Republic of Iraq
Ministry of Higher Education
and Scientific Research
University of Babylon
College of Engineering
Electrical Engineering Department

Channel Estimation Enhancement of Massive Multiple Input Multiple Output Wireless Communication System for Ideal and non-Ideal Hardware Transceivers

A Thesis

Submitted to the College of Engineering of the University of Babylon in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Engineering / Electrical Engineering / Electronics and Communications

By:
Ahmed Hussein Shatti Hamza

Supervised by:
Prof. Dr. Ehab A. Hussein

2023 A.D. 1444 A.H.

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

﴿ قَالُوا سُبْحَانَكَ لَا عِلْمَ لَنَا إِلَّا مَا

عَلَّمْتَنَا إِنَّكَ أَنتَ الْعَلِيمُ الْحَكِيمُ ﴾

صَدَقَ اللَّهُ الْعَلِيّ الْعَظِيم

البقرة (23)

Dedication

This work is dedicated to:
my mother and the memory of my father
my family, my wife and
my friends with love and respect

Acknowledgment

In the name of Allah, the Most Gracious and the Most Merciful First and above all, I praise ALLAH, the almighty, for providing me with this opportunity and granting me the capability to proceed with this thesis.

I want to express my thanks, respect, and gratitude to my supervisor “Prof. Dr Ehab A. Hussein”, who stood with me throughout the research and supported me in difficult times.

I warmly thank my parents, mother, and brothers for their material and spiritual support in all aspects of my life. I want to thank my wife for supporting and encouraging me during the study period.

Also, I would like to thank my second family in Electrical Engineering for their spiritual support during my PhD journey.

Ahmed Hussein Shatti AL-Isawi

26/June/2023

Supervisor's Certification

I certify that this thesis, entitled "Channel Estimation Enhancement of Massive Multiple Input Multiple Output Wireless Communication System for Ideal and non-Ideal Hardware Transceivers)", and submitted by the student "Ahmed Hussein Shatti" was prepared under my supervision at the Department of Electrical Engineering, College of Engineering, University of Babylon, as a part of the requirements for the Doctoral degree of Philosophy in Engineering / Electrical Engineering / Electronics and Communications.

Signature:
Supervisor: Prof. Dr. Ehab A. Hussein
Date: / /2023

I certify that the thesis mentioned above has been completed in Electrical Engineering at the College of Engineering at the University of Babylon.

Signature:
Head of Department: Prof. Dr. Qais Kareem Omran
Date: / /2023

Examining Committee's Certificate

We certify that we have read this thesis entitled "Channel Estimation Enhancement of Massive Multiple Input Multiple Output Wireless Communication System for Ideal and non-Ideal Hardware Transceivers)" and, as an examining committee, examined the student, "Ahmed Hussein Shatti", in its contents and that, in our opinion, it meets the standard of a thesis for the degree of Doctoral Philosophy in Engineering / Electrical Engineering / Electronics and Communications.

Signature:
Name: Prof. Dr. Samir Jasim Mohammed
(Chairman)
Date: / /2023

Signature:
Name: Prof. Dr. Bayan M. Sabbar
(Member)
Date: / /2023

Signature:
Name: Prof. Dr. Osama Qasim Jumah
(Member)
Date: / /2023

Signature:
Name: Prof. Dr. Saad Saffah Hassoon
(Member)
Date: / /2023

Signature:
Name: Prof. Dr. Ahmed A. Hamad
(Member)
Date: / /2023

Signature:
Name: Prof. Dr. Ehab A. Hussein
(Supervisor)
Date: / /2023

Approval of Head of Department
Signature:
Name:Prof. Dr. Qais Kareem Omran
Date: / /2023

Approval of the Dean of College
Signature:
Name:Prof. Dr. Laith Ali Abdul-Rahaim
Date: / /2023

List of Publications

1. A. H. Shatti and E. A. Hussein, ”Low-complex Bayesian estimator for imperfect channels in the massive multi-input multi-output system,” International Journal of Electrical and Computer Engineering (2088-8708), vol. 12, no. 6, 2022.
2. A. H. Shatti and E. A. R. Hussein, ”Massive multiple-input multiple-output channel estimation under hardware and channel impairments,” Bulletin of Electrical Engineering and Informatics, vol. 12, no. 1, pp. 215-223, 2023.
3. IEEE Accepted paper, A. H. Shatti and E. A. R. Hussein, “Massive MIMO Effective Channel Estimation Models with Nonlinear distortion and based-Deep Learning approach” in the 5th International Conference on Information Technology, Applied Mathematics and Statistics (ICITAMS 2023).

Abstract

Massive Multiple Input Multiple Output (MIMO) is one of the most cutting-edge technologies for the current and future generations of the wireless communication system. The reason behind that is the large number of antennas (M = 64 or more) at the Base Station (BS), which can potentially increase the network's data rate in terms of Spectral Efficiency (SE). Hence, this technique permits spatial multiplexing of dozens of User Equipments (UEs) in the network, making it more difficult to estimate the uplink channels accurately. However, the hardware impairments on each antenna's Radio Frequency (RF) chain will pose another challenge in channel estimation. It tends to impair the received signals and results in inaccurately estimated channels. Motivated by these facts, this thesis presents three channel estimation models that consider the hardware quality at the BS and UEs.

The first proposed channel estimation model considers ideal hardware on both the BS and UEs and leverages the spatial correlation characteristics of the uplink channels to reduce their computational complexity at the Bayesian Minimum Mean Squared Error (MMSE) estimator by truncating their full rank covariance matrices based on their eigenstructure decomposition. Two channels are used: the local scattering channel with a completely known covariance matrix at the BS and the practical channel scenario, where the BS lacks knowledge of the channel statistics. The results of this proposed model show a comparable channel quality in both cases while the computational complexity has reduced to about 30% of the calculations used at the Bayesian estimation method.

The second proposed model considers the non-ideal hardware transceivers with residual impairments and estimates channels in practical environments exploiting Bussgang's model to describe these impairments, then uses a convex optimization procedure to regularize the practical estimated channels and reduce

the error floor of these impairments. The results of this procedure are validated with the local scattering channels by measuring the normalized mean squared error (NMSE) versus the effective signal-to-noise ratio. The results are gotten in MATLAB R2020a and show a significant reduction (nearly by one order of magnitude) in the error floor compared to the conventional one, especially at a high signal-to-noise ratio (SNR) range between 20dB to 30dB.

The third estimation model investigates the effect of hardware nonlinearity issues on the estimated uplink channel in the Rician Fading environment. Third-order polynomials are used to simulate the joint impact of the nonlinearity on the BS and UEs hardware. Then, the nonlinearity distortions are learned using deep learning scenarios and leveraged to estimate the effective channels. Different deep learning models (based on models developed in the state-of-the-art literature in this thesis) are trained with different optimization algorithms, loss functions, modulating symbols, and polynomial types using the TensorFlow environment in Python version 3.9 installed on a Laptop with a graphic card of NVIDIA RTX2070 MAX-Q, 16GB RAM, and Core i7-10750H processor. The analytical results of the deep learning procedure are compared with the state-of-the-art estimators from Bayesian, which are either aware or unaware of the system's distortions. The simulation results demonstrate that the deep learning approach is not always improving the estimation quality, and under some optimization algorithms, it is the worst. Extremely importantly, it is found that the Adam optimizer with the Gaussian sending symbols and real coefficients memoryless polynomial showed a significant performance gain of about 6dB less in the NMSE than Bayesian estimators.

Table of Contents

List of Figures **XII**

List of Tables **XVIII**

List of Abbreviations **XIX**

List of Symbols **XXIII**

Chapter 1: Introduction **1**

1.1 History Glance and Motivation 1

1.2 Literature Review 3

1.2.1 Channel estimation with ideal transceivers 3

1.2.2 Channel estimation with non-ideal transceivers (Residual impact) 4

1.2.3 Channel estimation with non-ideal transceivers (Non-linear impact) 5

1.3 Problem Statements 6

1.4 Thesis Aims 7

1.5 Thesis Contribution 8

1.6 Thesis Organization 8

Chapter 2: Background Theory **10**

2.1 Massive MIMO Definition 10

2.1.1 Propagation channels and coherence block 11

2.1.2 Spatial channel correlations 13
2.1.3 Uplink transmission model 14
2.1.4 Spatial correlations impact 16
2.1.5 Local scattering model 18
2.2 Channel Estimation Concepts 21
2.2.1 Pilot uplink transmission model 21
2.2.2 Pilot book design 23
2.2.3 Bayesian MMSE channel estimate 24
2.2.4 Statistical properties of channel estimates 26
2.2.5 Correlation of channel estimates 27
2.2.6 Spatial correlations impact on channel estimation 28
2.2.7 Practical channel model 30
2.3 Hardware Impact on Channel Estimation 33
2.3.1 Residual impairments model (Bussgangs' model) 34
2.3.2 Uplink transmission model with residual impairments . 35
2.3.3 Bayesian channel estimates with residual impairments . 36
2.3.4 Error floor of residual impairments of single user channel 38
2.3.5 Non-linear hardware distortion 39
2.3.6 Uplink transmission model with nonlinear distortion . . 40
2.3.7 Bayesian channel estimates with nonlinear distortion . . 43
Chapter 3: Proposed Models **45**
3.1 Introduction 45
3.2 Part I: Channel Estimation with Ideal Transceivers 46
3.3 Proposed Model-1 46
3.3.1 Using local scattering channel model 47

3.3.2 Using practical channel model 49
3.3.3 Computational complexity 50
3.4 Part II: Channel Estimation with non-Ideal Transceivers 53
3.5 Proposed Model-2 53
3.5.1 Algorithm steps 54
3.6 Proposed Model-3 56
3.6.1 Deep learning based-estimation model 56
Chapter 4: Results and Discussions **60**
4.1 Results and Analysis of Proposed Model-1 60
4.1.1 Using local scattering channel model 60
4.1.2 Using practical channel model 68
4.1.3 Unique results 72
4.1.4 Computational complexity results 74
4.2 Results and Analysis of Proposed Model-2 75
4.2.1 Performance evaluation of practical channel estimation . 75
4.3 Results and Analysis of Proposed Model-3 78
4.3.1 Optimization algorithm and loss function effect 78
4.3.2 Modulating symbol and polynomial type effect 86
4.3.3 Number of users (K) effect 91
4.3.4 Comparison works 97
Chapter 5: Conclusions and Future Works **98**
5.1 Conclusions 98
5.2 Future Works 100
References **101**

Appendix A: Basic Fundamentals and Definitions **110**

A.1 Signal-to-Noise Ratio 110

A.1.1 Transmit power 110

A.1.2 Channel gain 110

A.1.3 Noise power 111

A.2 Complex Random Vector 112

A.2.1 Circular-symmetry Gaussian complex vector 112

A.2.2 PDF of circular-symmetry Gaussian complex vector . . 113

Appendix B: Principles of Cellular Network **114**

B.1 Cellular Network Definition 114

B.2 Spectral Efficiency (SE) Definition 116

B.3 Improving Methods of Uplink SE 119

B.3.1 Increasing UEs' signal power 121

B.3.2 Using array of antennas at the BS 124

B.3.3 Increasing number of UEs per cell 130

B.4 Channel State Information (CSI) 135

B.4.1 Pilot signaling 135

B.4.2 Time Division Duplex (TDD) and reciprocity 136

Appendix C: Principles of Estimation Theory **137**

C.1 Parameter Estimation 137

C.1.1 MMSE estimator of nonzero mean complex Gaussian vector 138

C.1.2 MMSE estimator of zero mean complex Gaussian vector 138

C.1.3 Linear MMSE estimator 140

C.1.4 Law of large numbers 141

List of Figures

2.1 Coherence block with T_c and B_c division 12

2.2 Uplink transmission in massive MIMO system 15

2.3 Local scattering model for spatially correlated channels 18

2.4 General block diagram of the uplink transmission system 42

3.1 Block diagram of the ideal hardware system 45

3.2 Block diagram of the non-ideal hardware system 45

3.3 Proposed model-1 of channel estimation 47

3.4 The deep neural network for estimating the effective channel elements. 57

3.5 The deep neural network for estimating the distortion correlation elements 57

4.1 NMSE as a function of effective SNR of spatial-rank=1 and full-rank=100 channels with Gaussian ASD=10^o and M=1, 10, 64, and 100 61

4.2 NMSE as a function of effective SNR of spatial-rank=10 and full-rank=100 channels with Gaussian ASD=10^o and M=1, 10, 64, and 100 62

4.3 NMSE as a function of effective SNR of spatial-rank=20 and full-rank=100 channels with Gaussian ASD=10^o and M=1, 10, 64, and 100 63
4.4 NMSE as a function of effective SNR of spatial-rank=40 and full-rank=100 channels with Gaussian ASD=10^o and M=1, 10, 64, and 100 63
4.5 NMSE as a function of effective SNR of spatial-rank=64 and full-rank=100 channels with Gaussian ASD=10^o and M=1, 10, 64, and 100 64
4.6 NMSE as a function of ASD, for full-rank=100 and spatial rank=1 using Gaussian, Uniform, and Laplace ASD distributions, M = 100 and SNR=10dB 65
4.7 NMSE as a function of ASD, for full-rank=100 and spatial rank=10 using Gaussian, Uniform, and Laplace ASD distributions, M = 100 and SNR=10dB 66
4.8 NMSE as a function of ASD, for full-rank=100 and spatial rank=20 using Gaussian, Uniform, and Laplace ASD distributions, M = 100 and SNR=10dB 66
4.9 NMSE as a function of ASD, for full-rank=100 and spatial rank=40 using Gaussian, Uniform, and Laplace ASD distributions, M = 100 and SNR=10dB 67
4.10 NMSE as a function of ASD, for full-rank=100 and spatial rank=64 using Gaussian, Uniform, and Laplace ASD distributions, M = 100 and SNR=10dB 67

4.11 NMSE as a function of ASD, for full-rank=100 and spatial rank=64 using Gaussian, Uniform, and Laplace ASD distributions, $M = 100, 64$ and SNR=10dB 68
4.12 NMSE as a function of sample numbers of practical channels with full rank=100 and successive spatial ranks r=20, 30, SETSVD=35 and optimal-SVHT=48, $M = 100$, Gaussian ASD= 10^o, and effective SNR=10dB 69
4.13 NMSE as a function of effective SNR of practical channels with full rank=100 and successive spatial ranks r=20, 30, SETSVD=35 and optimal-SVHT=48, $M = 100$, Gaussian ASD= 10^o, and effective SNR=10dB 71
4.14 Eliminating and keeping eigenvalues using the proposed SETSVD spatial rank 71
4.15 Eliminating and keeping eigenvalues using the optimal SVHT spatial rank 72
4.16 NMSE as a function of sample numbers of practical channels with full rank=100 and spatial ranks SETSVD=64 and optimal-SVHT=48, $M = 100$, Gaussian ASD= 10^o, and effective SNR=10dB 73
4.17 NMSE as a function of effective SNR of practical channels with full rank=100 and spatial ranks SETSVD=64 and optimal-SVHT=48, $M = 100$, Gaussian ASD= 10^o, and effective SNR=10dB 73
4.18 The normalized MSE of the channel estimation using the local scattering model with equal hardware quality at the transceiver ends. 75

