Jimma University

Technology Institute

Fuculty of Electrical and Computer Engineering

Communication Stream

Title: Performance Analysis of Non-Orthogonal Beam Division Multiple Access for Full Duplex based Massive-MIMO Systems

A thesis submitted to Jimma Institute of technology, School of graduate studies in partial fulfillment of the requirements for masters degree in communication Engineering.

By:
Henok Berhanu

Main Advisor: Dr. Kinde Anlay
Co-Advisor: Mr. Getachew Alemu (M.Sc.)

June 22, 2019
Jimma, Ethiopia

Jimma University

Technology Institute

Fuculty of Electrical and Computer Engineering

Communication Stream

Title: Performance Analysis of Non-Orthogonal Beam Division Multiple Access for Full Duplex based Massive-MIMO Systems

A thesis submitted to Jimma Institute of technology, School of graduate studies in partial fulfillment of the requirements for masters degree in communication Engineering.

By:
Henok Berhanu

Main Advisor: Dr. Kinde Anlay
Co-Advisor: Mr. Getachew Alemu (M.Sc.)

June 22, 2019
Jimma, Ethiopia

Declaration

I hereby declare that this thesis has not been and will not be submitted in whole or in part to another University for the award of any other degree.

Research proposal submitted by:

Henok Berhanu

____________________ ____________________

Signature Date

Approved by advisors:

Dr. Kinde Anlay (Main adviser):

____________________ ____________________

Signature Date

Getachew Alemu (Co-adviser)

____________________ ____________________

Signature Date

Approved by faculty of Electrical and Computer Engineering research proposal examination members.

1. ____________________ ____________ ______________

   Name Signature Date

2. ____________________ ____________ ______________

   Name Signature Date

3. ____________________ ____________ ______________

   Name Signature Date

Abstract

For the Full Duplex (FD) scheme based large-scale multi-input Multi-Output (MIMO) systems, a multi-user MIMO (MU-MIMO) non-orthogonal multiple access (NOMA) scheme, known as the beam division NOMA (BD-NOMA) is addressed for massive MIMO systems. Traditional MIMO in the spatial domain is analyzed using lens antenna array to form Beam Division Multiple Access scheme which has orthogonal resource allocation lacking optimization. The Power Domain-Non Orthogonal Multiple Access (PD-NOMA) provides different power allocation for near and far users by exploiting supper position coding at the base station followed by Successive Interference Cancellation (SIC) at mobile stations. To get technical advantage for resource optimization, BD-NOMA is formed by combining BDMA with PD-NOMA.

Millimeter wave communication is a promising technology for future wireless systems. One of the practical difficulty is to achieve its large-antenna gains with only limited number of radio frequency (RF) chains. To this end, a new lens antenna array enabled mm-wave MIMO is addressed. Using lens antenna array, a conventional spatial channel is transformed in to beam space channel in order to capture the channel sparsity at mm-wave frequencies. Accordingly, the dominant beams are selected from the sparse beam space channel to reduce the number of required RF chains by using interference-aware beam selection.

A precoding scheme based on the principle of wiener filter (WF) is designed to reduce Inter-Beam Interference (IBI) in the downlink. Additionally, to maximize the achievable sum rate, a dynamic (iterative) power allocation is proposed by solving the joint power optimization problem, that address not only intra-beam optimization, but also considers inter-beam optimization.

In this paper, an FD base station is analyzed with the Half-Duplex (HD) users. At this FD base station, downlink antennas affects the performance of uplink antennas creating a Self-Interference (SI) because of that both transmission and reception is conducted at the same time and frequency at the FD base station. To mitigate the effect of SI in the uplink, the paper uses two SI channel estimators assuming that the interference follows both Rician and Rayleigh fading channels. The first one is the Minimum Mean Square Error (MMSE) estimator, while the second is the Least Square (LS) estimator. After these estimators, an uplink spectral efficiency of proposed BD-NOMA is analyzed.

Simulation results showed that BD-NOMA scheme has around 10% spectral and energy efficiency increment over BD-OMA (BDMA). Traditional MIMO has higher spectral efficiency than BD-NOMA with a cost of enormous energy consumption. But it is proved that Traditional MIMO will have almost a zero energy efficiency as compared to proposed BD-NOMA.

Key Words : NOMA, BDMA, IBI, SIC, PD-NOMA, BD-NOMA, WF

Acknowledgment

In the first place, I would like to pass my gratitude to my advisers Dr. Kinde Anlay and Mr. Getachew Alemu for their supervision, advice, and guidance from the very early stage of this research as well as giving me extraordinary experiences throughout the work. Their involvement with their originality has triggered and nourished my intellectual maturity that I will benefit from, for a long time to come.

I gratefully acknowledge Dr. Emil Björnson and Mr. Vaibhav Kumar for their ideas, assistance and crucial contribution, which made an important factor for the accomplishment of this thesis. Above all and the most needed, they provided me unwavering encouragement and support in various ways.

Words fail me to express my appreciation to Mrs. Lidet Tewachew whose dedication, love and persistent confidence in me, has taken the load off my shoulder. I owe her for being unselfishly let her intelligence, passions, and ambitions collide with mine.

Finally, I would like to thank everybody who was important to the successful realization of the thesis, as well as expressing my apology that I could not mention personally one by one.

Contents

List of Figures

List of Tables

Acronyms

BS	Base Station
BDMA	Beam Division Multiple Access
BD-NOMA	Beam Division Non Orthogonal Multiple Access
CDMA	Code Division Multiple Access
EE	Energy Efficiency
FD	Full Duplex
FDD	Frequency Division Duplex
FDMA	Frequency Division Multiple Access
HD	Half Duplex
HARQ	Hybrid Automatic Repeat Request
IBI	Inter Beam Interference
LSAS	Large Scale Antenna Systems
LS	Least Square
LDS	Low-Density Spreading
MIMO	Multi-Input Multi-Output
MBB	Mobile Broadband
MF	Matched Filter
MMSE	Minimum Mean Square Error
MUSA	Multi-User Shared Access
NOMA	Non Orthogonal Multiple Access
OMA	Orthogonal Multiple Access
OFDMA	Orthogonal Frequency Division Multiple Access
PD-NOMA	Power Domain-Non Orthogonal Multiple Access
PDMA	Pattern Division Multiple Access
PBS	Per Beam Synchronization
RF	Radio Frequency
RSI	Residual Self Interference
RZF	Regularized Zero Forcing
SI	Self Interference
SIC	Successive Interference Cancellation
SE	Spectral Efficiency
SCMA	Spares-Code Multiple Access
SC-FDMA	Single Carrier-Frequency Division Multiple Access
TDD	Time Division Duplex
TDMA	Time Division Multiple Access
WF	Wiener Filter
ZF	Zero Forcing

Chapter 1

Introduction

1.1 Motivation

The fast growth of mobile Internet has propelled the 1000-fold data-traffic increase that is expected for 2020 [1]. Therefore, the major challenge in the future generation become spectral efficiency which could support mobile broadband (MBB)services like virtual reality and video conferencing.

Wireless communication networks in the next generation are required to have such qualities like, improved data rate, low latency, high capacity, high throughput, high power and spectral effciency [2]. in order to achieve these qualities in the future, there are five basic technologies which are expected to appear in the next generation of wireless communication (5G); these are Millimeter wave, Small cell, Massive MIMO, Beam-forming, and Full duplex technologies.

Among these pillars of the fifth generation wireless network, the author was interested to study about Massive MIMO or large scale antenna systems (LSAS), the Non-Orthogonal Beam Division Multiple Access (BD-NOMA) which encompasses the Beam-forming technology, and the Full Duplex technology.

Non-orthogonal access is a promising technology that can increase the system throughput and simultaneously serve massive connections. It allows multiple users to share time and frequency resources in the same spatial layer via simple linear superposition or code-domain multiplexing. so, in order to attain the high throughput of non-orthogonal multiple access proper scheduling and hybrid automatic repeat requests (HARQ) are essentially needed for both downlink and uplink non-orthogonal transmission. In this case, the transmission format of a radio link matches the instantaneous channel condition even in fast fading [1].

NOMA is a novel multiplexing scheme that improves spectrum efficiency by utilizing an additional new domain which is the power domain, which is not sufficiently utilized in previous systems. The major concept behind is that multi-user signals are superimposed at the transmitter with different power allocation. And in this access method, successive interference cancellation (SIC) technique should be applied at a near user who has better channel conditions, so as to remove a far user signal within the beam before detecting his own signal. The NOMA system can enhance throughput by sharing available resources between near and far users, unlike the conventional orthogonal multiple access (OMA) system in which resources are allocated orthogonally for each near and far users.

The second basic technology that should be included in the next generation wireless network is the Beam Division Multiple Access (BDMA) where the base station communicates with users across different orthogonal beams.

Combining NOMA which has different power allocation for near and far users directly with BDMA which has the same power allocation for all users might be difficult to think it in reality. So, in this thesis, it is tried to show the theoretical setups that makes NOMA work together with BDMA with simulations.

Massive MIMO or very large multiple-input multiple-output (MIMO) has massive number of antennas at base stations (BS) to serve hundreds of mobile users simultaneously. Via spatial multiplexing and directing power intently, Massive MIMO can greatly outperform the state-of-the-art cellular standards jointly in terms of spectral efficiency (SE) and energy efficiency (EE) [3].

5G and other future wireless systems promise the user access to services that require considerably higher data rates in spite of the limited wireless spectrum. This demand for higher data rates entails achieving superior performance and efficiency in terms of wireless resources utilization. Previous generations of mobile communication mainly depended on half-duplex transmission schemes, in which the transmitted and received signals are separated either in time domain (TDD) or in the frequency domain (FDD). The term "full-duplex" (FD) was traditionally used when the device had simultaneous bidirectional communication, in contrast to half-duplex (HD), which assumed time-division duplexing. Previously, use of the term full-duplex assumed utilizing a pair of frequencies to transmit and receive simultaneously. However, in recent years the term has carried a new concept: the device can transmit and receive at the same time and over the same frequency [1].

Theoretically, Full duplexing scheme can double the achievable sum-rate of half-duplex (HD) radios. However, practically the spectral efficiency gain of FD over HD is highly determined by the self-interference either on the base-station or on the mobile-station and cross-mode interference between simultaneous uplink and downlink channels.

1.2 Statement of the problem

Non-orthogonal multiple access (NOMA) which has different power allocation for near and far users will be analyzed with the beam division multiple access (BDMA) which has uniform orthogonal power allocation for all users. In the proposed BD-NOMA, since the beams are non-orthogonal, there will be inter-beam interference (IBI) in between the beams and there will also be intra-beam interference within the beam. There has to be a mechanism which mitigate this interference and both the inter-beam power optimization and intra-beam power optimization should be carried out to minimize the effect of these two interference.

The Traditional MIMO which has equal number of radio frequency (RF) chain with that of the base station antenna is difficult to analyze in mm-wave communications having several hundred antenna. This is because that there will be an enormous amount of energy which could be consumed without use due to each RF chain.

Since non-orthogonal beam division multiple access (BD-NOMA) is analyzed by assuming full duplex scheme using suitable channel model, Self-interference (SI), Co-channel interference (CCI), and cross-mode interference will highly affect the double spectral efficiency gain of FD-scheme over the HD-scheme.

1.3 Objectives of the research

1.3.1 General Objective

The main objective of this thesis work is to combine the energy efficient multiple access technique called NOMA with the BDMA to improve the capacity and spectral efficiency gain of multiple access techniques using FD scheme, by optimizing the available power and mitigating IBI and SI.

1.3.2 Specific objective

The specific objectives of this research are:

- To analyze the spectral and energy efficiency of BDMA, Traditional MIMO, and BD-NOMA in massive MIMO systems.
- To compare the spectral and energy efficiency of NOMA with BDMA scheme.
- To analyze the capacity advantage of BD-NOMA over the traditional-MIMO.
- To minimize the number of Radio Frequency (RF) chain in the traditional-MIMO for an optimum use of available power.
- To mitigate Inter Beam Interference by using linear precoding technique.
- To suppress Self-Interference by using SI channel estimation techniques at the FD base station antennas.
- To ensure iterative power optimization for both Inter and Intra beam scenarios.

1.4 Methodology

Firstly, different publications are reviewed concerning NOMA and BDMA by analyzing the capacity advantages and requirements of combining these two schemes. Since there is unequal power allocation to near and far users in NOMA, it might be difficult to realize it practically by combining it to BDMA which has equal power allocation to all users.

In the BDMA system, users access the network across different orthogonal beams. All the available spectrum is given to all users lacking optimization.

Here, after mitigating the IBI, BD-NOMA scheme is analyzed by assuming FD scheme, then at the FD base station antennas self-interference is suppressed after SI channel estimation and the capacity is analyzed and validated using MAT-LAB.

1.5 Scope of the thesis

There are three basic tasks that has been accomplished in this thesis, the first one is combining the individual performance of non-orthogonal multiple access in the power domain and beam division multiple access and analyze BD-NOMA to see its performance and capacity advantage over BDMA and Traditional-MIMO.

The second task that is to be addressed in this thesis was, mitigating the effect of Inter-Beam interference (IBI) which occur between non-orthogonal beams using linear precoding technique. Along with this concept, by using an iterative power allocation strategy, the inter and intra beam power optimization is analyzed and validated by using simulations. Here channel state information and user scheduling are assumed perfectly in the network.

The third task that has been addressed in this thesis is, analyzing the above BD-NOMA technique with the concept of Full-Duplex scheme. So, the performance gain of using full-duplex scheme with BD-NOMA is attained and validated by suppressing the effect of Self-Interference (SI) at the full duplex base station. Here again for the FD scheme, this paper analyzes only by suppressing the self-interference and after the SI channel estimator, the spectral efficiency is analyzed.

1.6 Significant of the thesis

On the next generation of wireless mobile communication, Non-orthogonal multiple access (NOMA), Beam Division Multiple Access (BDMA), Massive MIMO and Full Duplex (FD) scheme which is a modified version of the Half Duplex time division duplex technique are the major pillar technologies that are expected to appear so as to satisfy the needs of high network capacity, improved data rate, enhanced spectral efficiency.

The non-orthogonal multiple access is mainly characterized by different power allocation for near and far users and when these users access the base station across different beams by using efficient scheduling technique. This makes NOMA combine with BDMA which has an advantage of higher performance and capacity as compared to individual NOMA and BDMA. Users of this network will access the service in a full-duplex fashion which allows simultaneous transmission and reception of information over the same channel.

The other contribution that this paper contributes is that the Inter-Beam interference between non-orthogonal beams is mitigated. And again apart from this one, SI interference that prevents the two-fold spectral efficiency gain of FD over HD will be mitigated and the next generation wireless mobile network might be real including these technologies.

1.7 Organization of the Thesis

This thesis work contains six chapters. The first chapter is the introduction part which contains motivational overview, statement of the problem, objective, methodology, scope and significance of the thesis. The second chapter deals with technical background containing literature review and background of multiple access techniques described in terms of Orthogonal and Non-Orthogonal schemes that are mostly used in communication systems. Chapter three generally discuss about the Beam Division Non-Orthogonal Multiple Access (BD-NOMA). Chapter four discusses the BD-NOMA in the uplink scenario assuming Full Duplex scheme and the mechanisms of estimating and canceling the self interference is also discussed. Chapter five is about simulation results and discussions; while the last chapter contains conclusion and recommendation for future works.

Chapter 2

Technical background

2.1 Introduction

In communications systems, multiple access techniques play great role in allowing several users to access the available resource in an efficient way. It is known that resources (bandwidth) are very limited in communication systems and this finite resource has to be shared among multiple users by using efficient multiple access scheme. When there is more than one user to access such limited bandwidth, a multiple access scheme must used to share the available resource among users.

Several researches have been made to find efficient way of multiple access scheme from the early stage of communication systems. During the beginning times of multiple access schemes, the most widely and frequently used ways are sharing either time or frequency resources among users.

There are different generations of Multiple access techniques in communication network which are evolving from first generation to the fourth generation of wireless communication with advancements in capacity of handling larger number of users at a time.

The multiple access technique that were used in the first generation was FDMA (Frequency Division Multiple Access) where the overall bandwidth is divided in to different channel frequencies and each users are provided with a specific dedicated channels. This multiple access technique is prone to misuse of the scarce resource which is bandwidth due to dedicated channels assigned to each users and as the number of users increases, there will not be sufficient frequency channels to assign [1].

The second generation of wireless communication,GSM, uses TDMA (Time Division Multiple Access) where the available bandwidth is divided into time slots and each users access the available channel through synchronously assigned time slots. This type of multiple access may handle higher number of users, but there is delay due to guard time assigned between each time slots [4].

For the third generation, the multiple access used were Code Division Multiple Access (CDMA)and for the fourth generation, Orthogonal Frequency Division Multiple Access (OFDMA) was used. Most of these are orthogonal multiple-access(OMA) schemes, especially for the downlink transmission; in other words different users are allocated with orthogonal resources, either in time, frequency or code domain, in order to alleviate cross-user interference. In this way, multiplexing gain is achieved with reasonable complexity [1].

2.2 Orthogonal Multiple Access (OMA) schemes

Different orthogonal channels that are separated either in time or frequency are given for each user in Orthogonal Multiple Access schemes. Here because of the orthogonality of wave forms of each users, there will be low interference in between users which allows the receiver of OMA decode signal with less complexity because of orthogonality. The total available bandwidth **B** either in time or frequency is divided into **F** frequency channels between the **K** users to ensure orthogonality. Some of the examples of orthogonal multiple access schemes are, Frequency Division Multiple Access (FDMA), Time Division Multiple Access (TDMA), Orthogonal Frequency Division Multiple Access (OFDMA). Even though there is non-complex Multi-User Detection (MUD) at the receiver of OMA, but is difficult to achieve sum-rate capacity of wireless system [1].

2.2.1 Frequency Division Multiple Access (FDMA)

FDMA is the first analog multiple access scheme which were used in the first generation of wireless communication systems. FDMA assigns specific frequency channels to each available users out of the available bandwidth. Multiple users using separate frequency channels could access the same system without significant interference from other users with in the system. It is the simplest way of having multiple access scheme in a multi-user system [4].

In FDMA, the available system bandwidth **B** is divided into **F** equal frequency channels serving **K** users simultaneously where each user is allocated its individual channel. In between each assigned channels, there is a guard band frequency assuring the removal of interference in between each users [1].

The basic challenges of classical analog FDMA are the requirement of **K** modulators and demodulators at the base station to serve **K** users simultaneously. This will be very difficult to analyze mm-wave communication having several hundred antennas at the BS. FDMA assigns fixed channel to multiple users which makes it not flexible in handling users with varying transmission rate. FDMA also high amount of wastage of bandwidth where no sub-channel is reallocated to other users if it is not in use by the assigned user [1].

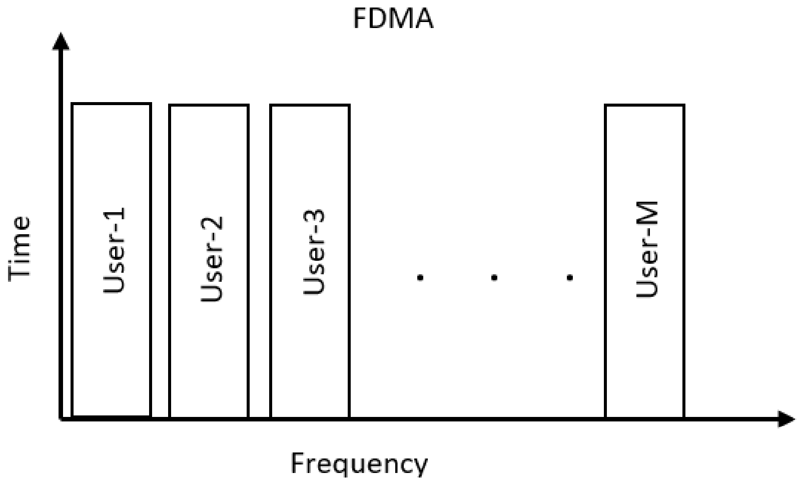

Figure 2.1: Orthogonal FDMA resource allocation [1]

2.2.2 Time Division Multiple Access (TDMA)

The era of digital communication has revived using specific time slots for each available users of a communication medium. This multiple access scheme is called Time Division Multiple Access. In TDMA, each users access the available resource through assigned orthogonal time slots. In TDMA, a single user could send a large data file within time slots of different periodical frames. Data from a single user always sits in the same time slot position of a frame, so at the receiver collects all information from that portion and rearranges the sent frames of each users information. TDMA has kept its dominance in wired and wireless systems for many years such as, the second generation (2G) Global System for Mobile Communications (GSM) and the 2.5G General Packet Radio Service (GPRS) adopted TDMA as their multiple access scheme [4].

2.2.3 Orthogonal Frequency Division Multiple Access (OFDMA)

In Orthogonal Frequency Division Multiple Access scheme, the signals to be transmitted are mapped onto several parallel orthogonal sub-carriers [1]. The Inter Channel Interference (ICI) between users will be mitigate by using guard intervals which are added to each OFDM symbol. Furthermore, the bandwidth of each sub-carrier is narrower than the coherence bandwidth of the channel, which ensures frequency-selective channel [1].

