



**JIMMA UNIVERSITY**

**JIMMA INSTITUTE OF TECHNOLOGY**

**SCHOOL OF GRADUATE STUDIES**

**FACULTY OF ELECTRICAL AND COMPUTER  
ENGINEERING**

**JOINT USER CLUSTERING AND PSEUDO RANDOM  
BASED PILOT DECONTAMINATION IN MASSIVE  
MIMO TDD SYSTEM**

By

Edemiale Tamiru

This thesis is submitted to School of Graduate Studies of Jimma University in  
partial fulfilment of the requirements for the degree of

Master of Science

in

Communication Engineering

March 2022

Jimma, Ethiopia

**JIMMA UNIVERSITY**  
**JIMMA INSTITUTE OF TECHNOLOGY**  
**SCHOOL OF GRADUATE STUDIES**  
**FACULTY OF ELECTRICAL AND COMPUTER**  
**ENGINEERING**

**JOINT USER CLUSTERING AND PSEUDO RANDOM  
BASED PILOT DECONTAMINATION IN MASSIVE  
MIMO TDD SYSTEM**

By  
Edemiale Tamiru

Advisor: Dr. Kinde Anlay

Co-Advisor: Mr. Getachew Alemu

Submission Date: November, 2022

# Declaration

I declare that this thesis work titled "Joint User Clustering and Pseudo Random Based Pilot Decontamination in Massive MIMO TDD System" has been presented in my own except I refer different works that are cited. Also this thesis has not been previously presented for any other certified or uncertified qualification by any one.

## THESIS SUBMITTED BY:

Edemialem Tamiru \_\_\_\_\_  
Signature Date

## APPROVED BY ADVISORS:

ADVISOR: \_\_\_\_\_  
Signature Date

CO-ADVISOR: \_\_\_\_\_  
Signature Date

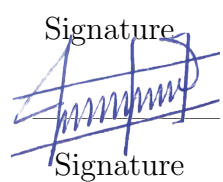
## APPROVED BY THE BOARD OF EXAMINERS :

Approved by faculty of Electrical and computer engineering research thesis Examination members

1. \_\_\_\_\_  
Chairman Signature Date

2. \_\_\_\_\_  
Examiner (Internal) Signature Date

*Fikreselam Gared(PhD)* \_\_\_\_\_  
3. \_\_\_\_\_  
Examiner (External) Signature Date

  
Signature

*09/03/2022*  
Date

# Acknowledgements

First and foremost, I thank God, the Almighty, for all of his blessings throughout my life and for providing me with the ability to proceed effectively in my studies. Second, I would like to convey my heartfelt gratitude to my advisors, Dr. Kinde Anlay and Mr. Getachew Alemu, for their valuable suggestions, advice, and continuous encouragement during my thesis. And I owe a huge sense of gratitude to everyone who helped me to complete my master's thesis in any way. It is also an honor for me to acknowledge Jimma University, which provided me with a scholarship for the duration of my master's studies. Finally, I want to offer my heartfelt gratitude to my wonderful parents and friends for their constant support.

# Abstract

Wireless communication technology is now progressing to 5G and beyond. The number of wireless connections is growing in line with the growing number of subscribers and demands. One of the promising technology to overcome this is the usage of massive Multiple Input Multiple Output systems (M-MIMO). However, M-MIMO system's performance suffers as a result of the random nature of the wireless channels. In Pilot based channel estimation there is a problem of pilot contamination due to the reuse of pilot sequences, which is the bottleneck for M-MIMO systems. This problem arises when the number of user terminals grows, they compute for the limited pilot sequence with minimum coherence interval in channel estimation. Therefore, users' channels to the BS need to be effectively estimated by applying proper pilot assignment method to achieve the performance of M-MIMO.

In this thesis, joint user clustering and pseudo random sequence based pilot decontamination for M-MIMO TDD system is proposed. The K-means clustering algorithm is used to cluster users in the cell depending on the severity of the problem affecting each user: cell center users and cell edge users. The same pilot sequences are reused to the cell center users and orthogonal pilot sequences are assigned to the cell edge users. Pseudo random sequences are used to improve the orthogonality of pilot sequence of cell center users of different cells. At the BS, the MMSE channel estimation has been used. Then, using the estimated channels, multi cell MMSE detectors have been used to detect data from users in the uplink.

The performance of the proposed system is analyzed based on the following metrics: NMSE, SINR, and target cell's SE. To simulate the results Matlab software has been used. The simulation results of the proposed system when compared to random pilot assignment, it shows that for cell edge users and cell center users the normalized mean square error(NMSE) of the estimated channel has improved by 7.76% and 19.021%, respectively. Furthermore, the proposed technique improves the target cell's SE by 8.714%.

*Keyword: Massive MIMO, Coherence Block, Ortogonal Pilot Sequence, Pseudo Random sequence, Pilot Contamination*

# Contents

<b>Declaration</b>	<b>i</b>
<b>Acknowledgements</b>	<b>ii</b>
<b>Abstract</b>	<b>iii</b>
<b>Contents</b>	<b>iv</b>
<b>List of Figures</b>	<b>vii</b>
<b>List of Tables</b>	<b>ix</b>
<b>Abbreviations</b>	<b>x</b>
<b>List of Symbols</b>	<b>xi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Background . . . . .	1
1.2 Statement of the Problem . . . . .	4
1.3 Objectives of the Thesis . . . . .	4
1.3.1 General Objective . . . . .	4
1.3.2 Specific Objectives . . . . .	4
1.4 Significance of the Thesis . . . . .	5
1.5 Methodology . . . . .	5
1.6 Scope of the Thesis . . . . .	6
1.7 Contributions of the Thesis . . . . .	6
1.8 Thesis Outline . . . . .	7
<b>2 Literature Review</b>	<b>8</b>
2.1 Literature Review . . . . .	8
<b>3 Overview of Massive MIMO System</b>	<b>13</b>
3.1 Massive MIMO . . . . .	13
3.1.1 Uplink Transmission in Massive MIMO . . . . .	14
3.1.2 Channel Coherence Block . . . . .	15
3.1.3 Channel Model . . . . .	18
3.1.3.1 Spatially Correlated Rayleigh fading Channel . . . . .	19

3.1.4	Pilot Sequence for Channel Estimation . . . . .	21
3.1.5	Channel Estimation . . . . .	22
3.1.5.1	Least Square channel estimation . . . . .	23
3.1.5.2	MMSE Channel Estimation Under correlated Rayleigh fading channel . . . . .	24
3.1.6	Characterization of Multi-antenna Channel . . . . .	25
3.1.6.1	Favorable Propagation . . . . .	25
3.1.6.2	Channel hardening . . . . .	26
3.1.7	Linear Detector processing in Uplink . . . . .	27
3.1.7.1	Maximum Ratio Combining (MRC) Detector . . . . .	28
3.1.7.2	Zero Forcing (ZF) processing . . . . .	28
3.1.7.3	MMSE Detector . . . . .	29
3.1.8	Pseudo Random Sequence . . . . .	29
3.1.8.1	Generation of Pseudo Random Sequence . . . . .	30
3.1.8.2	Linear Feedback Shift Register . . . . .	30
3.1.8.3	Correlation and Orthogonality . . . . .	32
3.2	Pilot Contamination in Massive MIMO System . . . . .	33
3.2.1	Uplink Pilot Transmission and Channel Estimation . . . . .	33
3.2.2	Uplink Data Transmission . . . . .	36
<b>4</b>	<b>System Model Under Spatially Correlated Rayleigh fading Channel</b>	<b>39</b>
4.1	The Proposed Pilot Decontamination Scheme . . . . .	39
4.1.1	User Grouping Algorithm . . . . .	40
4.2	The Pseudo Random Based Pilot for Center Users . . . . .	41
4.3	Channel Estimation . . . . .	44
4.3.1	Center Users Channel Estimate and MSE . . . . .	46
4.3.2	Edge Users Channel Estimate and MSE . . . . .	47
4.4	Uplink Users Data Detection . . . . .	48
4.5	System Computational Complexity . . . . .	51
<b>5</b>	<b>Result and Discussion</b>	<b>55</b>
5.1	Simulation Parameters and Assumption Made . . . . .	55
5.2	User Distribution in Each Cell and User Clustering . . . . .	56
5.3	Normalized Mean Square Error(NMSE) of Channel Estimate . . . . .	58
5.4	Commulative Distribution Function(CDF) of SINR and SE . . . . .	58
5.5	Cell Sum Spectral Efficiency(SE) Against BS Antennas . . . . .	60
5.6	Sum Spectral Efficiency(SE) Versus Average Transmit Power of Users . . . . .	62
5.7	Sum Spectral Efficiency(SE) Versus Coherence Block Length . . . . .	63
5.8	System Complexity Versus Number Of BS Antenna . . . . .	64
<b>6</b>	<b>Conclusion and Recommendation</b>	<b>66</b>
6.1	Conclusion . . . . .	66
6.2	Recommendations . . . . .	68

---

<b>Bibliography</b>	<b>69</b>
<b>Appendix</b>	<b>74</b>
<b>A</b>	<b>75</b>
A.1 The law of large number . . . . .	75
A.2 MMSE based Channel Estimation . . . . .	75
A.3 Proof of Equation (4.16) . . . . .	77
A.4 Basic of Hexagonal cell . . . . .	78



# List of Figures

Figure 1.1	The Uplink and Downlink channel model of multi-cell multi-user M-MIMO TDD system. . . . .	2
Figure 1.2	Speed improvement of wireless networks over the years [6]. . . . .	2
Figure 1.3	Intercell Interference in Pilot Transmission Phase for M-MIMO system[11]. . . . .	3
Figure 3.1	Block Diagram of Uplink Massive MIMO system for $l_{th}$ cell BS. . . . .	15
Figure 3.2	The Frequency Division duplexing (FDD) and Time Division duplexing (TDD) in a Coherence Block. . . . .	16
Figure 3.3	Block diagram of OFDM multiplexing with Coherence interval, wherein $N_c$ is number of sub carrier in coherence block and $N_s$ is the number of sub carriers. . . . .	17
Figure 3.4	Local Scattering model of k user in the jth cell, scattering is centered around the UE with a nominal angle of $\theta_{Nominal}$ degree with reference BS antenna and of $\theta_{deviation}$ angular standard deviation (ASD) of the multipath components. . . . .	20
Figure 3.5	TDD Based Transmission Protocol. . . . .	23
Figure 3.6	Block diagram of Uplink Data Detection in Massive MIMO. . . . .	27
Figure 3.7	Linear Feed Back Shift Register. . . . .	31
Figure 3.8	The problem of pilot contamination between two cells in massive MIMO. . . . .	34
Figure 4.1	Block Diagram of Pseudo Random Based Pilot Sequence. . . . .	42
Figure 5.1	Simulation Result of User Distribution Across L=7 cells using R=500m and K=10. Wherein the red crosses and yellow triangle denote users and BS respectively. . . . .	57
Figure 5.2	Simulation scenario of User Clustering in each Cell. Wherein the blue diamond and red circle represent the center and edge users respectively. . . . .	57
Figure 5.3	Simulation Result of NMSE versus BS Antenna with R=500m, K=10,L=7 and ASD = 10 degree . . . . .	59
Figure 5.4	Simulation Result of CDF of the Uplink SINR, with M = 192, K = 10 and ASD= 10 degree . . . . .	60
Figure 5.5	Simulation Result of CDF of the Uplink SE per UE, with M = 192, K = 10,and ASD = 10 degree . . . . .	61

---

Figure 5.6	Simulation Result of target cell Sum SE versus BS Antenna, with $K = 10$ , and with ASD = 10 degree . . . . .	62
Figure 5.7	Simulation Result of the sum SE [bps/Hz/cell] against average transmit power of users with $M = 192$ , $K = 10$ , $L=7$ and ASD = 10 degree . . . . .	63
Figure 5.8	Simulation Result of the sum SE [bps/Hz/cell] against Coherence Block Length with $M = 192$ , $K = 10$ , $L=7$ and ASD = 10 degree . . . . .	64
Figure 5.9	Simulation Result of System Complexity Versus Number Of BS Antenna with $K = 10$ , and $L=7$ . . . . .	65
Figure A.1	Diagram of Hexagonal cell in terms of distance . . . . .	79
Figure A.2	Hexagonal cell in terms of angular domain[49] . . . . .	79

# List of Tables

Table 4.1	The Computational Complexity Analysis	54
Table 5.1	Parameters Used for Simulation	56

# Abbreviations

<b>BS</b>	<b>B</b> ase <b>S</b> tation
<b>CSI</b>	<b>C</b> annel <b>S</b> tate <b>I</b> nformation
<b>FDD</b>	<b>F</b> requency <b>D</b> ivision <b>D</b> uplexing
<b>IID</b>	<b>I</b> ndependent <b>I</b> dentically <b>D</b> istributed
<b>LS</b>	<b>L</b> east <b>S</b> quare
<b>LFSR</b>	<b>L</b> inear <b>F</b> eedback <b>S</b> hift <b>R</b> egister
<b>MIMO</b>	<b>M</b> ultiple <b>I</b> nput <b>M</b> ultiple <b>O</b> utput
<b>M-MIMO</b>	<b>M</b> assive <b>M</b> ultiple <b>I</b> nput <b>M</b> ultiple <b>O</b> utput
<b>MRC</b>	<b>M</b> aximum <b>R</b> atio <b>C</b> ombining
<b>MMSE</b>	<b>M</b> inimum <b>M</b> ean <b>S</b> quare <b>E</b> rror
<b>MSE</b>	<b>M</b> ean <b>S</b> quare <b>E</b> rror
<b>MU-MIMO</b>	<b>M</b> ultiuser <b>M</b> ultiple <b>I</b> nput <b>M</b> ultiple <b>O</b> utput
<b>NMSE</b>	<b>N</b> ormalized <b>M</b> ean <b>S</b> quare <b>E</b> rror
<b>SDMA</b>	<b>S</b> patial <b>D</b> ivision <b>M</b> ultiple <b>A</b> ccess
<b>SE</b>	<b>S</b> pectral <b>E</b> fficiency
<b>SINR</b>	<b>S</b> ignal to <b>I</b> nterference plus <b>N</b> oise <b>R</b> atio
<b>SISO</b>	<b>S</b> ingle <b>I</b> nput <b>S</b> ingle <b>O</b> utput
<b>SU-MIMO</b>	<b>S</b> ingle user <b>M</b> ultiple <b>I</b> nput <b>M</b> ultiple <b>O</b> utput
<b>TDD</b>	<b>T</b> ime <b>D</b> ivision <b>D</b> uplexing
<b>UE</b>	<b>U</b> ser <b>E</b> quipment
<b>ZF</b>	<b>Z</b> ero <b>F</b> orcing

# List of Symbols

$\alpha$	The path loss exponent
$\beta_{jk}^l$	The large scale fading coefficient of user k in cell $j_{th}$
$\rho_u$	The average transmit power for each user in the uplink
$\tau_c$	The length of coherence block
$\tau_p$	The length of pilot sequence in a coherence block
$\theta_{ji}$	The cross correlation of the pseudo-random sequence between cell j and cell i
$\varphi_{jk}$	The pilot sequence of user k in cell j
$\Psi_{jk}^l$	, The covariance matrix of the received pilot signal
$\sigma_{sha}$	The standard deviation of the shadow fading
$a_p \in C^M$	The array responses of the multipath component
$B_c$	The Coherence bandwidth
$C_{jk}^l$	The covariance matrix of the channel $h_{jk}^l$
$d_H$	The antenna spacing
$g_{jK}^l$	The small-scale fading vector of user k in cell j to cell l th BS
$\hat{d}_{jk}$	The detected symbol of the kth user in cell j
$h_{jk}^l$	The channel vector of kth user of the $j_{th}$ cell to lth cell BS
$\hat{h}_{jk}^l$	The channel estimate for channel of user k in cell j to lth BS
$\tilde{h}_{jk}^l$	The estimation error for channel of user k in cell j to lth BS
$H_j^l$	The uplink channel matrix of the $j_{th}$ cell to $l_{th}$ cell BS
$K_{Cu}$	The number of center users
$K_{Eu}$	The number of edge users
<b>K</b>	Number of users in a cell
<b>M</b>	Number of antennas at BS

---

$P_{path}$	The multipath component
$r_{jk}^l$	The distance between user k in cell j to lth BS
$R_p^l$	The uplink received pilot signal at $l_{th}$ cell BS
$R$	The Cell radius
$SINR_{jk}^{UL}$	The uplink SINR of the user $K$ in cell j
$T_c$	The coherence time
$u_{jk}$	The uplink detector for the kth user in the jth cell
$z_{jk}^l$	Shadow fading coefficient of user k in cell j

# Chapter 1

## Introduction

### 1.1 Background

Now a days due to the increasing number of subscribers, the demand for higher data rates, coverage, spectral efficiency, capacity, reduced latency, etc. forced researchers to the evolution of 5G and beyond networks. Therefore, to address these demands, there are different technologies like millimeter wave, small cell and multiple input multiple out put (MIMO) communications are proposed [1][2]. The spectral efficiency (SE) and bandwidth (BW) are the two aspects, which are used to improve the throughput of wireless radio communication; though, the bandwidth is a scarce resource, so, improving the spectral efficiency is the way to maximize throughput. A well-known way to increase spectral efficiency is using multiple antennas at the transmitter and receiver [1], [3].

A massive multiple input multiple output (M-MIMO) system is one form of conventional MIMO. It has large base station (BS) antennas serving a number of users simultaneously. Therefore, M-MIMO technology is one of the most promising solutions to improve the aforementioned challenges. It greatly improves the spectral efficiency and energy efficiency of wireless communication, since each cell has a BS with a large number of antennas, that allows simultaneous utilization of resources by different users in the cell [4]. It uses Spatial division multiple access

(SDMA) to achieve a multiplexing gain by serving multiple users equipment (UE) on the same time-frequency resources. It has more BS antennas than UEs per cell to achieve efficient interference suppression. The improved energy efficiency obtained from the array gain provided by the large set of antennas. The main difference between M-MIMO and conventional Multiuser MIMO are: BS antennas are typically much larger than users, only the base station learns the channel, and last simple linear signal processing is used both on the uplink and on the downlink. Therefore, these features render M-MIMO scalable to the number of BS antennas. [5]. Figure 1.2 shows the speed improvement of wireless networks for M-MIMO

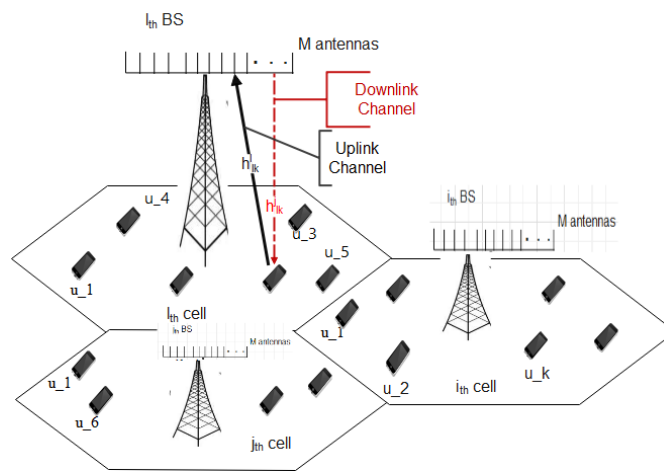


FIGURE 1.1: The Uplink and Downlink channel model of multi-cell multi-user M-MIMO TDD system.

over the years relative to single-input-single-output (SISO) systems, single user (SU-MIMO), and multiple users (MU-MIMO) networks [6].

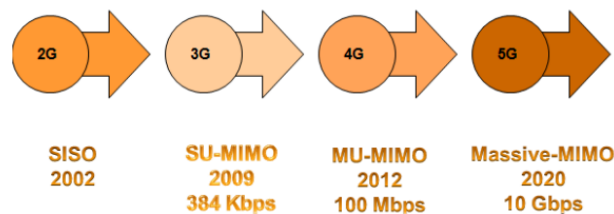


FIGURE 1.2: Speed improvement of wireless networks over the years [6].

The advantages of having numerous antennas at the BS become apparent when the BS is aware of the desired UE's channel response. As a result, obtaining precise channel state information is required for the base station to get those



benefits. Because the channel's coherence duration is limited [7], in the pilot based channel estimation there are a finite number of orthogonal pilots. Due to increasing number of served users, users in other cells must reuse these pilots. As a result, user's channel contamination occurs by a linear combination of the channels of other users who utilize the same pilots [8]. The same pilot sequences are reused across cells, results in increased inter-cell interference as shown in Figure 1.3.

The widespread usage of pseudo-random signal processing as a viable technique for many civilian and commercial applications has emerged from the expansion of digital mobile radio systems, the increasing demand for positioning systems, as well as advances in integrated circuit complexity. During the late 1980s, pseudo-random signal processing got considerable attention, particularly with the development of cellular mobile radio systems [9]. The potential to avoid or mitigate multipath propagation, minimize to interference, and the ability to share allocated bandwidth with other users are all significant elements of these processing systems for cellular radio. The linear feedback shift register [10] generates good autocorrelation and cross-correlation features to replicated random sequences.