4.19 The normalized MSE of the practical channel estimation using the sample covariance with equal hardware quality at the transceiver ends. 77
4.20 The normalized MSE of the optimized channel estimation using the sample covariance with equal hardware quality at the transceiver ends. 77
4.21 Effective channel estimates, loss: mean squared error, optimizer: Adam 80
4.22 Effective channel estimates, loss: mean squared error, optimizer: RMSprop 80
4.23 Effective channel estimates, loss: mean squared error, optimizer: AdaDelta 81
4.24 Effective channel estimates, loss: Huber, optimizer: Adam . . . 81
4.25 Effective channel estimates, loss: Huber, optimizer: RMSprop . 82
4.26 Effective channel estimates, loss: Huber, optimizer: AdaDelta . 82
4.27 Distortion correlation estimates, loss: mean squared error, optimizer: Adam 83
4.28 Distortion correlation estimates, loss: mean squared error, optimizer: RMSprop 83
4.29 Distortion correlation estimates, loss: mean squared error, optimizer: AdaDelta 84
4.30 Distortion correlation estimates, loss: Huber, optimizer: Adam . 84
4.31 Distortion correlation estimates, loss: Huber, optimizer: RMSprop 85
4.32 Distortion correlation estimates, loss: Huber, optimizer: AdaDelta 85
4.33 Effective channel estimates, Gaussian, quasi-memoryless 87

4.34 Effective channel estimates, QPSK, quasi-memoryless 87
4.35 Distortion correlation estimates, Gaussian, quasi-memoryless . 88
4.36 Distortion correlation estimates, QPSK, quasi-memoryless . . . 88
4.37 Effective channel estimates, Gaussian, memoryless 89
4.38 Effective channel estimates, QPSK, memoryless 89
4.39 distortion correlation estimates, Gaussian, memoryless 90
4.40 distortion correlation estimates, QPSK, memoryless 90
4.41 Effective channel estimates, Gaussian, quasi-memoryless, Adam optimizer, MSE loss function, K=20 92
4.42 Distortion correlation estimates, Gaussian, quasi-memoryless, Adam optimizer, MSE loss function, K=20 92
4.43 Effective channel estimates, Gaussian, quasi-memoryless, Adam optimizer, Huber loss function, K=20 93
4.44 Distortion correlation estimates, Gaussian, quasi-memoryless, Adam optimizer, Huber loss function, K=20 93
4.45 Effective channel estimates, Gaussian, memoryless, Adam optimizer, MSE loss function, K=30 94
4.46 Distortion correlation estimates, Gaussian, memoryless, Adam optimizer, MSE loss function, K=30 94
4.47 Effective channel estimates, QPSK, quasi-memoryless, Adam optimizer, MSE loss function, K=10, using Large DNN 95
4.48 Distortion correlation estimates, QPSK, quasi-memoryless, Adam optimizer, MSE loss function, K=10, using Large DNN 95
4.49 Effective channel estimates, QPSK, quasi-memoryless, Adam optimizer, MSE loss function, K=10, using Large dataset 96

4.50 Distortion correlation estimates, QPSK, quasi-memoryless, Adam optimizer, MSE loss function, K=10, using Large dataset 96

B.1 Cellular network topology 114

B.2 A general memoryless discrete channel 116

B.3 Single-Input Single-Output (SISO) system 117

B.4 Single-Input Single-Output (SISO) system with interference . . 118

B.5 Wyner model 120

B.6 Uplink SE as a function of SNR 123

B.7 LoS propagation-SIMO channel 125

B.8 NLoS propagation-SIMO channel 126

B.9 Uplink SE as a function of the number of antennas 128

B.10 Uplink sum SE as a function of the number of UEs using MR combining 133

B.11 Uplink sum SE as a function of the number of UEs M-MMSE combining 134

B.12 Pilot signalling from single-antenna to multiple antennas 135

B.13 Time Division Duplex (TDD) protocol 136

List of Tables

3.1 DNNs and single-cell scenario parameters 59

4.1 Standard and proposed Computational Complexity (CC) comparison for the MMSE estimator with successive spatial ranks . 74

4.2 DNN convergence: effective channel estimate case, $K = 10$, QPSK 79

4.3 DNN convergence: : distortion correlation estimate case, $K = 10$, QPSK 79

4.4 Comparison with Proposed Model-1 97

4.5 Comparison with Proposed Model-2 97

List of Abbreviations

3GPP	Third Generation Partnership Project
ADC	Analog-to-Digital Converter
ASD	Angular Standard Deviation
AWGN	Additive White Gaussian Noise
BPSK	Binary Phase-Shift Keying
BS	Base Station
CC	Computational Complexity
CDF	Cumulative Distribution Function
CSI	Channel State Information
dB	decibel
DFT	Discrete Fourier Transform
DA-LMMSE	Distortion-Aware LMMSE estimator
DuA-LMMSE	Distortion-unAware LMMSE estimator
EW-MMSE	Element-Wise MMSE
FBMC	Filter Bank Multi-Carrier
FDD	Frequency-Division Duplex
FLOPS	Floating Point Operations
GSM	Global System for Mobile Communications
i.i.d.	Independent and Identically Distributed
I/Q	In-Phase/Quadrature

LMMSE	Linear MMSE
LoS	Line-of-Sight
M-MMSE	Multicell Minimum Mean-Squared Error
MIMO	Multiple-Input Multiple-Output
MMSE	Minimum Mean-Squared Error
MR	Maximum Ratio
MSE	Mean-Squared Error
NLoS	Non-Line-of-Sight
NMSE	Normalized MSE
OFDM	Orthogonal Frequency-Division Multiplexing
QPSK	Quadrature Phase-Shift Keying
QoS	Quality of Service
SE	Spectral Efficiency
SETSVD	Successive Empirical Threshold SVD
SISO	Single-Input Single-Output
SIMO	Single-Input Multiple-Output
SNR	Signal-to-Noise Ratio
SVD	Singular Value Decomposition
SVHT	Singular Value Hard Threshold
TDD	Time Division Duplex
UE(s)	User Equipment(s)
UL	Uplink

List of symbols

The notations for this thesis are as follows: Matrices are represented using upper-case boldface letters (e.g., $\mathbf{X}, \mathbf{Y}$), while column vectors are denoted with lower-case boldface letters (e.g., $\mathbf{x}, \mathbf{y}$). Scalars are denoted using lower or upper-case italic letters (e.g., *x, y, X, Y*), and sets are represented using calligraphic letters (e.g., $\mathscr{X}, \mathscr{Y}$).

$\mathbb{C}^{N\times M}$	The set of complex-valued $N \times M$ matrices
$\mathbb{C}^{N}$	Short forms of the vector $\mathbb{C}^{N\times 1}$
$\mathbb{C}_{\mathscr{N}}(\mathbf{0},\mathbf{R})$	The circularly symmetric complex Gaussian with mean and covariance matrix $\mathbf{R}$
$\mathbb{C}_{\mathscr{N}}(\mathbf{x},\mathbf{R})$	The generalization of the circularly symmetric complex Gaussian
$\mathrm{diag}(\cdot)$	$\mathrm{diag}(x_1,\ldots,x_N)$ is a diagonal matrix with the scalars $x_1,\ldots,x_N$ on the diagonal,
$\det(\mathbf{X})$	Determinant of a square matrix $\mathbf{X}$
e	Euler's number ($e \approx 2.718281828$)
$\mathbb{E}\{x\}$	The expectation of a random variable x
$\mathbf{I}_M$	The $M \times M$ identity matrix
j	The imaginary unit

$\mathcal{N}(\mathbf{x}, \mathbf{R})$	The real Gaussian distribution with mean $\mathbf{x}$ and covariance matrix $\mathbf{R}$
$\mathbb{R}^N$	Short forms of the vector $\mathbb{R}^{N\times 1}$
$\mathbb{R}^{N\times M}$	The set of real-valued $N \times M$ matrices
$\text{rank}(\mathbf{X})$	Rank of $\mathbf{X}$ (i.e., number of non-zero singular values)
$\text{tr}(\mathbf{X})$	Trace of a square matrix $\mathbf{X}$
$U[a, b]$	Uniform distribution between a and b
$\mathbb{V}\{x\}$	The variance of a random variable x
$\{x \in \mathscr{S} : P\}$	The subset of $\mathscr{S}$ containing all members that satisfy a property P
$[\mathbf{x}]_i$	The ith element of a vector $\mathbf{x}$
$[\mathbf{X}]_{ij}$	The (i, j) th element of a matrix $\mathbf{X}$
$\mathbf{X}^\star$	The complex conjugate of $\mathbf{X}$
$\mathbf{X}^{\text{T}}$	The transpose of $\mathbf{X}$
$\mathbf{X}^{\text{H}}$	The conjugate transpose of $\mathbf{X}$
$\mathbf{X}^{-1}$	The inverse of a square matrix $\mathbf{X}$
$\mathbf{X}^{\frac{1}{2}}$	The square-root of a square matrix $\mathbf{X}$
$\lvert x \rvert$	Absolute value (or magnitude) of a scalar variable x
$\lVert \mathbf{x} \rVert$	The L_2-norm $\lVert \mathbf{x} \rVert = \sqrt{\sum_i \lvert [\mathbf{x}]_i \rvert^2}$ of a vector $\mathbf{x}$
$\lVert \mathbf{X} \rVert_F$	The Frobenius norm $\lVert \mathbf{X} \rVert_F = \sqrt{\sum_{i,j} \left\lvert [\mathbf{X}]_{ij} \right\rvert^2}$ of $\mathbf{X}$
$\lVert \mathbf{X} \rVert_2$	The spectral norm of $\mathbf{X}$
$\Re(x)$	Real part of x
$\Im(x)$	Imaginary part of x
0_M	The $M \times 1$ matrix (i.e., vector) with only zeros
$0_{N\times M}$	The $N \times M$ matrix with only zeros

Chapter 1

Introduction

1.1 History Glance and Motivation

Since the time of wired communication between any two predefined locations, the modern communications era has revolutionized the way in which terminal users communicate. Many services like smartphones, internet connections, and computers are wirelessly connected nowadays, thanks to the development of cellular network generations, local area networks, and satellite communication services. Behind the various applications of wireless connectivity, the need for such technologies becomes more urgent and vital to modern societies. The on-demand streaming media, augmented reality applications, home automation, and machine-to-machine communications are the most significant application examples that have already been witnessed today. Looking ahead to the coming fifteen years, new innovations in wireless connectivity may be found, which cannot be predicted at this time.

The cellular network is mainly designed to provide voice communication, and even nowadays, it is used to serve data transmission like on-demand video transmission, which increases the traffic data in the network. The aim of delivering reliable wireless services with a high data rate was born earlier, since the time of the Bell Labs postulations in 1947. These postulations were the seminal idea of the cellular network, where each cell has a Base Station (BS) with a fixed location and serves a set of User Equipments (UEs). Since Marconi's first wireless transmission invention in 1895, the researcher Martin Copper [1] has noticed that there is 32% annual growth in the rate of wireless connections for voice and data communication. This number doubled every 2.5 years and is

denoted by Copper's Law. Looking forward, the Ericsson Mobility Report has prepared a faster annual growth rate of 42% for the traffic load in mobile data between 2016 and 2022 [2]. Without any doubt, the demand for wireless connectivity will continue to grow in the near future as more and more electronic devices from social networks are connected. Two crucial questions are raised here. The first question is how the current wireless connectivity technology is harnessed to serve the on-time increasing demand and avoid the coming trouble in traffic data. Another critical question is how to investigate the Quality of Service (QoS) to the customers under such huge connections. Customers expect to receive constant QoS everywhere and at any time. Many industrial companies and academic researchers are working on designing a new era of wireless communication to keep up with this growth in connections.

The massive MIMO technology is one of the most promising technologies that is developed nowadays to serve many device connections and handles many orders of magnitude more than conventional wireless technologies [3, 4, 5, 6, 7]. This technology has significantly improved the traditional MIMO system by employing large-scale antennas at the base station and permitting spatial multiplexing of dozens of users in the network. Although this MIMO technology provides big advantages over the conventional MIMO system, it comes with many challenges. Accurate channel estimation for every UE in the cell is one of the most challenging problems that should be carefully considered.

However, the non-ideal hardware transceivers of the BS and UE will pose another challenge for channel estimation and result in inaccurately predicted channels in the network, especially in the practical environment. Most academic research assumes ideal hardware on each transmitter and receiver device and characterizes it as a linear filter. That is, the passband signal of the transmitter will be correctly modulated from its complex baseband samples, while the receiver will demodulate and sample its signal synchronously with the transmitter. However, in practice, the transceiver hardware is always non-ideal [8], and such classical operations in the ideal system are not satisfied.

This thesis focuses on the channel estimation models in the uplink (UL) direction (the transmission from UEs to the BS) of the massive MIMO system in Rayleigh and Rician fading environments. The present work discusses three

Bayesian channel estimation models considering the ideal and non-ideal hardware quality, where each model has a specific problem as follows: The first channel estimation model considers the computational complexity problem of the Bayesian MMSE estimate, while the inevitable error floor problem caused by transceivers' residual hardware impairments is considered in the second proposed model. The third channel estimation model deals with the nonlinear distortion effect by both transceivers of the BS and UEs and uses the deep learning approach to estimate the effective channels and then compares it with the state-of-the-art Bayesian estimators. The deep neural networks are trained with different optimization algorithms, loss functions, symbol modulations, and with two types of distorting polynomials.

1.2 Literature Review

1.2.1 Channel estimation with ideal transceivers

In almost all literature on massive MIMO systems with ideal hardware, perfectly known channel covariance matrices are often assumed [9, 10, 11, 12], which is debatable because in practice unknown large-dimensional covariance matrices exist [13], and result in higher computational complexity of the MMSE Bayesian estimator at the BS. Hence, promising approaches have been suggested by many researchers to estimate such practical large-dimensional matrices with lower computational complexity as follows:

In 2004, O. Ledoit et al., [14] proposed a very well estimator for large-scale covariance matrices, which would be more accurate than the usual estimator based on the sample covariance matrix method. The proposed estimator regularizes the covariance matrices asymptotically using the linear combination between the sample covariance matrix and the identity matrix.

In 2008, X. Mestre et al., [15] considered eigenstructure estimating of the covariance matrices using some tools from the random matrix theory. The proposed estimator has been consistent with the statistical covariance matrix size

and observation number.

In 2014, N. Shariati et al., [16] proposed new Polynomial Expansion Channel (*PEACH*) estimators from the Bayesian MMSE in second-order complexity. The proposed procedure has achieved near-optimal NMSE with a low polynomial degree and derived accurate and lower-complexity computations for the Bayesian estimator MMSE based on polynomial floating point operations reduction.

In 2016, E. Bjornson et al., [17] estimated the channels in a practical environment when the BS has imperfect knowledge about the statistical information of its received channels in the uplink direction. The proposed method proceeded by regularizing the estimated covariance matrices for an accurate channel estimate.

In 2018, Xiao Wei et al., [18] proposed a novel decomposition method called the factor analysis that decomposes the practical channel covariance matrix space into three interference-free subspaces to estimate the channels accurately and with low computational complexity.

In 2022, X. Zhang et al., [19] proposed a low-complexity MIMO detection method based on the Neumann series (NS). They leverage the statistical information of the Gram matrix and matrix-vector multiplication operations to compute an initialization matrix, which guarantees the convergence of NS expansion. The computational complexity of MIMO signal detection has decreased from $\mathcal{O}(BU^2+U^3)$ to $\mathcal{O}(BUN)$, where B, U, and N are the numbers of antennas at the base station, user equipment, and NS terms, respectively.

1.2.2 Channel estimation with non-ideal transceivers (Residual impact)

In practice, the transceiver hardware is always non-ideal, and many issues like nonlinear distortion in amplifiers, quantization error in Analog to Digital Converters (ADCs), and in-phase/quadrature imbalance in mixers exist. Most of the literature has taken a cost-quality trade-off policy when designing the system with M antennas at the BS since the cost of realizing such a system is M times larger than that needed for the conventional system with one antenna. Hence, lowering the quality of hardware components with a compensation technique to reduce the cost has interested many researchers, as follows:

In 2014, U. Gustavsson et al. [20] discussed the possibility of using cheap, compact, and lower power-consuming transceivers at the BS of the massive MIMO system. On the other hand, they simulated the hardware impairments with a well-simplified statistical model.

In 2014, E. Bjornson et al., [21] considered the hardware impairments incorporated between the BS transceivers and the UE and showed a finite ceiling on the accuracy of the estimated channels of each UE behind the existing impairments.