Practically, OFDMA is implemented by using an efficient computation of

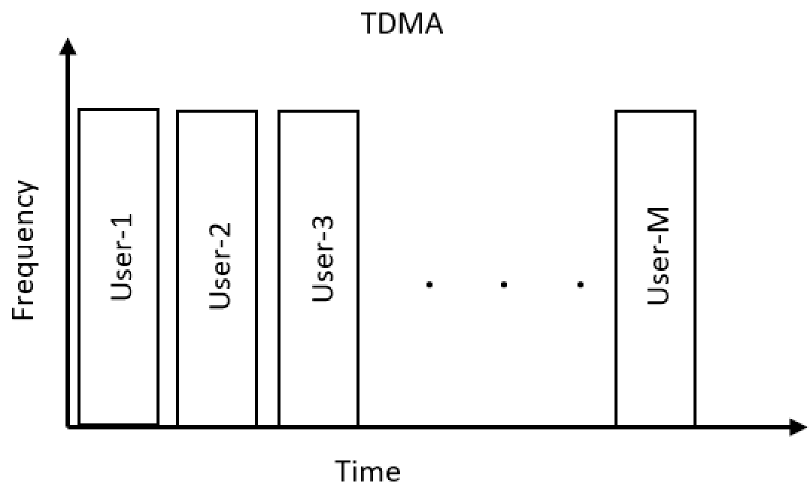

Figure 2.2: Orthogonal TDMA resource allocation [1]

Inverse Fast Fourier Transform (IFFT) for modulation and uses FFT for demodulation. OFDMA is a hybrid combination of FDMA and OFDM [4]. It is currently used in wireless LAN, WiMAX (IEEE 802.16), and LTE downlink systems. The system bandwidth is divided into many sub-channels and each user is allocated multiple dedicated sub-channels, allowing **K** users transmit simultaneously. The number of sub-carriers allocated to each user is flexible depending on its rate and Quality Of Service (QOS) requirement. There are two types of sub-carrier allocation to different users, it can be either adaptive or fixed [2]. Fixed sub-carrier allocation does not adapt to users channel conditions and remain unchanged throughout the communication session leading to a simpler implementation without incurring high overheard. Furthermore, the users can be allocated adjacent sub-carriers, which simplifies frequency and time synchronization on the expense of vulnerability to deep fading, or separated by more than the coherent bandwidth of the channel to exploit maximum frequency diversity, at the expense of a minimum separation between sub-carriers from different users requiring strict cross-user synchronization to avoid ICI [1].

Adaptive sub-carrier allocation dynamically allocates sub-carriers to users based on their channel condition so as to optimize some performance criteria. The main challenges with OFDMA is that it suffers from a high Peak-to-Average Power Ratio (PAPR) which leads to inefficient operation of power amplifiers [4]. This is especially critical in uplink where user transmit powers are limited Secondly, OFDM is very sensitive to errors in time and frequency synchronization which leads to frequency and phase offset causing ICI and ISI [1].

2.2.4 Single Carrier-Frequency Division Multiple Access (SC-FDMA)

Single Carrier-FDMA is a multiple access technique which was used in the uplink transmission of OFDM system of LTE in order to reduce the high Peak-to-Average Power Ratio (PAPR) leading to interference [1]. Due to this PAPR reduction capability of SC-FDMA in the uplink, the power consumption issue of OFDMA system is managed at the user terminals. In order to achieve minimum PAPR in SC-FDMA, consecutive resource allocation strategy should be implemented. Using Non-Orthogonal Multiple Access (NOMA) together with SC-FDMA, costs to have efficient scheduling between uplink users. But the problem is that in the uplink, the transmission power of a single user may limit the requirements of other user terminals with in the same group. This because that in the uplink, all users terminals signals are multiplexed together to share the same available resource.

In the SC-FDMA, if a user terminal reaches its maximum transmission power then the base station will stop allocating extra resource to that user specifically. But in the case of NOMA, since resources are allocated to a group of multiplexed users; there has to be efficient scheduling and grouping of users so that it will not be difficult for the BS to allocate resource satisfying each users requirements.

So that in order to use SC-FDMA together with NOMA, efficient user grouping and scheduling has to be implemented to maximize the achievable sum rare of wireless cellular network.

2.3 Non-Orthogonal Multiple Access (NOMA) schemes

The throughput and sum rate capacity of multiple access schemes could efficiently be maximized by using NOMA which is key scheme expected in the future generations of cellular communication [4]. NOMA allows multiple users to share time and frequency resources in the same spatial layer via simple linear superposition at the BS and Successive Interference Cancellation (SIC) at the user terminals. The interference in NOMA is made controllable by non-orthogonal resource allocation, at the cost of slightly increased receiver complexity, where SIC is employed. In OMA, although the orthogonally multiplexed users facilitates simple and interference-free Multi-User Detection (MUD) at receivers, it does not achieve the sumrate capacity of a wireless system.

2.3.1 Power Domain NOMA (PD-NOMA)

PD-NOMA is based on the principles of superposition coding where the transmitted signal is the sum of each users' signals in the power domain, as illustrated in the figure below. The scheme exploits the received power differences due the channel gain difference among users to solve the problem of detection ambiguity. i.e. the weaker users are allocated more power compared to the stronger users. Thus, the larger the difference in power level of the users' signals, the better the performance which outperforms the orthogonal schemes. As an example understanding of PD-NOMA, let us take two users located at different distance from the BS. Let user-1 locating near the BS has been allocated with a power of $\mathbf{P}_1$ and user-2 locating farther from the BS was allocated a power of $\mathbf{P}_2$. Since user-1 has better channel condition as compared to user-2, it is given lower power and the farther user will be given higher power as compared to user-1. Then the received super imposed signals of each user in the downlink and uplink of PD-NOMA system will be [5],

$$\begin{aligned} \mathbf{r}^{[\mathbf{DL}]} &= \mathbf{h}_{\mathbf{DL}}(\sqrt{\mathbf{P_1}}\mathbf{x_1} + \sqrt{\mathbf{P_2}}\mathbf{x_2}) + \mathbf{w}_{\mathbf{DL}} = \mathbf{h}_{\mathbf{DL}}\mathbf{x} + \mathbf{w}_{\mathbf{DL}} \\ \mathbf{r}^{[\mathbf{UL}]} &= \mathbf{h_1}\sqrt{\mathbf{P_1}}\mathbf{x_1} + \mathbf{h_2}\sqrt{\mathbf{P_2}}\mathbf{x_2} + \mathbf{w}_{\mathbf{UL}} \end{aligned} \tag{2.1}$$

Where x_1 are x_2 are the message signals of each users, h_1 and h_2 the channel matrix of user-1 and BS and user-2 and BS respectively. And w_{DL} and w_{UL} terms are the Additive White Gaussian Noise (AWGN).

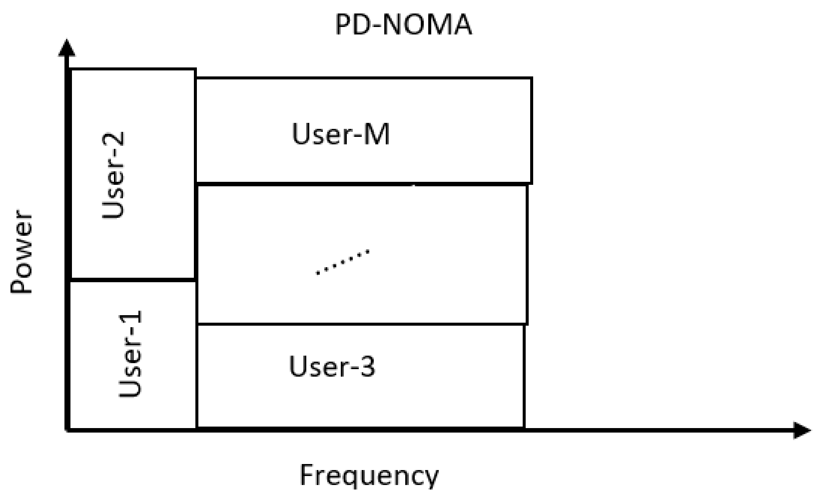

Figure 2.3: Power domain NOMA resource allocation [1]

In the above example of downlink PD-NOMA, the same supper imposed transmitted signal from the BS is received at both users. Then a multi-user signal separation technique called Successive Interference Cancellation (SIC) is implemented at each user terminals to eliminate inter-user interference. For SIC, the optimal order for decoding is in the order of the decreasing channel gain among near and far users from the BS, $|h_1|^2 > |h_2|^2$. So by using the decode and forward approach, user-2 does not perform interference cancellation since it comes first in the decoding order. User-1 first decodes user-2 signal x_2 and subtracts its component from received signal $r^{[DL]}$, then decodes x_1 without interference from x_2. In uplink PD-NOMA, the multiplexed signal from the users are received at the BS. The BS conducts SIC according to the descending order of channel gains, with the least power user detected interference free.

The main disadvantage of PD-NOMA is that although larger the difference in power level results in better the performance, the user throughput fairness is poor due to the rate of the weaker users becoming significantly lower than the stronger users as the power difference increases.

2.3.2 Code Division Multiple Access (CDMA)

CDMA, which is based on spread spectrum technology were firstly used for military applications before it is used for commercial applications. In a CDMA system the relatively narrowband users' information is spread into a much wider spectrum using a high clock (chip) rate [1]. The spreading codes have zero correlation with each other and if the spreading code is known to the receiver, it is possible to send multiple users' information on the same frequency spectrum without significant difficulty in detecting the desired signal at the receiver side. The signal from each user will have very low power and be seen by others as background noise. Therefore, as long as the total power of noise is less than a threshold, it is possible to detect the desired signal using the spreading code used to encode the signal at the transmitter. Using spread spectrum techniques, CDMA has become a dynamic channel allocation MA scheme that has no rigid channel allocation limitation for individual users. The number of users is also not fixed as in TDMA and FDMA, and a new user can be added to the system at any time. The upper limit for the maximum number of simultaneous users in the system using the same frequency spectrum is decided by the effect of total power of multi-user interference; thus, adding new users to a CDMA system will only cause graceful degradation of signal quality. That is why CDMA is a MA scheme that has no fixed maximum number of users which is different from TDMA and FDMA schemes [4].

The main disadvantage in CDMA is the bandwidth inefficiency due to bandwidth expansion. Additionally, the maximum number of users on a given channel with equal bit rate is limited by the number of sequences [4].

2.3.3 Low-Density Spreading (LDS)

A sparse spreading sequences are used in the place of the previous conventional dense spreading sequence in the case of Low-Density Spreading. LDS is an improved version of CDMA where the number of non-zero spreading sequences are much less than that of CDMA which leads to less cross-sequence interference. While orthogonal spreading sequences would significantly reduce the inter user interference, they are generally not designed for channel overloading. Another advantage is that LDS-CDMA can directly be converted to LDS-OFDM, where the chips are replaced by sub-carriers in OFDM [1].

2.3.4 Sparse-Code Multiple Access (SCMA)

Sparse-Code Multiple Access (SCMA) has similar analogy with that of the Low-Density Spreading, the bit streams are directly mapped to different sparse codewords in SCMA which makes it different and enhanced version of LDS spreads. In SCMA all coded bits of data streams are directly mapped to a codeword from a codebook which is built based on a multi-dimensional constellation. SCMA also has similar transmission model as the LTE transmission model where the joint design of multiplexing and spreading makes SCMA different from that of LTE [4].

A multiple access based on SCMA can be, Code domain non-orthogonal signal superposition where multiple symbols of different users are super imposed for transmission, Sparse spreading where the number of users symbol collision are minimized, and Multi-dimensional multiplexing which makes SCMA different from CDMA technique which uses linear spreading [1].

2.3.5 Multi-User Shared Access (MUSA)

MUSA is a new non-orthogonal multiple access scheme which operate in the code domain designed to minimize the scarce spectrum resource of wireless network. Multi-User Shared Access (MUSA) [1] is a NOMA spreading scheme where each users modulated data symbols are spread by a specially designed sequence that facilitates robust SIC implementation which makes it different from the sequences used by CDMA scheme. A technique called "Shared access" is used in MUSA to transmit each users spread symbols in the same radio resources using the principle of superposition in the code domain [1].

In MUSA, each user' symbols are spread by a multiple spreading sequence. This multiple spreading sequences constitute a pool from which each user can randomly pick one sequence. Different spreading sequences may also be used for different symbols for a single user. This may further improve the performance via interference averaging. Then, all spreading symbols are transmitted over the same time frequency resources. The spreading sequences should have low cross-correlation so as to have low interference. At the receiver, codeword-level Successive Interference Cancellation (SIC) is implemented to separate data of different users. The complexity of codeword-level SIC has less impact when it comes to uplink transmission, this is because that in the uplink as the receiver needs to decode the data for all users anyway [4].

2.3.6 Pattern-Division Multiple Access (PDMA)

The diversity gain of BS antennas are exploited in NOMA by using Pattern Division Multiple Access scheme which uses non-orthogonal patterns which is capable of increasing diversity gain by decreasing the interference among multiple users.

The diversity is not only actualized in the spatial domain, but also in time or frequency domain. The detection also uses a Message-Passing Algorithm (MPA) to compute the marginal functions of the global code constraint by iterative computation of a local code constraint. Although the codewords of PDMA do not have the low-density property in general, appropriate diversity order disparity can be observed where diversity order disparity leads to faster convergence of MPA. PDMA results in improved overall system capacity and error performance for each user, due to the diversity diversity gain either in time, frequency or spatial diversity [1].

2.4 Literature review

The performance comparison of non-orthogonal multiple access (NOMA) with that of orthogonal multiple access (OMA) has been validated different researchers. The performance of NOMA in 5G systems with randomly deployed users has been analyzed [5]. The channel model of this paper was analyzed using simulation and the result showed that NOMA which has different power allocation can achieve better outage performance than the orthogonal MA technique, under the condition that the users rates and power coefficients are carefully chosen.

Using partial channel state information at the BS, the performance of NOMA has also been addressed [6]. It addressed the downlink transmission with uniformly deployed users. This paper showed the performance of NOMA with CSI based on second order statistics (SOS) which close to that based on perfect CSI at low SNR, while the NOMA with SOS always outperforms NOMA with imperfect CSI. This paper addressed the performance of NOMA only for downlink scenario.

The performance of NOMA with one-bit feedback of its CSI to the base station has also been addressed [7]. They have derived a closed form expression for the common outage probability (COP), as well as the optimal diversity gains under short-term and long-term power constraints. Simulation results have been provided to demonstrate that the proposed NOMA schemes with one-bit feedback can outperform various existing multiple access schemes and achieve an outage performance close to the optimal one in many cases. Just like the previous papers, it addresses only downlink performance.

NOMA for 5G and beyond was discussed [8]. This paper is a survey which addressed with an emphasis of the following topics: the basic principles of NOMA, the amalgams of multiple antenna techniques and NOMA, the interplay of NOMA and cooperative communications, the resource control of NOMA, its coexistence with other key 5G techniques, and the implementation challenges and standardization. A survey like paper which tried to provide an overview on the rationales in incorporating massive multiple-input multiple-output (MIMO), non-orthogonal multiple access (NOMA), and interleave division multiple access (IDMA) in a unified framework [9].

Since power allocation has an enormous effect in NOMA, in order to suppress inter-user interferences and improve the achievable sum rate, lots of studies have been done to design power allocation in existing MIMO-NOMA systems. The early integration of NOMA and MIMO was investigated, where two users were considered in each beam with a random beamforming, and fixed power allocation schemes were utilized at the BS [10]. In addition, fixed power allocation strategies have also been considered in [11]. A coordinated frequency block-dependent inter-beam power allocation was proposed in [12] to generate distinct power levels for different beams. Magnus and Neudecher [13] considered equal power allocation for different groups, and intra-group power allocation has been optimized to maximize the achievable sum rate, where each group only included two single-antenna users. The intra-group power optimization has also been investigated in [14] and [15], where a convex optimization algorithm was utilized to obtain the closed-form solution to power allocation. In [16], a non-convex power allocation problem was formulated for MIMO NOMA systems, where only two users have been considered, and sub-optimal solutions were provided. Zhang et al.[17] investigated joint optimization of beamforming and power allocation, where channel uncertainties have been considered to maximize the worst-case achievable sum rate.

The Beam Division Multiple Access (BDMA) for massive MIMO downlink transmission has been discussed [18]. This paper deals with large scale antenna systems which have hundreds of antenna at the base station. The BDMA is an access method where an orthogonal beam is assigned for active users of the base station, i.e. users access the channel through different beams. They assumed a half duplex scheme which is the Frequency Division Duplexing (FDD) allowed the base station to have statistical channel state information by using feedback information from the mobile stations. They designed scheduling to separate users on non-overlapping beams. But again this paper accomplished only the downlink performance of BDMA since it is difficult to analyze both downlink and uplink performance only knowing the statistical CSI at the base station because of FDD scheme.

A paper also tried to show massive MIMO BDMA transmission with per-beam synchronization (PBS) in time and frequency for wideband massive MIMO over millimeter wave channels [19]. The proposed PBS can reduce channel delay and doppler spreads by a factor of the number of UT antennas compared with the conventional synchronization approaches.

BDMA in multi-cell massive MIMO communications for power allocation algorithm has also been addressed [20]. They have developed two power allocation algorithms: by Concave-Convex Procedure (CCCP) based power allocation and the deterministic equivalent (DE) based power allocation. Utilizing the deterministic equivalent, DE based power allocation algorithm can substantially reduce complexity. Simulation results demonstrated that the DE based power allocation algorithm exhibits a negligible performance loss compared with the CCCP based power allocation. Furthermore, both iterations in DE based algorithm can converge quickly.

All the papers, [21], [22], [11], [23], [3], and [24], discusses the Full Duplex (FD) technique existence either with NOMA or massive MIMO. Since in FD scheme, both user terminals and base stations uses the same frequency channel, it will double the spectral efficiency gain of FD over HD. These papers shown us that there are two basic interference that would decrease the two fold increase of SE of FD scheme. i.e. self-interference and cross-mode interference. Some of the papers have mitigated the effect of self-interference by using different techniques and validated their method using simulations.

In this paper,the Non-Orthogonal Multiple Access (NOMA) is combined with the Beam Division Multiple Access (BDMA) by assuming a FD scheme. The spectral and energy efficiency advantage of the Non-orthogonal Beam Division Multiple Access over the Tradional-MIMO and the Beam Division

Multiple Access is analyzed and validated. The Inter-Beam interference is mitigated by using wiener filter based linear precoding technique. To increase capacity, an iterative power allocation technique is used which also used for both Inter and Intra beam power optimization. Throughout the paper a FD based base station having several hundred antennas both for uplink and downlink transmission is used together with HD users. With an assumption of the FD scheme at the base station, there will be Self-Interference (SI) on the uplink antennas by the downlink antennas. In the uplink the paper considers two SI channel estimation techniques, the Minimum Mean Square Error (MMSE) and the Least Square (LS) channel estimators. By using this estimates an uplink spectral efficiency is analyzed mathematically and validated using MAT-LAB simulations.

Chapter 3

The Non-Orthogonal Beam Division Multiple Access

3.1 Introduction

This section considers a single-cell downlink mm-Wave communication system, where the base station (BS) is equipped with **M** antennas and $\mathbf{M}_{RF}$ radio frequency chains located in the center of the cell and **K** users having single antenna are simultaneously served by the BS.

3.2 Traditional MIMO in the Spatial Domain

As shown in the figure 3.1 below, for a traditional MIMO system in the spatial domain, the K x 1 received signal vector **r** for all **K** users in the downlink can be represented by [25]:

$$\mathbf{r} = \mathbf{H^H x} + \mathbf{w} \tag{3.1}$$

Where **H** is the channel matrix equal to $\mathbf{H} = [\mathrm{h}_1 , \mathrm{h}_2 , \mathrm{h}_3 ,..., \mathrm{h}_K]$ and $\mathbf{h}_K$ of size M x 1 is the M-dimensional channel vector between the BS and the K^{th} user which will be discussed in detail later, the transmitted signal, $\mathrm{x} = \mathrm{Gs} = \sum_{i=1}^{K} g_i s_i$, where s of size K x 1 is the original signal vector for all K users with normalized power $\mathrm{E(ss)}^H = \mathrm{I}_K$ and G of size M x K is the precoding matrix satisfying the total transmit power **P** as $\mathrm{E}[||\mathrm{x}||^2] = \mathrm{tr}(\mathrm{GG}^H \vartheta_s) \leq \mathrm{P}$ with ϑ_s denoting the diagonal correlation matrix of **s**. Finally $\mathbf{w} \sim \mathrm{CN}(0,\sigma^2 I_K)$ is the K x 1 Additive White Gaussian Noise (AWGN) vector [1].

From the figure below, we can easily understand that for a traditional MIMO systems, the number of required radio frequency chains is equal to the number of base station antennas.

Since the number of antennas for mm-Wave massive MIMO system is large which might be around 250, there should be some mechanism which could reduce the number of RF chains.

The channel matrix **H** governs the performance of the MU-MIMO link. Due to the highly directional and quasi-optical nature of propagation at mm Wave frequencies, LOS propagation is the predominant mode of propagation, which possibly a sparse set of single bounce multipath components [4]. Assuming that the LOS path exists for all mobile stations, let $\theta_{k,0}$, for k

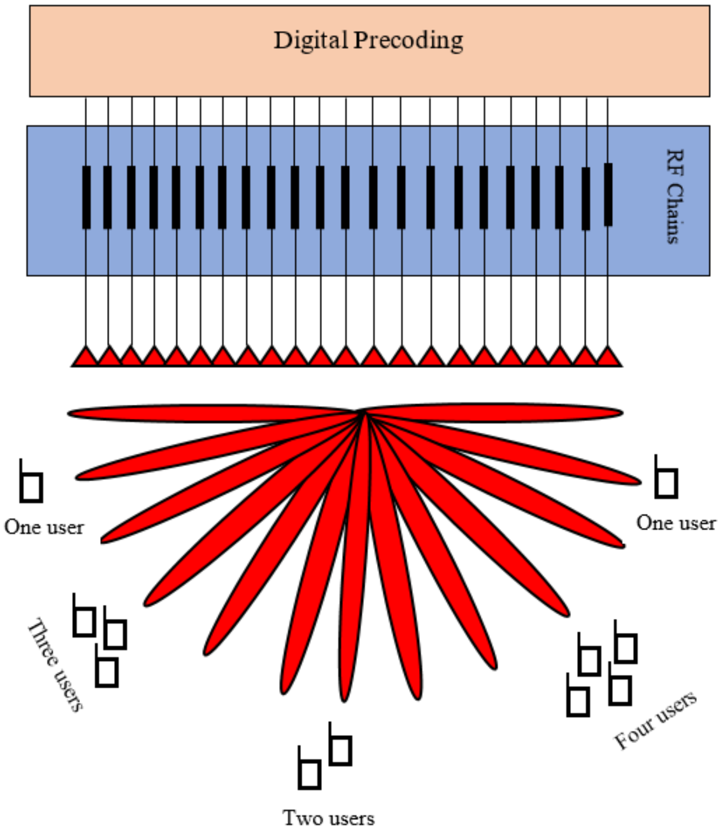

Figure 3.1: Traditional MIMO [10]

= 1,2,...,K, denote the LOS directions (spatial frequencies) for the K users. Then the LOS channel for the Kth user is h$_k$ = $\beta_{k,0}\, a(\theta_{k,0})$, where $\beta_{k,0}$ is the complex path loss or it can be denoted as the complex gain. In general, for a sparse multipath channels [25],

$$\mathbf{h_k} = \beta_{\mathbf{k,0}}\, \mathbf{a}(\theta_{\mathbf{k,0}}) + \sum_{\mathbf{l=1}}^{\mathbf{L}} \beta_{\mathbf{k,l}}\, \mathbf{a}(\theta_{\mathbf{k,l}}) \tag{3.2}$$

Where $\theta_{k,l}$ denote the path angles and $\beta_{k,l}$ represents the complex path losses associated with the different paths (non-line-of-sight paths) for the Kth user. And **L** is the total number of NLOS components, $a(\theta)$ is the M x 1 array steering vector.