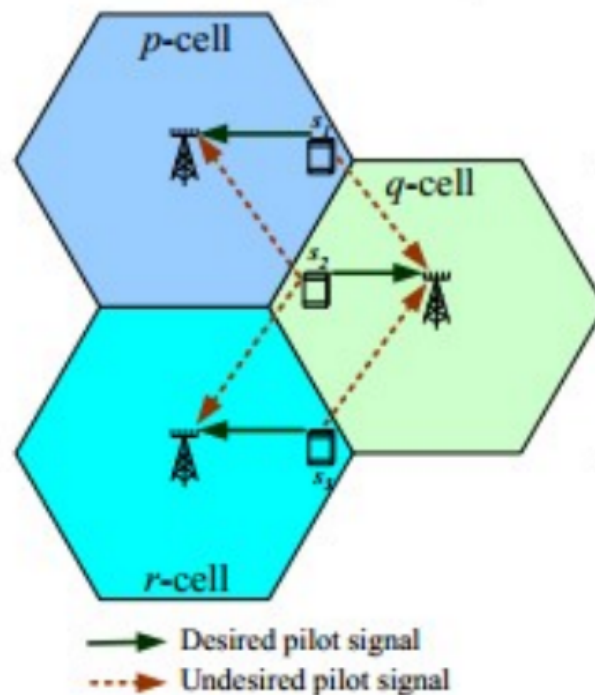


FIGURE 1.3: Intercell Interference in Pilot Transmission Phase for M-MIMO system[11].

## 1.2 Statement of the Problem

Massive MIMO systems use a very large number of antennas at the base station to simultaneously serve a large number of users. However, due to the short coherence interval, the number of orthogonal pilot sequences that can be assigned for channel estimation is limited. Furthermore, as the number of subscribers increases, this limited number of orthogonal pilot sequences can be re-used in the adjacent cell users. Due to these, in the uplink pilot transmission re-used pilot sequences can overlap with each other and results pilot contamination in M-MIMO. As a result of this issue, the SINR and the uplink transmission rates of M-MIMO is degraded, since the receive combining used at the BS are highly dependent on the estimates of the the channel state information.

## 1.3 Objectives of the Thesis

### 1.3.1 General Objective

The general objective of this thesis is reduction of pilot contamination by integrating user clustering algorithm and pseudo random sequences in massive MIMO TDD cellular networks.

### 1.3.2 Specific Objectives

The specific objectives of this thesis includes:

- To cluster users in each cell using K-means clustering algorithm.
- To apply pseudo random based pilot assignment for cell center users.
- To evaluate the NMSE in channel estimation for a random pilot assignment.

- To analyse the NMSE in pilot-based channel estimation using joint user clustering and pseudo-random sequence.
- To evaluate uplink SINR of users in a random pilot assignment.
- To evaluate uplink SINR of users in pilot-based channel estimation using user grouping and pseudo-random sequence.
- To analyse the uplink SE of each user and the target's cell sum SE for a random pilot assignment.
- To evaluate the uplink SE of each user and the target's cell sum SE for joint user clustering and pseudo-random sequence based pilot assignment.
- To analyse complexity of joint user clustering and pseudo-random sequence based pilot assignment.

## 1.4 Significance of the Thesis

M-MIMO is one of the promising 5G technologies, with hundreds to thousands of antennas installed at the base station. It improves energy efficiency and spectrum efficiency by serving multiple user terminals at the same time. However, in channel estimation phase there is a problem of pilot contamination in M-MIMO due to the reuse of pilot sequences. So, to improve the performance of M-MIMO the problem of pilot contamination has to be mitigated. Therefore, Joint user clustering and pseudo-random sequence based channel estimation were employed to reduce the effect of pilot contamination, hence improving M-MIMO performance.

## 1.5 Methodology

In this thesis, to fulfill the objectives outlined in the objectives section effectively, the first step begins with the acquisition of different literatures about pilot decontamination techniques in massive MIMO system. The second task in proposed

pilot decontamination scheme is clustering users as cell center and cell edge users using K-means clustering algorithm. The clustering depend on the severity of pilot contamination in each user. Then, cell center users pilot is scrambled by pseudo random sequence to reduce the correlation between pilots. The third task is system designing, which includes proposing the appropriate processing technique used and preparing the mathematical modeling. The MMSE Channel estimator is going to be used in channel estimation of each users, which minimizes the inter cell interference due to pilot contamination. Then the MSE of the channel , the uplink SINR and the SE of the target cell for the proposed system is analysed. After that, the mathematical model of the proposed and random pilot assignment system using the MATLAB simulation program is evaluated.

## 1.6 Scope of the Thesis

After collecting the necessary data, we performed a mathematical analysis of the effect of pilot contamination in M-MIMO TDD wireless communication for an urban macro cell by grouping users as center and edge users of the cell in the system and assigning the pilot sequence based on the grouping. The second intention of this thesis is to reduce the channel estimate error of the cell's center users by using a pseudo random sequence. As a result, Matlab software is used to evaluate the mitigation of the pilot contamination of the center users. The SINR of users and the SE of the target cell were simulated by Matlab software as the third step in this thesis.

## 1.7 Contributions of the Thesis

The contributions of the thesis is Summarized as follow:

- Adapting the K-means clustering algorithm in pilot decontamination of massive MIMO system.

- Integrating user clustering and pseudo random sequence
- Mitigating the problem of pilot contamination in a scenario of multi-user multi-cell massive MIMO system, with a more practical channel model of spatially correlated fading channel.
- Applying a more efficient multi-cell MMSE channel estimation and multi-cell MMSE detector at the BS.
- Demonstrating the performance of the proposed pilot decontamination system in terms of channel estimate error, SINR and SE using mathematical and numerical results.

## 1.8 Thesis Outline

There are six chapters in this thesis paper. Literature reviews of pilot decontamination techniques are covered in the second chapter. The third chapter presented M-MIMO system's theoretical basis and the issue of pilot contamination in a M-MIMO system, which is a problem that must be addressed. The proposed pilot decontamination system under correlated Rayleigh fading channel is included in Chapter four. The fifth chapter is the simulation results and discussion. The conclusion and recommendations are presented in the final chapter.

# Chapter 2

## Literature Review

### 2.1 Literature Review

In a M-MIMO TDD 5G system, good CSI helps to maximize network throughput by focusing of transmit power on the DL and collection of receive power on the UL [39]. The pilot signals that are transmitted in the uplink to estimate the channels can be contaminated because of the reuse of non-orthogonal pilot signals in a multi-cell system [40], which in turn reduces the spectrum efficiency and the achievable rates in the network [41]. Therefore, there are a variety of researches are done to mitigate the effect of pilot contamination in massive MIMO 5G wireless communication.

The authors in [42] mitigate the effect of severe pilot contamination of cell edge users in massive MIMO systems they propose Soft Pilot Reuse and Multi-Cell Block Diagonalization Precoding in uncorrelated block-fading Rayleigh fading channel model. First, users are grouped depending on users' large-scale fading coefficients and some user-grouping thresholds that are dependent on system configuration. For channel estimation, least-square was used and they employed matched filter detector at the BS to detect the uplink user's data. Finally, they evaluated their system with  $L=19$  hexagonal cell,  $K=10$  users per cell,  $R=500\text{m}$  cell radius

$M=128$ , and pilot overhead of 0.1. The simulation results show that the probability of distribution of uplink SINR of edge users improved than conventional scheme however when we have seen the uplink SINR cell center users are the same for both of the schemes. Moreover, the uplink achievable rate of cell edge users is improved in the SPR scheme but there is a loss for cell-centered users due to the increased pilot overhead in the SPR scheme. In addition, the average UL cell throughput of the SPR scheme improved with the increased number of BS antennas.

Asynchronous Pilot Scheduling (APS) method is proposed for massive MIMO TDD system to mitigate Pilot contamination in uncorrelated Rayleigh fading channel [43]. In this system, first users are grouped as cell center and cell edge users depending on the large-scale fading coefficient of each user and some threshold parameter. Then it used the asynchronous pilot transmission that used the principle of; after the edge users in all the cells send the pilots in the non-overlapped periods, the center users in all cells send the pilots synchronously to their corresponding BS. They used a least square channel estimator for channel estimation and matched filter to detect at the base station in the uplink data transmission.

Finally, they analyzed the effect of the number of BS antennas on the uplink user's data rate and the uplink cell throughput with  $L=7$  cells,  $K=8$  users,  $R=1600\text{m}$  cell radius, and the same 10dB of Up-link transmitting power for pilot and data. They obtained the data rate of edge users about 2-3bps/Hz/user larger than the conventional scheme when  $M$  is 128 and 256 due to edge users in different cells transmitting pilots in non-overlapped periods so the pilot contamination was mitigated. Furthermore, although the sum rate has been improved, the pilot contamination between the center users is still present.

In reference to [44], uplink asynchronous fractional pilots scheduling is proposed in Massive MIMO systems to reduce pilot contamination. Like paper [43] they agree with pilot contamination in massive MIMO is one of the main obstacles that limit its performance and users are grouped as cell center and cell edge users depending on the large-scale fading coefficient of each user and some threshold parameter. Unlike the previous asynchronous scheduling, since center users are

reused the same pilot sequence they employ time-shifting operation on the pilot transmissions of the center users of each cell, to reduce the pilot contamination due to the reuse of pilot for center users.

Finally, they evaluated their proposed model in terms of uplink sum rate against the number of BS antennas with  $L=7$  cell,  $K=15$  users,  $R=1000$  meter cell radius,  $T=196$  channel coherence time, and same 10dB of Up-link transmitting power for pilot and data. The simulation result verifies that uplink asynchronous fractional pilots' scheduling-based pilot decontamination scheme outperforms SPR and conventional-based pilot decontamination systems. However, there is still a problem of interference between the pilot and the data signal during transmission when the downlink transmit power is large for center users due to the time-shifting operation like the paper in [45]. Relative to the conventional asynchronous integer pilot scheduling, the simulation results show that it is preferred in improving spectral efficiency.

In [46] propose two algorithms separately; the first algorithm employed a path loss user grouping scheme in the mitigation of pilot contamination in a massive MIMO system under uncorrelated Rayleigh fading channel. In this scheme, first users in each cell are grouped as cell center users and cell edge users depending on path loss of each user and decision parameter of 0.9, and the then orthogonal pilot is assigned for cell edged user and the same pilot is reused for cell center user in the adjacent cells. The least Square channel estimator was employed to obtain an estimate of the uplink channel of each user in a TDD system and after channel estimation, the target BS utilizes an MF detector to detect the original data symbol transmitted by the target. Finally, they evaluated their system in terms of quality of channel estimate (NMSE), SINR, and achievable rate perusers against BS antenna with  $L=7$  cell,  $K=15$  users,  $R=1000$  meter cell radius, pilot length of 128, the pilot overhead of 0.05, and same 10dB of average uplink transmitting power for pilot and data of each user. They obtained an improvement of 13.83 dB and 4.68 dB in NMSE of edge users and center users respectively at 256 BS antenna over fully reuse pilot assignment. However, there is still pilot contamination that happened in cell center users due to the reuse of pilot sequences in adjacent cells.



The second algorithm depends on the fact that the problem of pilot contamination is due to the non-orthogonality of a pilot sequence of users used for channel estimation, they employed a pseudo-random code-based pilot decontamination scheme in each cell. They used an uncorrelated Rayleigh fading channel model, and an MMSE channel estimator was employed to estimate the channel. In addition, they assumed single-user mathematical analysis. The simulation result validates that the pseudorandom-based pilot design is effective when the normalized cross-correlation between the pseudo-random sequences used in each cell is small which is close to zero.

The authors in [47] proposed Joint distance-based user grouping and pilot assignment schemes to reduce the effect of pilot contamination in a Massive MIMO TDD system. In this scheme to find, the best user-grouping boundary they used a distance-based user-grouping algorithm that used exhaustive search to find the global optimal distance that maximizes the SINR of both cell edge and cell center users at the same time. After grouping users, they proposed a location-aware pilot assignment, which planned three pilot groups for adjacent cells and determined the user's distance to BS, the planned pilots assigned to the users. They used an uncorrelated Rayleigh fading channel model which does not demonstrate a practical scenario, and the least square channel estimator employed to estimate this channel at the BS. Furthermore, an MF detector is used to detect each user's data at the desired BS. Finally, they try to optimize pilot schedules in each group by maximizing the uplink rate of each user with auction-based optimization pilot assignment. In addition, they evaluate the proposed system in terms of the mean square error of the channel estimate against the number of BS antennas, the distribution of the uplink SINR, and the distribution of the uplink achievable sum-rate with  $M=192$ ,  $K=10$ , and  $R=500$ . The simulation result shows that the proposed system outperforms the SPR based [42] pilot decontamination. However, there is still pilot contamination happening for users assigned with the same pilot sequence. Moreover, in a location-aware pilot assignment as the number of users grow the total number of orthogonal pilot sequence (sum of three pilot groups)

needed increase, even though it is lower relative to the conventional orthogonal pilot assignment has a problem for the limited coherence interval available.

Generally, all papers agree with the pilot sequence has to be reused to compromise the condition of a limited pilot sequence with a large number of user terminals the question is how?

Therefore, in this paper, we used the K-means grouping algorithm to group users as cell center and cell edge users after that we try to address the effect of pilot contamination in center users, which are using the same set of pilot sequences as the neighboring cell and edge users of weak channel condition in correlated Rayleigh fading channel. To do so we used pseudo-random code for center users. It enhances the orthogonality of the pilot sequence for different cells by scrambling the pilot sequence. We assign different pseudo-random codes for different cells to improve the orthogonality of the pilot sequence.

# Chapter 3

## Overview of Massive MIMO System

### 3.1 Massive MIMO

A massive multiple input multiple output system is one form of conventional MIMO. The term MIMO was used to indicate circuits with multiple input and multiple output ports in their original context, and the term multiple input and multiple outputs were used to denote the signals that were entering the communication channel from the multiple antennas at transmitting side and signals received at the multiple antennas of the receiver side. Massive MIMO, as one of the core technologies of 5G, is key to meeting the high-performance requirements and does offer great promises for highly capable 5G with wider bandwidth, more connections, lower latency, and better reliability [12],[13]. Generally, massive MIMO is a physical-layer technology, that equips each BS with hundred to thousands number of antennas that can be used to multiplex many UEs spatially to communicate at the same time-frequency resource.

There are different benefits gained from M-MIMO, Some of them are: it has multiplexing gain due to the usage of SDMA, so The SE per cell can be improved than conventional cellular networks by orders of magnitude and it mitigate the

interference and signal attenuation by receive combining at the receiver and transmit precoding at transmitter [1]. M-MIMO has also better performance in terms of robustness and link reliability, since it used large number of antennas at the BS. Then the diversity gain is increased and fast fading, uncorrelated noise and multi-user interference averages out [3]. The energy efficiency of M-MIMO can also be improved, due to the array gain which focuses transmission power of BS into the direction where each user is located. It helps to improve the SNR by coherently combining the desired signal and cancellation of the interference at the receiver [14],[3].

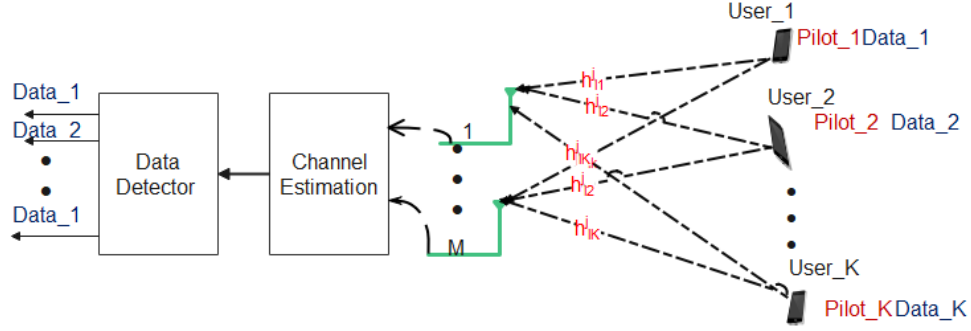
Even though M-MIMO has enormous benefits due to its large number of antennas at the BS, there are challenges faced like Unfavorable Propagation, Array implementation, hardware impairment, pilot contaminations, and others. Therefore, in this thesis, our main target is to address the pilot contamination problem.

### 3.1.1 Uplink Transmission in Massive MIMO

As shown in Figure 3.2 and 1.1, we used  $M$  number of antennas at BS providing service to  $K$  single-antenna users. If we represent the uplink data symbol of  $K$  users in cell  $j$  as  $D_j \in C^{K \times 1}$ , and  $H_j^l \in C^{M \times K}$  is the uplink channel matrix connecting the  $K$  users of the  $j$ th cell to  $l$ th cell BS then, In addition, if we consider the average transmit power of each user of this link  $\rho_u$ , which is equal for all users. Then, the uplink signal received at  $l$ th cell BS is expressed as  $R_u^l$ :

$$R_u^l = \sqrt{\rho_u} H_j^l D_j + N^l \quad (3.1)$$

Then, at the Decoding side, the  $l$ th cell BS will coherently detect the signals transmitted from  $K$  users by using the estimated channel state information (CSI) together with the received signal vector  $R_u^l$  of Equation (3.1) while CSI estimate can be obtained from the uplink pilot sequence. In the preceding sub-section, we discussed the detail of each block of Figure 3.2.

FIGURE 3.1: Block Diagram of Uplink Massive MIMO system for  $l_{th}$  cell BS.

### 3.1.2 Channel Coherence Block

Coherence block contains several time samples and subcarriers over which the stochastic channel response has become approximately constant and flat fading. Each coherence block contains  $\tau_c = B_c T_c$  complex-valued samples, wherein  $B_c$  is the coherence bandwidth and  $T_c$  is the coherence time and they depend on the carrier frequency, UE speed and delay spread [2]. As stated in [15], it considers  $d_1$  and  $d_2$  propagation path length for the two paths and if the receiver moves with velocity ( $V$ ), then the coherence time  $T_c$ , given as

$$T_c = \frac{\lambda}{4V} \quad (3.2)$$

In addition, the range of frequency over which the magnitude of the channel response is approximately flat is expressed with  $B_c$  and given as

$$B_c = \frac{C}{2|d_2 - d_1|} \quad (3.3)$$

Where  $|d_1 - d_2|$  is the path difference between the longest and the shortest propagation Paths, which is related to delay spread  $T_d = \frac{|d_2 - d_1|}{C}$ .

The length of the channel Coherence block determines the frequency range and the time duration over which the changing radio channel to be approximately constant. Knowing this range helps to determine the number of pilot overhead transmitted to update the knowledge of the channel in channel estimation. Therefore, the

length of coherence block in mobile radio channels has great practical importance in massive MIMO technology systems, which is one of the factors that determines the achievable spectral efficiency of the system and the maximum number of users that the system can serve simultaneously [16]. Therefore, it is of great importance to have a reasonable estimate of the duration of the coherence block of different channels.

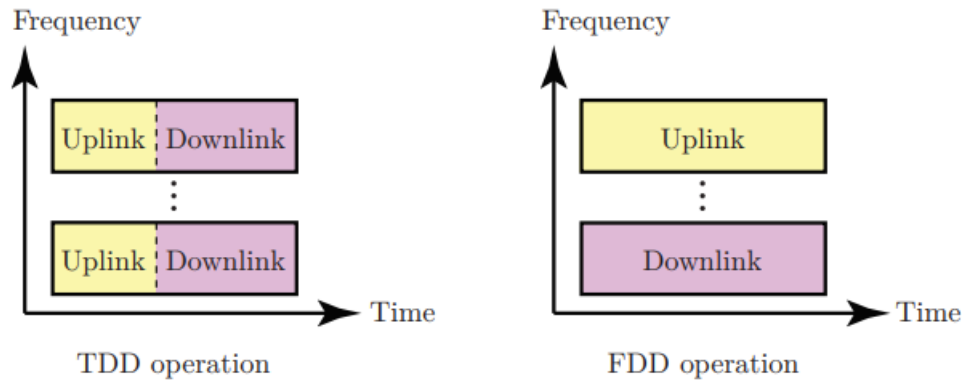


FIGURE 3.2: The Frequency Division duplexing (FDD) and Time Division duplexing (TDD) in a Coherence Block.

For the downlink, the link is from Base Station to user users. Moreover, for the Uplink, the users are transmitting towards the  $M$  antennas of the BS. Then, the channel matrix in the uplink denoted as  $H_u$ , has the size of  $M \times K$  whereas the transmitted signal is a  $K$  dimensional vector and, the noise and received signal at the BS, have the same length as the number of BS antenna  $M$ .

Orthogonal frequency-division multiplexing is one way of multiplexing scheme used in the communication system, which used FFT to decompose a single wideband carrier, which is a frequency-selective channel into many parallel channels that are a flat-fading channel to transport information. Each OFDM symbol contains a length of  $T_{ds}$  seconds duration for data transmission part and a length of  $T_{cp}$  seconds duration cyclic prefix. Moreover, it has a frequency separation between neighboring subcarriers given as  $B_s$  [5]. In addition, in a coherence block  $N_{slot}$  the number of OFDM symbols grouped in a slot in the time domain and  $N_c$  number of consecutive subcarriers grouped in the frequency domain, the duration of one slot is the product of  $N_{slot}$  and a single OFDM symbol duration in seconds. Moreover,

the coherence Bandwidth is the product of  $N_c$  and frequency separation between neighboring subcarriers ( $B_s$ ) as shown in Figure 3.3.