In 2017, S. Jacobsson et al., [22] proposed a novel channel estimation based on the Bussgang decomposition under the absence of the CSI information at the BS transceivers. They assumed low-resolution ADCs at the BS and quantified using 1-bit ADC with combining schemes, such as zero-forcing and maximum ratio, to achieve a reliable channel estimation in the multi-user MIMO system.

In 2018, C. Mollen et al., [23] analyzed the nonlinearity effects of the inexpensive amplifiers at the BS using a polynomial model. The proposed procedure derived the spatial correlation for the third-order dominant effect of the distortions at the BS antennas.

In 2018, Q. Zhang et al., [24] considered different types of hardware impairments with Rician channels in a multi-cell scenario. The proposed framework analyzes the network's achievable rate in the uplink direction.

1.2.3 Channel estimation with non-ideal transceivers (Nonlinear impact)

As mentioned earlier, different hardware issues have already existed in practical transceivers and tend to impair the received RF signals. However, the nonlinearity impact caused by amplifiers is the most detrimental issue that affects the signals in the massive MIMO system. Some of the literature has described the nonlinearity distortions using a stochastic additive representation [21, 25, 26], while some other papers model these impairments in assistance with the behavioral models for better describing these types of distortions as follows:

In 2002, R. Raich et al., [27] modeled the nonlinearity effects of the power amplifier as a memoryless, quasi-memoryless, and memory behavioral model depending on the long, short, or no memory response of the power amplifier.

In 2015, Y. Zou et al. [28] analyzed the power amplifier units' nonlinear effect on the BS's downlink signal using channel-aware combining schemes of a single-antenna UE.

In 2018, S. Jacobsson et al., [29] analyzed various hardware impairments like amplifier nonlinearities, local oscillator phase noise, and resolution error of ADCs with the aid of Bussgang decomposition. The combined effect of these distortions is considered for the BS-based OFDM receivers in a multiuser scenario.

In 2019, D. Rönnow et al., [30] rederived the Bussgang theorem for Gaussian signals in multiband transmitters, taking into account the effect of some parameters like observation bandwidth and correlated and uncorrelated input signals. The attenuation matrix, nonlinear distortion, and concurrent multiband are analyzed.

In 2019, S. R. Aghdam et al., [31] proposed a multiple-input, single-output (MISO) linear precoding scheme with a nonlinear power amplifier at the BS. The hardware impairments are analyzed using the Bussgang theorem and projected gradient descent.

In 2019, M. Cherif et al., [32] analyzed the performance of the massive MIMO system in the presence of the nonlinear effect under Rayleigh channel fading. Their proposed model is implemented with different receivers like zero-forcing (ZF), MMSE, and least squares (LS).

In 2022, Ö. Tugfe Demir et al., [33] analyzed the joint nonlinear impact of the hardware on the BS and UEs and derived a model for the effective uplink channels in the Rician scenario. Their work compares the Bayesian estimates with proposed deep-learning estimates for the derived uplink channels.

1.3 Problem Statements

Although the massive MIMO system provides big advantages over conventional MIMO systems, it is clear from the previous literature that the unknown large-dimensional covariance matrices and the different hardware impairments will pose challenging problems in estimating accurate channels in the massive

MIMO system. The following list will summarise these problems as follows:

1. Since the MMSE Bayesian estimator comes with high computational complexity, especially when it applies to the BS of the massive MIMO system with large-scale antennas, reducing such large computations while keeping the same quality of the estimated channels will be very efficient at the BS in terms of reducing the Floating Point Operations (FLOPs).

2. Since the residual impairments in the transceivers' hardware cause an inevitable error floor in the estimated channels; hence, reducing such error floor while serving the same quality of the estimated channels with low-quality hardware components at the transceivers is the main aim to save the cost of building such large system.

3. Since when the nonlinear effect of the hardware transceivers exists on the BS and UEs, deriving equations of the conventional Bayesian estimators will be very messy and computationally intensive; hence, applying the deep learning techniques as a channel estimation-based approach will be very effective to learn these distortions and leveraging it to estimate the uplink channels.

1.4 Thesis Aims

The aims of this thesis can be listed as follows:

1. Lowering the computational complexity of the Bayesian MMSE channel estimate by leveraging the spatial correlations characteristics of the uplink channels and exploiting the Singular Value Decomposition (SVD) method to factor their covariance matrices into their eigenstructure and truncating the strongest eigenvalues (strongest paths) then evaluating the MMSE estimator's floating point operations (FLOPs) of the full-rank and reduced-rank channels.

2. Reducing the error floor (caused by the residual impairments of the non-ideal hardware transceivers of the BS and UEs) for the practical estimated channels by regularizing the diagonal and off-diagonal elements of their

sample covariance matrices using convex optimization. On the other hand, applying the Bussgang decomposition to model the residual impairments at both transceivers.

3. Enhancing the channel estimation quality by exploiting the power of deep learning as an estimation approach to learn the transceivers' non-linear distortions of the BS and UEs and estimate the effective channels, and comparing the resultant estimated channels to the channels estimated by the state-of-the-art Bayesian estimators.

1.5 Thesis Contribution

The contributions of this thesis are as follows:

- Developing a channel estimation model with reduced antennas at the BS that has lower computational complexity and comparable performance to the M=100 antenna deployment.
- Developing a channel estimation model that shows the residual impact of the hardware distortions at the BS and UEs for practical uplink channels.
- Developing a deep learning-based-channel estimation model that shows the joint nonlinear impact of the hardware impairments at the transceivers of the BS and UEs.

1.6 Thesis Organization

This thesis is structured into four chapters after Chapter One as follows:

- **Chapter Two** presents the necessary background theory for this thesis. The definitions of the cellular network, the massive MIMO system, and the spatial correlation types of the fading channel are all discussed in detail. On the other hand, the essential underlines of the Bayesian channel estimation for the uplink channels are reviewed for massive MIMO systems in the case of the ideal and non-ideal hardware transceivers on the BS and UEs.

- **Chapter Three** presents the proposed models of channel estimation with ideal and non-ideal hardware transceivers. The first part of this chapter deals with the channel estimation model under ideal hardware components on the BS and UEs transceivers and deals with two types of channel models, the local scattering and the practical channel models. The second part of this chapter is subdivided into two main sub-parts that consider the channel estimation model with non-ideal hardware transceivers. The first sub-part analyzes the residual impact of the hardware impairments with the error floor problem, while the second sub-part uses the deep learning approach to learn the nonlinear distortion on the BS and UEs and estimate the effective channels.
- **Chapter Four** In this chapter, the results and discussions of the proposed models in chapter three are analyzed in detail.
- **Chapter Five** concludes the thesis work and suggests some future works.

Chapter 2

Background Theory

2.1 Massive MIMO Definition

Based on the concepts in Appendix (B), the highly efficient cellular network is the coverage tier network that provides a more achievable sum SE in [bit/sec/Hz/cell] and has the following characteristics [34]:

- It serves multiple UEs in each cell to provide an achievable multiplexing gain based on the SDMA scenario.
- It uses numerous antennas at the BS, more than the number of UEs in each cell, to achieve high degrees of freedom at the BS to separate between UEs and suppress the interference.
- It uses the TDD protocol to overcome the overhead problems and estimate the channel responses using only the uplink pilot signalling.

Depending on these guidelines, the massive MIMO network can be defined [35] as a cellular multi-carrier network with L cells and K UEs and a fixed j^{th} BS in each cell. The j^{th} BS is equipped with $M_j \gg 1$ antennas and operates in TDD mode with antenna-UE ratio $M/K > 1$ to communicate with K_j single-antenna UEs simultaneously during each sample on the coherence block. Each BS uses the Linear combining scheme to process the received signals in the uplink transmission or the linear precoding scheme to process the transmitted signals in the downlink transmission.

2.1.1 Propagation channels and coherence block

The wireless channel is a time and frequency-changing parameter and has a dispersive nature, i.e., the energy of the travelling signals will be spread out in space and received with longer symbol intervals than before. The bandwidth is characterised by the number of baseband symbols that can be transmitted through the channel. Hence, depending on the comparison between the time interval of the symbol and the dispersion time of the channel, there are two types of fading in the channel [36] as follows:

2.1.1.1 Flat fading channels

The channel of the flat fading is assumed to be constant during the entire transmission of symbols where a high symbol rate can be transmitted through this type of channel, but an inter-symbol-interference (ISI) will occur between the complex-valued symbols since the time dispersion of the channel will be larger than the time interval of each symbol.

2.1.1.2 Fast fading channels

The channel of the fast fading will change every time we transmit a symbol; i.e., only a low symbol rate can be transmitted through such channels. There is no ISI problem in this type of channel since the time interval of each sample is larger than the time dispersion of the channel.

In digital communication, large blocks of information need to be transmitted through the channels without ISI. The classic solution to this problem is to divide the channel's bandwidth into parallel flat-fading subchannels, each with sufficient rate and narrow bandwidth, so that the information can be transmitted without ISI and be decodable at the receiver. Different multi-carrier schemes have been used in the massive MIMO system, applying either the conventional OFDM [37] or another multicarrier scheme with a bank of filters (FBMC) [38].

2.1.1.3 Coherence time T_c and Coherence bandwidth B_c

In the wireless communication context, the coherence time T_c and coherence bandwidth B_c are very important parameters that describe the spectral and tem-

poral characteristics of the channel. The coherence time T_c is defined as the period over which the communication channel is characterized as stationary. The coherence bandwidth B_c is defined as the frequency range over which the communication channel exhibits frequency-selective fading [36].

2.1.1.4 Coherence block

It is defined as a multicarrier block with constant coherence time T_c, subdivided among complex-valued samples, and constant coherence bandwidth B_c, subdivided among subcarriers. Hence, the total samples per the coherence block will be given by $\tau_c = B_c T_c$ as shown in Fig. 2.1 [11]. The samples per coherence block can be subdivided as follows: For uplink pilot signalling, τ_p pilot samples are used, and for uplink and downlink data transmission, τ_u and τ_d are used, respectively. That is $\tau_c = \tau_p + \tau_u + \tau_d$, where the uplink and downlink samples are managed according to the network traffic status, while the number of pilot samples is a designing factor.

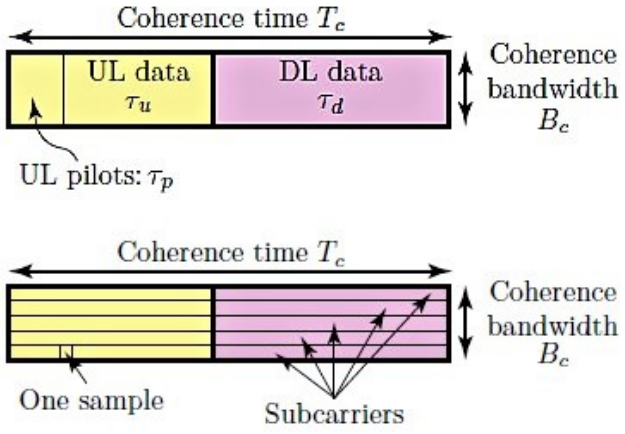

Figure 2.1: Coherence block with T_c and B_c division

Note that since each adjacent subcarriers in one coherence block are approximately equal and both have small time variations; hence, it is not needed to estimate the channel at each subcarrier and is estimated for one time only and regarded for all subcarriers. On the other hand, it is assumed that the subchannels in each coherence block are random and statistically independent with identical distributions to those subchannels in any other coherence block; hence, the description of such fading channels will be denoted by the stationary ergodic random process.

2.1.2 Spatial channel correlations

The channel responses that reach the BS are random variables and have two types of spatial correlation characteristics depending on the surrounding physical environment around the BS or the UEs. Suppose $\mathrm{h}_{lk}^{j} \in \mathbb{C}^{M_j}$ denotes the channel vector between the k^{th} user in cell l and the array antennas of the BS j in the uplink direction. In this case, the channel vector on the BS will be characterized by its magnitude and direction, which are also random variables. The dependency or not between the magnitude and direction of the channel vector at the BS will detect what the spatial correlation is, as follows:

- The channel response is said to be a spatially uncorrelated channel if the magnitude of the channel vector is independent of its direction, i.e., the BS will receive identical average channel gains of the multi-paths in all directions. The Rayleigh fading channel model in Appendix B in Eq.(B.20) is an example of spatially uncorrelated channels with rich scattering environments around the BS.

- The channel response is said to be a spatially correlated channel if the magnitude of the channel vector depends on its direction, i.e., the BS will receive strong channel gains in some directions and weak channel gains in others. The local scattering model [12], which will be described in Section 2.1.5, is an example of the spatially correlated channel model.

In practice, the channel responses are almost spatially correlated at the BS [39] due to the physical environment propagation, which makes some path directions with stronger gains than others. Hence, the spatially correlated channels can be modelled as follows:

$$\mathbf{h}_{lk}^{j} \sim \mathbb{C}_{\mathcal{N}}\left(0_{M_j}, \mathbf{R}_{lk}^{j}\right) \tag{2.1}$$

where the covariance matrix $\mathbf{R}_{lk}^{j} \in \mathbb{C}^{M_j \times M_j}$ defined as a positive semi-definite spatial correlation matrix [12]. For now, this covariance matrix is considered to be familiar (known) to the BS, but in practice, it should be estimated, as we will see in chapters 3 and 4. It describes the total large-scale propagation effect (i.e., the total channel gains on all the BS array antennas). Hence, the average

channel gain per each antenna can be given by

$$\beta_{lk}^{j} = \frac{1}{M_j} \mathrm{tr}\left(\mathbf{R}_{lk}^{j}\right) \tag{2.2}$$

where $\mathbf{tr}(.)$ denotes the trace function. Note that the covariance matrix of the spatially uncorrelated channel will be $\mathbf{R}_{lk}^{j} = \beta_{lk}^{j}\mathbf{I}_{M_j}$ which is a diagonal matrix. The eigenstructure of the covariance matrix $\mathbf{R}_{lk}^{j}$ describes the spatial characteristics of the channel response $\mathbf{h}_{lk}^{j}$, for example, the strongest eigenvalues in $\mathbf{R}_{lk}^{j}$ will determine the more likely spatial directions in $\mathbf{h}_{lk}^{j}$.

Remark 2.1.1 (Karhunen-Loeve Expansion of h) This remark shows the possibility of generating the channel vector $\mathbf{h}$ from the eigenstructure of $\mathbf{R}$ as follows: If $\mathbf{R} \in \mathbb{C}^{M\times M}$ is a positive semi-definite matrix, then it can be decomposed using the eigenvalue decomposition $\mathbf{R} = \mathbf{U}\mathbf{D}\mathbf{U}^{\mathrm{H}}$, where $\mathbf{U} \in \mathbb{C}^{M\times r}$ and $\mathbf{U}^{\mathrm{H}} \in \mathbb{C}^{r\times M}$ are the right and left eigenvectors of $\mathbf{R}$ respectively. $\mathbf{D} \in \mathbb{R}^{r\times r}$ given as a diagonal matrix that contains the eigenvalues of $\mathbf{R}$, where $r = \mathrm{rank}\,(\mathbf{R})$, and $\mathbf{U}^{\mathrm{H}}\mathbf{U} = \mathbf{I}_r$. Then, the Karhunen-Loeve expansion of $\mathbf{h}$ can be given by:

$$\mathbf{h} = \mathbf{R}^{\frac{1}{2}}\check{\mathrm{e}} = \mathbf{U}\mathbf{D}^{\frac{1}{2}}\mathbf{U}^{\mathrm{H}}\check{\mathrm{e}} \sim \mathbf{U}\mathbf{D}^{\frac{1}{2}}\mathbf{e} \tag{2.3}$$

such that $\check{\mathrm{e}} \sim \mathbb{C}_{\mathscr{N}}\,(0_M, \mathbf{I}_M)$ and $\mathrm{e} \sim \mathbb{C}_{\mathscr{N}}\,(0_r, \mathbf{I}_r)$ are white random variables with unity covariance matrices. The Karhunen model assumes that $\mathbf{U}\mathbf{D}^{\frac{1}{2}}\mathbf{e}$ and $\mathbf{h}$ have identical distributions. Note that the degrees of freedom at the BS will be decreased into $r \leq \mathrm{M}$ when using the spatial distribution of $\mathbf{U}\mathbf{D}^{\frac{1}{2}}\mathbf{e}$ instead of $\mathbf{h} \sim \mathbb{C}_{\mathscr{N}}\,(0_M, \mathbf{R}_M)$, and this gives a sense that there is a negative impact when using the Karhunen model for the spatially correlated channels. More details about the impact of the spatially correlated channels will be discussed in the following subsections. However, the practical channels may not have zero means or be completely Gaussian distributed as described here; instead, we used the above model for tractable analysis results.