For a critically spaced ULA, a plane wave in the direction of angle $\phi \in [-\pi/2, \pi/2]$ corresponds to a spatial frequency (direction), $\theta \in [-1/2, 1/2]$, given by [25]:

$$\theta = (\frac{\mathbf{d}}{\lambda})\mathbf{sin}(\phi) = \mathbf{0.5sin}(\phi) \tag{3.3}$$

Where θ is the physical direction, λ is the signal wavelength, and $\mathbf{d}$ is the antenna spacing satisfying d = $\lambda/2$ at mm-Wave frequencies. The corresponding M-dimensional array steering vector, a(θ), represents a plane-wave phase front associated with a point source in the direction θ and it is given by [25]:

$$\mathbf{a}(\theta) = \frac{\mathbf{1}}{\sqrt{\mathbf{M}}}(\mathbf{e}^{-\mathbf{j2}\pi\theta_{\mathbf{i}}}), \mathbf{i} \in \mathbf{I}(\mathbf{M}) \tag{3.4}$$

where,

$$\mathbf{I}(\mathbf{M}) = \mathbf{i} - (\mathbf{M} - \mathbf{1})/\mathbf{2}, \quad \text{for} \quad \mathbf{i} = \mathbf{0}, \mathbf{1}, \mathbf{2}, ..., \mathbf{M} - \mathbf{1} \tag{3.5}$$

Is a symmetric set of indices centered around zero.

3.3 Beam Division Multiple Access

In traditional MIMO systems as shown in the figure above, the received signal vector $\mathbf{r}$ = [r_1, r_2, ..., r_k]^T can be expressed as [26]:

$$\mathbf{r} = \mathbf{H^H Gs} + \mathbf{w} \tag{3.6}$$

Having $\mathbf{s}$ = [s_1, s_2, ..., s_k] is the K x 1 transmitted signal vector for all K users. $\mathbf{G}$ = [g_1, g_2, ..., g_k] is the M x K Wiener filter based linear precoding matrix with $||g_k||_2 = 1$ for k = 1,2, ..., K, and w is the noise vector following the distribution $\mathbf{CN}(0,\sigma^2 I_k)$. Having H as the M x K channel matrix, the channel vector h_k with size M x 1 has been stated in equation 3.2. Note that at mm-Wave frequencies, the complex path loss associated with NLOS component $\beta_{k,i}$ are typically 5 to 10 dB weaker than the complex gain $\beta_{k,0}$ of the LOS component [26].

As stated previously, in traditional MIMO systems, the number of required RF chains is equal to the number of BS antennas (M = M_{RF}), which is mostly large. For mm-Wave massive MIMO systems having several hundred antennas [10].

So, the direct application of massive MIMO at mm-Wave frequencies is prohibitive due to high hardware cost and energy consumption caused by RF chains [12].

In order to address this problem, the concept of beam division multiple access has been recently proposed, which can utilize lens antenna array to significantly reduce the number of required RF chains in mm-Wave massive MIMO systems.

So, by using lens antenna array, the channel in equation 3.2, which is in the spatial domain can be transformed in to the beam space model [13].

The mathematical function of the lens antenna array is based on the realization of the spatial Discrete Fourier Transform (DFT) with the M x M transform matrix **U** [10], which contains the array steering vectors of M-directions covering the entire space as shown below [10]:

$$\mathbf{U} = [\mathbf{a}(\bar{\theta_1}), \mathbf{a}(\bar{\theta_2}), ..., \mathbf{a}(\bar{\theta_M})]^{\mathbf{H}} \tag{3.7}$$

Where $\bar{\theta_m} = \frac{1}{M}$[m - (M+1)/2] for m = 1,2, ..., M are the pre-defined spatial directions. In fact, the transform matrix U is a unitary matrix satisfying, $U^H U = UU^H = I$. Then the overall received signal by all users will be represented as [10]:

$$\bar{\mathbf{r}} = \mathbf{H^H U^H G s} + \mathbf{w} = \bar{\mathbf{H}}^{\mathbf{H}} \mathbf{G s} + \mathbf{w} \tag{3.8}$$

Where $\bar{r}$ is the received signal vector in the beam space channel matrix $\bar{H}$ which is given by:

$$\bar{\mathbf{H}} = [\bar{\mathbf{h}}_\mathbf{1}, \bar{\mathbf{h}}_\mathbf{2}, ..., \bar{\mathbf{h}}_\mathbf{k}] = \mathbf{UH} = [\mathbf{U h_1}, \mathbf{U h_2}, ...\mathbf{U h_k}] \tag{3.9}$$

Where $\bar{h}_k$ is the beam space channel vector between the BS and the K^{th} user, which is the Fourier Transform of the spatial channel vector h_k in equation 3.2.

Just like the beam space channel described in [10], each rows of $\bar{H}$ corresponds to one beam, and all M rows correspond to M beams with spatial directions $\bar{\theta_1}, \bar{\theta_2}, ..., \bar{\theta_M}$, separately. In mm-Wave communications, since the number of dominant scatters is very limited, the number of NLOS components **L** is much smaller than the number of beams M [10]. Therefore the number of dominant elements of each beam space channel vector $\bar{h}_k$ is much smaller than M, indicating that the beam space channel matrix $\bar{H}$ has a sparse nature [14].

This sparse structure can be exploited to a dimension-reduced beam space MIMO systems without obvious performance loss by beam selection [26],[12].

More Generally, according to the sparse nature of the beam space channel matrix, only a small number of beams can be selected in order to simultaneously serve K users. Then the received signal vector in equation 3.6 can be written as:

$$\bar{\mathbf{r}} = \bar{\mathbf{H}}_\mathbf{r}^{\mathbf{H}} \mathbf{G_r s} + \mathbf{w} \tag{3.10}$$

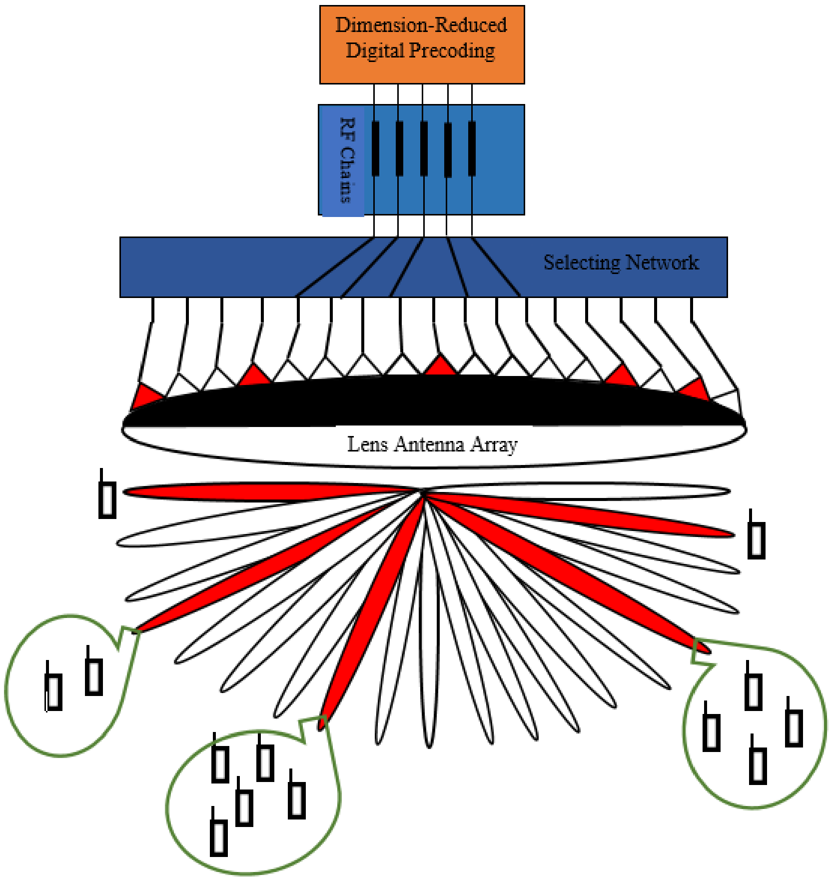

Figure 3.2: The beam division multiple access [26]

Where, $\bar{H}_r = \bar{H}(i,:)_{i\in\lambda}$ of size $|\lambda|$ x K is the dimension reduced beam space channel matrix including selected beams, and λ is the index set of selected beams. $\mathbf{G}_r$ of size $|\lambda|$ x K is the dimension reduced linear precoding matrix. Since the row dimension of G_r is much smaller than M, which is the row dimension of the original precoding matrix G, the number of RF chains can be significantly reduced and and we have $M_{RF} = |\lambda|$ [10].

As can be seen from the previously published papers on the beam division multiple access, there is no power optimization either among the beam or inside the beams. It means that all available power will be provided for all users within a specific beam equally (i.e, near and far users will be provided by the same power) and also the power allocated in between the beams is not optimized (i.e, beams having one user are provided with the same power with the beams having four or five users).

To break this fundamental barrier, this paper proposes a new mm-Wave transmission scheme that integrates Non-Orthogonal Multiple Access (NOMA) with the Beam Division Multiple Access scheme (BDMA) which will be discussed in detail later in the next section.

3.4 Beam Division Non-Orthogonal Multiple Access (BD-NOMA)

The Non-orthogonal multiple access has been proposed in the beam division multiple access for mm-Wave massive MIMO systems for further improvement of spectrum and energy efficiency.

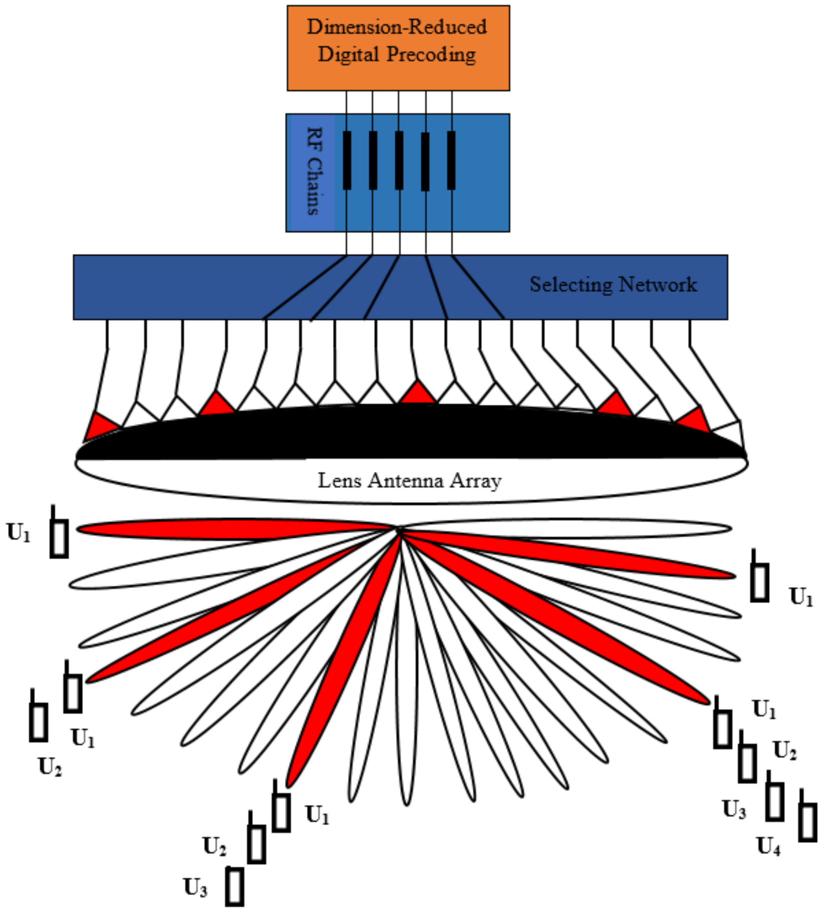

Figure 3.3: Beam Division Non-orthogonal Multiple Access (BD-NOMA) [13]

As it can be seen from figure 3.3, unlike existing beam division multiple access system, users of each beam, as per their distance from the base station, will be provided with different power. i.e., users which has better channel condition condition (near users) will be provided with less power than users having lower channel condition.

By having in mind that NOMA allows the overall available spectrum shared among users with superposition coding at the BS and successive interference cancellation (SIC) at the users, users selecting the same beam (conflicting users) will be served simultaneously using the same RF chain (the same beam) in the beam division non-orthogonal multiple access system.

Even though the number of selected beams is equal to the number of RF chain, the number of simultaneously served users can be larger than the number of RF chain, i.e., K $\geq M_{RF}$. Let S_m denote the set of users served by the m^{th} beam for m = 1,2,...,M_{RF} and assuring that users of the i^{th} beam is not the same as users of the j^{th} beam, $S_i \cap S_j = \phi$ for i $\neq$ j and $\sum_{m=1}^{M_{RF}} |S_m| = K$ [13].

The beam space channel vector of size $M_{RF} \times 1$ after beam selection between the base station and the t^{th} user in the m^{th} beam is denoted by $h_{t,m}$ which is analogous to equation 3.2 above.

And g_m of size $M_{RF} \times 1$ denotes the uniform linear array (ULA) precoding vector for users in the m^{th} beam.

In the Non-orthogonal beam division multiple access scheme, to pave the way for superposition coding at the transmitter and successive interference cancellation at the receiver, in this paper it is assumed that $||h_{1,m}^H g_m||_2 \geq ||h_{2,m}^H g_m||_2 \geq ... \geq ||h_{|s_m|,m}^H g_m||_2$ for m = 1,2, ..., M_{RF}. Having this in mind, the received signal vector of equation 3.8 can be modified in the non-orthogonal multiple access scenario as [13]:

$$\mathbf{\bar{r}} = \mathbf{H^H U^H G P s} + \mathbf{w} = \mathbf{\bar{H}^H G P s} + \mathbf{w} \tag{3.11}$$

Where P = diag$\{p\}$ includes the transmitted power for all K users where p = $[\sqrt{p_1}, \sqrt{p_2}, \sqrt{p_3}, ..., \sqrt{p_k}]$ satisfies $\sum_{k=1}^{K} p_k \leq P$, which is the maximum transmitted power at the BS.

So, the received signal $r_{t,m}$ at the t^{th} user in the m^{th} beam $(m = 1, 2, ..., M_{RF})$ and $(t = 1, 2, ..., |S_m|)$ can be given as:

$$\mathbf{r_{t,m}} = \mathbf{h_{t,m}^H} \sum_{\mathbf{j=1}}^{\mathbf{M_{RF}}} \sum_{\mathbf{i=1}}^{\mathbf{|S_j|}} \mathbf{g_j} \sqrt{\mathbf{p_{i,j}}} \mathbf{s_{i,j}} + \mathbf{w_{t,m}} \tag{3.12}$$

But from among those signals received at t^{th} user of the m^{th} beam, the required signal will be:

$$(\mathbf{r_{t,m}})_{\mathbf{required}} = \mathbf{h_{t,m}^H g_m} \sqrt{\mathbf{p_{t,m}}} \mathbf{s_{t,m}} \tag{3.13}$$

And the interference from other users of the m^{th} beam or the Intra beam interference at the t^{th} user will be:

$$(\mathbf{r_{t,m}})_{\mathbf{Intra-BI}} = \mathbf{h_{t,m}^{H} g_m} \sum_{\mathbf{i=1}}^{\mathbf{t-1}} \sqrt{\mathbf{p_{i,m}}} \mathbf{s_{i,m}} + \mathbf{h_{t,m}^{H} g_m} \sum_{\mathbf{i=t+1}}^{|\mathbf{s_m}|} \sqrt{\mathbf{p_{i,m}}} \mathbf{s_{i,m}} \tag{3.14}$$

And the interference from users of other or neighboring beams which is termed as the inter beam interference at the t^{th} user will be:

$$(\mathbf{r_{t,m}})_{\mathbf{Inter-BI}} = \mathbf{h_{t,m}^{H}} \sum_{\mathbf{j \neq m}} \sum_{\mathbf{i=1}}^{|\mathbf{s_j}|} \mathbf{g_j} \sqrt{\mathbf{p_{i,j}}} \mathbf{s_{i,j}} \tag{3.15}$$

Finally, the overall noise that will occur at t^{th} user of the m^{th} beam will be, $w_{t,m}$ which follows the distribution CN(0,σ^2).

Here, the precoding vectors should be carefully designed to mitigate inter beam interference which will be discussed in the later sections. Interference's with in the beam, the intra beam interference's which are caused by superposition coding at the BS can be mitigated by undergoing successive interference cancellation (SIC) according the increasing order of equivalent channel gains. i.e., it means that the t^{th} user in the m^{th} beam can remove the interference from the i^{th} user (for all i $>$ t) in the m^{th} beam by performing SIC.

Using NOMA in the beam division multiple access systems, allows higher number of users share the available spectrum with in a beam with respective of their distance from the base station. This makes it different from using BDMA alone having orthogonal power transfer for all users of a beam with a great lack of power optimization.

However, NOMA which have non-orthogonal power transfer cannot be guaranteed in practice to directly combine it with BDMA which have orthogonal transmission, this is because that users of one beam can suffer from interference's from other beams. So, power allocation among the non-orthogonal beams and between users, as well as precoding should be designed carefully to reduce interference's by maximizing the achievable sum rate.

3.5 Precoding Techniques in the Downlink Scenario

Here in this section, the precoding techniques will be discussed which will mitigate the effect of inter beam interference (IBI). Some of the linear precoding techniques include:

3.5.1 Matched Filter (MF)

The matched filter (MF) precoder is simply the conjugate transpose of the downlink channel matrix, i.e. [15],

$$\mathbf{G_{MF}} = \sqrt{\alpha}\mathbf{H^H} \tag{3.16}$$

Where α is a scaling factor to normalize signal power. MF precoder is also mostly known by its name maximum ratio transmission (MRT), which maximizes signal gain at the intended user [15]. It is the counterpart of the maximal ratio combining (MRC) receiver for uplink. The key performance parameters for single cell MF precoding, i.e., achievable sumrate and the total downlink transmit power, are discussed in [15].

With the increase of BS antennas, the channel vectors in **H** become closer to mutually orthogonal. As a result, the term $H^H H$ approaches a diagonal matrix [16], leading to the optimal solution. Consequently, MF precoding is near-optimal, as long as the number of BS antennas is much greater than the number of user terminals.

3.5.2 Zero Forcing (ZF)

Zero forcing precoding is another type of basic precoding technique, which eliminates the interference by transmitting the signal towards the intended user while nulling in the directions of other users. The ZF precoder is obtained by [16]:

$$\mathbf{G_{ZF}} = \sqrt{\alpha}\mathbf{H}(\mathbf{H^H H})^{\mathbf{-1}} \tag{3.17}$$

The term $H^H H$ forms a Gram matrix whose diagonal elements denote power imbalance among the channels, while the of-diagonal elements characterize mutual correlations between the channels. When highly correlated channels exist, ZF precoding decorrelates the channels at the price of losing channel capacity [17]. It is an optimal precoding scheme in the absence of additive noise. When additive noise is present, this precoding technique could amplify the noise together with the required signal.

3.5.3 Regularized Zero Forcing (RZF)

Regularized zero forcing precoder has been considered as the state-of-the-art linear precoder for MIMO wireless communication systems for its capability of trading off the advantages of MF and ZF precoders. The popularity of regularized zero forcing is also reflected by its alternative names such as eigenvalue-based beamforming [27], virtual signal-to-interference-noise ratio (SINR) Maximizing beamforming [25], transmit wiener filter [28], and signal-to-leakage-and-noise ratio (SLNR) maximizing beamforming [29]. The RZF precoding is given by:

$$\mathbf{G_{RZF}} = \sqrt{\alpha}\mathbf{H}(\mathbf{H^H H} + \mathbf{X} + \lambda \mathbf{I_k})^{\mathbf{-1}} \tag{3.18}$$

Which a ZF precoder regularized by a hermitian non-negative matrix $\mathbf{X}$ and a regularization factor λ.

The choices of X and λ are briefly discussed in [30]. If X = 0, then the above equations becomes an MF precoder when $\lambda \to \infty$ and a ZF precoder when $\lambda \to 0$. The extensions of the RZF technique with arbitrary user priorities are discussed in [31], where the RZF precoding matrix is modified to achieve optimality when the BS has the knowledge of statistical information of the user positions. This technique is referred to as position-aware RZF (PA-RZF) precoder. It turns out that PA-RZF is equivalent to the optimal linear precoder when the same SINR constraint and different path loss are imposed for all users. On the other hand, the commonly used RZF precoder becomes optimal only when the ratio between the SINR requirement and the average channel attenuation is the same for all users. Generally, the RZF precoder is obtained via minimizing the mean square error (MSE) between the transmitted and received symbols, which is thus also termed as minimum MSE (MMSE) precoder [32]. The computation of the RZF precoding matrix involves the inversion of a matrix with very large dimension, specially for large number of base station antennas and number of users.

