. DOI: 10.1109/JSAC.2013.130207.

[11] Y. Liang, G. Pan, and Z. D. Bai. "Asymptotic Performance of MMSE Receivers for Large Systems Using Random Matrix Theory". In: *IEEE Transactions on Information Theory* 53.11 (2007), pp. 4173–4190. ISSN: 0018-9448. DOI: 10.1109/TIT.2007.907497.

[12] M. Matthaiou et al. "Sum Rate Analysis of ZF Receivers in Distributed MIMO Systems". In: *IEEE Journal on Selected Areas in Communications* 31.2 (2013), pp. 180–191. ISSN: 0733-8716. DOI: 10.1109/JSAC.2013.130207.

[13] S. L. H. Nguyen and A. Ghrayeb. "Compressive sensing-based channel estimation for massive multiuser MIMO systems". In: *2013 IEEE Wireless Communications and Networking Conference (WCNC)*. 2013, pp. 2890–2895. DOI: 10.1109/WCNC.2013.6555020.

[14] L. Dai, Z. Wang, and Z. Yang. "Spectrally Efficient Time-Frequency Training OFDM for Mobile Large-Scale MIMO Systems". In: *IEEE Journal on Selected Areas in Communications* 31.2 (2013), pp. 251–263. ISSN: 0733-8716. DOI: 10.1109/JSAC.2013.130213.

[15] Z. Gao et al. "Spatially Common Sparsity Based Adaptive Channel Estimation and Feedback for FDD Massive MIMO". In: *IEEE Transactions on Signal Processing* 63.23 (2015), pp. 6169–6183. ISSN: 1053-587X. DOI: 10.1109/TSP.2015.2463260.

[16] P. Zhao, Z. Wang, and C. Sun. "Angular domain pilot design and channel estimation for FDD massive MIMO networks". In: *IEEE International Conference on Communications (ICC)*. 2017. DOI: 10.1109/ICC.2017.7996890.

[17] J. Choi et al. "Noncoherent Trellis Coded Quantization: A Practical Limited Feedback Technique for Massive MIMO Systems". In: *IEEE Transactions on Communications* 61.12 (2013), pp. 5016–5029. ISSN: 0090-6778. DOI: 10.1109/TCOMM.2013.111413.130379.

[18] J. Chen et al. "Precoder feedback versus channel feedback in massive MIMO under user cooperation". In: *2015 49th Asilomar Conference on Signals, Systems and Computers*. 2015, pp. 1449–1453. DOI: 10.1109/ACSSC.2015.7421384.

[19] S. Noh et al. "Pilot Beam Pattern Design for Channel Estimation in Massive MIMO Systems". In: *IEEE Journal of Selected Topics in Signal Processing* 8.5 (2014), pp. 787–801. ISSN: 1932-4553. DOI: 10.1109/JSTSP.2014.2327572.

[20] X. Rao and V. K. N. Lau. "Distributed Compressive CSIT Estimation and Feedback for FDD Multi-User Massive MIMO Systems". In: *IEEE Transactions on Signal Processing* 62.12 (2014), pp. 3261–3271. ISSN: 1053-587X. DOI: 10.1109/TSP.2014.2324991.

[21] Y. G. Lim and C. B. Chae. "Compressed channel feedback for correlated massive MIMO systems". In: *IEEE International Conference on Communications Workshops (ICC)*. 2014. DOI: 10.1109/ICCW.2014.6881223.

[22] M. S. Sim et al. "Compressed channel feedback for correlated massive MIMO systems". In: *Journal of Communications and Networks* 18.1 (2016), pp. 95–104. ISSN: 1229-2370. DOI: 10.1109/JCN.2016.000012.

[23] A. Ge et al. "Principal component analysis based limited feedback scheme for massive MIMO systems". In: *IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*. 2015. DOI: 10.1109/PIMRC.2015.7343318.