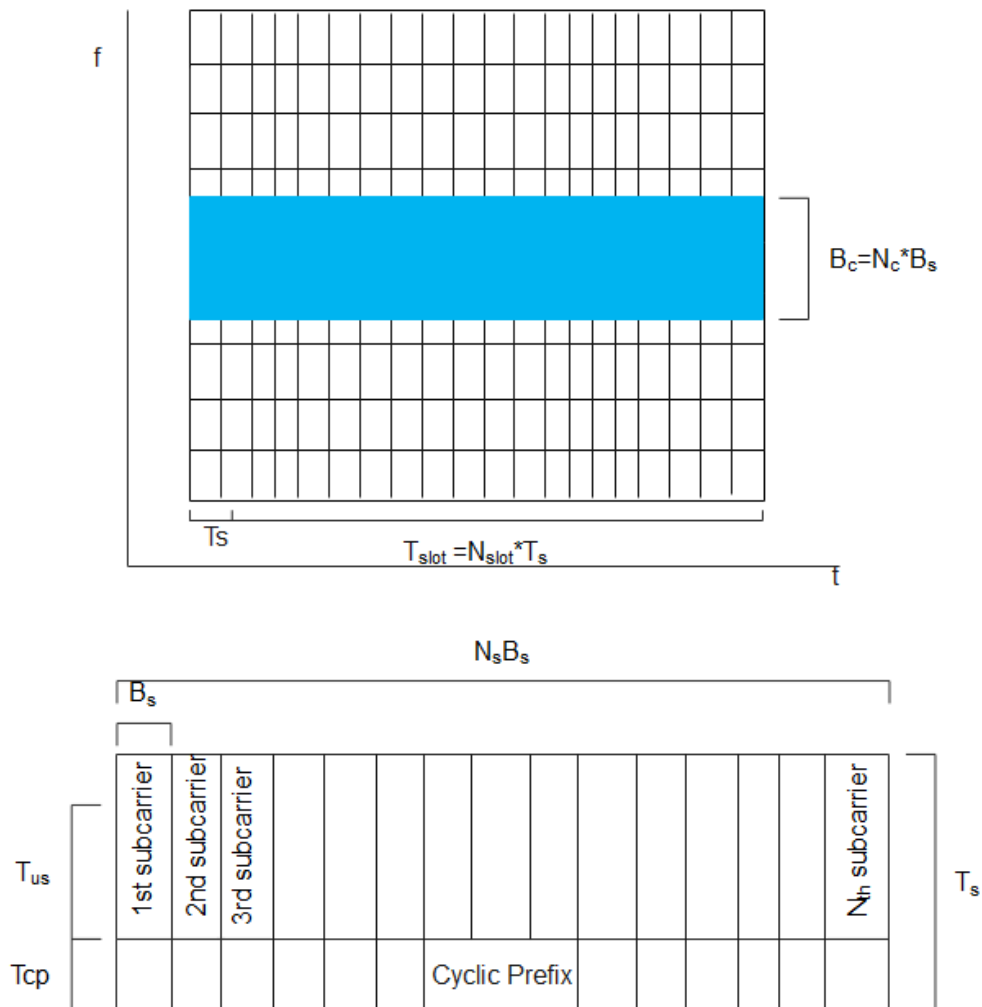


FIGURE 3.3: Block diagram of OFDM multiplexing with Coherence interval, wherein  $N_c$  is number of sub carrier in coherence block and  $N_s$  is the number of sub carriers.

Moreover, in this thesis, we assume that the duration of one slot  $T_{slot}$  is less than or equal to the coherence time given in Equation (3.2), to sustain the channel is flat fading. In addition, the Macro cell 3GPP channel model with 2GHz frequency used, having a subcarrier separation of 15KHz in this scenario it can support mobility of 135 km/h and  $2.5 \mu s$  delay spread [2], [17]. Therefore, from Equations (3.2) and (3.3) we have 200 kHz coherence bandwidth, 1 ms coherence time, and 14 consecutive subcarriers in a coherence bandwidth. Moreover, the coherence block that sustains flat fading has 200 samples.

### 3.1.3 Channel Model

The channel model of all users in the  $j_{th}$  cell's channel propagation matrix to the  $l_{th}$  cell BS [5] is:

$$H^l_j = G^l_j \sqrt{D^l_j} \quad (3.4)$$

Where in matrix's form

$$H^l_j = \begin{pmatrix} h_{l1j1} & h_{l1j2} & \cdots & h_{l1jk} \\ \vdots & \ddots & & \vdots \\ h_{lmj1} & h_{lmj2} & \cdots & h_{lmjk} \end{pmatrix}_{M \times K} \quad G^l_j = \begin{pmatrix} g_{l1j1} & g_{l1j2} & \cdots & g_{l1jk} \\ \vdots & \ddots & & \vdots \\ g_{lmj1} & g_{lmj2} & \cdots & g_{lmjk} \end{pmatrix}_{M \times K}$$

Moreover, the channel vector  $h^l_{jK} \in \mathbb{C}^{M \times 1}$  connecting the  $k_{th}$  user of the  $j_{th}$  cell to  $l_{th}$  cell BS antennas is written as:

$$h^l_{jK} = g^l_{jK} \sqrt{\beta^l_{jk}} \quad (3.5)$$

Where  $G^l_j = [g^l_{j1} g^l_{j2} \dots g^l_{jK}]$ ,  $g^l_{jK} \in \mathbb{C}^{M \times 1}$  is a small-scale fading, each vector is iid to each other given as  $g^l_{jK} \sim CN(0, I_M)$ .  $D^l_j = \text{diag}(\beta^l_{j1} \beta^l_{j2} \dots \beta^l_{jk})$  is a  $K \times K$  diagonal matrix that describes the  $j_{th}$  cell large-scale fading coefficient of each user to the  $l_{th}$  cell BS, and we assume that the large-scale fading coefficients are independent of the antenna index  $M$  of a given BS. Furthermore, these coefficients only change when a user's geographic location changes significantly. Moreover, the following is its mathematical equation:

$$\beta^l_{jk} = \frac{z^l_{jk}}{(r^l_{jk}/R)^\alpha} \quad (3.6)$$

Where,  $r^l_{jk}$  is the distance between the  $k_{th}$  user in cell  $j$  to BS  $l$ ,  $R$  cell radius,  $z^l_{jk}$  is shadow fading coefficient which obeys a lognormal distribution, that is,  $10 \log_{10}(z^l_{jk})$  follows zero-mean Gaussian distribution having standard deviation  $\sigma_{sha}$  and  $\alpha$  is path loss exponent.

Generally,  $h^l_{jk}$  a channel is a vector, which characterized by its unit vector direction  $\frac{h^l_{jk}}{\|h^l_{jk}\|}$  and magnitude  $\|h^l_{jk}\|^2$ . In a fading channel, the norm and the direction of a



channel are modeled as random variables. Then depending on the relation between them in paper [18], it is defined as follow:

Definition 1. A fading channel  $h_{jk}^l \in C^M$  is spatially uncorrelated if the channel gain  $\|h_{jk}^l\|^2$  and the channel direction  $\frac{h_{jk}^l}{\|h_{jk}^l\|}$  are independent random variables, and the channel direction is uniformly distributed over the unit-sphere. Otherwise, the channel is spatially correlated.

### 3.1.3.1 Spatially Correlated Rayleigh fading Channel

Rayleigh fading is one of the channel models, which used for modeling channel fading. Depending on definition 1, practical channels are spatially correlated. Although, the UEs are generally physically separated by multiple wavelengths so that their channels are well modeled as statistically uncorrelated [2],[18]. To model spatially correlated channels with none-line of sight path is using correlated Rayleigh fading:

$$h_{jk}^l \sim NC(0, C_{jk}^l) \quad (3.7)$$

Where  $C_{jk}^l \in C^{M \times M}$  is the covariance matrix of the channel which is positive semi-definite matrix. Macroscopic propagation characteristics like radiation patterns and the antenna gains represented with this matrix. In this thesis, we assumed  $C_{jk}^l \in C^{M \times M}$  is known at the base station since the time-scale changes is much larger than the coherence time interval of the small-scale fading [19]. In addition, To model small-scale fading we used the Gaussian distribution. The normalized trace  $\beta_{jk}^l = \frac{1}{M} \text{tr}(C_{jk}^l)$  is the average channel gain from an antenna at BS l to UE k in cell j, and  $\beta_{jk}^l$  is the large-scale fading coefficient.

The spatial correlation matrix ( $C_{jk}^l \in C^{M \times M}$ ) for non-LOS channel user k in cell j to lth BS equipped with ULA antenna is modeled using the Local Scattering model in this thesis. The azimuth angles to the UEs are used to parameterize the subspaces of correlation matrixes. The signal received at the BS is a multipath NLOS component superposition of  $P_{path}$ . Because the BS is located at an elevated geographical location and hence has no scatters in its near field, this model assumes

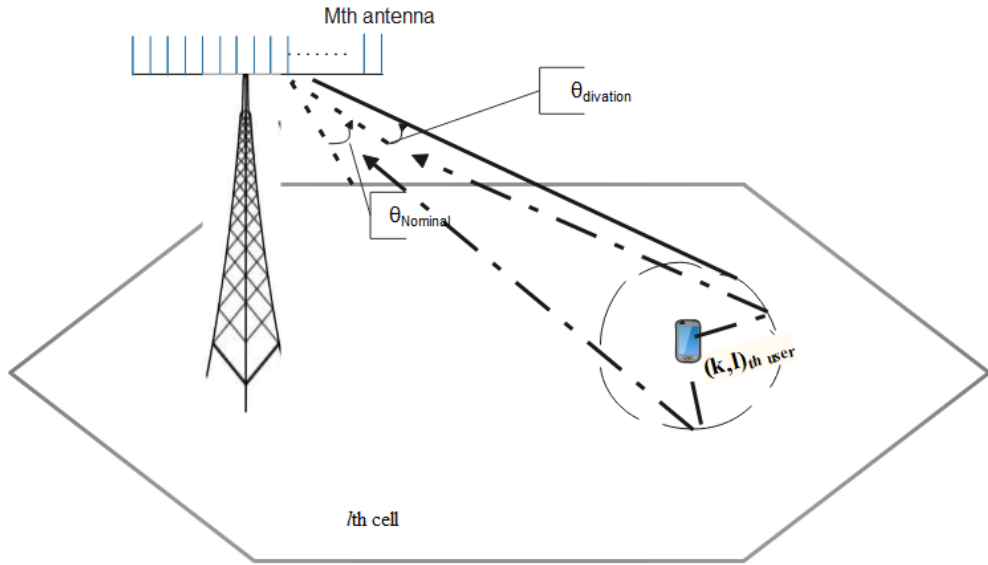


FIGURE 3.4: Local Scattering model of  $k$  user in the  $j$ th cell, scattering is centered around the UE with a nominal angle of  $\theta_{Nominal}$  degree with reference BS antenna and of  $\theta_{deviation}$  angular standard deviation (ASD) of the multipath components.

that the scattering is limited around the UE. As a result, each of the multipath components produces a plane wave that enters the array at a specific angle  $\theta$ , yielding array responses of  $a_p \in C^M$  as [2],[19],[20]

$$a_p = g_p [1 \quad e^{2\pi j d_H \sin(\theta)} \quad e^{4\pi j d_H \sin(\theta)} \quad \dots \quad e^{2\pi j d_H (M-1) \sin(\theta)}]^T \quad (3.8)$$

Where  $a_p \in C^M$  accounts for the gain and phase rotation for the  $p$ th path,  $d_H$  is the antenna spacing in the array and  $\theta$  is *i.i.d.* random variables with angular probability density function of  $f(\theta)$  which denote an angle of arbitrary multipath component and  $g_p$  are *i.i.d.* random variables with zero mean and the variance  $E\{|g_p|^2\}$ , represents the average gain of the  $p$ th path component. Then the channel vector between the user and the BS is  $h^l = \sum_{p=1}^{P_{path}} a_p$ . Then the total average gain of the multipath components is

$$\beta = \sum_{p=1}^{P_{path}} E\{|g_p|^2\} \quad (3.9)$$

Then from the central limit theorem when the number of multipath components

$P_{path}$  approaches infinity the channel model becomes  $h^l \rightarrow NC(0, C)$ . Where the correlation matrix is  $C = E\{h_{jk}^l h_{jk}^{lH}\}$ . If we consider a particular, set up the  $(l, m)_{th}$  element of C, then:

$$\begin{aligned} [C]_{l,m} &= E\{h_{jk}^l h_{jk}^{lH}\} = E\left\{\sum_{p=1}^{P_{path}} a_{pl} a_{pm}^H\right\} \\ &= \sum_{p=1}^{P_{path}} E\{|g_p|^2\} E\{e^{2\pi j d_H(l-1)\sin(\theta)} e^{-2\pi j d_H(m-1)\sin(\theta)}\} \\ &= \beta \int e^{2\pi j d_H(l-m)\sin(\theta)} f(\theta) d\theta \end{aligned} \quad (3.10)$$

Equation (3.10),  $[C]_{l,m}$  depends on the difference  $l - m$ , so it can be computed numerically for any angular distribution  $\theta$ . we assumed that all the multipath componets are from the user side, Then  $\theta = \theta_{Nominal} + \theta_{divation}$ . Where in  $\theta_{Nominal}$  is a deterministic nominal angle,  $\theta_{divation}$  is a random deviation distributed as  $\theta_{divation} \sim N(0, ASD^2)$  and ASD is the angular standard deviation measured in radians determines how large the deviations from the nominal angle are [21].

### 3.1.4 Pilot Sequence for Channel Estimation

As stated in subsection 3.1.2, in each independent frequency-flat channel coherence interval of  $\tau_C$  samples,  $\tau_P$  of the available samples are utilized to transmit the pilot sequence known at both ends of the link to estimate the channel at the BS. Therefore, in the design of the orthogonal pilots,  $\varphi$  all columns are mutually orthogonal ( $\varphi \varphi^H = \tau_P I_P$ ) and all elements in the pilot have the unit magnitude to keep a constant power level per sample [2].

In this thesis to generate a mutually orthogonal pilot sequence, we used the Zadoff-Chu sequence that satisfies the above condition of elements has unit norm and it has perfect periodic autocorrelation property. It is defined as [22].

$$Z_u = \left\{ Z_u(n) = e^{(-j\pi\mu\frac{n^2}{P})} \right\}$$

Where,  $n=0,1,\dots,P-1$ ,  $P$  is the length of the sequence and  $\mu \in \{1, 2, \dots, P-1\}$  is the root index of the pilot sequence.

Then, we assume the pilot sequence matrix of all users in the  $j$ th cell is:

$\varphi_j = [\varphi_{j1}, \varphi_{j2}, \dots, \varphi_{jk}]^T$   $j = 1, 2, \dots, L$  which has unit magnitude elements of  $K \times \tau$  dimension, implies  $\varphi_j \varphi_j^H = \tau I_K$ , where  $I_K$  is the identity matrix of order  $K \times K$ , then  $K$  pilot signals are required for a single cell for orthogonal pilot assignment. Where  $\tau$  is the pilot length. Each pilot sequence is of the form  $\varphi_{jk} = [\varphi_{jk1} \ \varphi_{jk2} \ \dots \ \varphi_{jk\tau}]$ . Furthermore, from the system model we have then, the uplink received pilot signal  $R_p^l \in C^{M^l \times \tau}$  at  $l$ th BS:

$$R_p^l = \sum_{j=1}^L \sqrt{\rho_j} H_j^l \varphi_j + N_p^l \quad (3.11)$$

Where  $\sqrt{\rho_j}$  is the transmitted pilot power of each user;  $H_j^l$  is an  $M \times K$  channel matrix between the  $K$  users and the  $l$ th base station antennas; and  $N_p^l \in C^{M^l \times \tau}$  is the additive noise at the  $l$ th base station.

### 3.1.5 Channel Estimation

There is no wired media in wireless communication between transmitter and receiver the channel must be estimated before any data transfer takes place, as mentioned in the previous section. The channel estimation technique then proceeds like this: To estimate the channel, each user in the UL broadcasts training sequences that are known to both ends. The estimated channel is then used at the BS in the uplink for data processing in receive combining at reception and in a downlink precoding. In a TDD system, channel estimation uses the same frequency resource with distinct time slots for uplink and downlink transmissions. It also applies Reciprocity characteristics, although owing to hardware chain mismatches, perfect channel reciprocity is unattainable in practice. This non-reciprocity, however, can be eliminated with appropriate calibration [23].

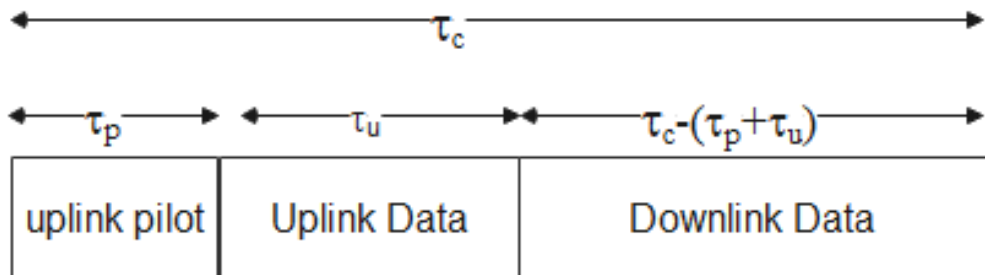


FIGURE 3.5: TDD Based Transmission Protocol.

In this thesis, TDD protocol is used due to its reciprocity property. The uplink and downlink transmission in this protocol is: Uplink Transmission: Each user transmits the pilot sequence, and the base station requires channel status information to detect the signals sent by the  $K$  users ( $k$  orthogonal pilot sequence). Based on the received pilot signals, the BS estimates the CSI. As a result, the channel estimation process in the TDD system utilizes a minimum of  $K$  pilot sequences, which is independent of the number of BS antennas. Downlink Transmission: The BS requires CSI to process signals from the BS to each user during the precoding process, and each user requires good channel estimates to detect the desired signals from their BS. The channel estimated at the BS in the uplink can be employed using the properties of channel reciprocity discussed above. In the downlink FDD system, however, at least  $M$  pilot sequences are required [24]. Furthermore, unlike the TDD system, FDD-based Channel Estimation uses distinct frequency resources for downlink and uplink transmissions, and the uplink and downlink channels are not reciprocal. Then CSI must be estimated at BS and on the users' side [24]. The two forms of pilot-based TDD channel estimation approaches, least square channel estimation and MMSE channel estimation techniques, are discussed in further detail in the preceding section. Furthermore, because  $M$  in Massive MIMO is large, TDD operation is preferred than FDD in Massive MIMO.

### 3.1.5.1 Least Square channel estimation

The goal of Least Square channel estimation is to reduce the squared error between the actual received signal,  $R_p^l$  and the estimated received signal. Which stated

[25],[26] as follows:

$$\hat{H}_j^l = \arg \min \|R_p^l - H_j^l \varphi_j^H\|^2 \quad (3.12)$$

Then the LS channel estimate of the channel between lth cell users to lth BS from Equation (3.12) become:

$$\hat{H}_j^l = \frac{1}{\sqrt{\rho_j}} R_p^l \varphi_j^H = H_j^l + \sum_{i \neq j} H_i^l + \frac{1}{\sqrt{\rho_j}} N_p^l \varphi_j^H \quad (3.13)$$

The LS estimate approach relies just on the received matrices  $R_p^l$  and does not require any prior information of the channel. As a result, the approach is robust to lack of CSI knowledge [27]. The channel estimate of the kth user in the jth cell is a linear combination of the channels of the users in the other cells with the same pilot sequence, as shown in Equation (3.13).

### 3.1.5.2 MMSE Channel Estimation Under correlated Rayleigh fading channel

The goal of MMSE estimation is to get the estimated value to be as close to the true value as possible. Because a channel is a realization of a random variable, Bayesian estimators are preferred. When estimating the CSI, the channel statistics are taken into account. The prior distribution of the channel vector  $h_{jK}^l \in \mathbb{C}^{M \times 1}$  connecting the kth user of the jth cell to the lth cell BS antennas is  $h_{jk}^l \sim NC(0, C_{jk}^l)$  for correlated Rayleigh fading channels. Then, using Equation (3.11) [27],[28], the MMSE estimate of the uplink channel  $h_{jK}^l$ :

$$\hat{h}_{jk}^l = E\{h_{jk}^l / R_p^l \varphi_{jk}^H\} = \sqrt{p_{jk}} \frac{C_{jk}^l R_p^l \varphi_{jk}^H}{\Psi_{jk}^l} \quad (3.14)$$

Where  $C_{jk}^l$  is the covariance matrices of the channel vector  $h_{jk}^l$ , and  $\Psi_{jk}^l$  is the covariance matrix of the received pilot signal  $R_p^l$ , and is given by

$$\Psi_{jk}^l = \sum_{i=1}^L \sum_{k=1}^K \rho_{ik} C_{ik}^l \varphi_{ik} \varphi_{jk}^H + \sigma^2 I_{Ml}$$

In addition, the estimation error of MMSE estimate of the kth user in the lth cell

BS  $\tilde{h}_{jk}^l$  become:

$$\tilde{h}_{jk}^l = h_{jk}^l - \hat{h}_{jk}^l \quad (3.15)$$

The correlation matrix of  $\tilde{h}_{jk}^l$  given by:  $CC_{jk}^l = E \left\{ \left( \tilde{h}_{jk}^l \right) \left( \tilde{h}_{jk}^l \right)^H \right\}$  Then, the MSE of the MMSE estimator is expressed as  $E \left\{ \left\| \tilde{h}_{jk}^l \right\|^2 \right\} = \text{tr}(CC_{jk}^l)$  indicating the estimated channel's estimation quality. The channels for each of the antennas for the specific user are statistically identical. The channel estimation error is uncorrelated with both the channel estimate and the pilot signal, the channel estimate  $\hat{h}_{jk}^l$  and estimation error  $\tilde{h}_{jk}^l$  are jointly Gaussian distributed. For BS 1 can estimate the channels vector of each user  $k$   $\hat{h}_{jk}^l$  directly from the uplink pilots so there is no need for BS cooperation in signaling between the BSs [2]. The uplink control channel can also be used to collect long-term channel statistics such as channel attenuation and the pilot allocation of all users to each BS. As a result, the M-MMSE technique [28] is fully scalable.