Table 3.1: Advantages and disadvantages of the MF, ZF, and RZF precoding techniques

Precoding techniques	**Advantages**	**Disadvantages**
MF	a. Near optimal performance if there are more BS antennas than users b. Low computational complexity c. Better performance at lower SNRs	I. Unable to achieve full diversity at high spectral efficiency II. Suffering from error floors for all positive multiplexing gains III. Having lower achievable rate in the case of less BS antennas IV. Not robust against inter-user interference
ZF	a. Low computational complexity as compared to RZF based on matrix inversion computation b. Higher power efficiency c. Better performance at high SNRs d. Decoupling a multi-user channel in to independent single-user channels e. Achieving a large portion of dirty paper coding capacity	I. Noise amplification and power penalty if the channel is highly correlated II. Unable to support too many users III. Medium complexity
RZF	a. Guaranteed optimality if the ratio between the SINR requirement and the average channel attenuation is the same for all users	I. Requiring matrix inverse calculation, leading to high complexity II. Suffering from error floors for all positive multiplexing gains

3.5.4 Complexity of precoding techniques

The computational complexity of the above three linear precoding schemes are analyzed based on only number of multiplications. This is because that other computations like addition and inverse computations are easy to implement in hardware. So, based on the complex multiplications, Regularized Zero Force precoding has almost the same number of multiplication with that of Zero Forcing precoding matrix. While the Matched Filter based precoding requires less number of multiplications resulting in lower complexity. The complexity is summarized in the table below.

Table 3.2: Computational complexity analysis of linear precoding schemes

Scheme	**Transmission multiplications**	**Computing precoding vectors multiplications**
RZF	$\mathbf{M} * (\mathbf{K^2})$	$\mathbf{2} * ((\mathbf{M^2}) * \mathbf{K})$
ZF	$\mathbf{M} * (\mathbf{K^2})$	$\mathbf{2} * ((\mathbf{M^2}) * \mathbf{K})$
MF	$\mathbf{M} * (\mathbf{K^2})$	-

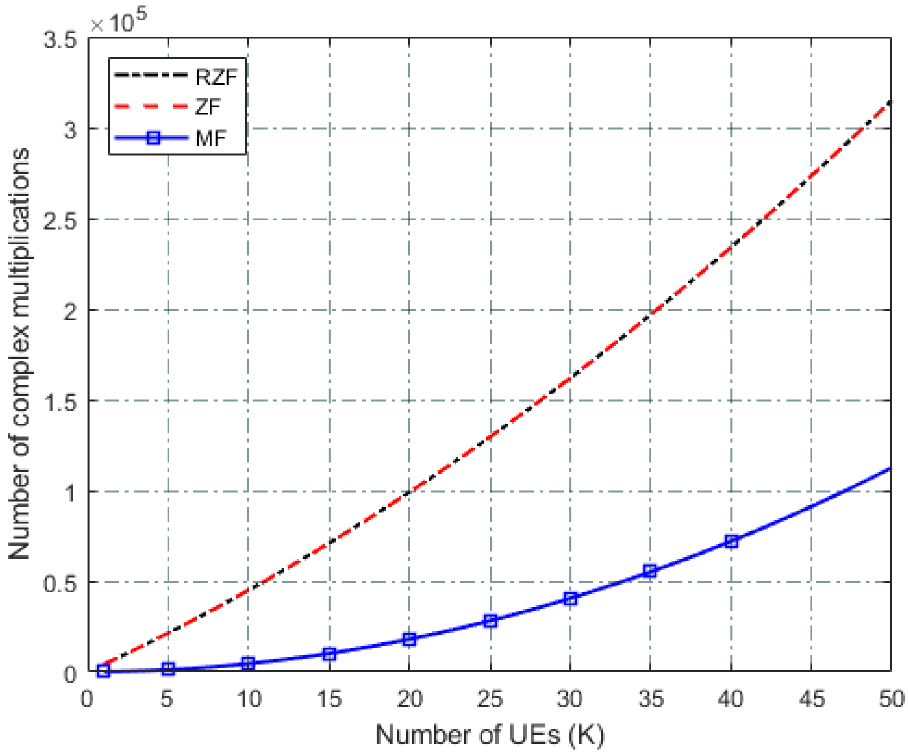

Figure 3.4: Complexity analysis of precoding techniques [25]

From among those linear precoding techniques, the regularized zero forcing or Wiener Filter based precoding technique is used because of its optimality as shown in the table 3.1 above. This precoding technique is realized by the pseudo-inverse of the beam space channel matrix for all users with in a specified beam. However, in the beam division non orthogonal multiple access scheme, the number of users is larger than the number of available beams, K $\geq M_{RF}$, which means that the pseudo-inverse of the beam space channel matrix of size $M_{RF} \times K$ does not exist. As a result, the conventional RZF precoding cannot be directly used.

In order to overcome this problem, an equivalent channel can be determined for each beam to generate the precoding vector. There are different methods which are used to generate the equivalent channel for each beam, like the Singular Value Decomposition (SVD) based equivalent channel and the strongest user based equivalent channel [10].

We mostly use the former technique, when Line-of-Sight (LOS) component does not exist in the multipath environment and the later technique is selected or preferred when the LOS component exists in the multipath environment and it is known that user near the BS has better channel condition from among other users with in a beam and uses the LOS component.

In this paper, since the concern is mm Wave massive MIMO where cells of the BS are assumed to have smaller size (i.e., BS are located with a small distance to each other), LOS components frequently exist with in a specific beam and this is the reason why strong user based equivalent channel is chosen [1].

3.5.5 Strong User Based Equivalent Channel

Since the amplitudes of non-LOS (NLOS) components are typically 5 to 10 dB weaker than the amplitudes of the LOS component [26], the LOS component can primarily characterize the multipath channel in mm Wave communications [10]. The channel matrix of the BD-NOMA has the sparse structure representing the directions of different users [26]. So, if the LOS component exists, the sparse beam space channel vectors of different users within the same beam are highly correlated [10].

Therefore, one of the channel vector of BD-NOMA for multiplexed users in the m^{th} beam can be regarded as the equivalent channel vector of the m^{th} beam.

Considering that the first user in each beam having the LOS component should perform successive Interference Cancellation (SIC) to decode all the other users signals in this beam, the channel vector of the first user in each beam as the equivalent channel vector of that particular beam has been selected. Therefore, the equivalent channel matrix of size $M_{RF} \times M_{RF}$ for all selected beams can be written as [13]:

$$\tilde{\mathbf{H}} = [\mathbf{h_{1,1}}, \mathbf{h_{1,2}}, ..., \mathbf{h_{1,M_{RF}}}] \tag{3.19}$$

Then, the precoding matrix of size $M_{RF} \times M_{RF}$ can be generated by [15]:

$$\tilde{\mathbf{G}} = \sqrt{\alpha}\tilde{\mathbf{H}}(\tilde{\mathbf{H}}^{\mathbf{H}}\tilde{\mathbf{H}} + \mathbf{X} + \lambda\mathbf{I_k})^{-\mathbf{1}} \tag{3.20}$$

As we discussed earlier, $\mathbf{X}$ is a hermitian non-negative matrix, λ is a regularization factor, and α is a scaling factor to normalize signal power [13].

$$\tilde{\mathbf{G}} = [\tilde{\mathbf{g}}_{\mathbf{1}}, \tilde{\mathbf{g}}_{\mathbf{2}}, ..., \tilde{\mathbf{g}}_{\mathbf{M_{RF}}}] \tag{3.21}$$

After normalizing the precoding vectors, the precoding vector for the m^{th} beam (m = 1, 2, ..., M_{RF}) can be written as:

$$\mathbf{g_m} = \frac{\tilde{\mathbf{g}}_{\mathbf{m}}}{||\tilde{\mathbf{g}}_{\mathbf{m}}||_{\mathbf{2}}} \tag{3.22}$$

So, here the first user in each beam can completely remove the inter-beam interference's, i.e. [13],

$$\frac{\mathbf{h_{1,j}^H g_m}}{||\mathbf{h_{1,j}^H g_m}||_\mathbf{2}} = \begin{cases} 0 & \text{for} \quad j \neq m \\ 1 & \text{for} \quad j = m \end{cases} \tag{3.23}$$

Where $1 \leq j, m \leq M_{RF}$. Thus, after performing SIC, the SINR at the first user in the m^{th} beam can be written as [13]:

$$\gamma_{\mathbf{1,m}} = \frac{||\mathbf{h_{1,m}^H g_m}||_\mathbf{2}^\mathbf{2} \mathbf{P_{1,m}}}{\sigma^\mathbf{2}} \tag{3.24}$$

3.6 Achievable Sum Rate and Power Allocation

3.6.1 Achievable Sum Rate

As it is discussed above, using NOMA with superposition coding and SIC in the m^{th} beam, the i^{th} (i > t) user's signal is detectable at the t^{th} user, knowing that it is detectable at itself [33], as the equivalent channel gain of the t^{th} user is larger than that of the i^{th} user, i.e., $||h_{1,m}^H g_m||_2 \geq ||h_{2,m}^H g_m||_2 \geq ... \geq ||h_{|s_m|,m}^H g_m||_2$ [13] as discussed before.

So it means that the t^{th} user can detect the i^{th} user's signals for $1 \leq t < i \leq |S_m|$, and then remove the detected signals (signals of users which are located farther from the t^{th} user starting from the BS) from its received signals successively.

As a result, the intra beam interference of equation 3.14 will be changed by removing the second term as [13]:

$$(\mathbf{\hat{r}_{t,m}})_{\mathbf{Intra-BI}} = \mathbf{h_{t,m}^H g_m} \sum_{\mathbf{i=1}}^{\mathbf{t-1}} \sqrt{\mathbf{p_{i,m}}} \mathbf{s_{i,m}} \tag{3.25}$$

So, the overall remaining received signal at the t^{th} user in the m^{th} beam can be written as:

$$(\mathbf{\hat{r}_{t,m}}) = (\mathbf{r_{t,m}})_{\mathbf{required}} + (\mathbf{\tilde{r}_{t,m}})_{\mathbf{Intra-BI}} + (\mathbf{r_{t,m}})_{\mathbf{Inter-BI}} + \mathbf{w_{t,m}} \tag{3.26}$$

Where, the $(r_{t,m})_{required}$ and $(r_{t,m})_{Inter-BI}$ are the required and the inter beam interference terms of equation 3.26 are already described in equations 3.13 and 3.15 respectively having in mind that $w_{t,m}$ is the noise term of the t^{th} user in the m^{th} beam.

According to the above equation, the signal to interference plus noise ratio (SINR) of the t^{th} user in the m^{th} beam can be expressed as [13]:

$$\psi_{\mathbf{t,m}} = \frac{||\mathbf{h}^{\mathbf{H}}_{\mathbf{t,m}}\mathbf{g}_{\mathbf{m}}||^{\mathbf{2}}_{\mathbf{2}}\mathbf{P}_{\mathbf{t,m}}}{\tau_{\mathbf{t,m}}} \tag{3.27}$$

Where,

$$\tau_{\mathbf{t,m}} = ||\mathbf{h}^{\mathbf{H}}_{\mathbf{t,m}}\mathbf{g}_{\mathbf{m}}||^{\mathbf{2}}_{\mathbf{2}}\sum_{\mathbf{i=1}}^{\mathbf{t-1}}\mathbf{p}_{\mathbf{i,m}} + \sum_{\mathbf{j\neq m}}||\mathbf{h}^{\mathbf{H}}_{\mathbf{t,m}}\mathbf{g}_{\mathbf{j}}||^{\mathbf{2}}_{\mathbf{2}}\sum_{\mathbf{i=1}}^{|\mathbf{s_j}|}\mathbf{p}_{\mathbf{i,j}} + \sigma^{\mathbf{2}} \tag{3.28}$$

The achievable rate at the t^{th} user in the m^{th} beam is given finally by:

$$\mathbf{R}_{\mathbf{t,m}} = \log_{\mathbf{2}}^{(\mathbf{1+SINR})} = \log_{\mathbf{2}}^{(\mathbf{1}+\psi_{\mathbf{t,m}})} \tag{3.29}$$

Finally, the achievable sum rate of the beam division non-orthogonal multiple access (BD-NOMA) can be given as [13]:

$$\mathbf{R}_{\mathbf{sum}} = \sum_{\mathbf{m=1}}^{\mathbf{M_{RF}}}\sum_{\mathbf{t=1}}^{|\mathbf{S_m}|}\mathbf{R}_{\mathbf{t,m}} \tag{3.30}$$

Having in mind that the first summation is for a number of selected beams from among $\mathbf{M}$ beams of the system, while the second summation is for the group of (number of) users located in a specific beam. The achievable sum rate of BD-NOMA can be improved by a careful design of the precoding and power optimization.

Generally, as it is described previously, this paper assumes that the BD-NOMA channel is known by the BS, which means that there is a perfect channel state information (CSI) known by the BS. Actually, efficient tools of compressive sensing can be utilized to reliably estimate the channel with low pilot overhead. This will be possible because of the sparsity of the beam space channel in mm-Wave massive MIMO systems [26]. And also the other assumption is that there is a perfect scheduling of users which paves the way for the power allocation as well as precoding to take place.

3.6.2 Power Allocation

Efficient power allocation strategy plays an important role on the performance of NOMA systems. This is because that the channel gain difference among users can be translated into multiplexing gains by superposition coding in NOMA scheme [25].

Having this in mind, the target is to improve the achievable sum rate of the BD-NOMA system by reducing both inter-beam interference and intra-beam interference by using efficient optimization technique. However, in most of the existing papers, there is fixed inter-beam power allocation and fixed number of users are allowed in each beam. In contrast to this, the BD-NOMA allows multiple users with multiple antennas in each beam and dynamic power allocation scheme is used to maximize the achievable sum rate by solving the joint power optimization problem, which includes not only the intra-beam power optimization, but also considers the inter-beam power optimization.

The power allocation problem can be formulated as [13]:

$$\mathbf{max}|_{\mathbf{P_{t,m}}} \sum_{\mathbf{m=1}}^{\mathbf{M_{RF}}} \sum_{\mathbf{t=1}}^{|\mathbf{S_m}|} \mathbf{R_{t,m}} \tag{3.31}$$

In order for this problem to be solved, there has to be three conditions such as:

$$\text{Condition}_{\mathbf{1}} : \mathbf{R_{t,m}} \geq \mathbf{R_{min}}, \forall_{\mathbf{m,t}} \tag{3.32}$$

$$\text{Condition}_{\mathbf{2}} : \mathbf{P_{t,m}} \geq \mathbf{0}, \forall_{\mathbf{m,t}} \tag{3.33}$$

$$\text{Condition}_{\mathbf{3}} : \sum_{\mathbf{m=1}}^{\mathbf{M_{RF}}} \sum_{\mathbf{t=1}}^{|\mathbf{S_m}|} \mathbf{P_{t,m}} \leq \mathbf{P} \tag{3.34}$$

Here, as discussed previously, $R_{t,m}$ is the achievable rate of the t^{th} user in the m^{th} beam. The first constraint shows that the data rate constraint for each user with R_{min} being the minimum data rate for each user. The second constraint shows that the power allocated to each user must be positive and the last constraint which is the transmitted power constraint with $\mathbf{P}$ being the maximum total transmitted power by the BS.

The optimization problem can also be written as [13]:

$$\mathbf{max}|_{\mathbf{P_{t,m}}} \sum_{\mathbf{m=1}}^{\mathbf{M_{RF}}} \sum_{\mathbf{t=1}}^{|\mathbf{S_m}|} \log_{\mathbf{2}}^{(\mathbf{1+SINR})} = \mathbf{max}|_{\mathbf{P_{t,m}}} \sum_{\mathbf{m=1}}^{\mathbf{M_{RF}}} \sum_{\mathbf{t=1}}^{|\mathbf{S_m}|} \log_{\mathbf{2}}^{(\mathbf{1+\psi_{t,m}})} \tag{3.35}$$

As we could observe, the above equation is a concave function that do not converge. So, the optimization problem is an NP-hard problem which means that it is a non-deterministic polynomial-time hard problem which could be difficult to obtain the closed form solution to the optimal power allocation problem of equation 3.35.

In order to solve this concave problem, an iterative optimization algorithm is proposed in this paper to realize power allocation. According to the extension of the matrix inversion lemma formula [34],

$$(\mathbf{A}+\mathbf{UBV})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}(\mathbf{I}+\mathbf{BVA}^{-1}\mathbf{U})^{-1}\mathbf{BVA}^{-1} \tag{3.36}$$

Where, A, U, B, and V denotes matrices of the correct conformable size. It can be inverted using blockwise matrix inversion technique.

Special case of the above expression is when **B** is the 1 x 1 unit matrix, and this identity reduces to the Sherman-Morrison formula. In the special case when **B** is the identity matrix **I**, the matrix $I + VA^{-1}U$ is known in numerical linear algebra and numerical partial differential equations as the capacitance matrix.

So, according to the above equation, we have [13]:

$$\begin{aligned}(\mathbf{1}+\psi_{\mathbf{t,m}})^{-1} &= (\mathbf{1}+\frac{||\mathbf{h^H_{t,m}g_m}||^2_2\mathbf{P_{t,m}}}{\tau_{\mathbf{t,m}}})^{-1}\\ &= \mathbf{1}-\frac{||\mathbf{h^H_{t,m}g_m}||^2_2\mathbf{P_{t,m}}}{\tau_{\mathbf{t,m}}}(\frac{||\mathbf{h^H_{t,m}g_m}||^2_2\mathbf{P_{t,m}}}{\tau_{\mathbf{t,m}}}+\mathbf{I})^{-1}\\ &= \mathbf{1}-||\mathbf{h^H_{t,m}g_m}||^2_2\mathbf{P_{t,m}}(\tau_{\mathbf{t,m}}+||\mathbf{h^H_{t,m}g_m}||^2_2\mathbf{P_{t,m}})^{-1}\end{aligned} \tag{3.37}$$

Where m = 1, 2, ..., M_{RF} and t = 1, 2, ..., $|S_m|$. As we can see, the expression above has similar form as the minimum mean square error (MMSE). Based on the following lemma of MMSE.

Lemma$_1$: Let **x** be n x 1 hidden random vector variable, and let **y** be a m x 1 known random vector variable (the measurement and observation), both of them not necessarily of the same dimension. An estimator $\hat{x}(y)$ of x is any function of the measurement y. The estimation error vector is given by $e = \hat{x} - x$ and its mean square error (MSE) is given by the trace of error covariance matrix [13], [25].

$$\begin{aligned}\mathbf{MSE} &= \mathbf{tr}\Big\{\mathbf{E}\big\{(\hat{\mathbf{x}}-\mathbf{x})(\hat{\mathbf{x}}-\mathbf{x})^{\mathbf{H}}\big\}\Big\}\\ &= \mathbf{E}\big\{(\hat{\mathbf{x}}-\mathbf{x})^{\mathbf{H}}(\hat{\mathbf{x}}-\mathbf{x})\big\}\end{aligned} \tag{3.38}$$

Where the expectation **E** is taken over both **x** and **y**. When x is a scalar variable, the MSE expression simplifies to $E\big\{(\hat{x}-x)^2\big\}$ [25]. Note that MSE can equivalently be defined in other ways, since

$$\begin{aligned} \mathbf{tr}\Big\{\mathbf{E}\{\mathbf{ee^H}\}\Big\} &= \mathbf{E}\Big\{\mathbf{tr}\{\mathbf{ee^H}\}\Big\} \\ &= \mathbf{E}\{\mathbf{e^H e}\} \\ &= \sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{E}\{\mathbf{e_i^2}\} \end{aligned} \tag{3.39}$$

The MMSE estimator is then defined as the estimator achieving minimal MSE [25].

$$\hat{\mathbf{x}}_{\mathbf{MMSE}}(\mathbf{y}) = \text{argmin}_{\hat{\mathbf{X}}} \quad \mathbf{MSE} \tag{3.40}$$

By relying on the above Lemma of MMSE estimator, we can solve for $\mathbf{s_{t,m}}$ from $\hat{\mathbf{r}}_{\mathbf{t,m}}$ of equation 3.26 and the detection problem can be formulated as [13]:

$$\mathbf{d^0_{t,m}} = \text{argmin}_{\mathbf{d_{t,m}}} \quad \mathbf{e_{t,m}} \tag{3.41}$$

Where,

$$\mathbf{e_{t,m}} = \mathbf{E}\Big\{|\mathbf{s_{t,m}} - \mathbf{d_{t,m}}\hat{\mathbf{r}}_{\mathbf{t,m}}|^{\mathbf{2}}\Big\} \tag{3.42}$$

is the Mean Square Error (MSE), and $\mathbf{d_{t,m}}$ is the channel equalization coefficient, and $\mathbf{d^0_{t,m}}$ is the optimal value of $\mathbf{d_{t,m}}$ in order to minimize the MSE.