2.1.3 Uplink transmission model

The uplink transmission model of the massive MIMO system can be illustrated by Fig.2.2 [11] where the uplink received signal at the BS can be written

as:

$$\begin{aligned}\mathbf{y}_j &= \sum_{l=1}^{L}\sum_{k=1}^{K_l}\mathbf{h}_{lk}^{j}s_{lk}+\mathbf{n}_j \\ &= \underbrace{\sum_{k=1}^{K_j}\mathbf{h}_{jk}^{j}s_{jk}}_{\text{“Desired signals”}} + \underbrace{\sum_{\substack{l=1\\ l\neq j}}^{L}\sum_{i=1}^{K_l}\mathbf{h}_{li}^{j}s_{li}}_{\text{“Inter-cell interference”}} + \underbrace{\mathbf{n}_j}_{\text{“Noise”}}\end{aligned} \tag{2.4}$$

where $\mathbf{n}_j \sim \mathbb{C}_{\mathcal{N}}\left(0_{M_j}, \sigma_{\text{UL}}^2\mathbf{I}_{M_j}\right)$ is the additive noise and is signal independent, $s_{li} \in \mathbb{C}$ is the complex-valued scalar signal that can be either data signal with complex Gaussian distribution or deterministic pilot signal. However, its power can be given by $p_{lk} = \mathbb{E}\{|s_{lk}|^2\}$ irrespective what it is. The channel responses, for now, are assumed to be known at the BS and are constant during the coherence block. During the uplink transmission of data signals, the BS j combines

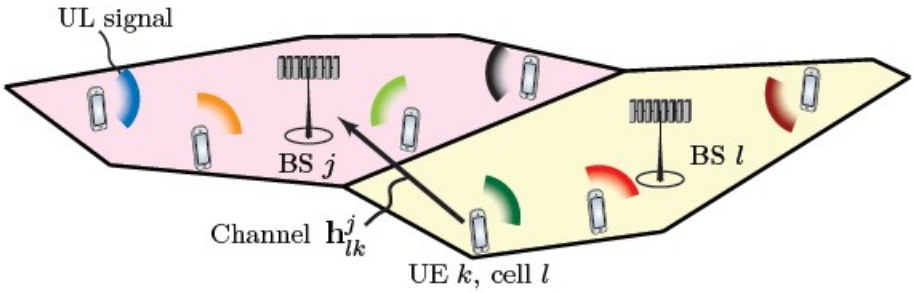

Figure 2.2: Uplink transmission in massive MIMO system

all received signals from the UEs in the network using the array antennas, and then coherently separates each desired UE by applying the combining vector $\mathbf{v}_{jk} \in \mathbb{C}^{M_j}$ of each UE to the received signal. The combining vector for each desired k^{th} UE is calculated based on its estimated channel during the pilot signalling period. The received signal after applying the combining vector will be given by:

$$\mathbf{v}_{jk}^{\text{H}}\mathbf{y}_j = \underbrace{\mathbf{v}_{jk}^{\text{H}}\mathbf{h}_{jk}^{j}s_{jk}}_{\text{“Desired signal”}} + \underbrace{\sum_{\substack{i=1\\ i\neq k}}^{K_j}\mathbf{v}_{jk}^{\text{H}}\mathbf{h}_{ji}^{j}s_{ji}}_{\text{“Intra-cell signals”}} + \underbrace{\sum_{\substack{l=1\\ l\neq j}}^{L}\sum_{i=1}^{K_l}\mathbf{v}_{jk}^{\text{H}}\mathbf{h}_{li}^{j}s_{li}}_{\text{“Inter-cell interference”}} + \underbrace{\mathrm{v}_{jk}^{\text{H}}\mathbf{n}_j}_{\text{“Noise”}} \tag{2.5}$$

2.1.4 Spatial correlations impact

To understand the impact of the spatially correlated channels [40] on the performance of the massive MIMO system, let us consider the uplink transmission model of a multiuser-single-cell scenario with channel responses of $\mathbf{h_k} \sim \mathbb{C}_{\mathscr{N}}(0_M, \mathbf{R}_k)$ for $k = 1, ..., K$, which assumed completely known at the BS.

First, let us assume that each UE in the cell has a different covariance matrix at the BS and the K independent channel covariance matrices [41] will be given by :

$$\mathbf{R}_k = K\mathbf{U}_k\mathbf{U}_k^{\mathrm{H}} \tag{2.6}$$

where $\mathrm{U}_k \in \mathbb{C}^{M \times M/K}$ is the k^{th} unitary matrix for any arbitrary UE in the cell, such that $\mathrm{U}_k^{\mathrm{H}}\mathrm{U}_k = \mathrm{I}_{M/K}$ for similar UEs, and $\mathbf{U}_k^{\mathrm{H}}\mathbf{U}_j = 0_{M/K \times M/K}$, for different UEs, i.e., $k \neq j$. The model in Eq.(2.6) implies that each user in the cell will have a spatial correlation channel with M/K instead of M strong spatial directions or degrees of freedom at the BS, i.e., there are M/K non-zero eigenvalues for each individual covariance matrix at the BS. Although the channels of UEs are random in nature, they will be mutually orthogonal since they live in mutually orthogonal sub-spaces. To investigate these facts, let's apply the Karhunen expansion for the total covariance model in Eq.(2.6); we obtain:

$$\mathbf{h}_k = \sqrt{K}\mathbf{U}_k\mathbf{e}_k \tag{2.7}$$

where $\mathrm{e}_k \sim \mathbb{C}_{\mathscr{N}}\left(0_{M/K}, \mathbf{I}_{M/K}\right)$. The received signal of the multiuser-single-cell scenario [41] will be given by:

$$\mathbf{y} = \sum_{i=1}^{K} \mathbf{h}_i s_i + \mathbf{n} \tag{2.8}$$

where s_i is the complex scalar signal with power p_i for $i = 1, ..., K$, and $\mathbf{n}$ is the additive noise at the BS. Now, the impact of having different spatial correlation matrices at the BS will be concluded as follows: Using the eigenspace $\mathbf{U}_k$ as a

combining vector [41] to detect the desired k^{th} UE, we get:

$$\begin{aligned}\mathbf{U}_k^{\mathrm{H}}\mathbf{y} &= \mathbf{U}_k^{\mathrm{H}}\left(\sum_{i=1}^{K}\mathbf{h}_i s_i + \mathbf{n}\right)\\ &= \sum_{i=1}^{K}\sqrt{K}\mathbf{U}_k^{\mathrm{H}}\mathbf{U}_i\mathbf{e}_i s_i + \mathbf{U}_k^{\mathrm{H}}\mathbf{n}\\ &= \sqrt{K}\mathbf{e}_k s_k + \check{\mathbf{n}}_k\end{aligned} \tag{2.9}$$

where $\check{\mathrm{n}}_k \sim \mathbb{C}_{\mathcal{N}}\left(0_{M/K}, \sigma_{\mathrm{UL}}^{\mathbf{2}}\mathbf{I}_{M/K}\right)$. The consequences in Eq.(2.9) show that the K multiuser channels are orthogonally separated into K of single-UE channels with M/K degrees of freedom for each UE at the BS; this is due to their structure of the spatial correlation. In other words, there is no interference between UEs at the BS, which is the main advantage of having spatially correlated channels. The average signal-to-noise ratio at each UE [41] will be:

$$\mathbb{E}\{\mathrm{SNR}_k\} = \mathbb{E}\left\{\frac{Kp_k\|\mathrm{e}_k\|^2}{\sigma_{\mathrm{UL}}^2}\right\} = \frac{Mp_k}{\sigma_{\mathrm{UL}}^2} \tag{2.10}$$

where the M in the expression Eq.(2.10) indicates that each user will experience full array gain in its signal power term.

Second, let us assume all UEs in the cell have similar spatial covariance matrices at the BS; hence, to detect any UE at the BS using its eigenspace U [41], we will get:

$$\begin{aligned}\mathbf{U}^{\mathrm{H}}\mathbf{y} &= \sum_{i=1}^{K}\sqrt{K}\mathbf{U}^{\mathrm{H}}\mathbf{U}\mathbf{e}_i s_i + \mathbf{U}^{\mathrm{H}}\mathbf{n}\\ &= \sum_{i=1}^{K}\sqrt{K}\mathbf{e}_i s_i + \check{\mathrm{n}}\end{aligned} \tag{2.11}$$

where $\check{\mathrm{n}} = \mathbf{U}^{\mathrm{H}}\mathbf{n}$. The results in Eq.(2.11) show the detrimental impact of using common covariance matrices among users since all K UEs will share only M/K spatial correlations at the BS, i.e., the degrees of freedom will be reduced to a common M/K effective antennas. This second scenario shows the negative side impact of having spatially correlated channels on the accuracy of the estimated channel.

In conclusion, the spatial correlation channels could result in a highly efficient massive MIMO system if all users have adequately distinct spatial covari-

ance matrices at the BS [42, 43]. But, in practice, it is hard to do that, and this case study may be used as an extreme example to show the effect of having spatial channels in the multiuser-single-cell scenario.

2.1.5 Local scattering model

Since the spatial correlation is a substantial identity for the scenario with several users in the cell, we will now create a spatial correlation model depending on the azimuth angles of the UEs to the BS. The main advantage of using such azimuth angles between serving users is to make separating the channels of all UEs more possible at the BS. The scenario of the single-UE single-cell [21] will be used here to develop the spatial model of the spatial correlation matrix $\mathbf{R} \in \mathbb{C}^{M \times M}$. The UE is assumed to be surrounded by local scattering objects, and there is no LOS path between the UE and the BS. On the other hand, the BS is assumed to be elevated, such that the received signals will be strong in some spatial directions and weak in others, as shown in Fig.2.3.

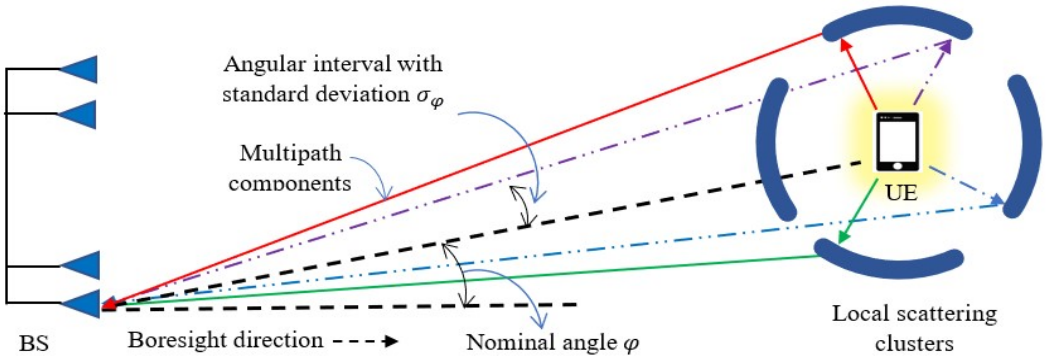

Figure 2.3: Local scattering model for spatially correlated channels

In other words, the received signal will be superimposed from many copies of the transmitted signal from multiple paths. That is, the reflected signal from each path will reach the BS antennas in a planer waveform from a particular angle to produce the array response $\mathbf{a}_n \in \mathbb{C}^M$ [21], like the case of LOS in Eq.(B.19) in Appendix B, as follows:

$$\mathbf{a}_n = g_n \left[\begin{array}{cccc} 1 & e^{2\pi \mathrm{j} d_\mathrm{H} \sin(\bar{\varphi}_n)} & \ldots & e^{2\pi \mathrm{j} d_\mathrm{H} (M-1) \sin(\bar{\varphi}_n)} \end{array} \right]^\mathrm{T}, \tag{2.12}$$

such that $\bar{\varphi}_n$ and $g_n \in \mathbb{C}$ are the azimuth angle and the gain from the individual path n to the BS array antennas, respectively, d_H is the spacing distance between

any two antennas. Thus, the channel response [21] for an N_{path} array responses at the BS will be given by:

$$\mathbf{h} = \sum_{n=1}^{N_{path}} \mathbf{a}_n \tag{2.13}$$

In general, the spatial angles $\bar{\varphi}_n$ and the gains g_n are usually assumed random variables with i.i.d probability density functions. Then, by following the central limit theorem, the channel response in Eq.(2.13) will take the distribution: $\mathbf{h} \sim \mathbb{C}_{\mathscr{N}}(0_M, \mathbf{R})$ as $N_{path} \to \infty$, where $\mathbf{R} = \mathbb{E}\{\sum_n a_n a_n^{\mathrm{H}}\}$ is the correlation matrix of the generated Rayleigh fading distribution. Note that, every $(l,m)^{th}$ element of the matrix $\mathbf{R}$ [12] can be given by:

$$\begin{aligned}[\mathbf{R}]_{l,m} &= \sum_{n=1}^{N_{\text{path}}} \mathbb{E}\left\{|g_n|^2\right\} \mathbb{E}\left\{e^{2\pi \mathrm{j} d_{\mathrm{H}}(l-1)\sin(\bar{\varphi}_n)} e^{-2\pi \mathrm{j} d_{\mathrm{H}}(m-1)\sin(\bar{\varphi}_n)}\right\} \\ &= \beta \int e^{2\pi \mathrm{j} d_{\mathrm{H}}(l-m)\sin(\bar{\varphi})} f(\bar{\varphi}) d\bar{\varphi}\end{aligned} \tag{2.14}$$

where $\mathbb{E}\{|g_n|^2\}$ is the average gain of any arbitrary path n, $f(\bar{\varphi})$ denotes the PDF of any arbitrary random angle $\bar{\varphi}$, and β is the total average gain of all multiple path components reached the BS antennas, it is given by $\beta = \sum_{n=1}^{N_{path}} \mathbb{E}\{|g_n|^2\}$.

Referred to Fig. 2.3, it is more suitable to imagine that the multiple path elements are emanating from the scatter's set surrounding the UE. This means that $\bar{\varphi} = \varphi + \delta$, where φ represents the nominal arrival angle from the UE to the BS and is deterministic, and δ denotes the random deviation from the φ angle. The distribution of the deviation angle δ has been considered in different ways in many works of literature as follows: in [44], the deviation distribution is taken as a Gaussian, i.e., $\delta \sim \mathscr{N}(0, \sigma_\varphi^2)$, as well as a Laplace distribution $\delta \sim \text{Lap}(0, \sigma_\varphi/\sqrt{2})$ used in [45], while a uniform distribution $\delta \sim \text{U}[-\sqrt{3}\sigma_\varphi, \sqrt{3}\sigma_\varphi]$ has been followed in [46]. The variable σ_φ denotes the Angular Standard Deviation (ASD) of the nominal angle φ and is a positive quantity measured by *radians*. The practical value of σ_φ is 10^o in the urban environments [47], while smaller or larger values are in the rural and hilly environments [**?**], respectively.