### 3.1.6 Characterization of Multi-antenna Channel

#### 3.1.6.1 Favorable Propagation

One of the features of multi-antenna channels is favorable propagation. When the number of BS antennas is large, the directions of the channel between the two UEs become asymptotically orthogonal using the law of large numbers given in the appendix part. As a result, this attribute makes it easier for the BS to conduct linear processing, which is addressed in the preceding section, and makes it easier for the BS to minimize interference between distinct user channels, improving the SE overall. As a result, large MIMO takes advantage of the fact that the channel vectors are nearly orthogonal. If the channels  $h_{jk}^l$  and  $h_{ik}^l$  are used together, they provide asymptotically favourable propagation if.

$$\frac{(h_{jk}^l)^H h_{ik}^l}{\sqrt{E\{\|h_{ik}^l\|^2\}E\{\|h_{jk}^l\|^2\}}} \rightarrow 0 \text{ as } M_j \rightarrow \infty \quad (3.16)$$

When a high number of BS antennas are employed, the inner product of the direction of the two-channel vector approaches zero, which helps to decrease interference between different users channels at the BS. Furthermore, a sufficient condition for asymptotic favourable propagation for correlated Rayleigh fading channel is that, the spectral norm of each respective spatial correlation matrices is bounded and the average channel gains of each channel stay strictly positive as the number of BS antennas increase.

### 3.1.6.2 Channel hardening

Wireless channel gain by nature fluctuates randomly due to small-scale fading, caused by microscopic changes in the propagation environments. As a result, one of the qualities of a multi-antenna channel is channel hardening, which makes this fading channel behave like a deterministic channel by reducing small scale fading. According to [2],[29] a propagation channel  $\mathbf{h}_{jk}^l$  provides asymptotic channel hardening if and only if

$$\frac{\|\mathbf{h}_{jk}^l\|^2}{\mathbb{E}\{\|\mathbf{h}_{jk}^l\|^2\}} \rightarrow 1 \text{ as } M^l \rightarrow \infty \quad (3.17)$$

The gain of  $\mathbf{h}_{jk}^l$  close to its predicted value when there are several antennas, according to this interpretation.

If this is the case, the channel will become deterministic, and the impact of fading will be reduced, resulting in increased communication performance. Furthermore, as the number of BS antennas increases,  $\beta_{jk}^l = \frac{1}{M} \text{tr}(C_{jk}^l)$  remains strictly positive, and the spectral norm of the spatial correlation matrix ( $\|C_{jk}^l\|_2$ ) is finite, this is a necessary requirement for asymptotic channel hardening for correlated Rayleigh fading channels [2].



### 3.1.7 Linear Detector processing in Uplink

The Detector is an  $M \times K$  dimension matrix, with the function of separating the intended user's data symbol from the received signal  $R_u^l$  [30]. The received signal  $R_u^l$  at the  $l$ th cell BS is the superposition of all users' data, according to Equation (3.1) in the uplink data transmission. The detector vector for each  $k$ th user is determined by their channel estimation vector. Then, from Figure 3.6, the general

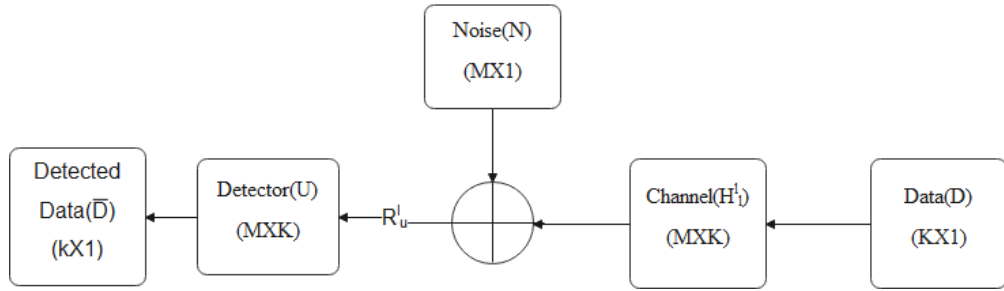


FIGURE 3.6: Block diagram of Uplink Data Detection in Massive MIMO.

form of the detected signal  $\hat{D}_j$  after the detector is expressed as:

$$\hat{D}_j = U_j^H R_u^l = \sum_{i=1}^L U_j^H \sqrt{\rho_u} H_i^l D_j + U_j^H N_l \quad (3.18)$$

To properly detect the data, the product term  $(U_j^H H_i^l)$  in Equation (3.18) must be simplified (one), and to detect the  $k$ th user data stream  $d_{jk}$ ,  $u_j^H$  is employed which is the detector's  $k$ th column, yielding:

$$\begin{aligned} \hat{d}_{jk} &= \sqrt{\rho_u} \sum_{i=1}^L \sum_{k=1}^K u_{jk}^H h_{ik}^l d_{ik} + u_{jk}^H n^l \\ &= \sqrt{\rho_u} u_{jk}^H h_{jk}^l d_{jk} + \sqrt{\rho_u} \sum_{u \neq k}^K u_{jk}^H h_{ju}^l d_{ju} + \sqrt{\rho_u} \sum_{i=1}^L \sum_{\substack{k=1 \\ i \neq j}}^K u_{jk}^H h_{ik}^l d_{ik} + u_{jk}^H n^l \end{aligned} \quad (3.19)$$

Then, from Equation (3.19) the SINR of the  $k$ th user for any detector  $u_{jk} \in \mathbb{C}^{M \times 1}$ , is given by [31]:

$$SINR_{jk}^{ULL} = \frac{\rho_u |\mathbb{E} \{u_{jk}^H h_{jk}^l d_{jk}\}|^2}{\rho_u \sum_{u \neq k}^K \mathbb{E} \{|u_{ju}^H h_{ju}^l d_{ju}\}|^2} + \rho_u \sum_{i=1}^L \sum_{k=1}^K \mathbb{E} \{|u_{jk}^H h_{ik}^l d_{ik}\}|^2} + \|u_{jk}\|^2 \quad i \neq j \quad (3.20)$$

As a result, the three types of linear detectors are given more emphasis in the preceding sub-section depending on how the detector matrix is chosen.

### 3.1.7.1 Maximum Ratio Combining (MRC) Detector

The simplest type of linear detector at the receiver is maximum ratio processing. The basic idea is that it detects the signal of interest and amplifies it while disregarding multiuser interference. However, it performs badly in interference-limited circumstances [32]. In this processing, each signal branch is multiplied by a weight factor proportional to the signal intensity. The general form of MRC detector is [33]:

$$U_j^{MRC} = \hat{H}_j^l \quad (3.21)$$

Then the achievable rate of the  $k$ th user in the  $l$ th cell using MRC detector expressed in [32] as:

$$R_{jk}^{MRC} = \log_2 (1 + SINR_{jk}^{MRC}) \quad (3.22)$$

### 3.1.7.2 Zero Forcing (ZF) processing

Another type of linear processing at the receiver is zero forcing, which is employed in the detection of signals in uplink massive MIMO. This technique is named from the fact that it examines multi-user interference while ignoring the effect of noise [34]. By projecting each stream onto the orthogonal complement of the inter-user interference, the multiuser interference is totally nulled out in this technique. When interference is significant compared to noise, ZF comes in useful. As a

result, it performs badly in noisy environments. The matrix of the ZF detector for the  $l$ th cell is [32]:

$$U_j^{ZF} = \hat{H}_j^l \left( \hat{H}_j^{lH} \hat{H}_j^l \right)^{-1} \quad (3.23)$$

Then, using the ZF detector, the  $k$ th user's achievable rate in the  $l$ th cell is expressed as:

$$R_{jk}^{ZF} = \log_2 \left( 1 + SINR_{jk}^{ZF} \right) \quad (3.24)$$

### 3.1.7.3 MMSE Detector

The third type of linear processing approach is the MMSE receiver. This Algorithm maximizes the received SINR at the receiver by minimizing the mean-square error between the estimate  $\hat{D}_j$  and the transmitted signal  $D$ . The minimum mean-squared error-combining vector [31] for the multi-cell scenario is:

$$U_{jk}^{MMSE} = \rho_{jk} \left( \sum_{i=1}^L \sum_{k=1}^K \rho_{ik} (\hat{h}_{ik}^l (\hat{h}_{ik}^l)^H + C_{ik}^l) + \sigma_{ul}^2 I_{M_j} \right)^{-1} \hat{h}_{jk}^l \quad (3.25)$$

Then, using the MMSE detector, the  $k$ th user's achievable rate in the  $l$ th cell is expressed as:

$$R_{jk}^{ZF} = \log_2 \left( 1 + SINR_{jk}^{MMSE} \right) \quad (3.26)$$

This method maximizes the instantaneous SINR while minimizing the MSE in data detection, which is the average squared distance between the desired signal and the processed received signal. When the distance between the BS and the UE is substantially more than the distance between the BS antennas, the average channel gain is considered to be the same for all BS antennas, which is an acceptable estimate.

### 3.1.8 Pseudo Random Sequence

The pseudo-random sequence is similar to the noise sequence in some extent. But pseudo-random sequence is a seemingly random with regular periodic binary

sequence. In terms of the behavior over time  $t$ , a deterministic signal with period  $T$  is predictable. Because this form of signal has some qualities in common with random signals, it is commonly referred to as a pseudo random signal [35]. When comparing pseudo random signals to true random signals, the autocorrelation qualities of individual signals and the cross-correlation properties of groups of signals are the most significant characteristics to consider. When the degree of similarity of a deterministic signal's qualities in terms of specific characteristics approaches that of a random signal, it is referred to as a pseudo random signal. Because the precise quantification in degree of resemblance depends on the specific application [9].

### 3.1.8.1 Generation of Pseudo Random Sequence

Pseudo-random codes are utilized in CDMA transmission because they have good auto-correlation and cross-correlation features [35]. The pseudo-random sequence is generated using linear feedback shift registers [36].

### 3.1.8.2 Linear Feedback Shift Register

A linear feedback shift register (LFSR) is a register that swaps its bits in a sequential manner, with the input bit based on its prior state. The seed value of the LFSR is defined, and because the action of this register is deterministic [37], the register's values are determined by its present state. A shift register is made using flip-flops for storing binary data. The coefficients  $S_0, S_1, S_2, \dots, S_n$ , which are an element of 0 and 1 of the characteristic polynomial ( $Q$ ), determine the LFSR feedback pattern, which is depicted in the picture by the taps or multipliers in the feedback channels. Linear XOR logic is used to add up the elements that were supplied back into the register.

The characteristic polynomial of generating a pseudo-random sequence is:

$$Q(r) = S_0 + S_1r + S_2r^2 + S_3r^3 + \dots + S_{n-1}r^{n-1} + S_nr^n \quad (3.27)$$

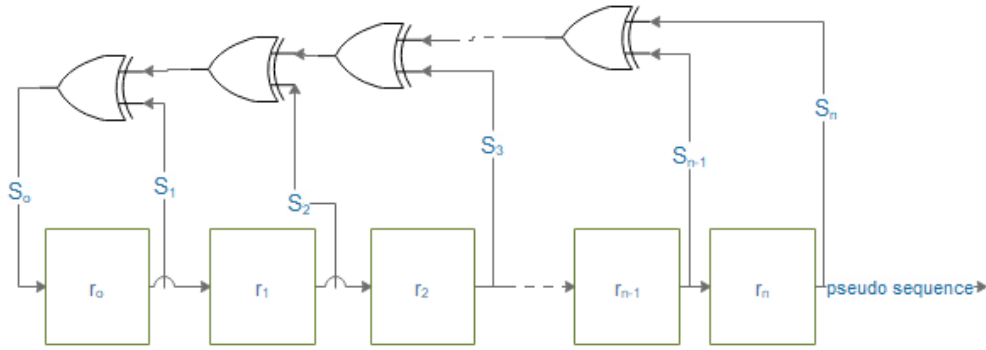


FIGURE 3.7: Linear Feed Back Shift Register.

Scrambling is a binary bit-level processing approach that is used to randomize the binary sequence to be sent [38]. Maximal-length sequences are L-stage linear shift register generators whose maximum possible period is the maximum possible period (m-sequences). The bipolar sequence is used to define the periodic auto-correlation function. The normalized cross-correlation value between two pseudo-random sequences  $a_j$  and  $b_j$  assigned for two separate base stations during the maximum feasible time of  $N = 2^L - 1$  is expressed as [9],[35]:

$$\theta_b(a_l, b_j) = \frac{1}{N} \sum_{n=1}^{N-1} a_l b_j \quad (3.28)$$

The following are the three criteria for determining whether a signal qualifies for pseudo randomness or not [9].

- a) Run property: it refers to a string of successive zeros or ones, so 1-runs and 0-runs alternate with the same number of 0-runs and 1-runs.
- b) The property of balance. In each period of a pseudo random sequence, the number of zeros differs by at most one from the number of ones.
- b) The criterion for correlation. The out-of-phase of a tiny constant was required for the periodic autocorrelation function.

In the information signal, one of the most significant functions of scrambling is to break up a regular pattern of long runs of zeros and ones. Impurities like pattern-dependent jitter and unstable clock recovery can be reduced in this way, lowering ISI [38].

### 3.1.8.3 Correlation and Orthogonality

In communication engineering, signal analysis, and mathematical science, the concept of orthogonality is widely utilized to develop an orthogonal expansion of signals or to provide a geometric representation of signals and noise. Two complex-valued signals  $X_n(t)$  and  $X_c(t)$  specified on their period are said to be orthogonal in signal analysis if the following condition holds [9],[37]:

$$\langle X_n(t), X_c(t) \rangle = \int_b^{b+T} X_n(t) X_c^*(t) dt = \begin{cases} k & \text{for } n = c \\ 0 & \text{for } n \neq c \end{cases} \quad (3.29)$$

Where  $b$  is any constant, and  $T$  is the function's period. And  $k = 1$  holds when the two signals  $X_n(t)$  and  $X_c(t)$  are orthonormal. As a result, the orthogonality condition between two signals  $x(t)$  and  $y(t)$  specified in Equation (3.29) is associated to their in-phase correlation value as follows:

$$\theta_{x,y}(0) = \begin{cases} k & \text{for } x(t) = y(t) \\ 0 & \text{for } x(t) \neq y(t) \end{cases} \quad (3.30)$$

The correlation functions for pseudo-random signal processing, on the other hand, are specified for any value of the discrete shift variable  $n$ . As an example, consider the two-periodic sequences  $\{X_i\}$  and  $\{Y_i\}$  of period  $N$ . If the in-phase periodic cross correlation value specified in Equation (??) is zero, the two sequences are considered to be orthogonal, [38]. Furthermore, if the following condition applies, the two periodic sequences are considered to be uncorrelated:

$$\theta_{x,y}(n) = \frac{1}{N} \sum_{i=1}^{N-1} x_i y_{i+n} = 0 \quad (3.31)$$

Where  $n$  is discrete shift variable.

## 3.2 Pilot Contamination in Massive MIMO System

In this thesis, a multiuser massive MIMO time division duplexing system with  $L$  number of cells and a single BS in its center,  $M$  number of co-located antenna, and  $K$  randomly located single-antenna user terminals is considered. We also assume that the channel is Rayleigh fading and that it is independent across the antennas. During the channel coherence time, the wireless channels are considered static. We assumed that all users in the same cell utilize orthogonal pilot, hence the number of orthogonal pilot sequences is  $\tau_p$  ( $K \leq \tau_p \leq KL$ ).

### 3.2.1 Uplink Pilot Transmission and Channel Estimation

The system runs in TDD duplexing mode, users in all cells send the pilot signal simultaneously, and the BS then estimates the uplink channel of each user based on the received pilot sequence. Upon the estimation of the uplink channel, the downlink channel is estimated using the channel reciprocity property. The uplink payload data is then transmitted from the user's side, and the BS detects the data signal using one of the detection algorithms as follows.

From subsection 3.1.4 we have a pilot matrix of  $K \times \tau$  dimension. The deterministic pilot assignment has been made when the UE connects to the BS. And from Equation (3.11) the received uplink pilot signals  $R_p^l \in C^{Ml \times \tau}$  at  $l$ th BS is:

$$R_p^l = \sum_{j=1}^L \sqrt{\rho_j} H_j^l \varphi_j + N_p^l \quad (3.32)$$

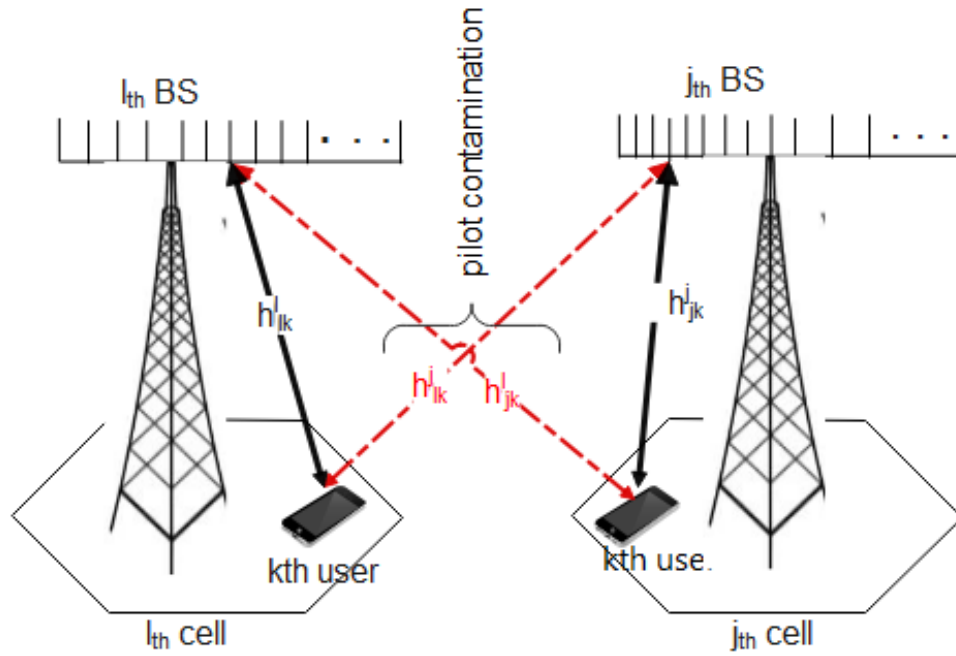


FIGURE 3.8: The problem of pilot contamination between two cells in massive MIMO.

Then from Equation (3.32) the pilot signal of user  $k$  in cell  $l$  at the  $l$ th BS, we can rewrite  $r^l_{lkp}$  as:

$$r^l_{jkp} = \underbrace{\sqrt{\rho_{jk}} h^l_{jk} \varphi_{jk}}_{\text{Desired user pilot}} + \underbrace{\sum_{\substack{k=1 \\ k \neq \kappa}}^K \sqrt{\rho_{jk}} h^l_{jk} \varphi_{jk}}_{\text{Intra user pilots}} + \underbrace{\sum_{i=1}^L \sum_{\substack{k=1 \\ i \neq j}}^K \sqrt{\rho_{ik}} h^l_{ik} \varphi_{ik}}_{\text{Inter user pilots}} + n^l \quad (3.33)$$

Where

$$r^l_{jkp} = \begin{pmatrix} r^l_{j1} \\ r^l_{j2} \\ \vdots \\ r^l_{jM} \end{pmatrix} \quad h^l_{jk} = \begin{pmatrix} hl1jk \\ hl2jk \\ \vdots \\ hLMjk \end{pmatrix} \quad \text{And } n^l_p = \begin{pmatrix} nl1 \\ nl2 \\ \vdots \\ nlM \end{pmatrix}$$

Where  $\sqrt{\rho_{jk}}$  is the uplink transmission power of user  $k$  in  $l$ th cell BS and  $n^l$  is the AWGN matrix Of order  $M \times \tau$  with iid elements as  $n^l \sim CN(0, \sigma_{UL}^2)$ .



The users sets reused the same pilot as UE  $k$  in cell  $l$  is defined as:

$$U_{ju} = \{(l, u) : \varphi_{ju} = \varphi_{lk}, l = 1, 2, \dots, L \text{ and } u = 1, \dots, K_l\} \quad (3.34)$$

The base station then estimates each user's channel response based on the observation of the received uplink signal  $r_{jkp}^l$  of Equation (3.33). As a result, in order for the BS to estimate the channel of a specific user  $k$ , the UE's pilot sequence ( $\varphi_{jk}$ ) must be known. It enables in channel estimation by correlating the received signal. In this section, The simplest LS channel estimator has been used to demonstrate the effect of pilot contamination in massive MIMO for non-orthogonal pilot sequence users during the channel estimation phase. Then, using the LS estimator stated in subsection 3.1.5.1 the LS estimate of the uplink channel  $h_{jk}^l$  using  $r_{jkp}^l$  and known pilot sequence  $\varphi_{jk}^H$  becomes:

$$\begin{aligned} \hat{h}_{jk}^l &= E\{h_{jk}^l / r_{jkp}^l \varphi_{jk}^H\} = \frac{1}{\sqrt{p_{jk}}} r_{jkp}^l \varphi_{jk}^H \\ &= h_{jk}^l \varphi_{jk} \varphi_{jk}^H + \sum_{\substack{k=1 \\ k \neq j}}^K h_{jk}^l \varphi_{jk} \varphi_{jk}^H + \sum_{i=1}^L \sum_{\substack{k=1 \\ i \neq j}}^K h_{ik}^l \varphi_{ik} \varphi_{jk}^H + \frac{1}{\sqrt{p_{jk}}} n^l \varphi_{jk}^H \\ &= \underbrace{\tau h_{jk}^l}_{\text{target user channel response}} + \underbrace{\sum_{\substack{(j,k) \subseteq U_{ju} \\ i \neq j}} \sum_k \tau h_{ik}^l}_{\text{other cells same pilot users channel}} + \underbrace{\frac{1}{\sqrt{p_{jk}}} n^l \varphi_{jk}^H}_{\text{Noise term}} \quad (3.35) \end{aligned}$$

From Equation (3.35), It is seen that the channel estimate of user  $k$  in cell  $l$  at the BS  $l$  is the superposition of the channel propagation matrix of other cells same pilot users to the target cell user  $k$  in addition to the target channel response and noise. That is the effect of pilot contamination due to non-orthogonal pilot users.