By substituting the expression for $\hat{\mathbf{r}}_{\mathbf{t,m}}$ of equation 3.26 in to equation 3.42, we have [13]:

$$\begin{aligned} \mathbf{e_{t,m}} &= \mathbf{E}\Big\{|\mathbf{s_{t,m}} - \mathbf{d_{t,m}}\hat{\mathbf{r}}_{\mathbf{t,m}}|^{\mathbf{2}}\Big\} \\ &= |\mathbf{1} - \mathbf{d_{t,m}}\sqrt{\mathbf{P_{t,m}}}\mathbf{h^H_{t,m}g_m}|^{\mathbf{2}} + |\mathbf{d_{t,m}}|^{\mathbf{2}}||\mathbf{h^H_{t,m}g_m}||^{\mathbf{2}}_{\mathbf{2}}\sum_{\mathbf{i=1}}^{\mathbf{t-1}}\mathbf{p_{i,m}} \\ &+ |\mathbf{d_{t,m}}|^{\mathbf{2}}\sum_{\mathbf{j\neq m}}||\mathbf{h^H_{t,m}g_j}||^{\mathbf{2}}_{\mathbf{2}}\sum_{\mathbf{i=1}}^{|\mathbf{s_j}|}\mathbf{p_{i,j}} + |\mathbf{d_{t,m}}|^{\mathbf{2}}\sigma^{\mathbf{2}} \end{aligned} \tag{3.43}$$

Where the full derivation of equation 3.43 is found in the appendix-2. Then by solving equation 3.41 based on equation 3.43, the optimal equalization coefficient $\mathbf{d^0_{t,m}}$ can be calculated by [13]:

$$\frac{\partial \mathbf{e_{t,m}}}{\partial \mathbf{d_{t,m}}}\bigg\|_{\mathbf{d^0_{t,m}}} = \mathbf{0} \tag{3.44}$$

Then, we have,

$$\mathbf{d^0_{t,m}} = (\sqrt{\mathbf{p_{t,m}}}\mathbf{h^H_{t,m}g_m})(\mathbf{p_{t,m}}||\mathbf{h^H_{t,m}g_m}||^2_2 + \tau_{\mathbf{t,m}})^{-1} \tag{3.45}$$

Again the full derivation of equation 3.45 is shown in appendix-3. By substituting the values of equation 3.45 in to equation 3.43 we have [13],

$$\begin{aligned}
\mathbf{e^0_{t,m}} &= \mathbf{E}\Big\{|\mathbf{s_{t,m}} - \mathbf{d^0_{t,m}}\hat{\mathbf{r}}_{\mathbf{t,m}}|^2\Big\} \\
&= |\mathbf{1} - \mathbf{d^0_{t,m}}\sqrt{\mathbf{P_{t,m}}}\mathbf{h^H_{t,m}g_m}|^2 + |\mathbf{d^0_{t,m}}|^2||\mathbf{h^H_{t,m}g_m}||^2_2\sum_{\mathbf{i=1}}^{\mathbf{t-1}}\mathbf{p_{i,m}} \\
&+ |\mathbf{d^0_{t,m}}|^2\sum_{\mathbf{j\neq m}}||\mathbf{h^H_{t,m}g_j}||^2_2\sum_{\mathbf{i=1}}^{|\mathbf{s_j}|}\mathbf{p_{i,j}} + |\mathbf{d^0_{t,m}}|^2\sigma^2 \\
&= \mathbf{1} - \mathbf{p_{t,m}}||\mathbf{h^H_{t,m}g_m}||^2_2(\mathbf{p_{t,m}}||\mathbf{h^H_{t,m}g_m}||^2_2 + \tau_{\mathbf{t,m}})^{-1}
\end{aligned} \tag{3.46}$$

As we could see equation 3.46, it shows that it is equal to the MMSE of equation of 3.37. We can write as [13],

$$(\mathbf{1} + \psi_{\mathbf{t,m}})^{-1} = \min_{\mathbf{d_{t,m}}} \quad \mathbf{e_{t,m}} \tag{3.47}$$

Then, the achievable rate of the t^{th} user in the m^{th} beam can be written as [13]:

$$\mathbf{R_{t,m}} = \log_{\mathbf{2}}^{(\mathbf{1}+\psi_{\mathbf{t,m}})} = \max_{\mathbf{d_{t,m}}}(-\log_{\mathbf{2}}^{\mathbf{e_{t,m}}}) \tag{3.48}$$

In order to remove the logarithmic function in the above equation, we can use the following proposition [34].

Proposition 1 : Let $\mathbf{f(a)} = -\frac{\mathbf{ab}}{\mathbf{ln2}} + \log_{\mathbf{2}}^{\mathbf{a}} + \frac{\mathbf{1}}{\mathbf{ln2}}$ and $\mathbf{a}$ be a positive real number, we have,

$$\max|_{\mathbf{a>0}} \quad \mathbf{f(a)} = -\log_{\mathbf{2}}^{\mathbf{b}} \tag{3.49}$$

Where, the optimal value of $\mathbf{a}$ is $\mathbf{a}^0 = \frac{1}{b}$.

Proof : The function $\mathbf{f(a)}$ is concave, and thus the maximum value of $\mathbf{f(a)}$ can be obtained by solving [34]:

$$\frac{\partial \mathbf{f(a)}}{\partial \mathbf{a}}\bigg\|_{\mathbf{a=a^0}} = \mathbf{0} \tag{3.50}$$

Then, we have $\mathbf{a}^0 = \frac{1}{b}$. By substituting a^0 in to $\mathbf{f(a)}$, the maximum value of $\mathbf{f(a)}$ is $-\log_2^b$.

By using proposition 1, the achievable sum rate of equation 3.48 can be written as [13]:

$$\mathbf{R_{t,m}} = \max|_{d_{t,m}} \max|_{a_{t,m}>0}(-\frac{\mathbf{a_{t,m}e_{t,m}}}{\mathbf{ln2}} + \log_{\mathbf{2}}^{\mathbf{a_{t,m}}} + \frac{\mathbf{1}}{\mathbf{ln2}}) \tag{3.51}$$

Finally, the objective function for the optimization problem in equation 3.35 has been transformed in to quadratic programming function as it can be shown below [13]:

$$\max|_{\{p_{t,m}\}} \sum_{m=1}^{M_{RF}} \sum_{t=1}^{|s_m|} \max|_{d_{t,m}} \max|_{a_{t,m}>0}(-\frac{\mathbf{a_{t,m}e_{t,m}}}{\mathbf{ln2}} + \log_{\mathbf{2}}^{\mathbf{a_{t,m}}} + \frac{\mathbf{1}}{\mathbf{ln2}}) \tag{3.52}$$

Then, in order to solve the re-arranged optimization problem in the above equation, this paper assumes to iteratively optimize $\{\mathbf{d}_{t,m}\}$, $\{\mathbf{a}_{t,m}\}$, and $\{\mathbf{p}_{t,m}\}$. Having this in mind, given the optimal power allocation solution $\{p_{t,m}^{y-1}\}$ in the (y-1)th iteration, the optimal solution of $\{d_{t,m}^{y}\}$ in the y^{th} iteration can be obtained according to equation 3.45 above as [13]:

$$\mathbf{d_{t,m}^{(y)}} = (\sqrt{\mathbf{p_{t,m}^{(y-1)}}}\mathbf{h_{t,m}^{H}g_m})(\mathbf{p_{t,m}^{(y-1)}}||\mathbf{h_{t,m}^{H}g_m}||_{\mathbf{2}}^{\mathbf{2}} + \tau_{\mathbf{t,m}}^{\mathbf{(y-1)}})^{-\mathbf{1}} \tag{3.53}$$

Where the noise term is given by [13],

$$\tau_{\mathbf{t,m}}^{\mathbf{(y-1)}} = ||\mathbf{h_{t,m}^{H}g_m}||_{\mathbf{2}}^{\mathbf{2}} \sum_{\mathbf{i=1}}^{\mathbf{t-1}} \mathbf{p_{i,m}^{(y-1)}} + \sum_{\mathbf{j \neq m}} ||\mathbf{h_{t,m}^{H}g_j}||_{\mathbf{2}}^{\mathbf{2}} \sum_{\mathbf{i=1}}^{\mathbf{|s_j|}} \mathbf{p_{i,j}^{(y-1)}} + \sigma^{\mathbf{2}} \tag{3.54}$$

As a result the corresponding MMSE expressed by equation 3.46 in the yth iteration can be obtained by [13]:

$$\mathbf{e_{t,m}^{0(y)}} = \mathbf{1} - \mathbf{p_{t,m}^{(y-1)}}||\mathbf{h_{t,m}^{H}g_m}||_{\mathbf{2}}^{\mathbf{2}}(\mathbf{p_{t,m}^{(y-1)}}||\mathbf{h_{t,m}^{H}g_m}||_{\mathbf{2}}^{\mathbf{2}} + \tau_{\mathbf{t,m}}^{\mathbf{(y-1)}})^{-\mathbf{1}} \tag{3.55}$$

Then according to proposition$_1$, the optimal solution of $\{a_{t,m}^{(y)}\}$ in the y^{th} iteration can be obtained by:

$$\mathbf{a_{t,m}^{(y)}} = \frac{\mathbf{1}}{\mathbf{e_{t,m}^{0(y)}}} \tag{3.56}$$

After getting the optimal values $\left\{d_{t,m}^{(y)}\right\}$ and $\left\{a_{t,m}^{(y)}\right\}$ in the y^{th} iteration, the optimal $\left\{p_{t,m}^{(y)}\right\}$ in the y^{th} iteration can be obtained by solving the following problem [13]:

$$\mathbf{min}|_{\mathbf{P_{t,m}^{(y)}}} \sum_{\mathbf{m=1}}^{\mathbf{M_{RF}}} \sum_{\mathbf{t=1}}^{|\mathbf{S_m}|} \mathbf{a_{t,m}^{(y)} e_{t,m}^{(y)}} \tag{3.57}$$

With the previous three constraints, such that:

$$\text{Condition}_\mathbf{1} : \mathbf{R_{t,m}} \geq \mathbf{R_{min}}, \quad \forall_{\mathbf{m,t}} \tag{3.58}$$

$$\text{Condition}_\mathbf{2} : \mathbf{P_{t,m}^{(y)}} \geq \mathbf{0}, \quad \forall_{\mathbf{m,t}} \tag{3.59}$$

$$\text{Condition}_\mathbf{3} : \sum_{\mathbf{m=1}}^{\mathbf{M_{RF}}} \sum_{\mathbf{t=1}}^{|\mathbf{S_m}|} \mathbf{P_{t,m}^{(y)}} \leq \mathbf{P} \tag{3.60}$$

Where,

$$\begin{aligned} \mathbf{e_{t,m}^{(y)}} &= |\mathbf{1} - \mathbf{d_{t,m}^{(y)}} \sqrt{\mathbf{P_{t,m}^{(y)}}} \mathbf{h_{t,m}^H g_m}|^\mathbf{2} + |\mathbf{d_{t,m}^{(y)}}|^\mathbf{2} ||\mathbf{h_{t,m}^H g_m}||_\mathbf{2}^\mathbf{2} \sum_{\mathbf{i=1}}^{\mathbf{t-1}} \mathbf{p_{i,m}^{(y)}} \\ &+ |\mathbf{d_{t,m}^{(y)}}|^\mathbf{2} \sum_{\mathbf{j \neq m}} ||\mathbf{h_{t,m}^H g_j}||_\mathbf{2}^\mathbf{2} \sum_{\mathbf{i=1}}^{|\mathbf{s_j}|} \mathbf{p_{i,j}^{(y)}} + |\mathbf{d_{t,m}^{(y)}}|^\mathbf{2} \sigma^\mathbf{2} \end{aligned} \tag{3.61}$$

To solve the convex optimization problem in equation 3.57, we had to see the following mathematical condition.

The Karush-Kuhn-Tucker (KKT) Condition [25]

Before directly entering in to KKT conditions, let us look at Lagrange multiplier. In mathematical optimization, the method of Lagrange multiplier is a good strategy for finding the local maxima and minima of a function subject to equality constraints (subject to the condition that one or more equations have to be satisfied exactly by the chosen values of the variables). For the case of only one constraint and only two choice variables considering the optimization problem, maximize f(x) subject to g(x) = 0, the Lagrange function is defined by [25]:

$$\mathbf{L(x, y, \lambda) = f(x, y) - \lambda g(x, y)} \tag{3.62}$$

Where, λ is the Lagrange multiplier.

Coming to KKT condition, since we use Lagrange multiplier to find solution for optimization problems contained to one or more equalities, for the case when we have inequality constraints, we need to extend the method of KKT conditions.

For a formulated problems below [25],

$$\mathbf{x}^* = \text{argmin}|_{\mathbf{x}} \quad \mathbf{f}(\mathbf{x}) \tag{3.63}$$

Such that, $h_i(x) = 0, \forall_{i=1,\ldots,m}$ and $g_i(x) \leq 0, \forall_{i=1,\ldots,n}$. The problem says that, find the solution that minimizes f(x), as long as all equalities, $h_i(x) = 0$ and all inequalities $g_i(x) \leq 0$ hold. So, in order to solve this problem, the KKT condition says that, it is easy to see that any equality or inequality constraint can be defined, so long as all terms are in the left side of the equation. The inequality conditions are added to the method of Lagrange multipliers in a similar way to the equalities: Put the cost function as well as the constraints in a single minimization problem, but multiply each equality constraint by a factor λ_i (the KKT multiplier). In the above example, there are **m** equalities and **n** inequalities. Hence the expression for the optimization problem becomes [25]:

$$\mathbf{x}^* = \mathbf{argmin}|_{\mathbf{x}} \quad \mathbf{L}(\mathbf{x}, \lambda, \mu) = \mathbf{argmin}|_{\mathbf{x}} \quad \mathbf{f}(\mathbf{x}) + \sum_{\mathbf{i=1}}^{\mathbf{m}} \lambda_{\mathbf{i}} \mathbf{h_i}(\mathbf{x}) + \sum_{\mathbf{i=1}}^{\mathbf{n}} \mu_{\mathbf{i}} \mathbf{g_i}(\mathbf{x}) \tag{3.64}$$

Where, $L(x, \lambda, \mu)$ is the Lagrangian and depends also on λ and μ, which are vectors of the multipliers.

Coming to the topic, the KKT condition together with Lagrangian multiplier shows as [13]:

$$\mathbf{L}(\mathbf{P}, \lambda, \mu) = \sum_{\mathbf{m=1}}^{\mathbf{M_{RF}}} \sum_{\mathbf{t=1}}^{|\mathbf{S_m}|} \mathbf{a}_{\mathbf{t,m}}^{(\mathbf{y})} \mathbf{e}_{\mathbf{t,m}}^{(\mathbf{y})} + \lambda\Big(\sum_{\mathbf{m=1}}^{\mathbf{M_{RF}}} \sum_{\mathbf{t=1}}^{|\mathbf{S_m}|} \mathbf{p}_{\mathbf{t,m}}^{(\mathbf{y})} - \mathbf{P}\Big) + \sum_{\mathbf{m=1}}^{\mathbf{M_{RF}}} \sum_{\mathbf{t=1}}^{|\mathbf{S_m}|} \mu_{\mathbf{t,m}} \theta_{\mathbf{t,m}} \tag{3.65}$$

Where, $\theta_{t,m}$ is the transformed version of the first condition of equation 3.31 ($R_{t,m} \geq R_{min}$) and the full derivation is again found in apendix-1.

$$\theta_{\mathbf{t,m}} = \mathbf{\Delta} ||\mathbf{h}_{\mathbf{t,m}}^{\mathbf{H}} \mathbf{g_m}||_{\mathbf{2}}^{\mathbf{2}} \sum_{\mathbf{i=1}}^{\mathbf{t-1}} \mathbf{p}_{\mathbf{i,m}}^{(\mathbf{y})} + \mathbf{\Delta} \sum_{\mathbf{j \neq m}} ||\mathbf{h}_{\mathbf{t,m}}^{\mathbf{H}} \mathbf{g_j}||_{\mathbf{2}}^{\mathbf{2}} \sum_{\mathbf{i=1}}^{|\mathbf{s_j}|} \mathbf{p}_{\mathbf{i,j}}^{(\mathbf{y})} - ||\mathbf{h}_{\mathbf{t,m}}^{\mathbf{H}} \mathbf{g_m}||_{\mathbf{2}}^{\mathbf{2}} \mathbf{p}_{\mathbf{t,m}}^{(\mathbf{y})} + \mathbf{\Delta} \sigma^{\mathbf{2}} \tag{3.66}$$

Where, $\lambda \geq 0$ and $\mu \geq 0 (m = 1, 2, ..., M_{RF}, t = 1, 2, ..., |s_m|)$. Then, by using the KKT condition [35], equation 3.57 can be obtained by the following three equations below [13]:

$$\frac{\partial \mathbf{L}}{\partial \mathbf{p_{t,m}}} = \mathbf{a_{t,m}^{(y)}}(-\mathbf{Re}(\mathbf{d_{t,m}^{(y)} h_{t,m}^H g_m})(\mathbf{p_{t,m}^{(y)}})^{-\frac{1}{2}} + |\mathbf{d_{t,m}^{(y)}}|^2 ||\mathbf{h_{t,m}^H g_m}||_2^2)$$
$$+ \sum_{\mathbf{u=t+1}}^{|\mathbf{s_m}|} \mathbf{a_{u,m}} |\mathbf{d_{u,m}^{(y)}}|^2 ||\mathbf{h_{t,m}^H g_m}||_2^2 + \sum_{\mathbf{v \neq m}} \sum_{\mathbf{u=1}}^{|\mathbf{s_v}|} \mathbf{a_{u,v}^{(y)}} |\mathbf{d_{u,v}^{(y)}}|^2 ||\mathbf{h_{u,v}^H g_m}||_2^2$$
$$+ \lambda - \mu_{\mathbf{t,m}} ||\mathbf{h_{t,m}^H g_m}||_2^2 + \sum_{\mathbf{u=t+1}}^{|\mathbf{s_m}|} \mu_{\mathbf{u,m}} \mathbf{\Delta} ||\mathbf{h_{u,m}^H g_m}||_2^2 + \sum_{\mathbf{v \neq m}} \sum_{\mathbf{u=1}}^{|\mathbf{s_v}|} \mu_{\mathbf{u,v}} \mathbf{\Delta} ||\mathbf{h_{u,v}^H g_m}||_2^2$$
$$= \mathbf{0} \tag{3.67}$$

$$\lambda(\sum_{\mathbf{m=1}}^{\mathbf{M_{RF}}} \sum_{\mathbf{t=1}}^{|\mathbf{S_m}|} \mathbf{P_{t,m}^{(y)}} - \mathbf{P}) = \mathbf{0} \tag{3.68}$$

$$\mu_{\mathbf{t,m}} \theta_{\mathbf{t,m}} = \mathbf{0}, \quad \forall_{\mathbf{m,t}} \tag{3.69}$$

Then from equation 3.68, we can obtain the optimal solution of $p_{t,m}^{(y)}$ as follows [13]:

$$\mathbf{p_{t,m}^{(y)}} = (\frac{\mathbf{a_{t,m}^{(y)} Re(d_{t,m}^{(y)} h_{t,m}^H g_m)}}{\zeta})^{\mathbf{2}} \tag{3.70}$$

Where,

$$\zeta = \sum_{\mathbf{u=t}}^{|\mathbf{s_m}|} \mathbf{a_{u,m}^{(y)}} |\mathbf{d_{u,m}^{(y)}}|^2 ||\mathbf{h_{t,m}^H g_m}||_2^2 + \sum_{\mathbf{v \neq m}} \sum_{\mathbf{u=1}}^{|\mathbf{s_v}|} \mathbf{a_{u,v}^{(y)}} |\mathbf{d_{u,v}^{(y)}}|^2 ||\mathbf{h_{u,v}^H g_m}||_2^2$$
$$+ \lambda - \mu_{\mathbf{t,m}} ||\mathbf{h_{t,m}^H g_m}||_2^2 + \sum_{\mathbf{u=t+1}}^{|\mathbf{s_m}|} \mu_{\mathbf{u,m}} \mathbf{\Delta} ||\mathbf{h_{u,m}^H g_m}||_2^2 + \sum_{\mathbf{v \neq m}} \sum_{\mathbf{u=1}}^{|\mathbf{s_v}|} \mu_{\mathbf{u,v}} \mathbf{\Delta} ||\mathbf{h_{u,v}^H g_m}||_2^2 \tag{3.71}$$

Here, even if f(a) is convex in equation 3.49, but the values $d_{t,m}^{(y)}$, $a_{t,m}^{(y)}$, and $p_{t,m}^{(y)}$ are the optimal solution in the y^{th} iteration therefore, iterarively updating the above values will increase or maintain the values of the main equation in 3.52.

3.6.3 Spectrum and Energy Efficiency Manipulations

When normalized bandwidth is considered, the spectrum efficiency can be defined as the achievable sum rate of equation 3.30.

While the energy efficiency denoted by Ω in the paper is defined as the ratio between the achievable sum rate R_{sum} and the total power consumption which is given by [13]:

$$\Omega = \frac{\mathbf{R_{sum}}}{\mathbf{P} + \mathbf{M_{RF}P_{RF}} + \mathbf{M_{RF}P_{S}} + \mathbf{P_{B}}} \tag{3.72}$$

Where, P is the total transmitted power, P_{RF} is the power consumed by each radio frequency chain, P_S is the power consumption of switch, and P_B is the base band power consumption. The energy efficiency is defined by the unit $bps/Hz/W$.

Chapter 4

FD-based Beam Division Non-Orthogonal Multiple Access (FD-BD-NOMA)

4.1 Introduction

The scarce spectrum resource has been demanded to have high data rate and throughput by the increasing number of users in the modern communication technologies like the 5^{th} generation. Among those, the MIMO communications technique has become a core technology to many wireless communication standards like LTE and WIMAX [25]. In the physical layer of wireless communication networks, MIMO technologies has been employed in both uplink and downlink transmissions [1]. Because of interference barriers and some practical limitations, the downlink and uplink channels are designed currently to operate in one dimension either in time or frequency domain. Having this in mind, there are two basic duplexing techniques that has been used in cellular communications. One is the time Time Division Duplex (TDD) scheme where both downlink and uplink channels are allocated with the same frequency but at different time slots to downlink and uplink channels. The other one is the Frequency Division Duplex (FDD) scheme where both downink and uplink channels access the resource at the same time, but with different frequencies for both downlink and uplink channels. This shows that the radio resources has not been optimally used in existing wireless communication systems and this is the main reason that motivated the author to design the FD-BD-NOMA system for the future wireless networks.

Full duplex transmissions have recently gained significant attention to further improve or even double the capacity of conventional half-duplex systems. The benefits of full duplex systems are of course brought by allowing the downlink and uplink channels to function at the same time and frequency [17]. Even if the gains of full duplex scheme can be easily foreseen, practical implementations has many challenges and there are a lot of technical problems which are still need to be solved. The crucial barrier in implementing full duplex systems is that the Self-Interference (SI) from the transmit antennas to the receive antennas either at the full duplex base station transceiver or at the full duplex users transceivers. More generally, the radiated power of the downlink channel interferes with its own desired received signals in the uplink channel [1]. So, the performance of full duplex systems depends on the capability of SI cancellation at the transceiver which is limited in practice.

As in the Beam Division Non-orthogonal Multiple Access scheme proposed in chapter three, this paper considers a scenario where a full duplex capable base station (FD-BS) communicates with the half-duplex users in both directions at the same time slot over the same frequency band.

This chapter mainly considers about techniques to estimate and cancel the residual self-interference that exist at the full-duplex base station.