2.1.5.1 Angular deviation impact

The angular interval of the deviation has a substantial impact on the channels to be spatially correlated or uncorrelated as follows: When the σ_φ is increased to cover all angles in the angular domain, the deviation distribution will be uniformly distributed between $-\pi \to \pi$. This leads to the array antenna at the BS receiving the signals in all directions between $-\pi \to \pi$, which makes the channel response at the BS close to (but not fully) the uncorrelated fading channel. This is because the ULA still have better resolutions in some spatial directions than others. It should be noted that the rank of the highly spatially correlated channel is small due to the large variations in its eigenvalues. This means that the spatial correlation characteristics increase as σ_φ is reduced and vice versa [21].

2.1.5.2 Approximate formula of $[\mathbf{R}]_{l,m}$

The closed formula of the spatial correlation matrix in Eq.(2.14) is presented here using Gaussian distributed angular interval $\delta \sim \mathcal{N}(0, \sigma_\varphi^2)$ with the identity of $\sin(\delta) \approx \delta$, and $\cos(\delta) \approx 1$, as follows [12]:

$$
\begin{aligned}
[\mathbf{R}]_{l,m} &= \beta \int_{-\infty}^{\infty} e^{2\pi \mathrm{j} d_\mathrm{H}(l-m)\sin(\varphi+\delta)} \frac{1}{\sqrt{2\pi}\sigma_\varphi} e^{-\frac{\delta^2}{2\sigma_\varphi^2}} d\delta \\
&\approx \beta \int_{-\infty}^{\infty} e^{2\pi \mathrm{j} d_\mathrm{H}(l-m)\sin(\varphi)} e^{2\pi \mathrm{j} d_\mathrm{H}(l-m)\cos(\varphi)\delta} \frac{1}{\sqrt{2\pi}\sigma_\varphi} e^{-\frac{\delta^2}{2\sigma_\varphi^2}} d\delta \\
&= \beta e^{2\pi \mathrm{j} d_\mathrm{H}(l-m)\sin(\varphi)} e^{-\frac{\sigma_\varphi^2}{2}(2\pi d_\mathrm{H}(l-m)\cos(\varphi))^2} \underbrace{\frac{1}{\sqrt{2\pi}\sigma_\varphi} \int_{-\infty}^{\infty} e^{-\frac{\left(\delta - 2\pi \mathrm{j}\sigma_\varphi^2 d_\mathrm{H}(l-m)\cos(\varphi)\right)^2}{2\sigma_\varphi^2}} d\delta}_{=1} \\
&= \beta e^{2\pi \mathrm{j} d_\mathrm{H}(l-m)\sin(\varphi)} e^{-\frac{\sigma_\varphi^2}{2}(2\pi d_\mathrm{H}(l-m)\cos(\varphi))^2}
\end{aligned}
\tag{2.15}
$$

where $\sin(\varphi+\delta) = \sin(\varphi)\cos(\delta) + \cos(\varphi)\sin(\delta) \approx \sin(\varphi) + \delta\cos(\varphi)$. This approximation will be used in this thesis to reduce the complexity of computations when simulating the correlation matrix $\mathbf{R}$. However, two insights come

with this approximate formula can be concluded here, as follows:

- Insight 1: when $\sigma_\varphi = 0$, the $(l,m)^{th}$ elements of $\mathbf{R}$ will be simplified into $[\mathbf{R}]_{l,m} = \beta e^{2\pi \mathrm{j} d_\mathrm{H}(l-m)\sin(\varphi)}$, which means that all the multiple paths will come from only the angle φ making $\mathbf{R}$ as highly spatially correlated with rank=1.

- Insight 2: when $\sigma_\varphi > 0$, the off-diagonal elements of the covariance matrix $\mathbf{R}$ will fall as $e^{-\frac{\sigma_\varphi^2}{2}(2\pi d_\mathrm{H}(l-m)\cos(\varphi))^2}$ to zero. This means that the correlation matrix $\mathbf{R}$ will be nearly spatially uncorrelated when σ_φ increases, making it with a full rank matrix. The following section will discuss how the channel responses are estimated depending on the knowledge of their correlation matrices of the UEs at the BS.

2.2 Channel Estimation Concepts

In this section, the process of the uplink channel estimation during the pilot transmission period will be discussed while considering the knowledge of the channel correlation matrices at the BS. The uplink model for transmitting pilots, the method of designing orthogonal pilot sequences, the MMSE Bayesian estimator, and the effect of the spatial channel correlation on the estimated channels will be discussed in detail.

2.2.1 Pilot uplink transmission model

In this part, we will use the same uplink transmission model in Eq.(2.4) to develop the pilot transmission model here [12, 48]. The scalar transmitted signal s_{jk} in Eq.(2.4) is now replaced by the UE's pilot sequence $\boldsymbol{\phi}_{jk}$. Recall that in subsection 2.1.1.4, τ_p samples are allocated for pilot signalling in the coherence block. Hence, if $\boldsymbol{\phi}_{jk} \in \mathbb{C}^{\tau_p}$ denotes the pilot sequence allocated for the user k in cell j, then the norm squared of the pilot sequence will be $||\boldsymbol{\phi}_{jk}||^2 = \boldsymbol{\phi}_{jk}^\mathrm{H}\boldsymbol{\phi}_{jk} = \tau_p$, which is designated to always give τ_p constant power. Now, assume the $\boldsymbol{\phi}_{jk}$ sequences are scaled by transmitting power $\sqrt{p_{jk}}$, i.e., $s_{jk} = \sqrt{p_{jk}}\boldsymbol{\phi}_{jk}$. Then,

the uplink model will [12] be given by:

$$\mathbf{Y}_j^p = \underbrace{\sum_{k=1}^{K_j} \sqrt{p_{jk}} \mathbf{h}_{jk}^j \boldsymbol{\phi}_{jk}^{\mathrm{T}}}_{\text{“Desired pilots”}} + \underbrace{\sum_{l=1}^{L} \sum_{i=1}^{K_l} \sqrt{p_{li}} \mathbf{h}_{li}^j \boldsymbol{\phi}_{li}^{\mathrm{T}}}_{\text{“Inter-cell pilots”}} + \underbrace{\mathbf{N}_j^p}_{\text{“Noise”}} \tag{2.16}$$

such that $\mathbf{Y}_j^p \in \mathbb{C}^{M_j \times \tau_p}$ is the observed signal at the BS j that will be used to estimate the channel responses $\mathbf{h}_{jk}^j$. $\mathbf{N}_j^p \in \mathbb{C}^{M_j \times \tau_p}$ is the additive noise which is independent of the transmitted signal.

To estimate the channel response $\mathbf{h}_{jk}^j$ for the UE k in cell j at the own BS in cell j, the pilot sequence of this user should be predefined first at the BS. Then, the BS can correlate the received signal in Eq.(2.16) with the predefined UE's pilot sequence $\boldsymbol{\phi}_{jk}$, which yields [12]:

$$\begin{aligned} \mathbf{y}_{jjk}^p &= \mathbf{Y}_j^p \boldsymbol{\phi}_{jk}^{\star} \\ &= \underbrace{\sqrt{p_{jk}} \mathbf{h}_{jk}^j \boldsymbol{\phi}_{jk}^{\mathrm{T}} \boldsymbol{\phi}_{jk}^{\star}}_{\text{“Desired pilot”}} + \underbrace{\sum_{\substack{i=1 \\ i \neq k}}^{K_j} \sqrt{p_{ji}} \mathbf{h}_{ji}^j \boldsymbol{\phi}_{ji}^{\mathrm{T}} \boldsymbol{\phi}_{jk}^{\star}}_{\text{“Intra-cell pilots”}} + \underbrace{\sum_{\substack{l=1 \\ l \neq j}}^{L} \sum_{i=1}^{K_l} \sqrt{p_{li}} \mathbf{h}_{li}^j \boldsymbol{\phi}_{li}^{\mathrm{T}} \boldsymbol{\phi}_{jk}^{\star}}_{\text{“Inter-cell pilots”}} + \underbrace{\mathbf{N}_j^p \boldsymbol{\phi}_{jk}^{\star}}_{\text{“Noise”}} . \end{aligned} \tag{2.17}$$

such that $\mathbf{y}_{jjk}^p \in \mathbb{C}^{M_j}$ denotes the processed signal of the observed pilot signal $\mathbf{Y}_j^p$. The terms that follow the desired term in Eq.(2.17) contain the inner product $\boldsymbol{\phi}_{(j,l)i}^{\mathrm{T}} \boldsymbol{\phi}_{jk}^{\star}$ between the pilot sequence of the desired UE and the pilot sequence of the interfering UE from either the own cell j or from another cell l. However, if the pilot sequences are orthogonal, then $\boldsymbol{\phi}_{(j,l)i}^{\mathrm{T}} \boldsymbol{\phi}_{jk}^{\star} = 0$; hence, the interfering user terms will vanish and not impact the estimation. However, several notes should be considered when designing pilot sequences for the cellular network [48] as follows:

- The pilot sequences is τ_p dimensional vectors; hence, the BS can only provide τ_p mutually orthogonal pilot sequences.
- The length of the pilot sequence τ_p should be less than the length of the coherence block τ_c. That means assigning orthogonal pilot sequences for all UEs will be impossible.
- Using a longer pilot sequence comes with the price of reducing the uplink

data transmission samples to pay. However, there is a rule of thumb [49] supposes that the pilot sequence length should be $\tau_p < \tau_c/2$.

- It is preferred to the cellular network in each cell to have a set of pilot sequences that satisfy the condition $\tau_p \geq maxK_l$ to make the BS able to allocate distinct pilot sequences to cover all UEs in the cell.

However, designing unique orthogonal pilot sequences for the UEs in each network cell is useful to suppress the strong intra-cell interference, but in practice, it is needed to reuse the pilots among other UEs in other cells. Let us define the following set:

$$\mathscr{P}_{jk} = \left\{ (l,i) : \phi_{li} = \phi_{jk}, \quad l = 1,\ldots,L, i = 1,\ldots,K_l \right\} \tag{2.18}$$

where $\mathscr{P}_{jk}$ is the pilots set of the users in cell j that can be reused for the UEs in cell l. Using the definition in Eq.(2.18), the processed signal in Eq.(2.17) is simplified to:

$$\mathbf{y}^p_{jjk} = \underbrace{\sqrt{p_{jk}}\tau_p \mathbf{h}^j_{jk}}_{\text{“Desired pilot”}} + \underbrace{\sum_{(l,i)\in\mathscr{P}_{jk}\setminus(j,k)} \sqrt{p_{li}}\tau_p \mathbf{h}^j_{li}}_{\text{“Interfering pilots”}} + \underbrace{\mathbf{N}^p_j \phi^\star_{jk}}_{\text{“Noise”}}. \tag{2.19}$$

where $\mathbf{N}^p_j \phi^\star_{jk} \sim \mathbb{C}_{\mathscr{N}}\left(0_{M_j}, \sigma^2_{\text{UL}} \tau_p \mathbf{I}_{M_j}\right)$, is the receiver noise and now it contains the processed gain τ_p, which is the result of $||\phi_{jk}||^2$. It should be noticed that the processed signal $\mathbf{y}^p_{jjk} = \mathbf{y}^p_{ilk}$ since they used the same set of pilot sequences, i.e., $(l,i),(j,k) \in \mathscr{P}_{jk}$.

2.2.2 Pilot book design

The pilot book Φ is a matrix with two conditions: First, all elements have unit magnitude, and Second, all columns are mutually orthogonal, that is $|[\Phi]_{i_1,i_2}| = 1$ for $(i_1,i_2 = 1,\ldots,\tau_p)$, and $\Phi^{\text{H}}\Phi = \tau_p \mathbf{I}_{\tau_p}$, respectively. The Walsh matrix [50] with dimensions $\tau_p \times \tau_p$ and -1 or +1 elements can be used as a pilot book because it satisfies the aforementioned conditions. It can be produced only for matrices with dimension $\tau_p = 2^n$ for $n = 0, 1, ...$, which gives the sense that each element in the matrix as it comes from the Binary-Phase-shift -keying (BPSK) constellation points. However, to develop pilot books with arbitrary dimensions

and with matrix elements depending on the pilot length, the Discrete Fourier Transform DFT matrix [12] can be used as follows:

$$\Phi = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \omega_{\tau_p} & \omega_{\tau_p}^2 & \dots & \omega_{\tau_p}^{\tau_p-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega_{\tau_p}^{\tau_p-1} & \omega_{\tau_p}^{2(\tau_p-1)} & \dots & \omega_{\tau_p}^{(\tau_p-1)(\tau_p-1)} \end{bmatrix} \tag{2.20}$$

where $\omega_{\tau_p} = e^{-\mathrm{j}2\pi/\tau_p}$, i.e., each element in the matrix depends on the pilot length τ_p; hence, it is τ_p-phase-shift keying (τ_p-PSK) constellation-like. In this thesis, the DFT pilot book matrix will be leveraged to provide predefined orthogonal pilot sequences for the UEs at the BS.

2.2.3 Bayesian MMSE channel estimate

In this section, the Bayesian minimum-mean-squared-error (MMSE) estimator of the channel response $\mathbf{h}_{lk}^{j}$ is developed based on the pilot-received signal in Eq.(2.1) and the pilot book design in previous sections. Since the Bayesian estimator considers the Probability Density Function (PDF) distribution of the unknown random variable, it will be used here for our channel response because it represents a realization of a random variable with known or partially known distributions. See Appendix (C.3) for more details about the fundamental of the estimation theory.

However, given the processed observation signal $\mathbf{y}_{jli}^{p}$ (the processed pilot received signal from all UEs in the network at the BS j) and needed to estimate the channel response $\mathbf{h}_{jk}^{j} \sim \mathbb{C}_{\mathcal{N}}(0_{M_j}, \mathbf{R}_{jk}^{j})$, then the MMSE estimator that minimizes the mean squared error $\mathbb{E}\{\|\mathbf{h}_{jk}^{j} - \hat{\mathbf{h}}_{jk}^{j}(\mathbf{y}_{jli}^{p})\|^2\}$ will take the formula [11]:

$$\hat{\mathbf{h}}_{jk}^{j} = \sqrt{p_{jk}}\mathbf{R}_{jk}^{j}\Psi_{li}^{j}\mathbf{y}_{jli}^{p} \tag{2.21}$$

where $\hat{\mathbf{h}}_{jk}^{j}$ is the channel estimate of the uplink channel between the UE k in cell j to the BS j, and $\Psi_{li}^{j} = \mathbb{E}\{\mathbf{y}_{jli}^{p}(\mathbf{y}_{jli}^{p})^{\mathrm{H}}\}/\tau_p$ is the inverse of the processed pilot signal correlation matrix normalized by τ_p, it accounts for the received pilot signals from all UEs in the network including the desired UE. However, it can

be given by the formula [11]:

$$\Psi_{li}^{j} = \left(\sum_{(l',i') \in \mathscr{P}_{lk}} p_{l'i'} \tau_p \mathbf{R}_{l'i'}^{j} + \sigma_{\mathrm{UL}}^2 \mathbf{I}_{M_j} \right)^{-1} \tag{2.22}$$

which takes into account the impact of the additive noise and the covariance matrices of all UEs that used the same pilot sequence as the k^{th} desired UE in cell j.

The estimation quality is measured by the mean squared error [11]:

$$\mathrm{MSE} = \mathbb{E}\left\{ \|\mathbf{h}_{jk}^{j} - \hat{\mathbf{h}}_{jk}^{j}(\mathbf{y}_{jli}^{p})\|^2 \right\} = \mathrm{tr}\left(\mathrm{C}_{jk}^{j} \right), \tag{2.23}$$

where a small MSE indicates good quality of channel estimation. The matrix $\mathrm{C}_{jk}^{j} = \mathbb{E}\{\tilde{\mathbf{h}}_{li}^{j}(\tilde{\mathbf{h}}_{li}^{j})^{\mathrm{H}}\}$ is the correlation matrix of the estimation error $\tilde{\mathbf{h}}_{li}^{j} = \mathbf{h}_{jk}^{j} - \hat{\mathbf{h}}_{jk}^{j}$, it can be given by [11]:

$$\mathbf{C}_{jk}^{j} = \mathbf{R}_{jk}^{j} - p_{jk} \tau_p \mathbf{R}_{jk}^{j} \Psi_{li}^{j} \mathbf{R}_{jk}^{j} \tag{2.24}$$

where $\mathbf{R}_{jk}^{j}$ is the covariance matrix of the desired channel that would be estimated. Note that the Bayesian estimator in Eq.(2.21) is linear in the sense that it multiplies the receive signal $\mathbf{y}_{jli}^{p}$ with two correlation matrices.