### 3.2.2 Uplink Data Transmission

If we represent the uplink data symbol of user  $k$  in cell  $j$  as,  $d_{jk}$  and it has a power of  $\rho_{jk} = E\{|d_{jk}|^2\}$ , which is a random data signal  $d_{jk} \sim CN(0, \rho_{jk})$ . Then, after the pilot transmission phase, all users in the cells send the uplink data signal to the BS using the same time-frequency resource. Then, the data signal received by the  $l$ th cell BS  $R_d^l \in C^{M^l}$  is:

$$\begin{aligned}
 R_d^l &= \sum_{i=1}^L \sqrt{\rho_i} H_i^l D_{ik} + N^l \\
 &= \sum_{k=1}^K \sqrt{\rho_{jk}} h_{jk}^l d_{jk}^l + \sum_{\substack{i=1 \\ i \neq j}}^L \sum_{k=1}^K \sqrt{\rho_{ik}} h_{ik}^l d_{ik}^l + n^l
 \end{aligned} \tag{3.36}$$

Which are the superposition of the desired signal  $j$ th cell users, the interfering cell signal and the noise of the corresponding cell. Then BS  $l$  selects the simplest MF receive combining vector ( $U_{jk} \in C^{M^l}$ ) and uses the channel estimation value of Equation (3.35), and the received data signal to separate the desired signal of the  $k$ th user from the interference part as:

$$\begin{aligned}
 \hat{d}_{jk}^l &= U_{jk}^H R_d^l \\
 &= \left( \hat{h}_{jk}^l \right)^H \left( \sum_{k=1}^K \sqrt{\rho_{jk}} h_{jk}^l d_{jk}^l + \sum_{\substack{i=1 \\ i \neq j}}^L \sum_{k=1}^K \sqrt{\rho_{ik}} h_{ik}^l d_{ik}^l + n^l \right)
 \end{aligned}$$

$$\begin{aligned}
&= \left( \begin{array}{c} \tau h_{jk}^l h^H + \sum_{\substack{(j,k) \in U_{lu} \\ i \neq j}}^K \tau h_{ik}^l h^H + \frac{1}{\sqrt{p_{jk}}} \varphi_{jk} n_l^H \end{array} \right) \\
&\times \left( \begin{array}{c} \sum_{k=1}^K \sqrt{\rho_{jk}} h_{jk}^l d_{jk}^l + \sum_{\substack{i=1 \\ i \neq j}}^L \sum_{k=1}^K \sqrt{\rho_{ik}} h_{ik}^l d_{ik}^l + n^l \end{array} \right) \\
&= \sqrt{\rho_{jk}} \tau h_{jk}^l h^H h_{jk}^l d_{jk}^l + \sum_{\substack{n=1 \\ n \neq k}}^K \tau \sqrt{\rho_{ln}} h_{jk}^l h^H h_{ln}^l d_{ln}^l \\
&+ \sum_{\substack{(i,k) \in U_{ju} \\ i \neq j}}^K \tau \sqrt{\rho_{iu}} h_{jk}^l h^H h_{iu}^l d_{ik}^l + h_{jk}^l h^H n^l \tag{3.37}
\end{aligned}$$

Which is the superposition of the desired signal, intracell interfere signal, inter-cell interferes signal, and the corresponding additive noise. Then, the uplink SINR of the  $k$ th user in the  $j$ th cell from Equation (3.37) is expressed as [42],[47]:

$$\begin{aligned}
SINR_{jk}^{UL} &= \frac{|\mathbb{E} \{ \tau \sqrt{\rho_{jk}} h_{jk}^l h^H h_{jk}^l d_{jk}^l \}|^2}{\left( \begin{array}{c} \sum_{\substack{n=1 \\ n \neq k}}^K E \left\{ |\tau \sqrt{\rho_{ln}} h_{jk}^l h^H h_{ln}^l d_{ln}^l|^2 \right\} \\ + \sum_{\substack{(i,k) \in U_{ju} \\ i \neq j}}^K E \left\{ |\tau \sqrt{\rho_{ik}} h_{jk}^l h^H h_{ik}^l d_{ik}^l|^2 \right\} + E \left\{ |h_{jk}^l h^H n^l|^2 \right\} \end{array} \right)} \tag{3.38}
\end{aligned}$$

When the number of BS antennas approaches infinity in an uncorrelated, Rayleigh fading channel, the intra-cell interference, and uncorrelated noise components decrease. The massive MIMO system's channel does progressive orthogonality, and the consequent expression of Equation (3.37) is [42],[47]:

$$\hat{d}_{jk}^l \approx M \tau \sqrt{\rho_{jk}} \beta_{jk}^l d_{jk}^l + M \tau \sum_{i \neq j}^L \sum_{k=1}^K \sqrt{\rho_{ik}} \beta_{ik}^l d_{ik}^l \tag{3.39}$$

Moreover, the uplink SINR of the  $k$ th user in the  $j$ th cell become:

$$\begin{aligned}
 SINR_{jk}^{UL} &\approx \frac{M\rho_{jk}\beta_{jk}^l{}^2}{M \sum_{\substack{(j,k) \in U_{iu} \\ i \neq j}} \rho_{ik}\beta_{ik}^l{}^2} \\
 &\approx \frac{\rho_{jk}\beta_{jk}^l{}^2}{\sum_{\substack{(j,k) \in U_{iu} \\ i \neq j}} \rho_{ik}\beta_{ik}^l{}^2} \tag{3.40}
 \end{aligned}$$

Where,  $U_{iu}$  represents the users in different cells using the same pilot as user  $k$  in cell  $j$ , and  $\sum_{(i,u) \in U_{iu}} \beta_{iu}^l{}^2$  is the pilot contamination.

Hence, the uplink average achievable rate  $R$  [47] as:

$$R_{jk}^{UL} = B \left( 1 - \frac{\tau_p}{\tau_c} \right) E \{ \log_2 (1 + SINR_{jk}^{UL}) \} \tag{3.41}$$

Where  $B$ , the system bandwidth.

$\tau_p$ , is the number of pilots used for channel estimation and

$\tau_c$ , is the total number of samples in a coherence block.

# Chapter 4

## System Model Under Spatially Correlated Rayleigh fading Channel

### 4.1 The Proposed Pilot Decontamination Scheme

Using the K-means Clustering Algorithm, users in the cells are clustered into cell center and cell edge user based on their distance from the BS. The SINR of the desired user is affected by other cells same pilot users large scale fading coefficient. The large scale fading coefficient has indirect effect with the distance of users from BS. Following the grouping of each user, the edge users are allocated an orthogonal pilot sequence, while the center users have reused the same pilot sequence for neighboring center users. Finally, we used a pseudo-random sequence to minimize the pilot contamination of center user pilots by reducing their correlation with adjacent center user pilots. The channel estimation and the uplink data detection is done for each clusters.

### 4.1.1 User Grouping Algorithm

The K-means clustering algorithm is used to group all of the users in the  $l$ th cell into two groups. The number of clusters needed, the input data set, and the specific data in each cluster were all taken into account by the K means clustering algorithm. And, using euclidean distance between them, it clusters each data into distinct clusters based on their correlation to the specific data.

---

#### Algorithm 1 User Clustering Algorithm

---

```

Input :  $R, L$ , and  $K$ 
Initialization :  $Is_{Edge} = \text{zeros}(K, L)$ ,  $check = 0$ ,  $K_c = \text{zeros}(1, L)$ ,  $K_e = \text{zeros}(1, L)$ , and,  $k = 2$ 
3: for  $j \leftarrow 1 \dots L$  do
    for  $l \leftarrow 1 \dots L$  do
        for  $k \leftarrow 1 \dots K$  do
6:         calculate  $dis(k, j, l)$ 
        end for
    end for
9: end for
    for  $j \leftarrow 1 \dots L$  do
         $x(i) \leftarrow dis(:, j, j)$ ,  $G_e = []$ ,  $G_c = []$ ,
12:      $x_c \leftarrow \min(dis(:, j, j))$ ,  $x_e \leftarrow \max(dis(:, l, l))$ 
        while  $check = 0$  do
            for  $i \leftarrow 1 \dots \text{length}(x)$  do
15:                 if  $abs(x(i) - x_c) > abs(x(i) - x_e)$  then
                     $G_e \leftarrow [x_e \ x(i)]$ ,  $Is\_Edge(i, l) \leftarrow 1$ 
                else
18:                  $G_c \leftarrow [x_c \ x(i)]$ 
                end if
            end for
21:              $x_{c1} \leftarrow \text{mean}(G_c)$ ,  $x_{e1} \leftarrow \text{mean}(G_e)$ 
            if  $(x_{c1} \sim x_c \ \&\& \ x_{e1} \sim x_e)$  then
                 $G_e = []$ ,  $G_c = []$ 
24:             else
                 $check \leftarrow 1$ 
            end if
27:         end while
    end for
    for  $i \leftarrow 1 \dots L$  do
30:      $K_e(1, i) = \text{nnz}(Is\_Edge(:, i))$ ,  $K_c(1, i) = K - K_e(1, i)$ 
    end for
outputs :  $Is\_Edge, K_c$  and  $K_e$ 

```

---

As a result, in Algorithm 1, this clustering algorithm uses two numbers of clusters, the distance of each user to the BS as input data, and the minimum and maximum distance users as two particular data in each clusters. The k-means clustering algorithm 1 can then be used to cluster the users in each cell into two clusters as:

$$G^{CU}_j, G^{EU}_j$$

Are center users group and edge user group respectively where ( $j = 1, 2, \dots, L$ ) then the number of center and edge users in the cluster are:

$$K_{Cu} = \max(G^{cu}_j) \quad (4.1)$$

$$K_{Eu} = \sum_{i=1}^L G^{Eu}_i \quad (4.2)$$

Wherein  $K_{Cu}$  and  $K_{Eu}$  is the number of center users and edge users in the system respectively. Therefore the total number of pilot sequences needed for channel estimation in this system is  $N_p$ , which is equivalent to  $\tau_p$ :

$$N_p = K_{Cu} + K_{Eu} \quad (4.3)$$

## 4.2 The Pseudo Random Based Pilot for Center Users

To use the output of the LFSR for channel estimation with the user pilot sequence, it must first be processed to a form that is appropriate for transmission, i.e. first it must pass through a  $l$  length rectangular window. Second, the window's output is modulated with QAM. Then the modulated sequence is then diagonalized. Finally, the scrambler sends the diagonalized pseudo random sequence to the user pilot in the center. In block diagram, the pseudo-random scrambling processes are as follows:

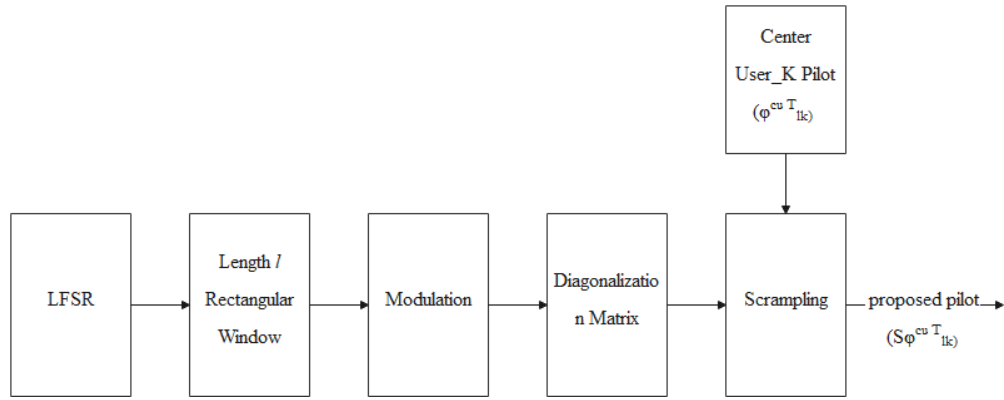


FIGURE 4.1: Block Diagram of Pseudo Random Based Pilot Sequence.

From the block diagram of 4.1 we have:  $LFSR_{out} = [p_1, p_2, p_3, p_4, \dots, p_N]$  At the output of LFSR. Then  $l$  length of the rectangular window used with zero centered is given as,

$$w(n) = \begin{cases} 1, & -\frac{l-1}{2} \leq n \leq \frac{l-1}{2} \\ 0, & otherwise \end{cases} \quad (4.4)$$

Then, the resulting  $l$  length of the output sequence is given as:

$$Rw_{out} = [p_1, p_2, p_3, p_4, \dots, p_l] \quad (4.5)$$

The  $Rw_{out}$  sequence was modulated with a QAM modulator, and the modulator's output was diagonalized, which is obtaining the eigenvalues of the resulting matrix. Then it used to scramble the center user pilot, and the resulting diagonal sequence expressed as:

$$S_j = \begin{pmatrix} A_1 & 0 & \dots & 0 \\ \vdots & \ddots & & \vdots \\ 0 & 0 & \dots & A_l \end{pmatrix}_{l \times l} \quad (4.6)$$

Finally, the diagonal matrix passes through the scrambler, which scrambles the center user pilot.  $\varphi^{cu}_{jKCU} = [\varphi_{jk1}, \varphi_{jk2}, \varphi_{jk3}, \dots, \varphi_{jkl}]$  with the diagonalized pseudo random matrix  $S_j$  of Equation (4.6). Then the resulting output of the scrambler



$\varphi^{Cu}_{KCU_j} \in l \times 1$  is:

$$\begin{aligned} \varphi^{Cu}_{KCU_j} &= S_j \varphi^{cu T}_{jKcu} = \begin{pmatrix} A_1 & 0 & \cdots & 0 \\ \vdots & \ddots & & \vdots \\ 0 & 0 & \cdots & A_l \end{pmatrix}^{l \times l} \begin{bmatrix} \varphi_{jk1} \\ \varphi_{jk2} \\ \varphi_{jk3} \\ \vdots \\ \varphi_{jkl} \end{bmatrix} \\ &= [A_1 \varphi_{jk1} \ A_2 \varphi_{jk2} \ A_3 \varphi_{jk3} \ \dots \ A_l \varphi_{jkl}]^T \end{aligned} \quad (4.7)$$

Therefore, the pilot sequences dedicated for all center users of the  $j_{th}$  cell becomes

$\varphi^{Cu}_j$ ,

$$\begin{aligned} \varphi^{Cu}_j &= [\varphi^{Cu}_{1j} \ \varphi^{Cu}_{2j} \ \dots \ \varphi^{Cu}_{KCU_j}] \\ &= \begin{bmatrix} A_1 \varphi_{j11} & A_1 \varphi_{j21} & \dots & A_1 \varphi_{jk1} \\ A_2 \varphi_{j12} & A_2 \varphi_{j22} & \dots & A_2 \varphi_{jk2} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ A_l \varphi_{j1l} & A_l \varphi_{j2l} & \dots & A_l \varphi_{jkl} \end{bmatrix} \end{aligned} \quad (4.8)$$

$$, j = 1, 2, \dots, L$$

Which has  $l \times KCU_j$  dimension,  $KCU_j$  is the total number of center users in the  $j_{th}$  cell. Moreover, the pilot sequence of edge users of the  $j_{th}$  cell is represented with  $\varphi^{Eu}_j$  and it has  $K_{EU} \times l$  dimension. For each edge users in each cell, the edge pilot sequence expressed in matrix form as:

$$\varphi^{Eu} = \begin{bmatrix} \varphi^{Eu}_1 \\ \varphi^{Eu}_2 \\ \vdots \\ \varphi^{Eu}_L \end{bmatrix} \quad (4.9)$$

Where  $\varphi^{Eu}_L$ , is the pilot sequence applied to edge user of Lth cell, and it has  $K_{Eu} \times l$  dimension.

Using Equation (4.8) and (4.9) the set of pilot sequences applied to the center and edge users of  $j_{th}$  cell are:

$$\varphi_j = \begin{bmatrix} \varphi^{Cu T_j} \\ \varphi^{Eu_j} \end{bmatrix} \quad j = 1, 2, \dots, L \quad (4.10)$$

Therefore, using Equation (4.9) and (4.10) all the available sets of pilot sequence used for the channel estimation phase become in matrix form;

$$\varphi = \begin{bmatrix} \varphi^{Cu T_1} \\ \vdots \\ \varphi^{Cu T_j} \\ \varphi^{Eu} \end{bmatrix} \quad j = 1, 2, \dots, L \quad (4.11)$$

### 4.3 Channel Estimation

We cluster users and their respective pilot sequence into two groups, and then we cluster the channel matrix of each user  $H^l_j$  as:

$$H^l_j \leftrightarrow [H^l_{j Cu} H^l_{j Eu}] \quad (4.12)$$

Where  $H^l_{j Cu} \in C^{M \times K_{Cu j}}$  is the channel matrix of all users of center users of the  $j_{th}$  cell to lth BS, and  $H^l_{j Eu} \in C^{M \times K_{Eu j}}$  is the channel matrix of all users of edge users of the  $j_{th}$  cell to lth BS,

Then the uplink received pilot signal at lth BS is:

$$\begin{aligned} R^l_p &= \sum_{i=1}^L \sqrt{\rho_j} H^l_i \varphi_i + N^l \\ &= R^{Cu l}_p + R^{Eu l}_p + N^l \end{aligned}$$

$$\begin{aligned}
&= \sqrt{\rho_p} \sum_{i=1}^L H^l_{iCu} \varphi^{CuT}_i + \sqrt{\rho_p} \sum_{i=1}^L H^l_{iEu} \varphi^{Eu}_i + N^l \\
&= \sqrt{\rho_p} \sum_{i=1}^L \sum_{k=1}^{K_{CUj}} h^l_{ikCu} \varphi^{CuT}_{K_i} + \sqrt{\rho_p} \sum_{i=1}^L \sum_{k=1}^{K_{EUi}} h^l_{ikEu} \varphi^{Eu}_{ik} + n^l \quad (4.13)
\end{aligned}$$

The pilot sequence is then received at the  $l$ th BS, and if needed, the channel of each user is estimated by correlating the received pilot signal ( $R_p^l$ ) with their respective pilot sequence. If the BS  $l$  wants to estimate the channel of the  $k$ th user, it correlates the signal to the pilot sequence.  $\varphi^{Cu}_{jk}^*$  as

$$\begin{aligned}
R^l_{pjk} &= R^l_p \varphi^{Cu}_{jk}^* \\
&= \sqrt{\rho_p} \sum_{i=1}^L \sum_{k=1}^{K_{CUi}} h^l_{ikCu} \varphi^{CuT}_{ik} \varphi^{Cu}_{jk}^* + \sqrt{\rho_p} \sum_{i=1}^L \sum_{k=1}^{K_{EUi}} h^l_{ikEu} \varphi^{Eu}_{ik} \varphi^{Cu}_{jk}^* \\
&\quad + n^l \varphi^{Cu}_{jk}^* \\
&= \sqrt{\rho_p} \sum_{i=1}^L \sum_{k=1}^{K_{CUi}} h^l_{ikCu} \varphi^{CuT}_{K_i} \varphi^{Cu}_{jk}^* + n^l \varphi^{Cu}_{jk}^* \\
&= \sqrt{\rho_p} \sum_{(i,k) \subseteq U_{ju}} \sum_k h^l_{ikCu} \overbrace{\varphi^{CuT}_{ki} \varphi^{Cu*}_{jk}}^{\varphi^H_{ik} S^T_j S^*_l \varphi^H_{jk}} + n^l \varphi^{Cu}_{jk}^* \quad (4.14) \\
&= \sqrt{\rho_p} \tau \sum_{(i,k) \subseteq U_{ju}} \sum_k h^l_{ikCu} \theta_{ji} + n^l \varphi^{Cu}_{jk}^*
\end{aligned}$$

Where,  $\varphi^H_{ik} S^T_j S^*_l \varphi^H_{jk} = \tau \theta_{ji}$  which is the cross correlation of the pseudo-random sequence and  $U_{ju}$  represents the users in different cells using the same pilot as user  $k$  in cell  $j$  assigned to cell  $j$  and user  $k$  as defined in Equation (3.34).