[24] W. Shen et al. "Channel Feedback Codebook Design for Millimeter-Wave Massive MIMO Systems Relying on Lens Antenna Array". In: *IEEE Wireless Communications Letters* 7.5 (2018), pp. 736–739. ISSN: 2162-2337. DOI: 10.1109/LWC.2018.2818130.

[25] W. Shen et al. "AoD-adaptive subspace codebook for channel feedback in FDD massive MIMO systems". In: *IEEE International Conference on Communications (ICC)*. 2017. DOI: 10.1109/ICC.2017.7997168.

[26] David Tse and Pramod Viswanath. *Fundamentals of Wireless Communication*. Cambridge University Press, 2005. DOI: 10.1017/CBO9780511807213.

[27] W. Santipach and M. L. Honig. "Asymptotic capacity of beamforming with limited feedback". In: *International Symposium onInformation Theory (ISIT)*. 2004. DOI: 10.1109/ISIT.2004.1365326.

[28] I. S. Dhillon et al. "Constructing Packings in Grassmannian Manifolds via Alternating Projection". In: *Experimental Mathematics* 17.1 (2008), pp. 9–35. DOI: 10.1080/10586458.2008.10129018.

[29] Mahmoud A. AlaaEldin, Karim G. Seddik, and Wessam Mesbah. "AoD-Adaptive Channel Feedback in FDD MassiveMIMO Systems with Multiple-Antenna Users". In: *IEEE Global Communications Conference (Globecom), Abu Dhabi, United Arab Emirates, Dec. 2018*. 2018.

[30] R. Schmidt. "Multiple emitter location and signal parameter estimation". In: *IEEE Transactions on Antennas and Propagation* 34.3 (1986), pp. 276–280. ISSN: 0018-926X. DOI: 10.1109/TAP.1986.1143830.

[31] H. Xie, F. Gao, and S. Jin. "An Overview of Low-Rank Channel Estimation for Massive MIMO Systems". In: *IEEE Access* 4 (2016), pp. 7313–7321. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2016.2623772.

[32] N. Ravindran and N. Jindal. "Limited feedback-based block diagonalization for the MIMO broadcast channel". In: *IEEE Journal on Selected Areas in Communications* 26.8 (2008), pp. 1473–1482. ISSN: 0733-8716. DOI: 10.1109/JSAC.2008.081013.

[33] G. G. Raleigh and J. M. Cioffi. "Spatio-temporal coding for wireless communication". In: *IEEE Transactions on Communications* 46.3 (1998), pp. 357–366. ISSN: 0090-6778. DOI: 10.1109/26.662641.

[34] D. J. Love and R. W. Heath. "Limited feedback unitary precoding for spatial multiplexing systems". In: *IEEE Transactions on Information Theory* 51.8 (2005), pp. 2967–2976. ISSN: 0018-9448. DOI: 10.1109/TIT.2005.850152.

[35] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt. "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels". In: *IEEE Transactions on Signal Processing* 52.2 (2004), pp. 461–471. ISSN: 1053-587X. DOI: 10.1109/TSP.2003.821107.

[36] V. Va, J. Choi, and R. W. Heath. "The Impact of Beamwidth on Temporal Channel Variation in Vehicular Channels and Its Implications". In: *IEEE Transactions on Vehicular Technology* 66.6 (2017), pp. 5014–5029. ISSN: 0018-9545. DOI: 10.1109/TVT.2016.2622164.

[37] T. S. Rappaport et al. "Millimeter Wave Mobile Communications for 5G Cellular: It Will Work!" In: *IEEE Access* 1 (2013), pp. 335–349. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2013.2260813.

[38] J. Max. "Quantizing for minimum distortion". In: *IRE Transactions on Information Theory* 6.1 (1960), pp. 7–12. ISSN: 0096-1000. DOI: 10.1109/TIT.1960.1057548.

[39] S. Lloyd. "Least squares quantization in PCM". In: *IEEE Transactions on Information Theory* 28.2 (1982), pp. 129–137. ISSN: 0018-9448. DOI: 10.1109/TIT.1982.1056489.