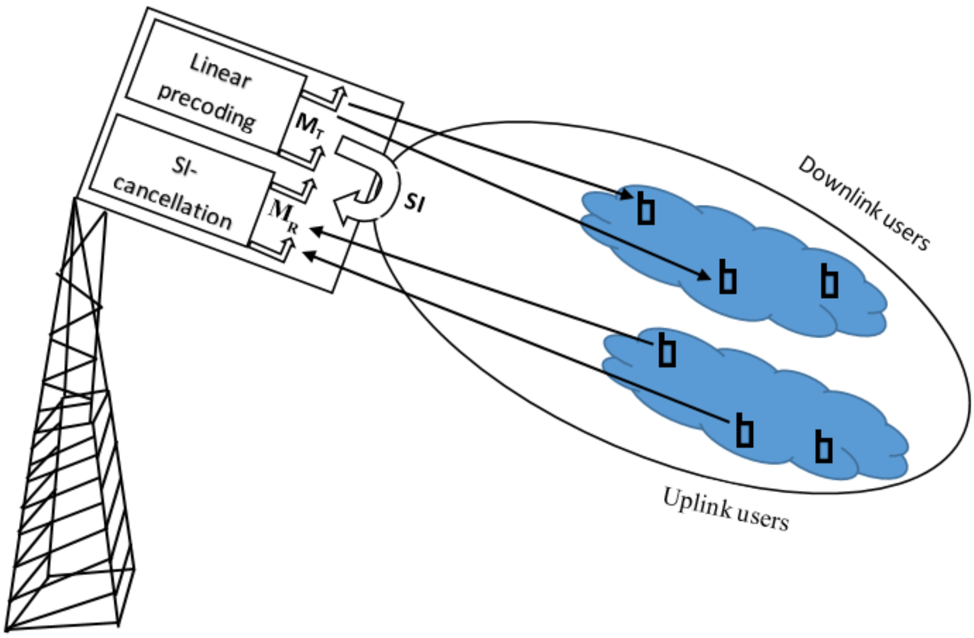

Figure 4.1: Sample for full-duplex BS communication with half-duplex users

It is known that the optimal transmit strategy for the downlink channels is achieved by non-linear precoding techniques like dirty paper coding (DPC), but it requires high complexity to implement and as it is used in chapter three, this paper uses linear beamforming technique which is Wiener Filter (WF) based precoding technique for the purpose of mitigating the inter-beam interference.

For the uplink channels, there are two residual self-interference (RSI) channel estimators used in this paper. Namely, Minimum Mean Square Error (MMSE) channel estimator and Least Square (LS) channel estimator which estimates the RSI from downlink BS antennas to uplink BS antennas [25].

The self-interference channel H_{SI} in between the downlink and uplink antennas of the BS is assumed to have both the LOS component and NLOS components as analyzed in equation 3.2.

The MMSE and LS channel estimators will be used for both Rician and

Rayleigh fading components. By using these estimates for maximum ratio (MR) combining, an uplink spectral efficiency is analyzed [25].

For the considered full duplex system, the problem of beam former design becomes more challenging since there still exists a small, but not negligible, amount of the SI between the transmit and receive antennas at the BS. It is also known that the SI level increases with the transmit power for any SI cancellation technique. Moreover, the difficulty of the design problem is increased further by the co-channel interference (CCI) caused by the users in the uplink channel to those in the downlink channel [1].

Analytic and simulation results are provided to illustrate the impact of these interference on uplink and downlink transmissions on different previous researches and the capacity advantage of the FD communication over the the HD communication is already verified. This paper tries to estimate the RSI channel and using these estimators, an uplink spectral efficiency of FD-BD-NOMA scheme will be analyzed and demonstrated for different channel components and for different estimators.

4.2 System Model

As usual this paper considers a single cell downlink and uplink mm-Wave communication system, where the full duplex BS is equipped with M_R antennas for the uplink and M_T antennas for the downlink transmission purposes with M_{RF} radio frequency chain either for downlink and uplink. The BS is located in the center of the cell and communicating with K half-duplex users having N antennas each, which are simultaneously served by the FD-base station.

As in the case of the downlink communication where the channel matrix is between uplink users and downlink antennas, here also it is assumed that the self-interference channel matrix H_{SI} has both LOS and NLOS components as shown in equation 3.2.

When the FD-BD-NOMA principle is used, all the downlink and uplink users are served simultaneously and in the joint uplink-downlink transmission, in the uplink the BS receives a signal in the m^{th} beam of the m_R^{th} antenna as follows [13],[25]:

$$\mathbf{\bar{r}_{m_R}} = \sum_{\mathbf{i=1}}^{|\mathbf{s_j}|} \sqrt{\mathbf{p_{i,m_R}^{\{U\}}}} \mathbf{h_{i,m_R}^{H}} \mathbf{S_{i,m_R}^{\{U\}}} + \mathbf{g_{m_T}} \mathbf{H_{SI}} \sum_{\mathbf{i=1}}^{|\mathbf{s_j}|} \sqrt{\mathbf{p_{SI}}} \mathbf{s_{i,m_T}^{\{D\}}} + \mathbf{w_{m_R}} \tag{4.1}$$

Where $\bar{r}_{m_R}$ is the total received signal at the m_R^{th} receive antennas of the BS. h_{i,m_R}^{H} stands for the uplink channel gain between the i^{th} user and the

m^{th} receiving beam. $p_{i,m_R}^{\{U\}}$ and p_{SI} are the transmission power of users in the m_R^{th} receive beam and the power of the residual self-interference respectively. $S_{i,m_R}^{\{U\}}$ denotes the uplink message of the i^{th} user of the m_R^{th} beam. While g_{m_T} stands for the precoding of the m_T^{th} downlink beam. H_{SI} is the residual self-interference channel gain of the m_T^{th} transmitting antenna on the m_R^{th} receiving antenna and this paper assumes that the self-interference channel follows noth rician and Rayleigh fading channels. $S_{i,m_T}^{\{D\}}$ is the supper-imposed message to the i^{th} user of the m^{th} beam in the downlink and finally w_{m_R} is the noise for the uplink transmission in the m_R^{th} receiving beam.

For the uplink transmission, statistical properties of minimum mean square error (MMSE)and least-square (LS) estimators for estimating the self interference channel is used. By using these estimates for maximum ratio combining (MRC), an uplink (UL) achievable spectral efficiency (SE) expressions are derived analyzed [25].

4.3 Residual Self-Interference Suppression (RSI Suppression)

The major task in the cancellation of RSI is to estimate RSI channel. In this paper, two estimation schemes are provided to estimate the RSI channel for the sake of performance comparison. These are the Least Square (LS) based estimator and Minimum Mean Square Error (MMSE) based estimator. Both estimators estimate the RSI channel by assuming the intended signal from the uplink users as an additive noise with in the specified beam space.

4.3.1 Least Square (LS) estimator

An estimate of the SI channel coefficient is obtained by LS algorithm by considering the intended signal channel coefficient as an additive noise as the first stage of the SI cancellation where the SI signal is known.

It means that this estimator is used when the characteristics of the SI matrix is known at the receiving antenna of a specific beam space.

By rearranging equation 4.1, we can write it as [25]:

$$\mathbf{\bar{r}_{m_R}} = \mathbf{X_{m_R}^{\{U\}} H_{m_R}} + \mathbf{X_{m_T}^{\{D\}} H_{SI}} + \mathbf{w_{m_R}} \tag{4.2}$$

Where $X_{m_R}^{\{U\}} = P_{m_R}^{\{U\}} S_{m_R}^{\{U\}}$ is the transmitted signal from all uplink users to the m_R^{th} uplink antenna in the m^{th} beam. While H_{m_R} is the channel matrix between the m_R^{th} antenna and uplink user of the m^{th} beam. $X_{m_T}^{\{D\}} =$

$g_{m_T} P_{SI} S_{m_T}^{\{D\}}$ is the transmitted signal from the m_T^{th} downlink antenna to the downlink users in m^{th} beam space. H_{SI} is the residual self interference channel gain of the m_T^{th} transmitting antenna on the m_R^{th} receiving antenna and w_{m_R} is the noise of the uplink transmission in the m_R^{th} receiving beam.

Then again for simplicity, we have to write equation 4.2 as shown below by considering the intended signal as noise [25],

$$\mathbf{\bar{r}_{m_R}} = \mathbf{X_{m_T}^{\{D\}} H_{SI}} + \mathbf{V} \tag{4.3}$$

With V $= X_{m_R}^{\{U\}} H_{m_R} + w_{m_R}$ and by using the LS algorithm, the expression of an SI channel is given by [25]:

$$\mathbf{\bar{H}_{SI_{LS}}} = (\mathbf{X_{m_T}^{\{D\}^H} X_{m_T}^{\{D\}}})^{-1} (\mathbf{X_{m_T}^{\{D\}^H}}) \mathbf{\bar{r}_{m_R}} \tag{4.4}$$

Then the estimate of the SI signal will be:

$$\mathbf{\bar{SI}_{LS}} = \mathbf{\bar{H}_{SI_{LS}} X_{m_T}^{\{D\}}}) \tag{4.5}$$

So that here, equation 4.5 is used to cancel the self-interference signal from the received data. Therefore, subtracting equation 4.5 from 4.2 and an interference free signal will be obtained as follows using the LS algorithm.

$$\mathbf{\widehat{r}_{m_{R_{LS}}}} = \mathbf{\bar{r}_{m_R}} - \mathbf{\bar{SI}_{LS}} = \mathbf{V} \tag{4.6}$$

4.3.2 Minimum Mean Square Error Channel Estimation (MMSE) estimator

By using the MMSE estimator, the linear channel estimate of the residual self interference will be given by [25],

$$\mathbf{\bar{H}_{SI_{MMSE}}} = \left[\left(\left(\mathbf{E}\{\mathbf{H_{SI} H_{SI}^H}\} \right)^{-1} + \frac{\mathbf{1}}{\sigma_{\mathbf{n}}^{\mathbf{2}} + \sigma_{\mathbf{s}}^{\mathbf{2}}} \mathbf{X_{m_T}^{\{D\}^H} X_{m_T}^{\{D\}}} \right)^{-1} \frac{\mathbf{1}}{\sigma_{\mathbf{n}}^{\mathbf{2}} + \sigma_{\mathbf{s}}^{\mathbf{2}}} \mathbf{X_{m_T}^{\{D\}^H}} \right] \mathbf{\bar{r}_{m_R}} \tag{4.7}$$

Where, $E\{.\}$ denotes statistical expectation and σ_n^2 and σ_s^2 are the variances of the thermal noise and the intended signal, respectively. While the latter needs the knowledge of the second-order statistics of the SI channel, it enjoys substantially lower channel estimation error than the LS estimator.

In the case of additive white noise the MMSE channel estimate of the above equation is simplified to the equation below [25]:

$$\mathbf{\bar{\bar{H}}_{SI_{MMSE}}} = [\frac{\sigma_{\mathbf{n}}^{\mathbf{2}}}{\sigma_{\mathbf{s}}^{\mathbf{2}}} + \mathbf{X_{m_T}^{\{D\}^H} X_{m_T}^{\{D\}}}]^{-1} \mathbf{X_{m_T}^{\{D\}^H}} \mathbf{\bar{r}_{m_R}} \tag{4.8}$$

The estimate of the SI signal using MMSE will be given by:

$$\mathbf{\bar{SI}_{MMSE}} = \mathbf{\bar{\bar{H}}_{SI_{MMSE}} X_{m_T}^{\{D\}}}) \tag{4.9}$$

Then according to equation 4.6, the interference free signal after this estimator will be given by [25]:

$$\begin{aligned}
\widehat{\mathbf{r}}_{\mathbf{m_{R_{MMSE}}}} &= \mathbf{\bar{r}_{m_R}} - \mathbf{\bar{SI}_{MMSE}} \\
&= \mathbf{X_{m_T}^{\{D\}} H_{SI}} + \mathbf{V} - \frac{\mathbf{X_{m_T}^{\{D\}^H} X_{m_T}^{\{D\}^2} H_{SI}}}{\frac{\sigma_n^2}{\sigma_s^2} + \mathbf{X_{m_T}^{\{D\}^H} X_{m_T}^{\{D\}}}} \\
&= \mathbf{X_{m_R}^{\{U\}} H_{m_R}} + \mathbf{w_{m_R}} + \mathbf{X_{m_T}^{\{D\}} H_{SI}} - \frac{\mathbf{X_{m_T}^{\{D\}^H} X_{m_T}^{\{D\}^2} H_{SI}}}{\frac{\sigma_n^2}{\sigma_s^2} + \mathbf{X_{m_T}^{\{D\}^H} X_{m_T}^{\{D\}}}}
\end{aligned} \tag{4.10}$$

4.4 Achievable Sum Rate and spectral efficiency of uplink transmission

The interference free received signal at the m_R^{th} receive antenna after LS estimator is given by [13]:

$$\widehat{\mathbf{r}}_{\mathbf{m_{R_{LS}}}} = \mathbf{V} = \mathbf{X_{m_R}^{\{U\}} H_{m_R}} + \mathbf{w_{m_R}} \tag{4.11}$$

We know that $X_{m_R}^{\{U\}} = P_{m_R}^{\{U\}} S_{m_R}^{\{U\}}$, replacing in the above equation gives us [13]:

$$\widehat{\mathbf{r}}_{\mathbf{m_{R_{LS}}}} = \mathbf{V} = \mathbf{P_{m_R}^{\{U\}} S_{m_R}^{\{U\}} H_{m_R}} + \sigma^{\mathbf{2}} \tag{4.12}$$

The signal to noise ratio of the received signal in the m^{th} beam of the m_R^{th} antenna using the least square estimator is given by [13]:

$$\psi_{\mathbf{m_{R_{LS}}}}^{\mathbf{\{U\}}} = \frac{||\mathbf{H_{m_R}}||_{\mathbf{2}}^{\mathbf{2}} \mathbf{P_{m_R}^{\{U\}}}}{\sigma^{\mathbf{2}}} \tag{4.13}$$

The achievable rate in the m^{th} beam of the m_R^{th} receive antenna is given finally by [13]:

$$\mathbf{R_{m_{R_{LS}}}} = \log_{\mathbf{2}}^{(\mathbf{1+SNR})} = \log_{\mathbf{2}}^{(\mathbf{1}+\psi_{\mathbf{m_{R_{LS}}}}^{\mathbf{\{U\}}})} \tag{4.14}$$

Finally, the achievable sum rate of the beam division non-orthogonal multiple access (BD-NOMA) for the uplink transmission after LS estimator can be given as [13]:

$$\mathbf{R_{sum_{LS}}} = \sum_{\mathbf{m=1}}^{\mathbf{M_{RF}}} \mathbf{R_{m_{R_{LS}}}} \tag{4.15}$$

For the case of MMSE estimation the overall received signal after the estimator will be given by the following formula [25]:

$$\begin{aligned}\widehat{\mathbf{r}}_{\mathbf{m_{R_{MMSE}}}} &= \mathbf{X}^{\{\mathbf{U}\}}_{\mathbf{m_R}}\mathbf{H_{m_R}} + \mathbf{w_{m_R}} + \mathbf{X}^{\{\mathbf{D}\}}_{\mathbf{m_T}}\mathbf{H_{SI}} - \frac{\mathbf{X}^{\{\mathbf{D}\}^{\mathbf{H}}}_{\mathbf{m_T}}\mathbf{X}^{\{\mathbf{D}\}^{\mathbf{2}}}_{\mathbf{m_T}}\mathbf{H_{SI}}}{\frac{\sigma^2_{\mathbf{n}}}{\sigma^2_{\mathbf{s}}} + \mathbf{X}^{\{\mathbf{D}\}^{\mathbf{H}}}_{\mathbf{m_T}}\mathbf{X}^{\{\mathbf{D}\}}_{\mathbf{m_T}}} \\ &= \mathbf{P}^{\{\mathbf{U}\}}_{\mathbf{m_R}}\mathbf{S}^{\{\mathbf{U}\}}_{\mathbf{m_R}}\mathbf{H_{m_R}} + \sigma^{\mathbf{2}} + \mathbf{g_{m_T}}\mathbf{P_{SI}}\mathbf{S}^{\{\mathbf{D}\}}_{\mathbf{m_T}}\mathbf{H_{SI}}\left(\mathbf{1} - \frac{\left(\mathbf{g_{m_T}}\mathbf{P_{SI}}\mathbf{S}^{\{\mathbf{D}\}}_{\mathbf{m_T}}\right)\left(\mathbf{g_{m_T}}\mathbf{P_{SI}}\mathbf{S}^{\{\mathbf{D}\}}_{\mathbf{m_T}}\right)^{\mathbf{2}}}{\frac{\sigma^2_{\mathbf{n}}}{\sigma^2_{\mathbf{s}}} + \left(\mathbf{g_{m_T}}\mathbf{P_{SI}}\mathbf{S}^{\{\mathbf{D}\}}_{\mathbf{m_T}}\right)\left(\mathbf{g_{m_T}}\mathbf{P_{SI}}\mathbf{S}^{\{\mathbf{D}\}}_{\mathbf{m_T}}\right)}\right)\end{aligned} \tag{4.16}$$

The signal to noise ratio of the received signal in the m^{th} beam of the m_R^{th} antenna using the MMSE estimator is given by:

$$\psi^{\{\mathbf{U}\}}_{\mathbf{m_{R_{MMSE}}}} = \frac{||\mathbf{H_{m_R}}||^{\mathbf{2}}_{\mathbf{2}}\mathbf{P}^{\{\mathbf{U}\}}_{\mathbf{m_R}}}{\rho_{\mathbf{m_R}}} \tag{4.17}$$

Where ρ_{m_R} is the noise plus interference term which is given by [13]:

$$\rho_{\mathbf{m_R}} = \sigma^{\mathbf{2}} + ||\mathbf{g_{m_T}}\mathbf{H_{SI}}||^{\mathbf{2}}_{\mathbf{2}}\mathbf{P_{SI}}\left(\mathbf{1} - \frac{\left(||\mathbf{g_{m_T}}||^{\mathbf{2}}_{\mathbf{2}}\mathbf{P_{SI}}\right)\left(||\mathbf{g_{m_T}}||^{\mathbf{2}}_{\mathbf{2}}\mathbf{P_{SI}}\right)^{\mathbf{2}}}{\frac{\sigma^2_{\mathbf{n}}}{\sigma^2_{\mathbf{s}}} + \left(||\mathbf{g_{m_T}}||^{\mathbf{2}}_{\mathbf{2}}\mathbf{P_{SI}}\right)\left(||\mathbf{g_{m_T}}||^{\mathbf{2}}_{\mathbf{2}}\mathbf{P_{SI}}\right)}\right) \tag{4.18}$$

The achievable rate in the m^{th} beam of the m_R^{th} receive antenna is given finally by:

$$\mathbf{R_{m_{R_{MMSE}}}} = \log_{\mathbf{2}}^{(\mathbf{1+SINR})} = \log_{\mathbf{2}}^{(\mathbf{1}+\psi^{\{\mathbf{U}\}}_{\mathbf{m_{R_{MMSE}}}})} \tag{4.19}$$

The achievable sum rate of the beam division non-orthogonal multiple access (BD-NOMA) for the uplink transmission after MMSE estimator can be given as [13]:

$$\mathbf{R_{sum_{MMSE}}} = \sum_{\mathbf{m=1}}^{\mathbf{M_{RF}}} \mathbf{R_{m_{R_{MMSE}}}} \tag{4.20}$$

Chapter 5

Results and Discussions

Here, the simulation results are provided to validate the spectral and energy efficiency advantage of the BD-NOMA system over the normal beam division multiple access (BDMA) system. The provided simulations are based on typical downlink mm-Wave massive MIMO system where the BS is equipped with an ULA of $\mathbf{M} = 256$ antennas (with $\mathbf{M}_T = 128$ for downlink transmission and $\mathbf{M}_R = 128$ for uplink reception) and communicates with K single antenna users. The total transmitted power is set as P = 32 mW (15 dBm) [10]. One LoS component and L = 3 NLoS components are assumed for all users' channels in the downlink transmission.

The channel parameters of the K^{th} user is considered as follows, a) $\beta_{k,0} \sim$ CN(0,1), $\beta_{k,l} \sim CN(0, 10^{-1})$ for 1 ≤ l ≤ L; b) $\theta_{k,0}$ and $\theta_{k,l}$ for 1 ≤ l ≤ L will follow the uniform distribution within $[-1/2, 1/2]$ [10].

5.1 Downlink Scenario

The following simulation for downlink transmission shows the comparison of three different schemes: Traditional-MIMO, BD-NOMA, and BDMA (BD-OMA) which is the Beam Division Orthogonal Multiple Access. In the case of Traditional-MIMO, Each downlink antennas are connected to one Radio Frequency (RF) chain which means that the number of RF chains are equal to the number of available downlink antennas. Where BD-OMA is performed when users of the same beam are allocated with orthogonal frequency resources. BD-NOMA combines BD-OMA and NOMA in power domain. In both BD-NOMA and BD-OMA schemes, the number of radio-frequency chain is less than the number of downlink users, this is possible by using Interference Aware (IA) beam selection technique which briefly discussed in [10].

The strong user based equivalent channel is used and the iterative power allocation technique discussed in chapter three are used to mitigate interference. The Wiener filter (WF) or the Regularized Zero Forcing (RZF) based precoding technique is used to mitigate the inter-beam interference in the downlink transmission.

The Spectral Efficiency comparison against the number of users is shown in Fig. 5.1, where SNR is set as 10 dB. We can see from the simulation results that with the increasing of the number of users K , the efficiency gap between the BD-OMA and BD-NOMA becomes larger. This is because the larger the number of users, the larger the probability that the same beam is selected for different users. As a result, existing BD-OMA will suffer from an obvious performance loss, while the BD-NOMA can still perform well due to the use of NOMA. For the traditional-MIMO scheme, it is obvious that this scheme can achieve higher spectral efficiency. This is because that all the available resources are allocated for all users without beam selection with a cost of enormous amount of energy due to unused RF chains.