Definition 2.2.1 (Effective SNR) The effective signal-to-noise ratio can be defined only during pilot transmission since the scaling power p_{jk} spans all the pilot sequence elements of ϕ_{jk}^{j}; hence, the product $(p_{jk}\tau_p)$ will appear in the effective SNR [11] as follows:

$$\mathrm{SNR}_{jk}^{p} = \frac{p_{jk} \tau_p \beta_{jk}^{j}}{\sigma_{\mathrm{UL}}^2}. \tag{2.25}$$

where $\beta_{jk}^{j} = \frac{1}{M_j} \mathrm{tr}(\mathbf{R}_{jK}^{j})$ is the average channel gain at each antenna of the BS. Note that the τ_p here is termed as *pilot processing gain* and is concluded within the SNR formula, and that is why the SNR is denoted by effective SNR. The τ_p processing gain is preferable to achieve high-quality estimation even for UEs who have limited power or/and weak channels.

2.2.4 Statistical properties of channel estimates

Since the channel responses $\mathbf{h}_{lk}^{j}$ are assumed random, and each of which represents a realization of an ergodic random process, their channel estimates $\hat{\mathbf{h}}_{li}^{j}$ will also be random. It follows that the estimation error $\tilde{\mathbf{h}}_{lk}^{j}$ will also be random and is independent of the channel estimate $\hat{\mathbf{h}}_{li}^{j}$. The distributions of $\hat{\mathbf{h}}_{li}^{j}$ and $\tilde{\mathbf{h}}_{lk}^{j}$ for any arbitrary UE i in cell l can be given by [11]:

$$\begin{aligned} \hat{\mathbf{h}}_{li}^{j} &\sim \mathbb{C}_{\mathcal{N}}\left(\mathbf{0}_{M_j}, \mathbf{R}_{li}^{j} - \mathbf{C}_{li}^{j}\right) \\ \tilde{\mathbf{h}}_{li}^{j} &\sim \mathbb{C}_{\mathcal{N}}\left(\mathbf{0}_{M_j}, \mathbf{C}_{li}^{j}\right). \end{aligned} \tag{2.26}$$

The norm square of the channel estimate $\mathbb{E}\{\|\hat{\mathbf{h}}_{li}^{j}\|^2\} = \mathrm{tr}(\mathbf{R}_{li}^{j} - \mathbf{C}_{li}^{j})$, can be used to measure the quality of the channel estimation, it can be rewritten as follows [11]:

$$\mathbb{E}\{\|\hat{\mathbf{h}}_{li}^{j}\|^2\} = \mathrm{tr}(\mathbf{R}_{li}^{j}) - \mathrm{tr}(\mathbf{C}_{li}^{j}) \tag{2.27}$$

It is noticed in Eq.(2.27) that the norm square is inversely proportional to the $\mathrm{tr}(\mathbf{C}_{li}^{j}) = \mathrm{MSE}$, i.e., the norm square will be large (i.e., we have good channel estimate) if the MSE is small and vice versa. In a unique case, at $\mathrm{tr}(\mathbf{C}_{li}^{j}) = 0$, the channels will be estimated without error, i.e., we have perfect channel estimation; hence, $\mathbb{E}\{\|\hat{\mathbf{h}}_{li}^{j}\|^2\} = \mathbb{E}\{\|\mathbf{h}_{li}^{j}\|^2\} = \mathrm{tr}(\mathbf{R}_{li}^{j})$. But, this will not happen in practice since there is always an error.

2.2.4.1 Important Observations

Referred to the channel estimate expression in Eq.(2.21), some observations can be concluded as follows:

- The channel estimate in Eq.(2.21) can be leveraged to estimate the channels from all UE in the network, i.e., this expression can be used with the multiuser-multicell scenario, which is the case in practice.
- By using the Wyner model with a multiuser scenario at the BS j, the channel estimate of any UE k in the own cell j is $\hat{\mathbf{h}}_{jk}^{j}$, and the channel estimate of any UE i from another cell l is $\hat{\mathbf{h}}_{li}^{j}$. If the UE i use the same pilot sequence of the UE k in cell j, then they will have the same processing signal

at the BS, since $\phi_{li} = \phi_{jk}$, i.e., $\Psi^j_{li} = \Psi^j_{jk}$, which yields $\mathbf{y}^p_{jli} = \mathbf{y}^p_{jjk}$. This means that both received signals will be multiplied by the same (Psi) matrix, and only the scaling power is different. Following Eq.(2.21), both channel estimates will be related as follows:

$$\hat{\mathbf{h}}^j_{li} = \frac{\sqrt{p_{li}}}{\sqrt{p_{jk}}} \mathbf{R}^j_{li} \left(\mathbf{R}^j_{jk}\right)^{-1} \hat{\mathbf{h}}^j_{jk} \tag{2.28}$$

This indicates that both of the channel estimates are highly correlated at the BS but still linearly independent vectors since $\mathbf{R}^j_{jk}$ and $\mathbf{R}^j_{li}$ are spatially correlated, and they are not scalar quantities.

Hence, If the channels are spatially uncorrelated, i.e., the covariance matrices given by $\mathbf{R}^j_{jk} = \beta^j_{jk}$, and $\mathbf{R}^j_{li} = \beta^j_{li}$ respectively; then both channel estimates will differ only by scaling factor and become linearly dependent vectors. This is the worst case in the channel estimation since the BS can not be able to separate the channels of the serving UEs. This gives us a sense that why the spatially correlated channels are more useful and can contribute to more degrees of freedom at the BS. The following corollary will measure the correlation between the channel estimates at the BS.

2.2.5 Correlation of channel estimates

To measure how the channel estimates are related to each other, the following correlation matrix and correlation coefficient are very crucial to be used here, as follows: The correlation matrix is given by [11]:

$$\mathbb{E}\{\hat{\mathbf{h}}^j_{jk}(\hat{\mathbf{h}}^j_{li})^{\mathrm{H}}\} = \begin{cases} \sqrt{p_{li}p_{jk}}\tau_p \mathbf{R}^j_{jk}\Psi^j_{li}\mathbf{R}^j_{li} & (l,i) \in \mathscr{P}_{jk} \\ 0_{M_j \times M_j} & (l,i) \notin \mathscr{P}_{jk}. \end{cases} \tag{2.29}$$

and the correlation coefficient is given by:

$$\frac{\mathbb{E}\{(\hat{\mathbf{h}}^j_{li})^{\mathrm{H}}\hat{\mathbf{h}}^j_{jk}\}}{\sqrt{\mathbb{E}\{\|\hat{\mathbf{h}}^j_{jk}\|^2\}\mathbb{E}\{\|\hat{\mathbf{h}}^j_{li}\|^2\}}} = \begin{cases} \frac{\mathrm{tr}\left(\mathbf{R}^j_{li}\mathbf{R}^j_{jk}\Psi^j_{li}\right)}{\sqrt{\mathrm{tr}\left(\mathbf{R}^j_{jk}\mathbf{R}^j_{jk}\Psi^j_{li}\right)\mathrm{tr}\left(\mathbf{R}^j_{li}\mathbf{R}^j_{li}\Psi^j_{li}\right)}} & (l,i) \in \mathscr{P}_{jk} \\ 0 & (l,i) \notin \mathscr{P}_{jk} \end{cases} \tag{2.30}$$

where $\mathbb{E}\{(\hat{\mathbf{h}}^j_{li})^{\mathrm{H}}\hat{\mathbf{h}}^j_{jk}\} = 0$ for all users that have unique pilot sequences, i.e.,

when $(l,i) \notin \mathscr{P}_{jk}$, this means that each UE in the network has a distinct (independent) received signal at the BS (i.e., $\mathbb{E}\{(\mathbf{y}^p_{jjk}(\mathbf{y}^p_{jli})^{\mathrm{H}}) = 0$). Otherwise, the users that have identical pilot sequences in another cell will contaminate the UEs in the desired cell [51, 52] since they have identical dependent received signals at the BS, i.e., $\mathbb{E}\{(\mathbf{y}^p_{jjk}(\mathbf{y}^p_{jli})^{\mathrm{H}}) \neq 0$.

Remember that the good Bayesian estimate is the one that minimizes the mean squared error between the estimated and desired channels and can be defined as follows:

$$\mathbb{E}\{\|\mathbf{h}^j_{li} - \hat{\mathbf{h}}^j_{li}\|^2\} = \mathbb{E}\{\|\tilde{\mathbf{h}}^j_{li}\|^2\} = \mathbb{E}\{\mathrm{tr}(\tilde{\mathbf{h}}^j_{li}(\tilde{\mathbf{h}}^j_{li})^{\mathrm{H}})\} = \mathrm{tr}(\mathbf{C}^j_{li}) \tag{2.31}$$

However, to measure the per-antenna estimation error, the suited metric to use is the normalized MSE given by [11]:

$$\mathrm{NMSE}^j_{li} = \frac{\mathrm{tr}(\mathbf{C}^j_{li})}{\mathrm{tr}(\mathbf{R}^j_{li})} \tag{2.32}$$

which can also be used to compare the quality of estimation between any two different channel estimation schemes.

2.2.6 Spatial correlations impact on channel estimation

To exemplify the impact of the spatial correlation on the estimated channel by the Bayesian MMSE estimator in Eq.(2.21), the single-cell single-UE scenario is used here with a unique pilot sequence of the desired user. Recall that the channel between the UE and BS is assumed as a correlated Rayleigh fading channel and given by $\mathbf{h} \sim \mathbb{C}_{\mathscr{N}}(0_M, \mathbf{R})$. According to linear algebra properties, the eigenstructure decomposition of the covariance matrix $\mathbf{R} \in \mathbb{C}^{M\times M}$ can be given by $\mathbf{R} = \mathbf{U}\mathbf{D}\mathbf{U}^{\mathrm{H}}$, where $\mathbf{U} \in \mathbb{C}^{M\times r}$ and $\mathbf{U}^{\mathrm{H}} \in \mathbb{C}^{r\times M}$ are the left and right eigendirections of the matrix $\mathbf{R}$, respectively, $r = \mathrm{rank}(\mathbf{R})$, and $\mathbf{D} \in \mathbb{R}^{r\times r}$ is a diagonal matrix contains decreasing eigenvalues of $\mathbf{R}$, i.e., $\mathbf{D} = \mathrm{diag}(\lambda_1, ..., \lambda_M)$, where $\lambda_1 \geq \lambda_2, ..., \geq \lambda_M$. The single user channel estimate that uses the Bayesian estimator in Eq.(2.21) can be given by [53]:

$$\hat{\mathbf{h}}_{\mathrm{MMSE}}(\mathbf{y}) = q^{\star}\mathbf{R}\left(|q|^2\mathbf{R} + \mathbf{S}\right)^{-1}\mathbf{y} \tag{2.33}$$

See Appendix C.3.2, where $\mathbf{y}$ is the received signal processed inside the BS, q denotes a scalar pilot sample transmitted by the single-antenna UE in the uplink direction, $\mathbf{R}$ is the channel covariance matrix, and $\mathbf{S} = \sigma^2_{\text{UL}}\mathbf{I}_M$ is the noise covariance matrix. However, in comparison with the expression in Eq.(2.21), the Ψ correlation matrix of the single UE will be equal to $\left(|q|^2\mathbf{R}+\mathbf{S}\right)^{-1}$, which now includes the desired UE's covariance matrix and the impact of the additive noise only without interference. Now, by considering all samples of the τ_p pilot sequence with a p transmitted power spans all the samples, the MMSE channel estimate in Eq.(2.33) can be rewritten as [53]:

$$\hat{\mathbf{h}}_{\text{MMSE}}(\mathbf{y}) = \sqrt{p}\tau_p\mathbf{R}\left(p\tau_p\mathbf{R}+\sigma^2_{\text{UL}}\mathbf{I}_M\right)^{-1}\mathbf{y} \tag{2.34}$$

However, to demonstrate the influence of the spatial correlation properties, let us consider the error correlation matrix given in Eq.(2.24), but now for a single UE [11] as follows :

$$\begin{aligned}\mathbf{C} &= \mathbf{R} - p\tau_p\mathbf{R}\left(p\tau_p\mathbf{R}+\sigma^2_{\text{UL}}\mathbf{I}_M\right)^{-1}\mathbf{R} \\ &= \mathbf{U}\left(\mathbf{D} - p\tau_p\mathbf{D}\left(p\tau_p\mathbf{D}+\sigma^2_{\text{UL}}\mathbf{I}_M\right)^{-1}\mathbf{D}\right)\mathbf{U}^{\text{H}} \\ &= \mathbf{U}\,\text{diag}\left(\lambda_1 - \frac{p\tau_p\lambda_1^2}{p\tau_p\lambda_1+\sigma^2_{\text{UL}}},\ldots,\lambda_M - \frac{p\tau_p\lambda_M^2}{p\tau_p\lambda_M+\sigma^2_{\text{UL}}}\right)\mathbf{U}^{\text{H}}\end{aligned} \tag{2.35}$$

where $\mathbf{I}_M = \mathbf{U}\mathbf{U}^{\text{H}} = \mathbf{U}^{\text{H}}\mathbf{U} = \mathbf{U}^{-1}\mathbf{U}$ since $\mathbf{U}$ is a square unitary matrix contains independent orthogonal columns. The expression in Eq.(2.35) represents the eigenstructure of the error correlation matrix with eigenvectors contained in $\mathbf{U}$ and M eigenvalues are given by:

$$\text{Variance Error} = \lambda_m - \frac{p\tau_p\lambda_m^2}{p\tau_p\lambda_m+\sigma^2_{\text{UL}}} = \frac{\sigma^2_{\text{UL}}\lambda_m}{p\tau_p\lambda_m+\sigma^2_{\text{UL}}} = \frac{\lambda_m}{\text{SNR}^p\frac{\lambda_m}{\beta}+1} \tag{2.36}$$

where $\beta = \frac{1}{M}\sum_{n=1}^{M}\lambda_n$ is the average channel gain per antenna element, and SNR^p is the effective signal to noise ratio given in Eq.(2.25).

Several insights can be concluded from the spatial characteristics in Eqs.(2.35) and (2.36) can be listed as follows:

- The correlation matrices **R** and **C** have the same eigendirections, and only the difference in the eigenvalues, which is smaller in **C** since they are normalized by the term $(\mathrm{SNR}^p \frac{\lambda_m}{\beta} + 1)$.
- The eigenvalues of **R** describe the channel gain in each eigendirection, while those in **C** define the variance of estimation error in each eigendirection.
- Since the eigenvalues of **C** are inversely proportional to the effective SNR^p; hence, it can decrease the variance of the error to zero when the SNR^p goes to infinity.
- The strong eigendirection of **R** is the eigenvector with a large eigenvalue and small normalized variance of error given by:

$$\text{Normalized Variance Error} = \left(\frac{\lambda_m}{\mathrm{SNR}^p \frac{\lambda_m}{\beta} + 1} \right) / \lambda_m = \frac{1}{\mathrm{SNR}^p \frac{\lambda_m}{\beta} + 1} \tag{2.37}$$

- It can independently estimate the eigenvectors with a large eigenvalue since they have high effective SNR^p.