### 4.3.1 Center Users Channel Estimate and MSE

The MMSE channel estimator is a good estimator in terms of accuracy. Then, the MMSE estimate of the uplink center users channel  $h^l_{jkCu}$  is:

$$\hat{h}^l_{jkCu} = E\{h^l_{jkCu}/R^l_{pjk}\} = \sqrt{p_{jk}} \frac{C^l_{jk} R^l_{pjk}}{\Psi^l_{jk}} \quad (4.15)$$

$\Psi^l_{jk}$ , is the covariance matrix of the received pilot signal  $R^l_{pjk}$ , and is given by

$$\begin{aligned} E\{R^l_{pjk} R^{lH}_{pjk}\}/\tau &= \Psi^l_{jk} = \sum_{i=1}^L \sum_{k=1}^K \rho_p \varphi^{CuT}_{jk} \varphi^{Cu*}_{ik} C^l_{ik} \varphi^{CuT}_{ik} \varphi^{Cu*}_{jk} + \sigma^2 I_{Ml} \\ &= \sum_{(i,k) \subseteq U_{ju}} \sum_k \rho_p \tau \theta^2_{ji} C^l_{ik} + \sigma^2 I_{Ml} \end{aligned}$$

Then, the MSE of the estimated channel is given as,

$$\begin{aligned} MSE &= E\left\{\left\| h^l_{jkCu} - \hat{h}^l_{jkCu} \right\|^2\right\} \\ &= E\{tr\{(h^l_{jkCu} - \hat{h}^l_{jkCu})(h^l_{jkCu} - \hat{h}^l_{jkCu})^H\}\} \\ &= tr\{E\{h^l_{jkCu} h^{lH}_{jkCu}\} - E\{h^l_{jkCu} \hat{h}^{lH}_{jkCu}\} - E\{\hat{h}^l_{jkCu} h^{lH}_{jkCu}\} + E\{\hat{h}^l_{jkCu} \hat{h}^{lH}_{jkCu}\}\} \\ &= tr\{C^l_{jk} - \underbrace{\sqrt{p_{jk}} E\{h^l_{jkCu} R^l_{pjk}\}}_{(1)} \Psi^{-H}_{jk} C^{lH}_{jk} - \underbrace{\sqrt{p_{jk}} C^l_{jk} \Psi^{-1}_{jk} E\{R^l_{pjk} h^{lH}_{jkCu}\}}_{(2)} \\ &\quad - \underbrace{p_{jk} C^l_{jk} \Psi^{-1}_{jk} E\{R^l_{pjk} R^{lH}_{pjk}\}}_{(3)} \Psi^{-H}_{jk} C^{lH}_{jk}\} \end{aligned} \quad (4.16)$$

After that using (4.16) and (4.18) the expression of MSE becomes;

$$\begin{aligned}
MSE &= tr\{C^l_{jk} - p_{jk}\varphi^{CuT}_{jk}\varphi^{Cu*}_{jk}C^l_{jk}\Psi^{-H}_{jk}C^{lH}_{jk} - p_{jk}C^l_{jk}\Psi^{-1}_{jk}C^l_{jk}\overbrace{\varphi^{CuT}_{jk}\varphi^{Cu*}_{jk}}^I \\
&\quad - p_{jk}C^l_{jk}\Psi^{-1}_{jk}\overbrace{\left(\sum_{i=1}^L\sum_{k=1}^K\rho_{ik}\varphi^{CuT}_{ik}\varphi^{Cu*}_{ik}C^l_{ik}\varphi^{CuT}_{ik}\varphi^{Cu*}_{jk} + \sigma^2IMl\right)}^I\Psi^{-H}_{jk}C^{lH}_{jk}\} \\
&= tr\{C^l_{jk} - p_{jk}C^l_{jk}\Psi^{-1}_{jk}C^l_{jk}\} \\
&= tr\{C^l_{jk} - p_{jk}C^l_{jk}\left(\sum_{i=1}^L\sum_{k=1}^K\rho_{ik}\overbrace{\varphi^{CuT}_{jk}\varphi^{Cu*}_{ik}}^{\varphi_{jk}S^T_lS^*_i\varphi^H_{ik}}C^l_{ik}\overbrace{\varphi^{CuT}_{ik}\varphi^{Cu*}_{jk}}^{\varphi^H_{ik}S^T_iS^*_j\varphi^H_{jk}} + \sigma^2IMl\right)^{-1}C^l_{jk}\} \\
&= tr\{C^l_{jk} - p_{jk}C^l_{jk}\left(\sum_{(i,k)\subseteq U_{ju}}\sum_k\rho_{ik}\tau\theta^2_{ji}C^l_{ik} + \sigma^2IMl\right)^{-1}C^l_{jk}\} \quad (4.17)
\end{aligned}$$

From Equation (4.17), we can see that the channel estimate error of the center users are reduced by decreasing the cross correlation of the pseudo random sequence  $(\theta_{ji})$  used by different cells in to zero value (uncorrelated). And also for numerical purpose the NMSE of the estimated channel is used in this thesis, wherein normalization has been done by the correlation of the channel.

### 4.3.2 Edge Users Channel Estimate and MSE

Moreover, when the BS wants to estimate the channel of edge user  $(h^l_{jkEu})$ , it correlates the received pilot signal  $(R^l_p)$  with their respective pilot sequence  $(\varphi^{Eu}_{jkH})$ , then using the orthogonality principle of center user pilot and edge user pilot  $R^l_{pjk}$

given as,

$$\begin{aligned}
R_{pjk}^l &= R_p^l \varphi_{jk}^{Eu} H \\
&= \sqrt{\rho_p} \sum_{i=1}^L \sum_{k=1}^{K_{Cu}^i} h_{ikCu}^l \varphi_{ik}^{CuT} \varphi_{jk}^{Eu} H + \sqrt{\rho_p} \sum_{i=1}^L \sum_{k=1}^{K_{Eu}^j} h_{ikEu}^l \varphi_{ik}^{Eu} \varphi_{jk}^{Eu} H \\
&\quad + n^l \varphi_{jk}^{Eu} H \\
&= \sqrt{\rho_p} \sum_{i=1}^L \sum_{k=1}^{K_{Eu}^i} h_{ikEu}^l \varphi_{ik}^{Eu} \varphi_{jk}^{Eu} H + n^l \varphi_{jk}^{Eu} H
\end{aligned} \tag{4.18}$$

Then, the MMSE estimate of the edge users channel  $\hat{h}_{jkEu}^l$  given as:

$$\hat{h}_{jkEu}^l = E\{h_{jkEu}^l / R_{pjk}^l\} = \sqrt{p_{jk}} \frac{C_{jk}^l R_{pjk}^l}{\Psi_{jk}^l} \tag{4.19}$$

Where,  $\Psi_{jk}^l = \rho_l \tau C_{jk}^l + \sigma^2 I M l$

In addition, its MSE of the channel estimate becomes:

$$MSE_{Eu} = tr\{C_{jk}^l - p_{jk} C_{jk}^l (\rho_{jk} C_{jk}^l + \sigma^2 I M l) C_{jk}^{lH}\} \tag{4.20}$$

Therefore, the entire channel estimate of the uplink channel of  $j_{th}$  cell users is expressed in matrix form as:

$$\hat{H}_j^l \leftrightarrow [\hat{H}_{jCu}^l \hat{H}_{jEu}^l] \tag{4.21}$$

## 4.4 Uplink Users Data Detection

Suppose  $d^{Cu}_j = [d^{Cu}_{j1}, d^{Cu}_{j2}, \dots, d^{Cu}_{jK_{Cu}_j}]^T$  represents the data vector transmitted from the  $K_{Cu}_j$  center users in the  $j_{th}$  cell and  $d^{Eu}_j = [d^{Eu}_{j1}, d^{Eu}_{j2}, \dots, d^{Eu}_{jK_{Eu}_j}]^T$  represents the data vector transmitted from the  $K_{Eu}_j$  edge users in the  $j_{th}$  cell, then we can express the data signal of all users in the  $j_{th}$  cell as  $D_j \leftrightarrow [d^{CuT}_j d^{EuT}_j]^T$

$j=1,2,\dots,L$ . Then the received data signal at the  $l$ th BS can be represented as,  $R_d^l$

$$\begin{aligned}
R_d^l &= \sum_{i=1}^L \sqrt{\rho_i} H_i^l D_i + N^l \\
&= \sum_{k=1}^{K^{Cu_j}} \sqrt{\rho_{jk}} h_{jk}^{Cu} d_{jk}^{Cu} + \sum_{k=1}^{K^{Eu_j}} \sqrt{\rho_{jk}} h_{jk}^{Eu} d_{jk}^{Eu} + \sum_{\substack{i=1 \\ i \neq j}}^L \sum_{k=1}^K \sqrt{\rho_{ik}} h_{ik}^l d_{ik}^l + n^l
\end{aligned} \tag{4.22}$$

Where,  $D_j$  denotes the symbol vector transmitted from the  $K_j$  users in the  $j_{th}$  cell, and  $N^l$  is the corresponding AWGN.

The  $l$ th cell BS used MMSE detector in conjugate with the channel estimate obtained by Equation (4.21) to detect the uplink transmitted data signal for each user as follows. The detected symbol vector for the  $K^{Cu_j}$  users in the  $j$ th cell for the Center users can be given as:

$$\begin{aligned}
\hat{d}_{jk}^{Cu_l} &= u^{CuH}_{jk} R_d^l \\
&= u^{CuH}_{jk} \left( \begin{array}{c} \sum_{k=1}^{K^{Cu}} \sqrt{\rho_{jk}} h_{jk}^{Cu} d_{jk}^{Cu} + \sum_{k=1}^{K^{Eu}} \sqrt{\rho_{jk}} h_{jk}^{Eu} d_{jk}^{Eu} \\ + \sum_{\substack{i=1 \\ i \neq j}}^L \sum_{k=1}^K \sqrt{\rho_{ik}} h_{ik}^l d_{ik}^l + n^l \end{array} \right) \\
&= \sqrt{\rho_{jk}} u^{CuH}_{jk} h_{jk}^{Cu} d_{jk}^{Cu} + \sum_{\substack{k=1 \\ k \neq j}}^{K^{Cu_i}} \sqrt{\rho_{jk}} u^{CuH}_{jk} h_{jk}^{Cu} d_{jk}^{Cu} \\
&\quad + \sum_{k=1}^{K^{Eu}} \sqrt{\rho_{jk}} u^{CuH}_{jk} h_{jk}^{Eu} d_{jk}^{Eu} + \sum_{\substack{i=1 \\ i \neq j}}^L \sum_{k=1}^K \sqrt{\rho_{ik}} u^{CuH}_{jk} h_{ik}^l d_{ik}^l + u^{CuH}_{jk} n^l
\end{aligned} \tag{4.23}$$

Moreover, from Equation (4.23) the uplink SINR of a user  $K^{Cu}_j$  is given as:

$$SINR^{UL Cu}_{jk} = \frac{|\mathbb{E} \{ \sqrt{\rho_{jk}} u^{CuH}_{jk} h^l_{jkCu} d^{Cu}_{jk} \}|^2}{\left( \begin{array}{l} \sum_{k=1}^{K^{Cu}} \rho_{jk} E \left\{ |u^{CuH}_{jk} h^l_{jkCu} d^{Cu}_{jk}|^2 \right\} \\ k \neq k \\ + \sum_{k=1}^{K^{Eu}} \rho_{jk} E \left\{ |u^{CuH}_{jk} h^l_{jkEu} d^{Eu}_{jk}|^2 \right\} \\ + \sum_{i=1}^L \sum_{k=1}^K \rho_{ik} E \left\{ |u^{CuH}_{jk} h^l_{ik} d^l_{ik}|^2 \right\} + \|u^{CuH}_{jk}\|^2 \\ i \neq j \end{array} \right)} \quad (4.24)$$

For the Edge users the detected symbol vector for the  $K^{Eu}_j$  users in the  $j$ th cell can be obtained as;

$$\begin{aligned} \hat{d}_{jk}^{Eu} &= u^{EuH}_{jk} R^l_d \\ &= u^{EuH}_{jk} \left( \begin{array}{l} \sum_{k=1}^{K^{Cu}} \sqrt{\rho_{jk}} h^l_{jkCu} d^{Cu}_{jk} + \sum_{k=1}^{K^{Eu}} \sqrt{\rho_{jk}} h^l_{jkEu} d^{Eu}_{jk} \\ + \sum_{i=1}^L \sum_{k=1}^K \sqrt{\rho_{ik}} h^l_{ik} d^l_{ik} + n^l \\ i \neq j \end{array} \right) \\ &= \sqrt{\rho_{jk}} u^{EuH}_{jk} h^l_{jkEu} d^{Eu}_{jk} + \sum_{k=1}^{K^{Cu}} \sqrt{\rho_{jk}} u^{EuH}_{jk} h^l_{jkCu} d^{Cu}_{jk} \\ &+ \sum_{\substack{k=1 \\ k \neq k}}^{K^{Eu}} \sqrt{\rho_{jk}} u^{EuH}_{jk} h^l_{jkEu} d^{Eu}_{jk} + \sum_{\substack{i=1 \\ i \neq j}}^L \sum_{k=1}^K \sqrt{\rho_{ik}} u^{EuH}_{jk} h^l_{ik} d^l_{ik} + u^{EuH}_{jk} n^l \end{aligned} \quad (4.25)$$



Moreover, from Equation (4.25) the uplink SINR of the user  $K^{Eu}_j$  is given as:

$$SINR^{UL Eu}_{jk} = \frac{|\mathbb{E} \{ \sqrt{\rho_{jk}} u^{EuH}_{jk} h^l_{jkEu} d^{Eu}_{jk} \}|^2}{\left( \begin{array}{l} \sum_{k=1}^{K^{Eu}} \rho_{jk} E \left\{ |u^{EuH}_{jk} h^l_{jkEu} d^{Eu}_{jk}|^2 \right\} \\ k \neq k \\ + \sum_{k=1}^{K^{Cu}} \rho_{jk} E \left\{ |u^{EuH}_{jk} h^l_{jkCu} d^{Cu}_{jk}|^2 \right\} \\ + \sum_{i=1}^L \sum_{k=1}^K \rho_{ik} E \left\{ |u^{EuH}_{jk} h^l_{ik} d^l_{ik}|^2 \right\} + ||u^{EuH}_{jk}||^2 \\ i \neq j \end{array} \right)} \quad (4.26)$$

Hence, The achievable uplink SE formulated as:

$$SE = (1 - \mu_0) E \{ \log_2 (1 + SINR) \} \quad (4.27)$$

Where  $\mu_0 = \frac{\tau_p}{\tau_c}$  is the spectral efficiency loss for channel estimation.

In addition, the uplink target cell sum SE is obtained as:

$$SE_{jsum} = \sum_{i=1}^K SE_i \quad (4.28)$$

## 4.5 System Computational Complexity

In this section, the computational complexity of joint user clustering and pseudo random based pilot assignment in massive MIMO TDD cellular system is analysed. Basically, the complexity analysis of the system has been done for three parts: The first part is the channel estimation phase and the second part is the uplink data detection phase and the last part is the K-means clustering part. Then in each part, the complexity analysis has been done in terms of multiplications and division in complex form.

Generally, for each part there are: matrix to matrix multiplication, matrix addition and matrix pseudo inverse operation. The computational complexity are

examined once in each coherence block. For complexity analysis, let's consider two matrices  $X \in C^{A \times B}$  and  $Y \in C^{B \times C}$ : Then the product of the two matrices  $XY$  requires  $ABC$  number of complex multiplication and  $(B-1)AC$  number of additions. and if matrix  $X$  is a hermitian symmetric matrix then the complex multiplication requirement of the matrix multiplication of  $XX^H$  is given as:

$$\left( A + \frac{A^2 - A}{2} \right) * B \quad (4.29)$$

Since for symmetric matrix:  $X=X^T$  [51]. For a positive definite matrix  $X \in C^{A \times A}$  and  $Y \in C^{A \times B}$  to compute the product  $X^{-1}Y$ , it has been required  $A^2B$  complex multiplications and  $A$  number of complex division [2]. first the inverse of ' $X$ ' matrix has to be computed: i.e  $X^{-1}$  using  $LDL^H$  technique it requires  $\frac{A^3-A}{3}$  complex multiplications [52].

Therefore, In the channel estimation phase each users transmit their pilot sequence to the serving BS. For the cell center users in pseudo random based pilot scrambling of Equation (4.7) it required  $\tau_p$  number of complex multiplications, Then the uplink pilots of all cell users ( $R_p^l C^{M \times \tau_p}$ ) are received at the  $l$ th BS as stated in Equation (4.13) it constitute  $H_j^l \in C^{M \times K}$ ,  $\varphi_j \in C^{K \times \tau_p}$  and  $N_p^l \in C^{M \times \tau_p}$  matrices. Then it requires  $LMK\tau_p$  number of complex multiplications and  $(K-1)M\tau_p + ML\tau_p$  number of additions. And also for computation of Equation (4.14) which is correlating the received pilot signal  $R_p^l C^{M \times \tau_p}$  with the hermitian of the specific user pilot sequence  $\varphi_{kl} \in C^{1 \times \tau_p}$  has been took  $2M\tau_p$  complex multiplications and  $2(\tau_p - 1)M$  number of additions.

The MMSE channel estimate per user of Equation(4.15) is the multiplication of  $C_{llk}^l C^{M \times M}$ ,  $\Psi_{lk}^l C^{M \times M}$  and  $R_{plk}^l C^{M \times 1}$  matrices. Then the number of complex multiplication required is expressed as:

$$\frac{M^3 + M^2}{2} + \frac{M^3 - M}{3} + M^3 + M^2 \quad (4.30)$$

The second part is the computational analysis of uplink data reception and detection using MMSE detection technique at the  $l$ th BS. The uplink data reception i.e

Equation(4.24), is the multiplication of  $D^j l \in C^{K \times 1}$ , and  $H_j^l \in C^{M \times K}$  matrices. Then it required  $MKL\tau_u$  number of complex multiplications. Then the detected users data by MMSE receive combining vector for each users expressed in Equation (4.25) and (4.27) at the  $l$ th BS is the multiplication of MMSE combining vector  $u_{lk}^H C^{1 \times M}$  and uplink data matrix  $R_d^l C^{M \times \tau_u}$ . Then it required  $MKL\tau_u$  number of complex multiplications. And also the number of complex multiplications required to compute the MMSE combining vector for the specific user defined in Equation (3.25)  $u_{lk}^{MMSE} C^{M \times 1}$  given as:

$$\frac{M^3 - M}{3} + M^2 + \left(\frac{M^3 + M^2}{2}\right)LK + 2M \quad (4.31)$$

The third part of the system complexity analysis is the K-means clustering part. As stated in [54] the space complexity of this algorithm is expressed in big O notation, which has  $O((K+2)L)$ .

Then table 4.1 summarise the computational complexity of the proposed joint user clustering and pseudo random based pilot assignment. It has been analysed in two parts: the channel estimation phase and the uplink data detection phase. For the two parts the complexity has been done in terms of complex multiplication and complex division.

TABLE 4.1: The Computational Complexity Analysis

	Multiplication	Division
Channel Estimation	$\tau_p + LMK\tau_p + 2M\tau_p + \frac{11}{6}M^3 + \frac{3}{2}M^2 - \frac{1}{3}M$	—
Uplink Data Detection	$\frac{1}{3}M^3 + \frac{1}{3}M^3LK + M^2 + \frac{1}{2}M^2LK + 2MLK\tau_u + \frac{5}{3}M$	$M$

# Chapter 5

## Result and Discussion

### 5.1 Simulation Parameters and Assumption Made

This chapter analyzes the performance of proposed joint user clustering and pseudo random based pilot assignment and random pilot assignment in massive MIMO multi-cell multi-user TDD cellular system. Furthermore, the 3GPP urban Macro cell channel model is used, with a 2GHz carrier frequency [17] and a regular hexagonal form to represent each cell. From the channel model, we used a cell radius of 500 m and a minimum distance to BS of 30 m [17]. As shown in Figure 5.1, we evaluated L number of cells with a single BS (each BS used M antenna) at each cell, placed in the cell's center and servicing K number of single antenna users distributed randomly in each cell.

We also assumed that there is no correlation at the BS antenna array due to the half-wavelength antenna spacing. For urban regions, we employed the Angular Standard Deviation (ASD) of the multipath components of spatially correlated channels of 10 degrees[48] and the Normalized Cross Correlation of the pseudo random sequence of 0.3[46]. The proposed and random pilot assignment systems are simulated using simulation parameters from Table 5.1. Matlab software has been used for simulation of User Grouping, the NMSE of channel estimation, the CDF of SINR, the CDF of spectral efficiency of each user, and the target cell sum

SE. 5.1 Presented the details of basic parameters used in the simulation to assess the proposed system's performance.

TABLE 5.1: Parameters Used for Simulation

Parameters	Symbols	Values
Cell radius	$R$	500m
Cell Shape	—	hexagonal
Number of antennas at each BS	$M$	$10 \leq M \leq 400$
Number of users in each cell	$K$	10
Average transmit power at each users	$\rho_{ul}$	10dBm
Average transmit power at BS	$\rho_{dl}$	12dBm
Path loss exponent	$\alpha$	3.76
Log normal shadowing fading	$\gamma_{sh}$	8dB
Carrier frequency	$f_c$	2GHz
System Bandwidth	$B$	20MHz
Samples per coherence block	$\tau_c$	200
pilot sequences for Random	$\tau_p$	$K < \tau_p < \tau_c, 15$

## 5.2 User Distribution in Each Cell and User Clustering

As shown in Figure 5.1, each user in a hexagonal cell with a radius of 500m is uniformly spread around the serving BS. So, using the parameters in Table 5.1, we generate users at random across the cell, each with an x and y position. The red crosses in each cell represent users, while the yellow triangle in the cell's center represents the serving BS. After users are distributed, using the K-means clustering algorithms users in the cell are grouped as a center and edge users. As shown in Figure 5.2, once users are grouped, the blue diamond shape represents center users, while the red circle shape represents edge users.