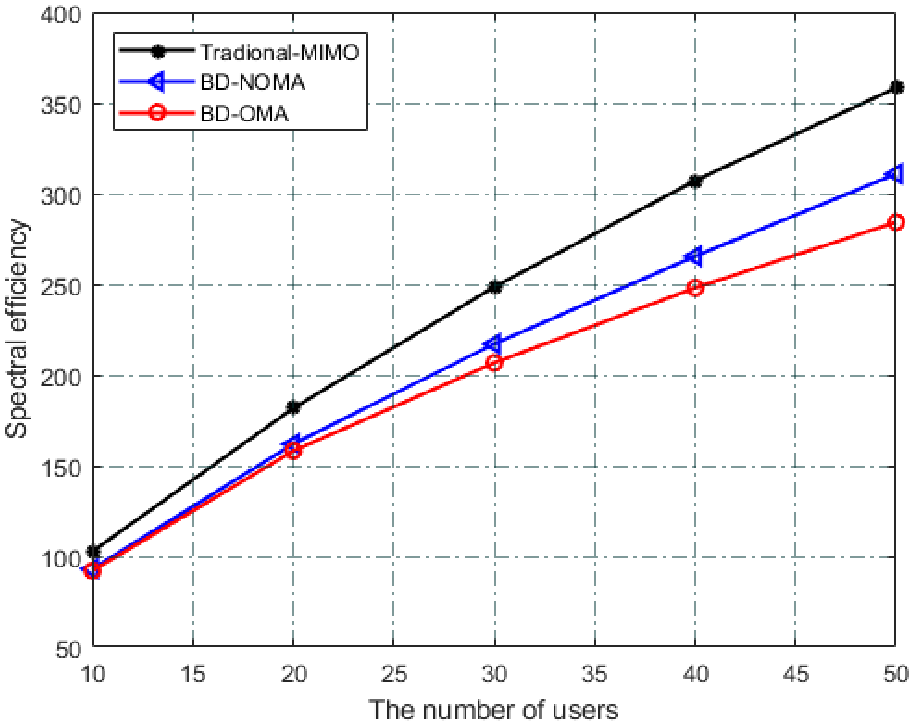

Figure 5.1: Downlink spectrum efficiency Vs number of users

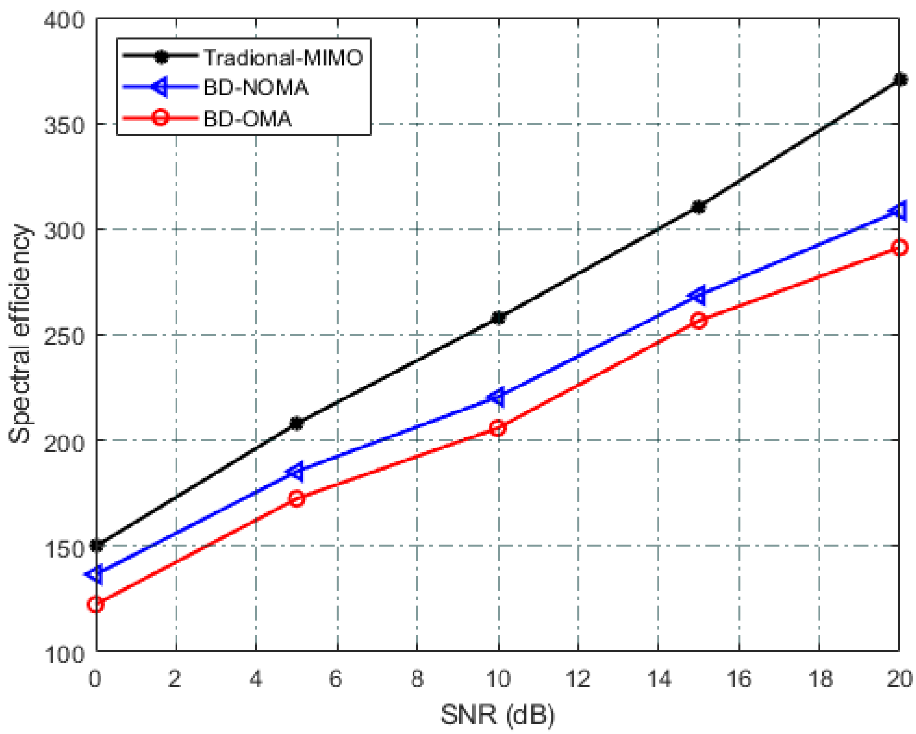

Figure 5.2: Downlink spectrum efficiency Vs SNR

Fig. 5.2 below shows the spectrum efficiency against SNR. Where the number of users is K = 32. We can find that the BD-NOMA can achieve higher spectrum efficiency than that of BD-OMA since BD-NOMA comprise NOMA in the power domain which allows to serve multiple users in each beam.

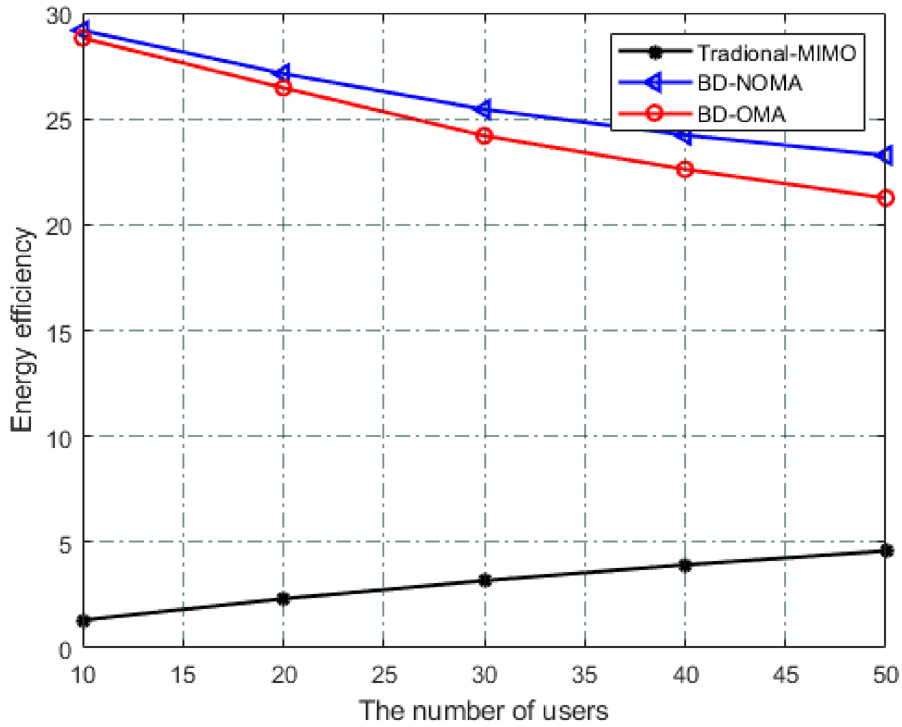

Figure 5.3: Downlink energy efficiency Vs number of users

The performance comparison in terms of energy efficiency against the number of users is shown in Fig. 5.3 below, where SNR is set as 10 dB. We can see that the energy efficiency of the BD-NOMA scheme is higher than both the traditional-MIMO and BD-OMA schemes even the number of users is very large. For the purpose of simulation, we adopt the typical values P_{RF} = 300 mW, P_s = 5 mW, and P_B = 200 mW [10].

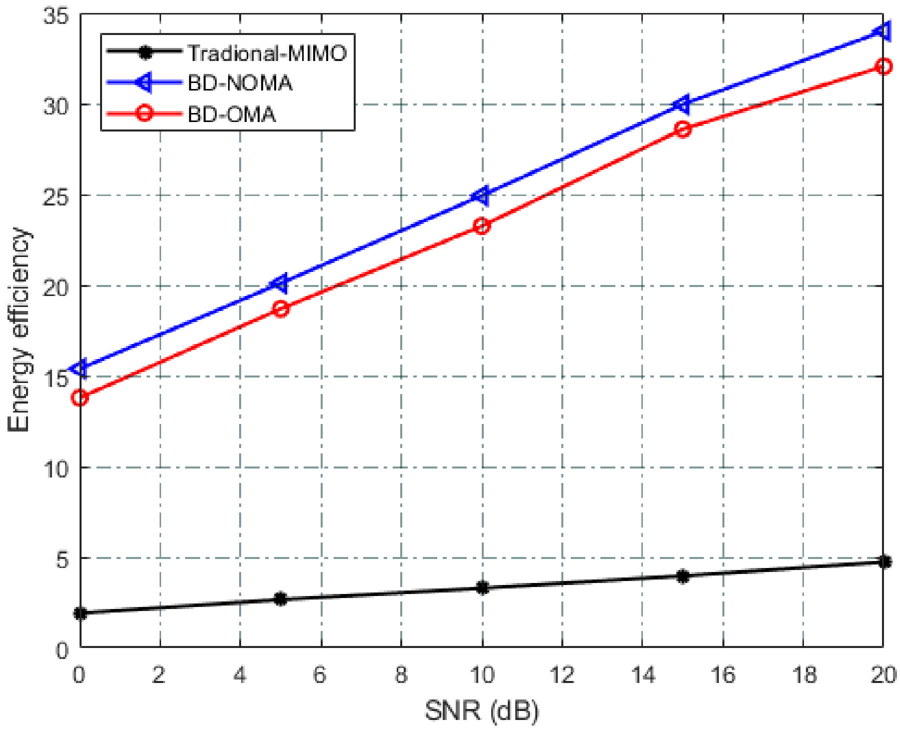

Figure 5.4: Downlink energy efficiency Vs SNR

Fig. 5.4 above shows the energy efficiency against SNR, where the number of users is also K = 32. Again we can observe that the BD-NOMA can achieve higher energy efficiency than both traditional-MIMO and BD-OMA schemes. More specifically it is observed that the BD-NOMA scheme has around 10% energy efficiency improvement as compared to BD-OMA.

The other thing that we can observe in both figure 5.3 and 5.4, both BD-NOMA and BD-OMA schemes has much higher energy efficiency than that of traditional-MIMO scheme where the number of RF chains is equal to the number of antennas available at the BS; this leads the tradional-MIMO consume a very high energy for each RF chain. On the other side, the number of RF chain is much smaller than the number of antennas in both BD-NOMA and BD-OMA schemes.

5.2 Uplink scenario

5.2.1 Computational complexity analysis

Since basic linear algebra operations, such as matrix-matrix multiplications, have a well-defined structure and can thus be implemented efficiently in hardware. However, the computational complexity can be a bottleneck when large matrices need to be manipulated. The exact complexity of a matrix operation depends strongly on the hardware implementation, including the bit width (i.e., the number of binary digits used to represent a number) and the data type (e.g., floating point or fixed point) [25].

In this section, the number of complex complex multiplications that are needed in the computations are analyzed, while the complexity of additions/subtractions is neglected since these operations are much easier to implement in hardware [25]. The computational complexities of MMSE and LS estimators are summarized in the table 5.1 below.

Table 5.1: Computational complexity analysis of LS and MMSE estimators

Scheme	**Number of multiplications**
MMSE	$\mathbf{2} * [(\mathbf{M^2}) * \mathbf{K} + (\mathbf{M^2}) * \mathbf{K}]$
LS	$((\mathbf{M^2}) * \mathbf{K} + (\mathbf{M^2}) * \mathbf{K})$

As can be seen from the table, the complexity of MMSE is higher than the complexity of LS estimator. This is because that MMSE estimator uses the second order statistics of the self-interference channel. By using the above table the complexity of MMSE and LS estimators are analyzed using matlab as follows.

After the complexity analysis, an uplink spectral efficiency expressions derived in chapter four are validated by simulations. There are two schemes which are assumed in the simulations. One is the MMSE based residual channel estimator when the self-interference follows both rician and Rayleigh fading channels and other is the LS estimator again in both rician and Rayleigh fading channel scenarios.

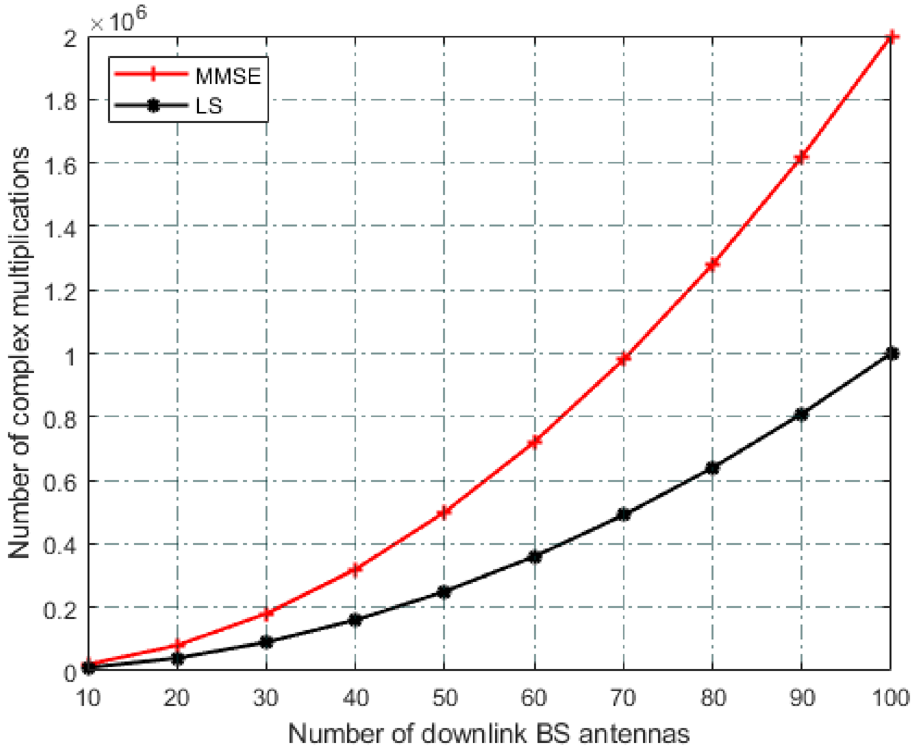

Figure 5.5: Complexity analysis of MMSE and LS estimators [25]

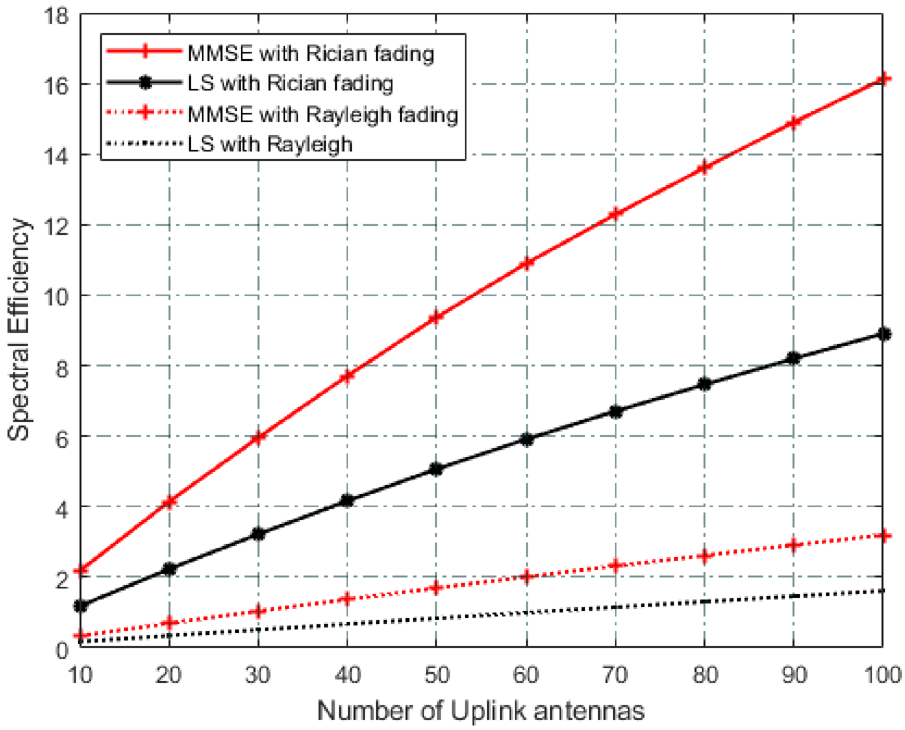

Figure 5.6: Uplink spectrum efficiency Vs number of uplink BS antennas for different channel estimators

Figure 5.6 above shows the uplink spectral efficiency versus number of uplink antennas by using MR-combining based on either the MMSE or the LS estimators. As we can see from the figure, MMSE estimator with rician fading component outperforms the LS-estimator using either of the two fading channels. This is because that since the downlink antenna of the BS are located near to the uplink antennas, the Line-Of-Sight (LOS) component of the residual self-interference can be easily grasped by the uplink antennas rather than the Non-Line-Of-Sight (NLOS) multi-path component.

In the figure 5.7, the paper considers a scenario where the self-interference only has the LOS component following rician fading which not possible in practical scenarios. Here again the MMSE estimator outperforms the LS-estimator since the MMSE utilizes the second order statistics of the residual self-interference.

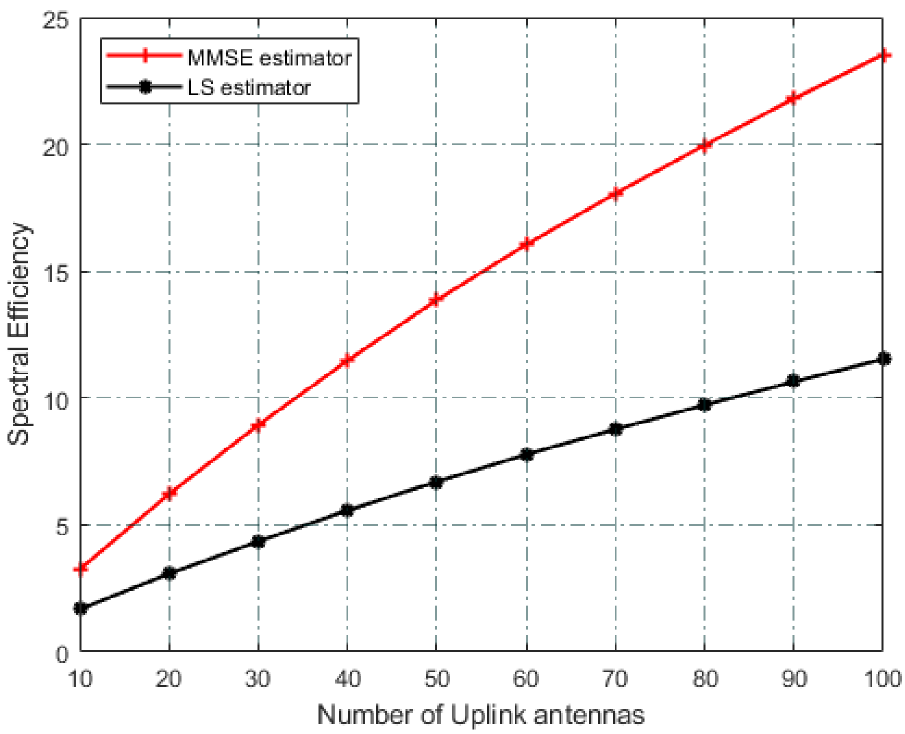

Figure 5.7: Uplink spectral efficiency Vs number of uplink BS antennas for difference channel estimators when the SI has only LOS component

Chapter 6

Conclusion and Recommendation

6.1 Conclusion

This paper studied the uplink and downlink spectral and energy efficiency efficiency of a FD based BD-NOMA scheme by integrating NOMA with BD-OMA inorder to overcome the fundamental limit of existing BDMA that users are allocated orthogonal resources regardless of their distance from the base station. The number of users of a beam is larger than the number of RF chain in the proposed scheme.

In the downlink scenario, wiener filter based precoding technique is used together with the strong user based equivalent channel determined for each beam for the sake of mitigating the Inter-Beam-Interference.

Further, the sum rate is maximized by jointly optimizing the power allocated to all users using the principle of NOMA in the power domain.

The spectrum efficiency of uplink transmission is analyzed after the self-interference is estimated and canceled using MMSE and LS channel estimators at the FD base station antennas.

Simulation results shown that BD-NOMA scheme has better spectral and energy efficiency compared to BD-OMA. The traditional-MIMO scheme may have higher spectral efficiency than that of BD-NOMA. This is because that there is no beam selection where each available antennas are directly connected to one RF chain. However, because of the high number of RF chains in the traditional-MIMO scheme, very high power will be consumed and as result the energy efficiency of this scheme will fall down compared to both BD-NOMA and BD-OMA scheme.

6.2 Recommendation

The following few promising future works are recommended by the author.

- This paper was done by assuming perfect channel state information at the base station antennas. The channel state information could be known by using tools of compressive sensing techniques which could be one of the interesting area to investigate.

- In the uplink scenario, to mitigate the self interference, this paper considers linear channel estimators (MMSE and LS). But there are different analog and digital cancellation techniques techniques which has high performance of canceling SI which increase the capacity advantage of FD scheme over the HD scheme by a considerable amount.

- This paper tries to address both inter and intra beam interference mitigation techniques in the downlink and self-self interference mitigation techniques in the uplink scenario. However, still there are two basic interferences; namely the co-channel (inter-user) interference between uplink and downlink users of the same beam since they use the same channel. The other one is the cross-mode interference that could occur between uplink and downlink channels before it reaches the receiver antennas. These two interferences are hampering the two fold spectral efficiency advantage of the FD scheme over the HD scheme and it will be good if one could find solution for this.

- The paper didn't use any multiplexer/modulator which are assumed in 5G. Analyzing BD-NOMA using the modulators assumed in 5G will help to idealize this work in the practical setups.

Bibliography

[1] FaLong Luo, Ph.D., IEEE Fellow Charlie (Jianzhong) Zhang, Ph.D., IEEE Fellow, "Signal Processing for 5G," John Wiley and Sons, Ltd, 2016.

[2] Akhil Gupta, Student Member, IEEE, Rakesh Kumar Jha, Senior Member, IEEE, "A Survey of 5G Network: Architecture and Emerging Technologies," IEEE Access, vol. 3, pp. 1206-1232, Jul. 2015.