Based on these important intuitions, reducing the eigenvalues of **R** into an arbitrary rank has a major impact on reducing the error variance in **C** and making the channel more spatially correlated. However, to demonstrate this effect of the spatial characteristics on the estimated channel, two important parameters that affect the spatial correlation are discussed here, which are the effective SNR^p and the angular standard deviation ASD (σ_φ). The parameter ASD can take different distributions, such as the Gaussian, the Uniform, or the Laplace distribution, see Section 2.1.5. The following Lemma of the SVD eigenstructure decomposition illustrates how the truncation procedure is used in our proposed model with successive arbitrary ranks of the covariance matrix.

2.2.7 Practical channel model

The Bayesian MMSE estimator in Eq.(2.34) is the channel vector that is used to minimize the second-order moment of the estimation error $\mathbb{E}\{\|\tilde{\mathrm{h}}\|^2 = \|\mathrm{h} -$

$\hat{\mathrm{h}}\|^2\}$, where $\tilde{\mathrm{h}}$ and $\hat{\mathrm{h}}$ are the error and the channel estimate vectors, respectively. They are independent random vectors with complex Gaussian distributions $\tilde{\mathrm{h}} \sim \mathbb{C}_{\mathscr{N}}(\mathbf{0}, \mathbf{R} - \tilde{\mathbf{R}})$, and $\hat{\mathrm{h}} \sim \mathbb{C}_{\mathscr{N}}(\mathbf{0}, \hat{\mathbf{R}})$, where $\hat{\mathbf{R}} = \mathbb{E}\{\hat{\mathrm{h}}(\hat{\mathrm{h}})^{\mathrm{H}}\}$ denotes the statistical information of the channel estimate vector, it can be written as:

$$\hat{\mathbf{R}} = \sqrt{p}\tau_p \mathbf{R} \left(p\tau_p \mathbf{R} + \sigma_{\mathrm{UL}}^2 \mathbf{I}_M \right)^{-1} \mathbf{R} \tag{2.38}$$

Note that the covariance matrix of the channel estimator $\hat{\mathrm{h}}$ mainly depends on the channel's statistic at the BS. It means that the BS can determine the MMSE estimator $\hat{\mathrm{h}}$ only when it knows the covariance matrix $\mathbf{R}$. However, the BS has no prior knowledge of that statistic in practice.

Hence, to estimate the channel vector $\mathrm{h} = [h_1, ..., h_M]^{\mathrm{T}}$, the BS needs to see many observations from this vector in different coherence blocks, which are assumed constant during the entire system bandwidth. Imagine the BS had acquired N random observations, given by the set of vectors $\mathrm{h}[1], ..., \mathrm{h}[N]$. Each arbitrary m^{th} BS antennae will see N observation elements representing the channel realizations from each channel vector at that antennae. Hence, for a given m^{th} antennae index at the BS, and according to the large numbers law (see appendix C.3.4), it can compute the sample variance ($\sum_{n=1}^{N} \frac{1}{N} |h_m[n]|^2$), which converges to the actual value ($\mathbb{E}\{|h_m|^2\}$) at that antennae when the number of observations goes to infinity. Note that the sample variance element is decaying as $1/N$; this yields that the standard deviation at each antenna element is decayed as $1/\sqrt{N}$. This means that only a few observations are required to estimate useful sample variance. The same procedure can be followed if we want to estimate the covariance matrix $\mathbf{R}$, which will be denoted by the sample covariance matrix of $\mathbf{R}$ [17] and can be given by:

$$\hat{\mathbf{R}}_{\text{“sample”}} = \frac{1}{N} \sum_{n=1}^{N} \mathbf{h}[n](\mathbf{h}[n])^{\mathrm{H}}. \tag{2.39}$$

According to the large numbers law, each individual element of $\hat{\mathbf{R}}_{\text{sample}}$ converges to its corresponding element in the true covariance matrix $\mathbf{R}$. But, it is very difficult to get aligned elements in both matrices since there is always an error occurring inside the M^2 entries of the sample covariance matrix. This is critical for the channel estimation topic since the Bayesian MMSE estimate

relies on the eigenvectors and eigenvalues of such matrices. However, some literature paper in [14, 16, 11] works on these errors and regularises the sample covariance matrix to make the estimation in this method robust for massive MIMO. They estimate the sample covariance matrix as a convex optimization between the diagonal matrix [17]

$$\hat{\mathbf{R}}_{\text{“diagonal”}} = \begin{bmatrix} \frac{1}{N}\sum_{n=1}^{N}|h_1[n]|^2 & & \\ & \ddots & \\ & & \frac{1}{N}\sum_{n=1}^{N}|h_M[n]|^2 \end{bmatrix} \tag{2.40}$$

and the sample covariance matrix as follows:

$$\hat{\mathbf{R}}(c) = c\hat{\mathbf{R}}_{\text{sample}} + (1-c)\hat{\mathbf{R}}_{\text{diagonal}}\,, \tag{2.41}$$

where c denotes the regularization factor, which is a random value between 0 and 1; one can optimize this parameter to obtain a robust estimation. However, due to this regularization method, full-rank covariance matrices are obtained even if the number of observations is less than the antennas-array number.

Now, to evaluate the performance of the Bayesian estimator on the estimated channels by the sample covariance method, the normalized mean squared error (NMSE) will be utilized here as a powerful performance metric to measure the quality of the channels. Hence, for the Bayesian estimator given by $\mathbf{h} = \mathbf{B}\mathbf{Y}^p\phi^*$ the **NMSE** will be given by [17]:

$$\text{NMSE}(\mathbf{B}) = 1 - \frac{\sqrt{p}\tau_p\Re(\text{tr}(\mathbf{R}\mathbf{B})) - \tau_p\,\text{tr}\left(\mathbf{B}\left(p\tau_p\mathbf{R} + \sigma_{\text{UL}}^2\mathbf{I}_M\right)\mathbf{B}^{\text{H}}\right)}{\text{tr}(\mathbf{R})}, \tag{2.42}$$

such that $\mathbf{B}$ prescribes the type of the utilized estimator; it is given by: $\mathbf{B} = \sqrt{p}\tau_p\mathbf{R}\left(p\tau_p\mathbf{R} + \sigma_{\text{UL}}^2\mathbf{I}_M\right)^{-1}$ for the MMSE in Eq.(2.34), whereas the expression $\mathbf{B}(c) = \sqrt{p}\tau_p\hat{\mathbf{R}}(c)(p\tau_p\hat{\mathbf{R}}(c) + \sigma_{\text{UL}}^2\mathbf{I}_M)^{-1}$ can be used if the regularized covariance matrices applied. Hence, the MMSE estimator in Eq.(2.34) can be rewritten in the form:

$$\hat{\mathbf{h}}_{\text{MMSE}}(\mathbf{y}, c) = \sqrt{p}\tau_p\hat{\mathbf{R}}(c)\left(p\tau_p\hat{\mathbf{R}}(c) + \sigma_{\text{UL}}^2\mathbf{I}_M\right)^{-1}\mathbf{y} \tag{2.43}$$

2.3 Hardware Impact on Channel Estimation

The channel estimation with perfect transceivers on the BS and UEs is well-discussed in almost literature. The case differs in practical systems, i.e., non-ideal hardware transceivers exist, and the signals are inevitably impaired [8]. In a massive MIMO system, non-ideal transceivers with hardware impairments can substantially impact the system's performance. Several issues in practical transceivers can arise either by the nonlinearity in amplifiers, quantization error in ADCs, or the in-phase/quadrature imbalance in mixers [8]. However, the major challenge posed by the non-ideal hardware impairments is the distortion in the received signals on each antenna's RF chain, which may result in a mismatch between the transmitted and received signals. Although some of the literature like [20, 21, 54] have analysed the general impact of these impairments, a residual impairment is still without compensation, while other literature like [23, 33] have dealt with each issue separately.

Since one RF chain per antenna is needed to design the massive MIMO system, the cost of implementing such a system will be very expensive. That is, if the BS has M_j antennas, then M_j copies of the equipment at each RF chain are required. However, to save costs, one can think of reducing the hardware quality of each component and following a cost-quality trade-off policy when implementing such a large system. Different modelling systems with analogue and/or digital algorithms, such as in [55, 56], and the references therein, have been issued to compensate for these impairments and reduce their impact on the performance of the communication system. But none of these algorithms has completely fixed these problems in practice. In other words, *residual impairments* [26] still exist in many of these modelling systems, raised either by inaccuracies in the design or by the devasting behaviour of certain hardware imperfections. Hence, this part would be concentrated on the residual distortions of the hardware issues and their effects in the channel estimation model, but first, we need to model these issues to generalise their effect on the massive MIMO system; hence, the Bussgang decomposition will be used to model these impairments in the following section.

2.3.1 Residual impairments model (Bussgangs' model)

Bussgang theorem [57] supposes that if the complex Gaussian random variable like $s \sim \mathbb{C}_{\mathcal{N}}(0,p)$ is fed into a distorting device with a non-linear memoryless function $g(.)$, then the output of this device is no longer be Gaussian, that is if $y = g(s)$, then y is a non-Gaussian random variable, and both s and y can be cross-correlated as follows:

$$y = \frac{\mathbb{E}\{ys^*\}}{p}s + \mu \tag{2.44}$$

where the term $\frac{\mathbb{E}\{ys^*\}}{p}$ is a constant ratio that defines the correlation ratio between the output y and input variable s. The additive factor μ is the distortion term that defines the impact of the hardware device on the input s. Based on the Bussgang analysis, if we assumed $\mathbb{E}\{ys^*\}/\mathbb{E}\{s^2\} = \sqrt{\rho}$ as a constant factor, and using this notation in Eq.(2.44), we can now generalize a consistent and analytical input-output model that fully characterizes the detrimental effect of residual defects over RF signals as follows:

$$y = \sqrt{\rho}\, s + \mu \tag{2.45}$$

where s and y are assumed to have the same p power. On the other hand, the additive term power is calculated from Eq.(2.45) as follows: $\{|\mu|^2\} = \mathbb{E}\{|y|^2\} - \rho\mathbb{E}\{|s|^2\} \rightarrow \{|\mu|^2\} = (1-\rho)p$. This means that the power of μ is related to the p power of the input with a scaling factor $(1-\rho)$. Thus, the additive distortion term is a power-dependent factor, unlike the classical additive noise σ^2, which is insensitive by input power p. However, the constant factor $\rho \in (0,1]$ describes the aspect of hardware quality and is used to measure the level of impairment of hardware devices at the transmitter and receiver as follows: when $\rho = 1$, the output will be the same as the input, that is, $y = s$, i.e., the distortion term $(\{|\mu|^2\} = 0)$, which is the case of ideal hardware. Otherwise, when $\rho = 0$, the distortion term will be Gaussian distributed as the input, i.e., $\mu \sim \mathbb{C}_{\mathcal{N}}(0,p)$. The latter case represents the worst-case condition since the input s is completely distorted at the output of the non-linear device.

2.3.2 Uplink transmission model with residual impairments

In this section, the joint impact of the hardware impairments of the BS and UE transceivers will be analyzed in the uplink transmission as follows: let the transmitted signal by arbitrary user equipment k located within arbitrary cell j given by $s_{jk} \sim \mathbb{C}_{\mathcal{N}}(0, p_{jk})$. Then, by following the model in Eq.(2.45), the circular-complex Gaussian s_{jk} will be defected through the non-ideal equipment of the user device and then transmitted over the channel as $\sqrt{\rho_t^{\mathrm{UE}}} s_{jk} + \mu_{jk}^{\mathrm{UE}}$ instead of s_{jk}. The scaling factor ρ_t^{UE} denotes the quality of the UE device, and for tractable notation, it is assumed to be the same for all UEs in the network. The additive factor μ_{jk}^{UE} determines the transmitter hardware distortion, which can have a distribution $\mu_{jk}^{\mathrm{UE}} \sim \mathbb{C}_{\mathcal{N}}(0, (1-\rho_t^{\mathrm{UE}}) p_{jk})$. In general, the received planner array signals that reach the M_j antennas at the BS j from all UEs in the network are as follows [11]:

$$\breve{\mathbf{y}}_j = \sum_{\ell=1}^{L} \sum_{i=1}^{K_\ell} \mathbf{h}_{\ell i}^{j} \left(\sqrt{\rho_t^{UE}} s_{\ell i} + \mu_{\ell i}^{UE} \right) \tag{2.46}$$

where L indicates the total cell number and K_ℓ denotes the maximum number of users residing in each cell. However, for a particular group of realization's channel $\{\mathbf{h}_{\ell i}^{j}\}$ within an arbitrary coherence block, the planner signal $\breve{\mathbf{y}}_j$ represents a complex Gaussian signal that has zero mean and a conditional correlation matrix which can be given by [11]:

$$\mathbb{E}\left\{ \breve{\mathbf{y}}_j \breve{\mathbf{y}}_j^H \mid \left\{ \mathbf{h}_{\ell i}^{j} \right\} \right\} = \sum_{\ell=1}^{L} \sum_{i=1}^{K_\ell} p_{\ell i} \mathbf{h}_{\ell i}^{j} \left(\mathbf{h}_{\ell i}^{j} \right)^H. \tag{2.47}$$

However, since $\mathbb{E}\{|\sqrt{\rho_t^{\mathrm{UE}}} s_{jk} + \mu_{jk}^{\mathrm{UE}}|^2\} = p_{\ell i}$, it could also now apply the model in Eq.(2.45) for the BS side to determine the effect of its hardware distortions. Thus, the signal $\breve{\mathbf{y}}_j$ is replaced by $\sqrt{\rho_t^{\mathrm{BS}}} \breve{\mathbf{y}}_j + \mu_{jk}^{\mathrm{BS}}$, where $\rho_t^{\mathrm{BS}} \in (0, 1]$ denotes the quality of the hardware of the BS j which is also assumed to be the same for all BSs, and $\mu_{jk}^{\mathrm{BS}} \in \mathbb{C}^{M_j}$ is the hardware distortion term of the BS j. For analytical convenience, it has been assumed that the distortion terms between the different RF chains that are attached to the M_j receive antennas are independent. In general, the marginal signal $\mathbf{y}_j \in \mathbb{C}^{M_j}$ received from the UL transmission at

BS j can be modelled as [11]:

$$\mathbf{y}_j = \sqrt{\rho_r^{\mathrm{BS}}} \left(\sum_{\ell=1}^{L} \sum_{i=1}^{K_\ell} \mathbf{h}_{\ell i}^{j} \left(\sqrt{\rho_t^{\mathrm{UE}}} s_{\ell i} + \mu_{\ell i}^{\mathrm{UE}} \right) \right) + \mu_j^{\mathrm{BS}} + \mathbf{n}_j. \tag{2.48}$$

where $\mathbf{n}_j$ is the receiver noise that has been added at the BS j. The signal $\mathbf{y}_j$ represents the complex baseband signal and will be used later in the coming subsection.