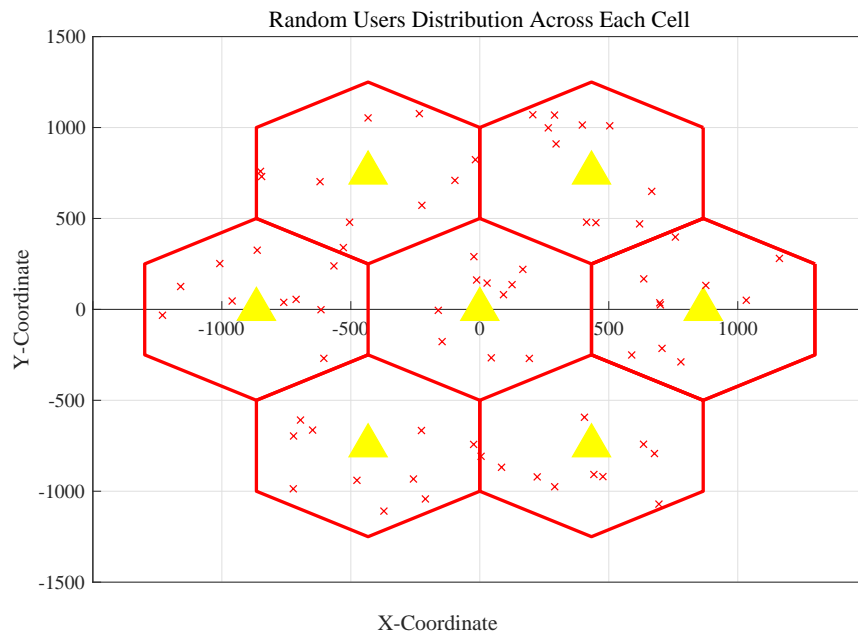


FIGURE 5.1: Simulation Result of User Distribution Across  $L=7$  cells using  $R=500\text{m}$  and  $K=10$ . Wherein the red crosses and yellow triangle denote users and BS respectively.

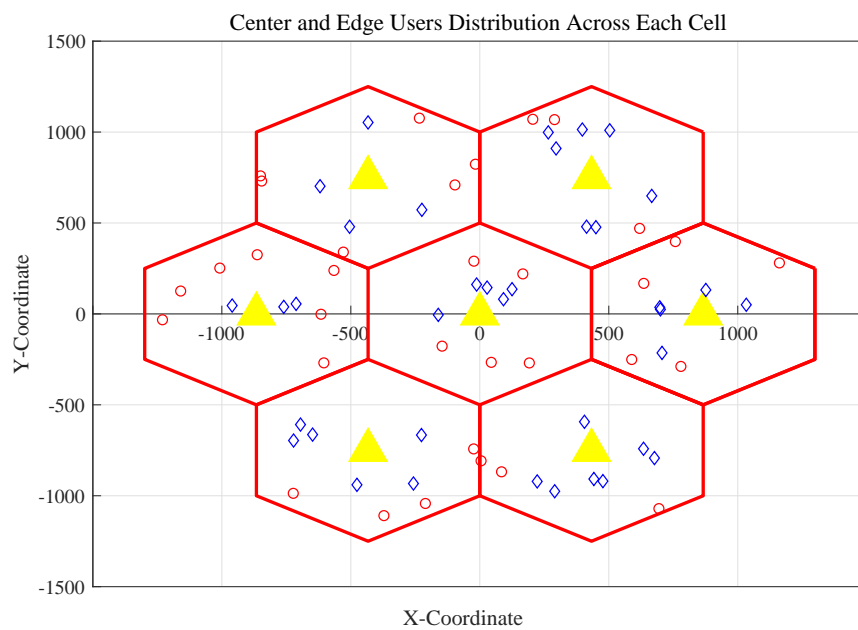


FIGURE 5.2: Simulation scenario of User Clustering in each Cell. Wherein the blue diamond and red circle represent the center and edge users respectively.

### 5.3 Normalized Mean Square Error(NMSE) of Channel Estimate

Figure 5.3 shows the simulation results of the proposed and random pilot assignment systems' normalized mean square error versus BS antennas for a 50-channel realization. As the number of BS antennas increases, the NMSE for both systems reduces in a decreasing trend, as seen in the figure. For both center and edge users, the proposed technique outperforms the random pilot assignment. This is because the proposed system takes into account the users' channel state by classifying them as center or edge users and assigning orthogonal pilot to edge users while reusing pilot for adjacent center users. In addition, as mentioned in section 5.1, pseudo-random code with a normalized cross-correlation of 0.3 [46] was used to reduce the correlation of this center user pilot.

As a result, the influence of pilot contamination is reduced, and we obtain a channel estimate with decreased estimation error. The NMSE is increased by 3.07 dB and 9.63 dB for edge and center users, respectively, using a specific BS antenna 210. As a result, the suggested system improves users' channel estimates by mitigating pilot contamination due to system pilot reuse.

### 5.4 Commulative Distribution Function(CDF) of SINR and SE

Figures 5.4 and 5.5 show the probability distribution function (CDF) of the SINR and the spectral efficiency of each user in each cell, respectively. In the simulation, 192 BS antennas are utilized to serve 10 users in each cell, with the rest of the parameters remaining unchanged from table 5.1. As demonstrated in Figure 5.4, the SINR of edge users in the proposed system is better than in the random pilot assignment, because the pilots assigned to each edge user in the proposed system are orthogonal to each other. As a result, their channel estimate improved, the



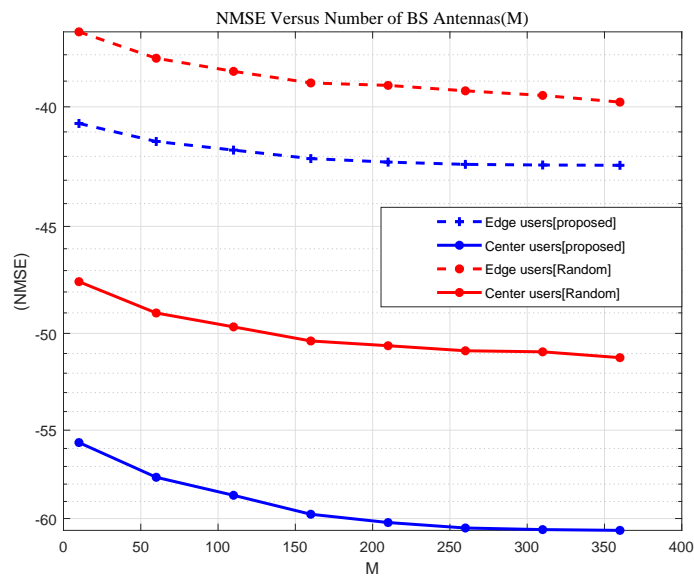


FIGURE 5.3: Simulation Result of NMSE versus BS Antenna with  $R=500\text{m}$ ,  $K=10, L=7$  and  $ASD = 10$  degree

interference part in the SINR denominator owing to pilot contamination decreased, and their SINR improved. Furthermore, the proposed system's center users have improved SINR, despite the fact that the center pilot sequences are reused in each cell. This is because the pseudo random sequence used in the proposed system has good cross correlation property, which reduces the correlation of center users pilot as defined in Equation (4.17). As a result, it enhance the channel estimate of the users due to the effect of pilot contamination.

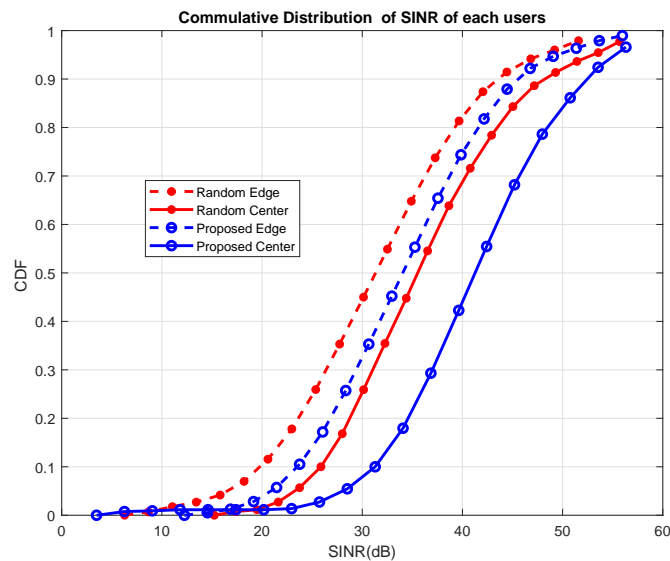


FIGURE 5.4: Simulation Result of CDF of the Uplink SINR, with  $M = 192$ ,  $K = 10$  and  $ASD = 10$  degree

As seen from Figure 5.5, The SE of center users for the proposed system is improved over the random pilot assignment, this is due to the gain of SINR term is greater than the reduction in the Prelog factor in the SE expression of (4.27). In addition, as compared to the random pilot assignment, 15% of edge users in the proposed system have their SE reduced. This is because, as mentioned in Equation (4.3), the proposed system required more pilot sequence for the channel estimation phase than the standard system. As a result, the increase in channel estimate does not outweigh the decrease in prelog factor in SE expression for such users. In general, as shown in the graph, 15% of edge users have some SE loss, 25% of edge users have the same SE, and 60% of edge users have improved SE for the proposed system when compared to the random pilot assignment.

## 5.5 Cell Sum Spectral Efficiency(SE) Against BS Antennas

Figure 5.6 shows the simulation result of the Sum SE versus the number of BS antennas. As can be seen in the graph, as the number of antennas at the BS increases, the sum SE of the target cell increases for both proposed and random pilot assignment. However, it is noticeable that as the number of BS antennas

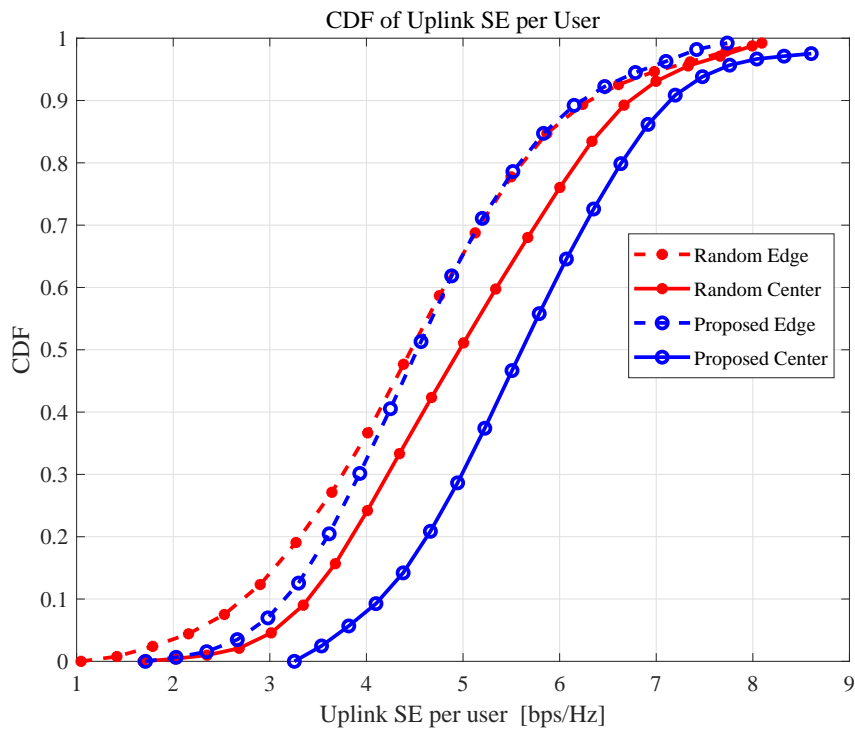


FIGURE 5.5: Simulation Result of CDF of the Uplink SE per UE, with  $M = 192$ ,  $K = 10$ , and  $ASD = 10$  degree

increases, the proposed system's target Cell sum SE outperforms the random pilot assignment from the graph. For example, when we use a BS antenna with  $M=210$ , the proposed system outperforms the random pilot assignment by 4.18bps/Hz/cell. This is because the effect of pilot contamination has been reduced by efficiently assigning pilot sequence to users based on the severity of the problem affecting each user.

Furthermore, the figure shows that when the number of BS antennas increases, the improvement gap between the two systems widens. Our proposed system outperforms the random pilot assignment by 4.566 bps/Hz/cell at  $M=370$  BS antenna, and this improvement is increasing in comparison to  $M=210$  BS antenna. This is due to the fact that performance improvement with increasing BS antenna in massive MIMO is limited by the effect of pilot contamination.

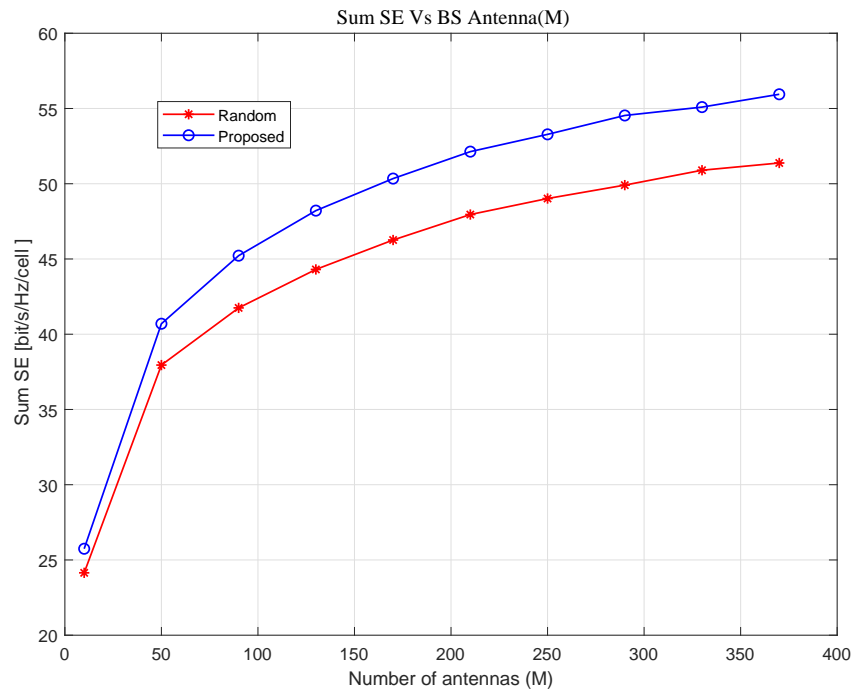


FIGURE 5.6: Simulation Result of target cell Sum SE versus BS Antenna, with  $K = 10$ , and with ASD = 10 degree

## 5.6 Sum Spectral Efficiency(SE) Versus Average Transmit Power of Users

The trend of target cell Sum SE versus average transmit power of each user is shown in Figure 5.7. As it can be observed in the graph, as the average transmit power of each user increases, the sum SE of the target cell increases for both proposed and random pilot assignment. Furthermore, as shown in the graph, the proposed system's Sum SE is higher than random pilot assignment, because as the average transmit power of the desired users increases, so does the average transmit power of the interference.

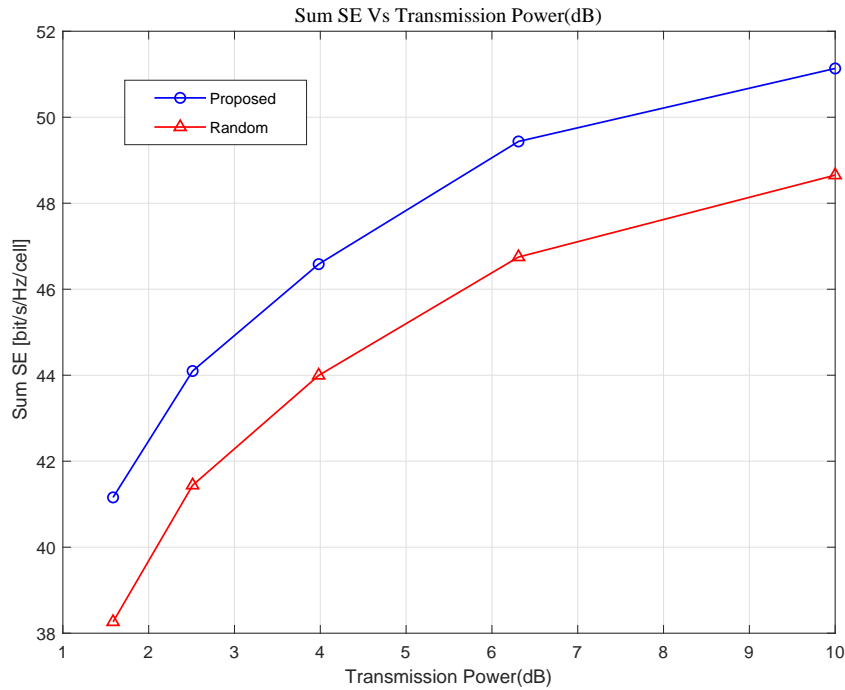


FIGURE 5.7: Simulation Result of the sum SE [bps/Hz/cell] against average transmit power of users with  $M = 192$ ,  $K = 10$ ,  $L=7$  and  $ASD = 10$  degree

## 5.7 Sum Spectral Efficiency(SE) Versus Coherence Block Length

Figure 5.8 shows the simulation result of the Sum SE vs the coherence block length. As seen in the graph, as the length of coherence block increase then the target cell sum spectral efficiency also increase for both of proposed and random pilot assignment techniques. The Sum SE increment with coherence length is due to the increment in the prelog factor  $(1 - \frac{\tau_p}{\tau_c})$  in Equation (4.29). From Equation (3.33) the coherence block length has an indirect relation to the mobility of the users, as the mobility of users increase the interval in which the channel approximately flat is short and vis versa. For a coherence block length of 200 supports a mobility of 135Km/h, and then as the coherence block length increases to 1000 the mobility which the channel supports decreased. However, it is noticeable that as the number of BS antennas increases, the proposed system's target Cell sum SE outperforms the random pilot assignment from the graph.

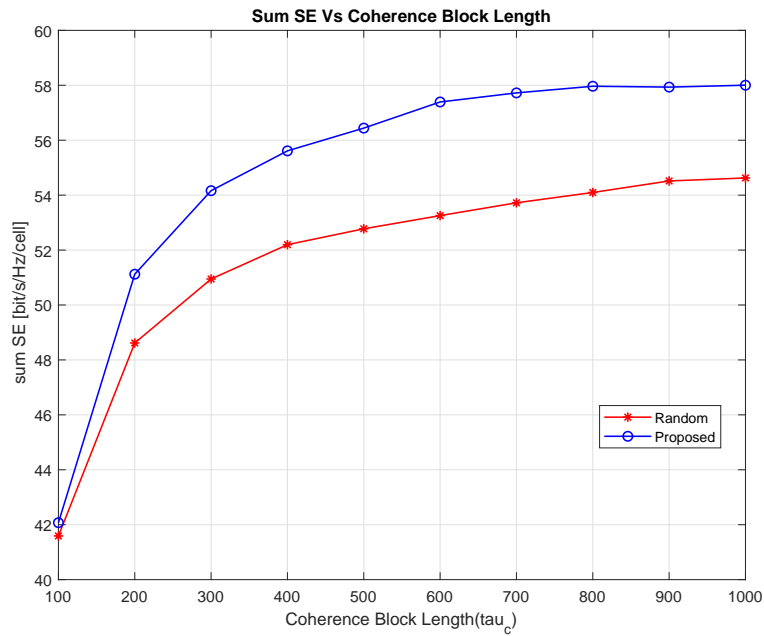


FIGURE 5.8: Simulation Result of the sum SE [bps/Hz/cell] against Coherence Block Length with  $M = 192$ ,  $K = 10$ ,  $L=7$  and ASD = 10 degree

## 5.8 System Complexity Versus Number Of BS Antenna

Figure 5.9 shows the simulation result of the system complexity Vs the Number Of BS Antenna. The system complexity is expressed in terms of number of complex multiplication, it is clearly seen from the graph that AS the number of base station antennas increase the number of complex multiplication required is also increase. There fore there is always trade off system sum spectral efficiency with system complexity.

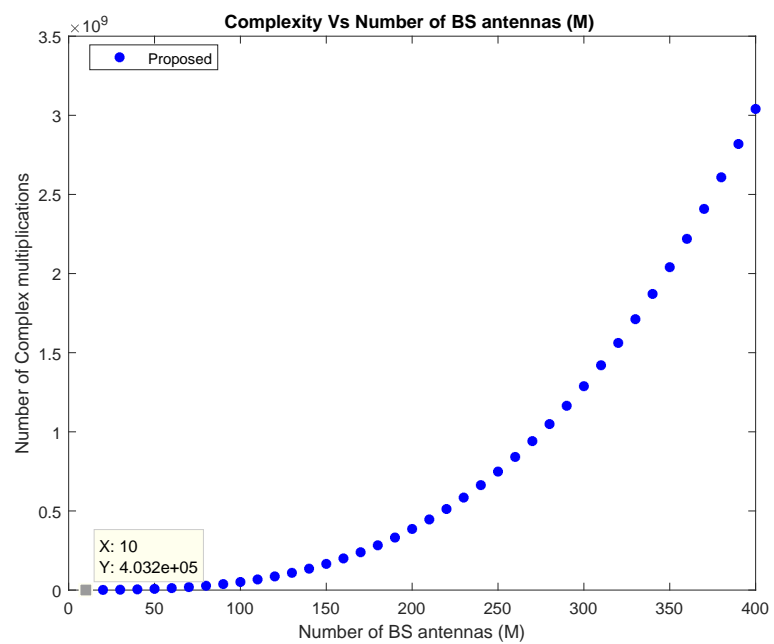


FIGURE 5.9: Simulation Result of System Complexity Versus Number Of BS Antenna with  $K = 10$ , and  $L=7$ .

# Chapter 6

## Conclusion and Recommendation

### 6.1 Conclusion

Massive MIMO is one of the promising 5G and beyond technologies, due to its higher spectrum efficiency. Although, it has this advantages due to the limited number of available pilots in a coherence interval pilots are reused in adjacent cells in channel estimation. This results in pilot contamination that degrades the estimation of channel state information at the BS. In addition, this problem weakens SINR of the uplink users that depends on estimate of CSI, and also it degrades the uplink Spectral Efficiency(SE)that depends on the estimated channel at the BS. For this existing and hot problems various solutions has been suggested by many researchers. Each of the the existing works relay on uncorrelated Rayleigh fading channel.