[3] Arman Shojaeifard, Member, IEEE, Kai-Kit Wong, Fellow, IEEE, Marco Di Renzo, Senior Member, IEEE, Gan Zheng, Senior Member, IEEE, Khairi Ashour Hamdi, Senior Member, IEEE, Jie Tang, Member, IEEE, "Massive MIMO-Enabled Full-Duplex Cellular Networks," IEEE Tran. on Commun., vol. 65, no. 11, pp. 4734-4750 Nov. 2017.

[4] Theodre S. Rappaport, Shu Sun, Rimma Mayzus, Hang Zhao, Yaniv Azar, Kevin Wang, George N. Wong, Jocelyn K. Schulz, Mathew Samimi, and Felix Gutierrez, "Millimeter Wave Mobile Communications for 5G Cellular: It Will Work!," IEEE Access, vol. 1, pp. 335-349, April 2013.

[5] Zhiguo Ding, Member, IEEE, Zheng Yang, Pingzhi Fan, Senior Member, IEEE, and H. Vincent Poor, Fellow, IEEE, "On the Performance of Non-Orthogonal Multiple Access in 5G Systems with Randomly Deployed Users," IEEE Signal Process. Lett., vol. 21, no. 12, pp. 1501-1505, Dec. 2014.

[6] Zheng Yang,Student Member, IEEE, Zhiguo Ding,Senior Member, IEEE, Pingzhi Fan,Fellow, IEEE, and George K. Karagiannidis,Fellow, IEEE, "On the Performance of Non-orthogonal Multiple Access Systems With Partial Channel Information," IEEE Tran. on Commun., vol. 64, no. 2,pp. 654-667, Feb. 2016.

[7] Peng Xu, Yi Yuan, Zhiguo Ding, Senior Member, IEEE, Xuchu Dai, and Robert Schober, Fellow, IEEE, "On the Outage Performance of Non-Orthogonal Multiple Access with One-Bit Feedback," IEEE Tran. on Commun., vol. 15, no. 10, pp. 6716-6730, Oct. 2016.

[8] Yuanwei Liu, Member IEEE, Zhijin Qin, Member IEEE, Maged eLkashLan, Member IEEE, Zhiguo ding, Member IEEE, AruMugaM NaLLanathan, Fellow IEEE, and Lajos hanZo ,Fellow IEEE, "Non-orthogonal Multiple Access for 5G and Beyond," Proceedings of the IEEE, vol. 105, no. 12, pp. 2347-2381, Dec. 2017.

[9] Chongbin Xu, Yang hu, Chulong Liang, Junjie ma, and Li ping, (Fellow, IEEE), "Massive MIMO, Non-Orthogonal Multiple Access and Interleave Division Multiple Access," IEEE Access, vol. 5, pp. 14728-14748, Jul. 2017.

[10] Xinyu Gao, Linglong Dai, Zhijie Chen, Zhaocheng Wang, and Zhijun Zhang, "Near-Optimal Beam Selection for Beamspace MmWave Massive MIMO Systems," IEEE Communications letters, VOL. 20, NO. 5, May 2016.

[11] Zhiguo Ding, Senior Member, IEEE, Pingzhi Fan, Fellow, IEEE, and H. Vincent Poor Fellow, IEEE, "On the Coexistence between Full-Duplex and NOMA," IEEE Wireless Commun. Lett., (Early Access), pp. 1-4, Mar. 2018.

[12] Pierluigi V. Amadori, Christos Masouros, "Low RF-Complexity Millimeter-Wave Beamspace-MIMO Systems by Beam Selection," IEEE transactions on communications, VOL. 63, NO. 6, June 2015.

[13] Bichai Wang, Linglong Dai, Zhaocheng Wang, Ning Ge, "Spectrum and Energy-Efficient Beamspace MIMO-NOMA for Millimeter-Wave Communications Using Lens Antenna Array," IEEE Journal on selected areas in communications, VOL. 35, NO. 10, October 2017.

[14] Yong Zeng, Rui Zhang, "Millimeter Wave MIMO with Lens Antenna Array: A New Path Division Multiplexing Paradigm," IEEE Transactions on Communications, pp. 1-15, 2016.

[15] Tebe Parfait , Yujun Kuang , Kponyo Jerry, "Performance Analysis and Comparison of ZF and MRT Based Downlink Massive MIMO Systems," IEEE transaction, pp. 383-389, 2014.

[16] un Zhu, Robert Schober, and Vijay K. Bhargava, "Secure Downlink Transmission in Massive MIMO System with Zero-Forcing Precoding," European Wireless, pp. 195-200, 2014.

[17] Xiang Gao, Ove Edfors, Fredrik Rusek, and Fredrik Tufvesson, "Linear pre-coding performance in measured very-large MIMO channels," IEEE Access, 2014.

[18] Chen Sun, Xiqi Gao, Shi Jin, Michail Matthaiou, Zhi Ding, and Chengshan Xiao, "Beam Division Multiple Access for Massive MIMO Downlink Transmission," IEEE ICC 2015-Wireless Commun. Symp., pp. 1970-1975, Sep. 2015.

[19] Li You, Xiqi Gao, Geoffrey Ye Li, Xiang-Gen Xia, and Ni Ma, "Millimeter-Wave/Terahertz Massive MIMO BDMA Transmission with Per-Beam Synchronization," IEEE ICC 2017-Wireless Commun. Symp., Jul. 2017.

[20] Chen Sun, Student Member, IEEE, Xiqi Gao, Fellow, IEEE, and Zhi Ding, Fellow, IEEE, "BDMA in Multi-Cell Massive MIMO Communications: Power Allocation Algorithms," IEEE Tran. on Signal Proc., vol. 65, no. 11, pp. 2962-2974, Jun 2017.

[21] Chih-Wei Jen, Hung-Wei Yang, Hsun-Wei Chan, Hung-Chih Liu, Henry Lopez, Chun-Yi Liu, Shyh-Jye Jou, "Digital Self-Interference Cancellation for OFDM Full-Duplex Transmission in 60 GHz Band," Circuits and Systems (ISCAS), IEEE Intern. Symp., May 2018.

[22] Hesham ElSawy, Ahmad AlAmmouri, Osama Amin, and Mohamed-Slim Alouini, "Can Uplink Transmissions Survive in Full-duplex Cellular Environments?," European Wireless, pp. 276-281, June 2016.

[23] Itsikiantsoa Randrianantenaina, Student Member, IEEE, Hayssam Dahrouj, Senior Member, IEEE, Hesham ElSawy, Member, IEEE, and Mohamed-Slim Alouini, Fellow, IEEE, "Interference Management in Full-Duplex Cellular Networks with Partial Spectrum Overlap," IEEE Access, vol. 5, pp. 7567-7583, March 2017.

[24] Radwa Sultan, Lingyang Song, Karim G. Seddik,and ZhuHan, "Full Duplex in Massive MIMO Systems: Analysis and Feasibility," Globecom Workshops (GC Wkshps), IEEE, February 2017.

[25] Emil Björnson, Jakob Hoydis and Luca Sanguinetti (2017), "Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency", Foundations and Trends in Signal Processing: Vol. 11, No. 3-4, pp 154–655. DOI 10.1561/2000000093.

[26] Akbar Sayeed and John Brady, "Beamspace MIMO for High-Dimensional Multiuser Communication at Millimeter-Wave Frequencies," Globecom - Wireless Communications Symposium, pp. 3679-3684, 2013.

[27] Rickard Stridh, Mats Bengtsson, and B. Ottersten, "System Evaluation of Optimal Downlink Beamforming with Congestion Control in Wireless Communication," IEEE transactions on wireless communications, VOL. 5, NO. 4, April 2006.

[28] Michael Joham, Wolfgang Utschick, Josef A. Nossek, "Linear Transmit Processing in MIMO Communications Systems," IEEE transactions on signal processing, VOL. 53, NO. 8, August 2005.

[29] Mirette Sadek, Alireza Tarighat, and Ali H. Sayed, "A Leakage-Based Precoding Scheme for Downlink Multi-User MIMO Channels," IEEE Transactions on wireless communications, VOL. 6, NO. 5, May 2007.

[30] Jakob Hoydis, Stephan ten Brink, M. erouane Debbah, "Massive MIMO in the UL/DL of Cellular Networks: How Many Antennas Do We Need?," IEEE Journal on selected areas in communications, VOL. 31, NO. 2, February 2013.

[31] Luca Sanguinetti, Emil B., Merouane Debbah, and Aris L. Moustakas, "Optimal Linear Precoding in Multi-User MIMO Systems: A Large System Analysis," Globecom - Wireless Communications Symposium, pp. 3922-3927, 2014.

[32] Lu Lu, Geoffrey Ye Li, A. Lee Swindlehurst, Alexei Ashikhmin, Rui Zhang, "An Overview of Massive MIMO: Benefits and Challenges," IEEE Journal of selected topics in signal processing, VOL. 8, NO. 5, October 2014.

[33] Yuya Saito, Yoshihisa Kishiyama, Anass Benjebbour, Takehiro Nakamura, and Kenichi Higuchi, "Non-Orthogonal Multiple Access (NOMA) for Cellular Future Radio Access," IEEE Access, 2013.

[34] Qi Zhang, Quanzhong Li, , and Jiayin Qin, "Robust Beamforming for Non-Orthogonal Multiple Access Systems in MISO Channels," IEEE Transactions on vehicular technology, pp. 1-5, 2016.

[35] S. Boyd and L. Vandenberghe, "Convex Optimization," IEEE Transactions on automatic control, VOL. 51, NO. 11, November 2006.

[36] Yong I. Choi, Jae W. Lee, Minjoong Rim, and Chung G. Kang, "On the Performance of Beam Division Nonorthogonal Multiple Access for FDD-based Large-scale Multi-user MIMO Systems," IEEE Transactions on Wireless Communications, vol. 16, no. 8, pp. 5077-5089, Aug. 2017.

[37] Yan Sun, Student Member, IEEE, Derrick Wing Kwan Ng, Senior Member, IEEE, Jun Zhu, Member, IEEE, and Robert Schober, Fellow, IEEE, "Robust and Secure Resource Allocation for Full-Duplex MISO Multicarrier NOMA Systems," IEEE Tran. on Commun., (Early Access), Apr. 2018.

[38] Guangxu Zhu, Student Member, IEEE, Kaibin Huang, Senior Member, IEEE, Vincent K. N. Lau, Fellow, IEEE, Bin Xia, Senior Member, IEEE, Xiaofan Li Member, IEEE, and Sha Zhang, "Hybrid Beamforming via the Kronecker Decomposition for the Millimeter-Wave Massive MIMO Systems," IEEE Journal on Selected Areas in Commun., vol. 35, no. 9 pp. 2097-2114, Sept. 2017.

[39] Pankaj Kumar Dalela, Pramod Bhave, Pushpender Yadav, Anshul Yadav and Vipin Tyagi, "Beam Division Multiple Access (BDMA) and Modulation Formats for 5G: Heir of OFDM?," Information Networking (ICOIN), 2018 Intern. Conf., pp. 450–455, Apr. 2018.

[40] Akbar M. Sayeed and Nader Behdad, "Continuous aperture phased MIMO: a new architecture for optimum line-of-sight links," IEEE Antennas and Propagation Symposium, pp. 293–296, 2011.

Appendix

Appendix-1

The non-linear equation of condition_1 of equation 3.33 can be simplified and converted in to linear inequality as as follows:

$$\mathbf{R_{t,m} \geq R_{min}}$$

Replacing equation 3.30 we get,

$$\log_{\mathbf{2}}^{\mathbf{(1+SINR)}} \geq \mathbf{R_{min}}$$

$$\log_{\mathbf{2}}^{(\mathbf{1}+\psi_{\mathbf{t,m}})} \geq \mathbf{R_{min}}$$

Replacing equation 3.28 we get,

$$\log_{\mathbf{2}}^{(\mathbf{1}+\frac{||\mathbf{h_{t,m}^H g_m}||_\mathbf{2}^\mathbf{2}\mathbf{P_{t,m}}}{\tau_{\mathbf{t,m}}})} \geq \mathbf{R_{min}}$$

$$\mathbf{2^{R_{min}}} \geq \mathbf{1} + \frac{||\mathbf{h_{t,m}^H g_m}||_\mathbf{2}^\mathbf{2}\mathbf{P_{t,m}}}{\tau_{\mathbf{t,m}}}$$

$$\mathbf{2^{R_{min}}} - \mathbf{1} \geq \frac{||\mathbf{h_{t,m}^H g_m}||_\mathbf{2}^\mathbf{2}\mathbf{P_{t,m}}}{\tau_{\mathbf{t,m}}}$$

Let us denote $2^{R_{min}} - 1$ to be Δ for simplicity,

$$\mathbf{\Delta} \geq \frac{||\mathbf{h_{t,m}^H g_m}||_\mathbf{2}^\mathbf{2}\mathbf{P_{t,m}}}{\tau_{\mathbf{t,m}}}$$

$$\mathbf{\Delta}\tau_{\mathbf{t,m}} \geq ||\mathbf{h_{t,m}^H g_m}||_\mathbf{2}^\mathbf{2}\mathbf{P_{t,m}}$$

From equation 3.29, replacing $\tau_{t,m}$ we have,

$$\mathbf{\Delta}||\mathbf{h_{t,m}^H g_m}||_\mathbf{2}^\mathbf{2}\sum_{\mathbf{i=1}}^{\mathbf{t-1}}\mathbf{p_{i,m}} + \mathbf{\Delta}\sum_{\mathbf{j\neq m}}||\mathbf{h_{t,m}^H g_j}||_\mathbf{2}^\mathbf{2}\sum_{\mathbf{i=1}}^{|\mathbf{s_j}|}\mathbf{p_{i,j}} + \mathbf{\Delta}\sigma^\mathbf{2} \geq ||\mathbf{h_{t,m}^H g_m}||_\mathbf{2}^\mathbf{2}\mathbf{P_{t,m}}$$

$$\mathbf{\Delta}||\mathbf{h_{t,m}^H g_m}||_\mathbf{2}^\mathbf{2}\sum_{\mathbf{i=1}}^{\mathbf{t-1}}\mathbf{p_{i,m}} + \mathbf{\Delta}\sum_{\mathbf{j\neq m}}||\mathbf{h_{t,m}^H g_j}||_\mathbf{2}^\mathbf{2}\sum_{\mathbf{i=1}}^{|\mathbf{s_j}|}\mathbf{p_{i,j}} - ||\mathbf{h_{t,m}^H g_m}||_\mathbf{2}^\mathbf{2}\mathbf{P_{t,m}} \geq -\mathbf{\Delta}\sigma^\mathbf{2}$$

This is the equation used in both equation 3.66 and 3.67 as $\theta_{t,m}$,

$$\theta_{\mathbf{t,m}} = \mathbf{\Delta}||\mathbf{h_{t,m}^H g_m}||_\mathbf{2}^\mathbf{2}\sum_{\mathbf{i=1}}^{\mathbf{t-1}}\mathbf{p_{i,m}^{(y)}} + \mathbf{\Delta}\sum_{\mathbf{j\neq m}}||\mathbf{h_{t,m}^H g_j}||_\mathbf{2}^\mathbf{2}\sum_{\mathbf{i=1}}^{|\mathbf{s_j}|}\mathbf{p_{i,j}^{(y)}} - ||\mathbf{h_{t,m}^H g_m}||_\mathbf{2}^\mathbf{2}\mathbf{p_{t,m}^{(y)}} + \mathbf{\Delta}\sigma^\mathbf{2} \tag{6.1}$$

Appendix-2

The Mean Square Error (MSE) of equation 3.43 and 3.44 can be solved as shown below.

$$\mathbf{e_{t,m}} = \mathbf{E}\Big\{|\mathbf{s_{t,m}} - \mathbf{d_{t,m}}[(\mathbf{r_{t,m}})_{\mathbf{required}} + (\tilde{\mathbf{r}}_{\mathbf{t,m}})_{\mathbf{Intra-BI}} + (\mathbf{r_{t,m}})_{\mathbf{Inter-BI}} + \mathbf{w_{t,m}}]|^{\mathbf{2}}\Big\}$$

$$\mathbf{e_{t,m}} = \mathbf{E}\Big\{|\mathbf{s_{t,m}} - \mathbf{d_{t,m}}[\mathbf{h^H_{t,m}g_m}\sqrt{\mathbf{p_{t,m}}}\mathbf{s_{t,m}} + \mathbf{h^H_{t,m}g_m}\sum_{\mathbf{i=1}}^{\mathbf{t-1}}\sqrt{\mathbf{p_{i,m}}}\mathbf{s_{i,m}} + \mathbf{h^H_{t,m}}\sum_{\mathbf{j\neq m}}\sum_{\mathbf{i=1}}^{|\mathbf{s_j}|}\mathbf{g_j}\sqrt{\mathbf{p_{i,j}}}\mathbf{s_{i,j}}\Big\}$$
$$+ \mathbf{w_{t,m}}]|^{\mathbf{2}}$$
$$= |\mathbf{1} - \mathbf{d_{t,m}}\sqrt{\mathbf{P_{t,m}}}\mathbf{h^H_{t,m}g_m}|^{\mathbf{2}} + |\mathbf{d_{t,m}}|^{\mathbf{2}}||\mathbf{h^H_{t,m}g_m}||^{\mathbf{2}}_{\mathbf{2}}\sum_{\mathbf{i=1}}^{\mathbf{t-1}}\mathbf{p_{i,m}} + |\mathbf{d_{t,m}}|^{\mathbf{2}}\sum_{\mathbf{j\neq m}}||\mathbf{h^H_{t,m}g_j}||^{\mathbf{2}}_{\mathbf{2}}\sum_{\mathbf{i=1}}^{|\mathbf{s_j}|}\mathbf{p_{i,j}}$$
$$+ |\mathbf{d_{t,m}}|^{\mathbf{2}}\sigma^{\mathbf{2}} \tag{6.2}$$

Appendix-3

The partial derivation of equation 3.45 of the optimal equalization coefficient $d^0_{t,m}$ can be given by,

$$\frac{\partial \mathbf{e_{t,m}}}{\partial \mathbf{d_{t,m}}}\Bigg\|_{\mathbf{d^0_{t,m}}} = \mathbf{0} \tag{6.3}$$

We know that $e_{t,m}$ of equation 6.2 is given by,

$$\begin{aligned}\mathbf{e_{t,m}} &= \mathbf{E}\Big\{|\mathbf{s_{t,m}} - \mathbf{d_{t,m}}\hat{\mathbf{r}}_{\mathbf{t,m}}|^{\mathbf{2}}\Big\} \\ &= |\mathbf{1} - \mathbf{d_{t,m}}\sqrt{\mathbf{P_{t,m}}}\mathbf{h^H_{t,m}g_m}|^{\mathbf{2}} + |\mathbf{d_{t,m}}|^{\mathbf{2}}||\mathbf{h^H_{t,m}g_m}||^{\mathbf{2}}_{\mathbf{2}}\sum_{\mathbf{i=1}}^{\mathbf{t-1}}\mathbf{p_{i,m}} \\ &+ |\mathbf{d_{t,m}}|^{\mathbf{2}}\sum_{\mathbf{j\neq m}}||\mathbf{h^H_{t,m}g_j}||^{\mathbf{2}}_{\mathbf{2}}\sum_{\mathbf{i=1}}^{|\mathbf{s_j}|}\mathbf{p_{i,j}} + |\mathbf{d_{t,m}}|^{\mathbf{2}}\sigma^{\mathbf{2}}\end{aligned} \tag{6.4}$$

Then after derivation we have,

$$\mathbf{0} = -\mathbf{d_{t,m}}\sqrt{\mathbf{P_{t,m}}}\mathbf{h^H_{t,m}g_m} + \mathbf{p_{t,m}}||\mathbf{h^H_{t,m}g_m}||^2_2 + ||\mathbf{h^H_{t,m}g_m}||^2_2 \sum_{\mathbf{i=1}}^{\mathbf{t-1}} \mathbf{p_{i,m}} + \sum_{\mathbf{j \neq m}} ||\mathbf{h^H_{t,m}g_j}||^2_2 \sum_{\mathbf{i=1}}^{|\mathbf{s_j}|} \mathbf{p_{i,j}} + \sigma^2 \tag{6.5}$$

$$\mathbf{d^0_{t,m}}\sqrt{\mathbf{P_{t,m}}}\mathbf{h^H_{t,m}g_m} = \mathbf{p_{t,m}}||\mathbf{h^H_{t,m}g_m}||^2_2 + ||\mathbf{h^H_{t,m}g_m}||^2_2 \sum_{\mathbf{i=1}}^{\mathbf{t-1}} \mathbf{p_{i,m}} + \sum_{\mathbf{j \neq m}} ||\mathbf{h^H_{t,m}g_j}||^2_2 \sum_{\mathbf{i=1}}^{|\mathbf{s_j}|} \mathbf{p_{i,j}} + \sigma^2 \tag{6.6}$$

After derivation of $e_{t,m}$ with respect to $d_{t,m}$, and replacing the value of $\tau_{t,m}$ of equation 3.29, we get,

$$\tau_{\mathbf{t,m}} = ||\mathbf{h^H_{t,m}g_m}||^2_2 \sum_{\mathbf{i=1}}^{\mathbf{t-1}} \mathbf{p_{i,m}} + \sum_{\mathbf{j \neq m}} ||\mathbf{h^H_{t,m}g_j}||^2_2 \sum_{\mathbf{i=1}}^{|\mathbf{s_j}|} \mathbf{p_{i,j}} + \sigma^2 \tag{6.7}$$

$$\mathbf{d^0_{t,m}}\sqrt{\mathbf{P_{t,m}}}\mathbf{h^H_{t,m}g_m} = [\mathbf{p_{t,m}}||\mathbf{h^H_{t,m}g_m}||^2_2 + \tau_{\mathbf{t,m}}] \tag{6.8}$$

Then, we have,

$$\mathbf{d^0_{t,m}} = (\sqrt{\mathbf{p_{t,m}}}\mathbf{h^H_{t,m}g_m})(\mathbf{p_{t,m}}||\mathbf{h^H_{t,m}g_m}||^2_2 + \tau_{\mathbf{t,m}})^{-1} \tag{6.9}$$