In this thesis, to minimize pilot contamination effect users in the cell are clustered depending on the severity of the problem affecting each user. The K-means clustering algorithm has been used to group users to cell centred and cell edge users. Then orthogonal pilots are assigned for cell edge users because they are more vulnerable to the effect, while cell center users reuse the same pilot sequence with the adjacent cells. The pseudo random sequence employed for cell center users is effective at reducing correlation between identical pilot users, resulting in



a good channel estimation. The mathematical and numerical results of the proposed system have been evaluated for the following performance metrics: NMSE, SINR, and SE under a spatially correlated Rayleigh fading channel.

From the numerical results the proposed system has shown a 7.76% and 19.021% improvement in normalized mean square error(NMSE) for cell Edge users and cell center users respectively, as compared to random pilot assignment. In addition, In terms of target cell sum SE, the proposed system outperforms random pilot assignment by 8.714%. As a result, we conclude that the proposed system pilot decontamination scheme outperforms random pilot assignment.

## 6.2 Recommendations

We recommend taking into account power control of uplink pilot and payload transmit power of users for further improvement of joint user clustering and pseudo-random based pilot decontamination scheme, which is power control schemes that optimize the SE of users, since we assumed equal transmit power for all users in our system. The last point we recommend is that if there is an optimization function that relates the number of users in a cell, the number of BS antennas, and the channel coherence length, to maximize the SE of the system.

# Bibliography

- [1] Borges, D.; Montezuma, P.; Dinis, R.; Beko, M. Massive MIMO Techniques for 5G and Beyond—Opportunities and Challenges. *Electronics* 2021, 10, 1667. <https://doi.org/10.3390/electronics10141667>.
- [2] Emil Bjrnson, 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 154655.
- [3] F. Rusek et al., "Scaling Up MIMO: Opportunities and Challenges with Very Large Arrays," in *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40-60, Jan. 2013.
- [4] M. Pappa , C. Ramesh , Madhushri N. Kumar, "Performance comparison of massive MIMO and conventional MIMO using channel parameters", *International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, 2017.
- [5] Marzetta, T. L. (2016). *Fundamentals of massive MIMO*. Cambridge University Press.
- [6] Hassan N, Fernando X. Massive MIMO wireless networks: An overview. *Electronics*. 2017 Sep; 6(3):63.
- [7] H. Ngo, E. Larsson, and T. Marzetta, "The multicell multiuser mimo uplink with very large antenna arrays and a finitedimensional channel," *IEEE Trans. Commun.*, vol. 61, no. 6, pp. 23502361, Jun. 2013.

- 
- [8] M. Boudaya, I. Kammoun, M. Siala, "Efficient pilot design based on the divide and conquer approach for pilot contamination mitigation in massive MIMO" IEEE conference paper, 2019.
- [9] Pseudo Random Signal Processing Theory and Application by Hans-Jurgen Zepernick, Adolf Finger (z-lib.org)
- [10] Topuzolu A., Winterhof A. (2006) PSEUDORANDOM SEQUENCES. In: Garcia A., Stichtenoth H. (eds) Topics in Geometry, Coding Theory and Cryptography. Algebra and Applications, vol 6. Springer, Dordrecht.
- [11] S. Memon, Z. Chen, F. Yin, "Pilot Decontamination in Multi-cell Massive MIMO Systems" ICCIP '16, November 26-29, 2016.
- [12] Chataut R, Akl R. Massive MIMO systems for 5G and beyond networks—overview, recent trends, challenges, and future research direction. *Sensors*. 2020 Jan; 20(10):2753.
- [13] 5G Massive MIMO Network Application, ZTE corporation, Hi-tech Road South, ShenZhen,P.R. China; NO.55,
- [14] J. Mohinder, Space-Time Codes and MIMO Systems, Artech House, 2004.
- [15] Costello, D. J. (2009). Fundamentals of wireless communication (tse, d. and viswanath, P.)[Book review]. *IEEE Transactions on Information Theory*, 55(2), 919-920.
- [16] Torres, R.P.; Pérez, J.R. A Lower Bound for the Coherence Block Length in Mobile Radio Channels. *Electronics* 2021, 10, 398.
- [17] 3GPP TR 25.942 V5.3.0 (2004-06). "Technical Specification Group Radio Access Networks; Radio Frequency (RF) system scenarios (Release 5)". Tech. rep.
- [18] Sanguinetti, L., Björnson, E. and Hoydis, J., 2019. Toward massive MIMO 2.0: Understanding spatial correlation, interference suppression, and pilot contamination. *IEEE Transactions on Communications*, 68(1), pp.232-257.

- 
- [19] You L, Gao X, Xia XG, Ma N, Peng Y. Pilot reuse for massive MIMO transmission over spatially correlated Rayleigh fading channels. *IEEE Transactions on Wireless Communications*. 2015 Feb 19;14(6):3352-66.
- [20] Tse, D, and Viswanath, P (2005). *Fundamentals of wireless communication*. Cambridge university press.
- [21] Yin, Haifan, et al. "A coordinated approach to channel estimation in large-scale multiple-antenna systems." *IEEE Journal on selected areas in communications* 31.2 (2013): 264-273.
- [22] Ali S, Chen Z, Yin F. Design of orthogonal uplink pilot sequences for TDD massive MIMO under pilot contamination. *Journal of Communications*. 2017 Jan;12(1):40-8.
- [23] R. Rogalin, O. Y. Bursalioglu, H. C. Papadopoulos, G. Caire, and A. F. Molisch, Hardware-impairment compensation for enabling distributed largescale MIMO in Proc. ITA Workshop, San Diego, CA, USA, 2013.
- [24] Xie H, Gao F, Jin S, Fang J, Liang YC. Channel estimation for TDD/FDD massive MIMO systems with channel covariance computing. *IEEE Transactions on Wireless Communications*. 2018 Apr 9;17(6):4206-18.
- [25] Kay, S. M. (1993). *Fundamentals of statistical signal processing*. Prentice Hall PTR.
- [26] De Figueiredo FA, Cardoso FA, Moerman I, Fraidenraich G. Channel estimation for massive MIMO TDD systems assuming pilot contamination and frequency selective fading. *IEEE access*. 2017 Sep 7;5:17733-41.
- [27] Cheng, Q. (2016). "Novel channel estimation methods under pilot contamination in Massive MIMO".
- [28] Li X, Bjornson E, Larsson EG, Zhou S, Wang J. A multi-cell MMSE detector for massive MIMO systems and new large system analysis. In *2015 IEEE global communications conference (GLOBECOM) 2015 Dec 6* (pp. 1-6). IEEE.

- [29] Chen Z, Bjornson E. Channel hardening and favorable propagation in cell-free massive MIMO with stochastic geometry. *IEEE Transactions on Communications*. 2018 Jun 11; 66(11):5205-19.
- [30] Chen JC. A low complexity data detection algorithm for uplink multiuser massive MIMO systems. *IEEE Journal on Selected Areas in Communications*. 2017 Jun 1;35(8):1701-14.
- [31] Krishnan N, Yates RD, Mandayam NB. Uplink linear receivers for multi-cell multiuser MIMO with pilot contamination: Large system analysis. *IEEE Transactions on Wireless Communications*. 2014 Apr 29;13(8):4360-73.
- [32] Muaayed AR. Massive MIMO system: an overview. *International journal of open information technologies*. 2017;5(2).
- [33] Beiranvand J, Meghdadi H. Analytical performance evaluation of MRC receivers in massive MIMO systems. *IEEE Access*. 2018 Sep 26;6:53226-34.
- [34] Altamirano CD, Minango J, Mora HC, De Almeida C. BER evaluation of linear detectors in massive MIMO systems under imperfect channel estimation effects. *IEEE access*. 2019 Nov 29;7:174482-94.
- [35] B. G. Lee and S. C. Kim, *Scrambling Techniques for CDMA Communications*. Kluwer Academic Publishers, Boston, MA, 2001.
- [36] R. Mahajan, K. Devi, D. Bagai, Review on standered and Multibit Linear Feedback shift Register, 2018 JETIR December 2018, Volume 5, Issue 12
- [37] Dey D, Giri D, Jana B, Maitra T, Mohapatra RN. Linear-feedback shift register-based multi-ant cellular automation and chaotic map-based image encryption. *Security and Privacy* 2018;1:e52. <https://doi.org/10.1002/spy2.52>
- [38] S. C. Kim and B. G. Lee, "Scrambling Techniques for Digital Transmission". Springer-Verlag, New York, 1994.
- [39] Elijah, O.; Leow, C.Y.; Rahman, T.A.; Nunoo, S.; Iliya, S.Z. "A Comprehensive Survey of Pilot Contamination in Massive MIMO5G System." *IEEE Commun. Surv. Tutor*. 2015, 18, 905923.

- [40] C. Zhang and G. Zeng, "Pilot contamination reduction scheme in massive MIMO multi-cell TDD systems," *Journal of Computational Information Systems*, 2014.
- [41] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination problem in multi-cell TDD systems," *IEEE International Symposium in Information Theory*, 2009.
- [42] Zhu X, Wang Z, Qian C, Dai L, Chen J, Chen S, Hanzo L. Soft pilot reuse and multicell block diagonalization precoding for massive MIMO systems. *IEEE Transactions on Vehicular Technology*. 2015 Jun 17; 65(5):3285-98.
- [43] Z. Zhou, D. Wang, and Z. Wang, "Asynchronous pilots scheduling in massive MIMO systems," *IEEE International Conference on Communication*, 2017.
- [44] R.Zhou, Y.FU, "Uplink Asynchronous Fractional Pilots Scheduling in Massive MIMO System" 18th IEEE International Conference on Communication Technology, 2018.
- [45] M. Wan, P. A. Martin, P. J. Smith, "pilot contamination reduction using time-shifted pilots in finite massive MIMO system", *Vehicular Technology conference*, 2014.
- [46] O. A. Saraereh , I. Khan 2 , B. M.Lee 3,\* and A. Tahat, "Efficient Pilot Decontamination Schemes in 5G Massive MIMO Systems," *Electronics* 2019, 8(1), 55.
- [47] Kalantarinejad N, Abbasi-Moghadam D. "Joint distance-based user grouping and pilot assignment schemes for pilot decontamination in massive MIMO systems". *Int J Commun Syst*. 2019; e4216.
- [48] Pedersen, K. I., P. E. Mogensen, and B. H. Fleury. 1997. "Power azimuth spectrum in outdoor environments". *Electronics Lett*. 33(18): 1583–1584.
- [49] Amit Patel (2021). Hexagonal grid (<https://www.redblobgames.com/grids/hexagons/>), Red Blob Games. Retrieved April 2, 2021.

- 
- [50] Daniel J. Benjamin, Matthew Rabin, Collin Raymond, A Model of Non-belief in the Law of Large Numbers, *Journal of the European Economic Association*, Volume 14, Issue 2, 1 April 2016, Pages 515–544, <https://doi.org/10.1111/jeea.12139>
- [51] C. R. Johnson (1970) Positive Definite Matrices, *The American Mathematical Monthly*, 77:3, 259-264, DOI: 10.1080/00029890.1970.11992465
- [52] Tylavsky, D. J., Sohie, G. R. L. (1986). Generalization of the matrix inversion lemma. *Proceedings of the IEEE*, 74(7), 1050–1052. doi:10.1109/proc.1986.13587
- [53] Ingemarsson C, Gustafsson O. On fixed-point implementation of symmetric matrix inversion. In 2015 European Conference on Circuit Theory and Design (ECCTD) 2015 Aug 24 (pp. 1-4). IEEE.
- [54] Firdaus S, Uddin MA. A survey on clustering algorithms and complexity analysis. *International Journal of Computer Science Issues (IJCSI)*. 2015 Mar 1;12(2):62.



# Appendix A

## A.1 The law of large number

The rule of large numbers states that sample measurements tend to the real for a random variable with a large number of realizations defining a set of experiments. Then for any pair of random vectors  $x \in C^M$  and  $y \in C^M$ , [50]. :

$$\lim_{M \rightarrow \infty} \frac{1}{M} x^H x \cong \sigma_x^2 \quad (\text{A.1})$$

$$\lim_{M \rightarrow \infty} \frac{1}{M} x^H y \cong 0 \quad (\text{A.2})$$

## A.2 MMSE based Channel Estimation

We assume the channel model as:

$$y = hx + n \quad (\text{A.3})$$

Where,  $h$  is the channel vector,  $x$  is the signal transmitted and  $n$  is AWGN noise. The MMSE estimator is the Bayesian estimator that minimizes the MSE as defined below,

$$E \left\{ \left\| h - \hat{h} \right\|^2 \right\} \quad (\text{A.4})$$

Where in,  $\hat{h} = ay + b$  and a and b are constants selected to minimize MSE of the channel estimate

$$\begin{aligned}
MSE &= E \{ \| h - (ay + b) \|^2 \} \\
&= tr(E \{ (h - ay - b)(h - ay - b)^H \}) \\
&= tr(E \{ hh^H - hy^H a^H - hb^H - ayh^H + ayy^H a^H + ayb^H - bh^H + by^H a^H + bb^H \}) \\
&= tr(E \{ hh^H \} - E \{ hy^H \} a^H - E \{ h \} b^H - aE \{ yh^H \} + aE \{ yy^H \} a^H \\
&\quad + aE \{ y \} b^H - bE \{ h^H \} + bE \{ y^H \} a^H + bb^H) \\
&= tr(E \{ hh^H \} - E \{ hy^H \} a^H - aE \{ yh^H \} + aE \{ yy^H \} a^H) \\
&\quad + tr(aE \{ y \} b^H - E \{ h \} b^H - bE \{ h^H \} + bE \{ y^H \} a^H + bb^H) \\
&= tr(E \{ hh^H \} - E \{ hy^H \} a^H - aE \{ yh^H \} \\
&\quad + aE \{ yy^H \} a^H) + E \{ \| b - E \{ h \} + aE \{ y \} \|^2 \}
\end{aligned} \tag{A.5}$$

Then by choosing the minimum of a and b, MSE is minimized as,  $b_{min} = E \{ h \} - aE \{ y \}$  Then the MSE expression becomes,

$$\begin{aligned}
MSE &= tr(E \{ hh^H \} - E \{ hy^H \} a^H - aE \{ yh^H \} + aE \{ yy^H \} a^H) \\
&= tr(C_{hh} - C_{hy} a^H - aC_{yh} + aC_{yy} a^H)
\end{aligned} \tag{A.6}$$

Where,  $C_{hh}$  is the autocorrelation of the channel (h),  $C_{hy}$  the cross-correlation of the channel and the received signal (y), and  $C_{yy}$  is the autocorrelation of y. Using the rule of completing the square method for matrix, it results;

$$MSE = tr(C_{hh}) + tr((aC_{yy} - C_{hy}) C_{yy}^{-1} (aC_{yy} - C_{hy})^H) - tr(C_{hy} C_{yy}^{-1} C_{hy}^H) \tag{A.7}$$

The result of MSE is then minimized by selecting a value of  $a = C_{hy}C_{yy}^{-1}$ , substituting this value then finally the mean square error of the channel become:

$$MSE = tr(C_{hh}) - tr(C_{hy}C_{yy}^{-1}C_{hy}^H) \quad (A.8)$$

And, the estimation error covariance matrix become,

$$CC = C_{hh} - C_{hy}C_{yy}^{-1}C_{hy}^H \quad (A.9)$$

Finally, the MMSE channel estimation becomes,

$$\begin{aligned} \hat{h} &= ay + b \\ &= (C_{hy}C_{yy}^{-1})y + E\{h\} - (C_{hy}C_{yy}^{-1})E\{y\} \end{aligned} \quad (A.10)$$

### A.3 Proof of Equation (4.16)

From equation (4.14), we have  $R_{pjk}^l$  and it has Hermitian conjugate of  $R_{pjk}^{lH}$  given as:

$$R_{pjk}^l = \sqrt{\rho_p} \sum_{i=1}^L \sum_{k=1}^{K_{CUi}} h_{ikCu}^l \varphi^{CuT} K_i \varphi_{jk}^{Cu*} + n^l \varphi_{jk}^{Cu*} \quad (A.11)$$

$$R_{pjk}^{lH} = \sqrt{\rho_p} \sum_{i=1}^L \sum_{k=1}^{K_{CUi}} \varphi_{jk}^{CuT} \varphi_{K_i}^{Cu*} h_{ikCu}^{lH} + \varphi_{jk}^{CuT} n^{lH}$$

Then from Equation (4.16) the expressions (1), (2), and (3) are reduced as follow:

$$\begin{aligned}
(1) \quad \underbrace{E\{h_{jkCu}^l R_{pjk}^{lH}\}}_{(1)} &= \sqrt{\rho_p} \sum_{i=1}^L \sum_{k=1}^{K_{Cu}^i} \varphi_{jk}^{Cu}{}^T \varphi_{K_i}^{Cu*} E\{h_{jkCu}^l h_{ikCu}^{lH}\} \\
&+ \varphi_{jk}^{Cu}{}^T E\{h_{jkCu}^l n^{lH}\} \\
&= \sqrt{\rho_{jk}} \varphi_{jk}^{Cu}{}^T \varphi_{jk}^{Cu*} C_{jk}^l \\
(2) \quad \underbrace{E\{R_{pjk}^l h_{jkCu}^{lH}\}}_{(2)} &= \sqrt{\rho_p} \sum_{i=1}^L \sum_{k=1}^{K_{Cu}^i} E\{h_{ikCu}^l h_{jkCu}^{lH}\} \varphi_{K_i}^{Cu}{}^T \varphi_{jk}^{Cu*} \\
&+ E\{n^l h_{jkCu}^{lH}\} \varphi_{jk}^{Cu*} \\
&= \sqrt{\rho_{jk}} C_{jk}^l \varphi_{jk}^{Cu}{}^T \varphi_{jk}^{Cu*} \\
(3) \quad \underbrace{E\{R_{pjk}^l R_{pjk}^{lH}\}}_{(3)} &= \rho_p \sum_{i=1}^L \sum_{k=1}^{K_{Cu}^i} \varphi_{jk}^{Cu}{}^T \varphi_{K_j}^{Cu*} E\{h_{ikCu}^l h_{ikCu}^{lH}\} \varphi_{K_i}^{Cu}{}^T \varphi_{jk}^{Cu*} \\
&+ E\{n^l \varphi_{jk}^{Cu*} \varphi_{jk}^{Cu}{}^T n^{lH}\} \\
&= \sum_{i=1}^L \sum_{k=1}^K \rho_{ik} \varphi_{jk}^{Cu}{}^T \varphi_{ik}^{Cu*} C_{ik}^l \varphi_{ik}^{Cu}{}^T \varphi_{jk}^{Cu*} + \sigma^2 IMl
\end{aligned} \tag{A.12}$$

## A.4 Basic of Hexagonal cell

The individual element of the cellular concept is known as a cell since it uses a hexagonal form to split coverage of the region. If  $R$  is the radius of a cell, then the distance between two neighboring cells from their respective centers in terms of their radius is equivalent to  $R\sqrt{3}$ , as shown in Figure A.1.

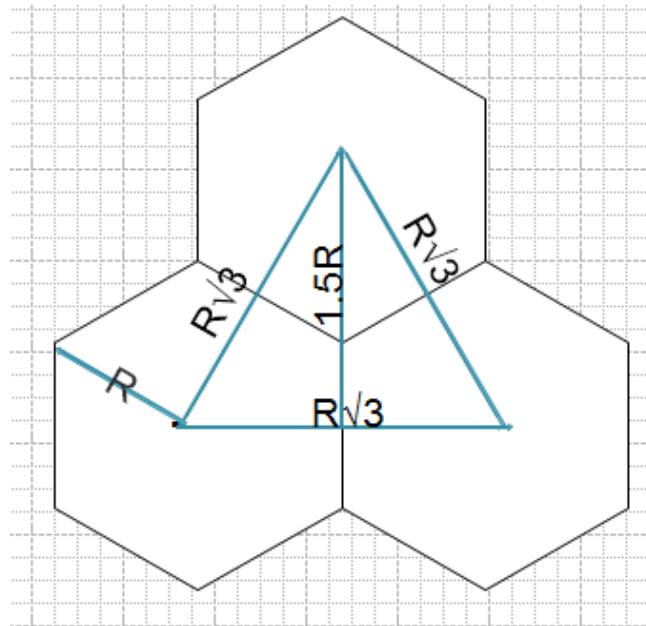


FIGURE A.1: Diagram of Hexagonal cell in terms of distance

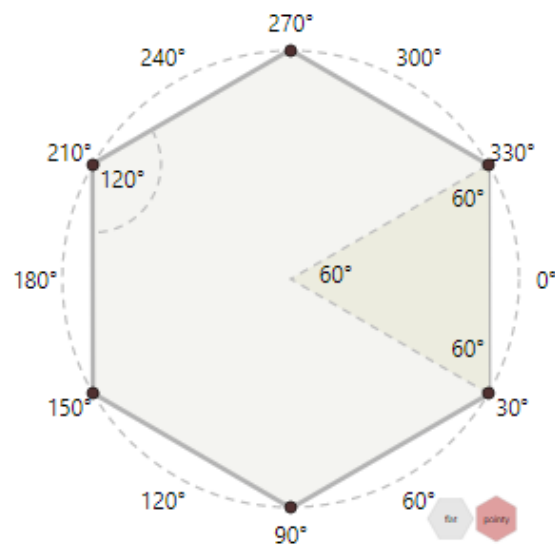


FIGURE A.2: Hexagonal cell in terms of angular domain[49]