2.3.3 Bayesian channel estimates with residual impairments

The uplink system model of a scalar symbol transmitting given in Eq.(2.48) will be leveraged here to derive an uplink transmission model for transmitting a sequence of pilot symbols, as follows: let $s_{\ell i} = \sqrt{p_{\ell i}} \phi_{\ell i} \in \mathbb{C}^{\tau_p}$, indicates the transmitted pilot vector employed by the user i in the cell ℓ; hence, the UL transmission model in Eq.(2.48) over τ_p transmission instances can be rewritten as follows [11]:

$$\mathbf{Y}_j^p = \sqrt{\rho_r^{\mathrm{BS}}} \left(\sum_{\ell=1}^{L} \sum_{i=1}^{K_\ell} \mathbf{h}_{\ell i}^{j} \left(\sqrt{p_{\ell i} \rho_t^{\mathrm{UE}}} \phi_{\ell i}^{T} + \left(\mu_{\ell i}^{\mathrm{UE}} \right)^{T} \right) \right) + \mathbf{G}_j^{\mathrm{BS}} + \mathbf{N}_j^p \tag{2.49}$$

such that $\mathbf{N}_j^p \in \mathbb{C}^{M_j \times \tau_p}$ is the additive noise on the M_j receive antennas, each element is independent and identically distributed as $\sim \mathbb{C}_{\mathcal{N}}(0, \sigma_{\mathrm{UL}}^2)$, the term $\mathbf{G}_j^{\mathrm{BS}} \in \mathbb{C}^{M_j \times \tau_p}$ is the receiver hardware distortion matrix, each column of which takes the same distribution as μ_j^{BS} in Eq.(2.48), while $\mu_{\ell i}^{\mathrm{UE}} \in \mathbb{C}^{\tau_p}$ denotes the transmitter impairments, which includes τ_p transmitting realizations from $\mu_{\ell i}^{\mathrm{UE}}$ in Eq.(2.48). However, to predict the channel of an i^{th} user in an arbitrary cell ℓ at the BS j, the BS first correlates the signal $\mathbf{Y}_j^p$ with the desired UE's pilot sequence $\phi_{\ell i}$, yielding the processed signal $\mathbf{y}_{j\ell i}^p = \mathbf{Y}_j^p \phi_{\ell i}^*$. For instant, to estimate the k^{th} user's channel from arbitrary cell j at BS j, the result signal will be

written as follows [11]:

$$\mathbf{y}^p_{jjk} = \mathbf{Y}^p_j \boldsymbol{\phi}^*_{jk} = \underbrace{\sqrt{p_{jk}\rho_t^{\mathrm{UE}}\rho_r^{\mathrm{BS}}}\tau_p \mathbf{h}^j_{jk}}_{\text{“Desired pilot”}} + \underbrace{\sum_{(\ell,i)\in\mathscr{P}_{jk}\setminus(j,k)} \sqrt{p_{\ell i}\rho_t^{\mathrm{UE}}\rho_r^{\mathrm{BS}}}\tau_p \mathbf{h}^j_{\ell i}}_{\text{“interfering pilots”}} + \underbrace{\sum_{\ell=1}^{L}\sum_{i=1}^{K_\ell} \sqrt{\rho_r^{\mathrm{BS}}}\mathbf{h}^j_{\ell i}\left(\mu^{\mathrm{UE}}_{\ell i}\right)^T \boldsymbol{\phi}^*_{jk}}_{\text{“Transmitter distortion”}} + \underbrace{\mathbf{G}^{\mathrm{BS}}_j \boldsymbol{\phi}^*_{jk}}_{\text{“Receiver distortion”}} + \underbrace{\mathbf{N}^p_j \boldsymbol{\phi}^*_{jk}}_{\text{“Noise”}} \tag{2.50}$$

where $\tau_p = \boldsymbol{\phi}^T_{\ell i}\boldsymbol{\phi}^*_{jk}$ when $\boldsymbol{\phi}_{\ell i} = \boldsymbol{\phi}_{jk}$. The set $\mathscr{P}_{jk}$ is used for all users that are using pilot sequences as the k^{th} desired user sequence $\boldsymbol{\phi}_{jk}$. Since $\|\boldsymbol{\phi}_{jk}\|^2 = \tau_p$. The last three terms in Eq.(2.50) will be distributed as follows: $\mathbf{N}^p_j\boldsymbol{\phi}^*_{jk} \sim \mathbb{C}_{\mathscr{N}}(0, \sigma^2_{\mathrm{UL}}\tau_p\mathbf{I}_{M_j})$, $(\mu^{\mathrm{UE}}_{\ell i})^T\boldsymbol{\phi}^*_{jk} \sim \mathbb{C}_{\mathscr{N}}(0, \tau_p(1-\rho_t^{\mathrm{UE}})p_{jk})$, and the receiver distortion term $\mathbf{G}^{\mathrm{BS}}_j\boldsymbol{\phi}^*_{jk} \sim \mathbb{C}_{\mathscr{N}}(0_{M_j}, \tau_p\mathbf{D}_{j,\{\mathbf{h}\}})$ for given channel realizations. It is seen from Eq.(2.50) that the processed signal $\mathbf{y}^p_{jjk}$ is affected by the joint distortion from the transmitter and receiver hardware of all signalling transmissions in the entire network. Since these distortions are almost surely non-orthogonal to their pilot sequences, this implies that there is pilot contamination between every two UEs in the network. However, according to the Bussgang theorem, the processed signal $\mathbf{y}^p_{jjk}$ in Eq.(2.50) represents a non-Gaussian signal. Therefore, a sub-optimal estimator will be used here instead of the exact one. In other words, a linear minimum mean squared error (LMMSE) estimator with a little performance loss will be used instead of the optimal MMSE estimator. In general, based on the processed signal $\mathbf{y}^p_{jjk} = \mathbf{Y}^p_j\boldsymbol{\phi}_{\ell i}$ the linear LMMSE estimator [11] $\hat{\mathbf{h}}^j_{\ell i}$ can be given by:

$$\hat{\mathbf{h}}^j_{\ell i} = \sqrt{p_{\ell i}\rho_t^{\mathrm{UE}}\rho_r^{\mathrm{BS}}}\mathbf{R}^j_{\ell i}\Psi^j_{\ell i}\mathbf{y}^p_{j\ell i} \tag{2.51}$$

where $\mathbf{R}^j_{\ell i} \in \mathbb{C}^{M_j\times M_j}$ denotes the desired covariance matrix of $\mathbf{h}^j_{\ell i}$, and $\Psi^j_{\ell i}$ refers to the overall covariance matrix, which includes the effect by the additive receiver noise plus hardware distortions by all UEs during pilot signalling, it is

written as follows [11]:

$$\Psi_{\ell i}^{j} = \left(\sum_{(\ell,i)\in\mathscr{P}_{jk}\setminus(j,k)} p_{\ell i}\rho_t^{\mathrm{UE}}\rho_r^{\mathrm{BS}}\tau_p \mathbf{R}_{\ell i}^{j} + \sigma_{\mathrm{UL}}^2 \mathbf{I}_{M_j} + \sum_{\ell=1}^{L}\sum_{i=1}^{K_\ell} p_{\ell i}\left(1-\rho_t^{\mathrm{UE}}\right)\rho_r^{\mathrm{BS}}\mathbf{R}_{\ell i}^{j} + \sum_{\ell=1}^{L}\sum_{i=1}^{K_\ell} p_{\ell i}\left(1-\rho_r^{\mathrm{BS}}\right)\mathbf{D}_{\mathbf{R}_{\ell i}^{j}} \right)^{-1} \tag{2.52}$$

where

$$\mathbf{D}_{\mathbf{R}_{\ell i}^{j}} = \mathrm{diag}\left(\left[\mathbf{R}_{\ell i}^{j}\right]_{11}, \ldots, \left[\mathbf{R}_{\ell i}^{j}\right]_{M_j M_j} \right) \tag{2.53}$$

represents a diagonal matrix that contains diagonal elements from $\mathbf{R}_{\ell i}^{j}$. However, the estimator that gives a minimum mean squared error MSE $\mathbb{E}\{\|\mathbf{h}_{\ell i}^{j} - \hat{\mathbf{h}}_{\ell i}^{j}\|^2\}$ is the best estimator that can be used, where $\mathbf{h}_{\ell i}^{j} - \hat{\mathbf{h}}_{\ell i}^{j} = \tilde{\mathbf{h}}_{\ell i}^{j}$ is the estimation error with a covariance matrix $\mathbf{C}_{\ell i}^{j} = \mathbb{E}\{\tilde{\mathbf{h}}_{\ell i}^{j}(\tilde{\mathbf{h}}_{\ell i}^{j})^{\mathrm{H}}\}$, and can be written as [11]:

$$\mathbf{C}_{\ell i}^{j} = \mathbf{R}_{\ell i}^{j} - p_{\ell i}\rho_t^{\mathrm{UE}}\rho_r^{\mathrm{BS}}\tau_p \mathbf{R}_{\ell i}^{j}\Psi_{\ell i}^{j}\mathbf{R}_{\ell i}^{j} \tag{2.54}$$

we can also calculate the covariance matrix for the linear estimator $\hat{\mathbf{h}}_{\ell i}^{j}$ as follows:

$$\mathbb{E}\{\hat{\mathbf{h}}_{\ell i}^{j}(\hat{\mathbf{h}}_{\ell i}^{j})^{\mathrm{H}}\} = \mathbf{R}_{\ell i}^{j} - \mathbf{C}_{\ell i}^{j} = p_{\ell i}\rho_t^{\mathrm{UE}}\rho_r^{\mathrm{BS}}\tau_p \mathbf{R}_{\ell i}^{j}\Psi_{\ell i}^{j}\mathbf{R}_{\ell i}^{j} \tag{2.55}$$

It is noticed from Eq.(2.55) that the linear estimator covariance matrix is fully dependent on the second-order statistics of all UEs in the network, which means that the BS can determine the LMMSE estimator $\hat{\mathbf{h}}_{\ell i}^{j}$ only when it knows the correlation matrices $\mathbf{R}_{\ell i}^{j}$ and $\Psi_{\ell i}^{j}$ including the scaling quantity $p_{\ell i}\rho_t^{\mathrm{UE}}\rho_r^{\mathrm{BS}}\tau_p$. However, the BS in practice has imperfect or no prior knowledge of these statistics and needs to be estimated.

2.3.4 Error floor of residual impairments of single user channel

Let the uplink channel between the BS and single user equipment is denoted by $\mathbf{h} \sim \mathbb{C}_{\mathscr{N}}(0_M,\ \mathbf{R})$, where $\mathbf{R} \in \mathbb{C}^{M\times M}$, is the semi-definite covariance matrix. Following Eq.(2.54), the correlation matrix of the error for a single user at the BS will be given by:

$$\mathbf{C} = \mathbf{R} - p\ \rho_t^{\mathrm{UE}}\rho_r^{\mathrm{BS}}\tau_p \mathbf{R}\Psi\mathbf{R} \tag{2.56}$$

where

$$\Psi = \left(p \left(1 + \rho_t^{\mathrm{UE}} (\tau_p - 1) \right) \rho_r^{\mathrm{BS}} \mathbf{R} + p \left(1 - \rho_r^{\mathrm{BS}} \right) \mathbf{D}_{\mathbf{R}} + \sigma_{\mathrm{UL}}^2 \mathbf{I}_M \right)^{-1}. \quad (2.57)$$

The Ψ expression in Eq.(2.57) describes the distortion effect of the transceiver's hardware of both the BS and the single UE plus noise effect, which differs from the ideal hardware case that given at $\rho_t^{\mathrm{UE}} = \rho_r^{\mathrm{BS}} = 1$. Sub. Eq.(2.57) in Eq.(2.56) and take the limits when p goes to infinity, the correlation matrix of the error becomes [11]:

$$\mathbf{C} = \mathbf{R} - \mathbf{R} \left(\frac{1 + \rho_t^{\mathrm{UE}} (\tau_p - 1)}{\rho_t^{\mathrm{UE}} \tau_p} \mathbf{R} + \frac{\left(1 - \rho_r^{\mathrm{BS}} \right)}{\rho_t^{\mathrm{UE}} \rho_r^{\mathrm{BS}} \tau_p} \mathbf{D}_{\mathbf{R}} \right)^{-1} \mathbf{R}. \quad (2.58)$$

For a special case of $\mathbf{R} = \beta \mathbf{I}_M$, the correlation matrix of the error in Eq.(2.58) [11] simplifies to:

$$\mathbf{C} = \beta \mathbf{I}_M - \frac{\beta^2}{\frac{1 + \rho_t^{\mathrm{UE}} (\tau_p - 1)}{\rho_t^{\mathrm{UE}} \tau_p} \beta + \frac{(1 - \rho_r^{\mathrm{BS}})}{\rho_t^{\mathrm{UE}} \rho_r^{\mathrm{BS}} \tau_p} \beta} \mathbf{I}_M = \frac{\beta (1 - \rho_t^{\mathrm{UE}} \rho_r^{\mathrm{BS}})}{1 + \rho_t^{\mathrm{UE}} \rho_r^{\mathrm{BS}} (\tau_p - 1)} \mathbf{I}_M \quad (2.59)$$

Note the term in front of $\mathbf{I}_M$ in Eq.(2.59); it represents the joint impact of the error floor term that results from the residual hardware imperfections at the BS and UE. As seen in Eq.(2.59), it is constrained by the parameters of ρ_t^{UE}, ρ_r^{BS}, τ_p, and β. Hence, one can use these parameters to optimize the error floor at the BS.

2.3.5 Non-linear hardware distortion

This subsection tends to model the joint impact of the non-linear hardware distortion of the BS and UEs on the uplink transmission from K UEs in a single-cell scenario in the Rician fading environment. Some literature has modelled the impairments as a stochastic model [21, 25, 24], while others used a behavioural model [33, 23, 30]. A quasi-memoryless polynomial model is one of the behavioural models often used to describe the problems of nonlinear systems [29, 58]. However, although the nonlinear distortions are not independent between antennas at the BS, one can treat them as an independent colour noise and derive a distortion-aware Bayesian estimator [54]. Hence, in this subsec-

tion, the third-order distorting polynomial derivations at the BS and UEs are reviewed based on the literature [33]. On the other hand, two linear minimum mean squared error (LMMSE) Bayesian estimators aware and unaware of the hardware distortion are derived and evaluated for the effective channel between one antenna element from the M-array BS antennas and the single-antenna UE.

2.3.6 Uplink transmission model with nonlinear distortion

This work considers a single-cell scenario with a practical environment model from the 3GPP [59]. We have considered the Urban Micro-cell environment with large-scale fading parameters to describe the Rician fading channel model in [[59], Sections 1.2.1 and 1.2.2]. The BS with M antennas serves K single-antenna UEs in the cell. A constant block fading is assumed during the channel estimation process. Each channel between a single antenna and a single UE is defined as a scalar complex symbol during each independent transmitting element of the coherence block. Hence, during the information transmitting phase, the reached planner signal at the M array BS antennas is given by [39]:

$$\mathbf{r} = \sum_{k=1}^{K} \mathbf{g}_k s_k = \mathbf{G}\mathbf{s}, \tag{2.60}$$

where $\mathbf{r} = [r_1, \ldots, r_M]^T \in \mathbb{C}^M$ is the received vector at the M-antennas array, $\mathbf{s} = [s_1, \ldots, s_K]^T \in \mathbb{C}^K$, is the transmitted vector of symbols from all K UEs where $s_k \in \mathbb{C}$, and $\mathbf{G} = [\mathbf{g}_1, \ldots, \mathbf{g}_K] \in \mathbb{C}^{M \times K}$, stands for the uplink channels matrix, such that $\mathbf{g}_k = [g_{k1}, \ldots, g_{kM}]^T \in \mathbb{C}^M$, denotes the generated channel vector between the k^{th} user and the M array of the BS antennas when transmitting one complex-valued symbol by the UE. This channel vector is assumed to be a spatially uncorrelated Rician fading channel as follows [39]:

$$\mathbf{g}_k \sim \mathbb{C}_{\mathcal{N}} (\bar{\mathbf{g}}_k, \beta_k \mathbf{I}_M) \tag{2.61}$$

with $\bar{\mathbf{g}}_k$, is the LOS part and $\beta_k \mathbf{I}_M$, denotes the spatially uncorrelated covariance matrix that describes the large-scale fading effect of the Rician environment. It is proved in [8] that the third-order term of the polynomial distortion effect dominates the nonlinear behavioural model. Depending on this outcome, the

following subsections will describe the third-order distorting polynomial on the UE and the BS as follows:

2.3.6.1 UE distorting polynomial

Here, the memoryless quasi-polynomial with third-order is exploited to model the hardware distortion impact of the UE on the transmitted symbol [23] as follows:

$$s_k = \sqrt{\eta_k}\left(\tilde{u}_0 \varsigma_k + \tilde{u}_1 \left|\varsigma_k\right|^2 \varsigma_k\right) \tag{2.62}$$

where s_k is the distorted symbol that would be transmitted with a scaling factor $\sqrt{\eta_k}$. This factor used to maintain the sending power of the signal always equal$\mathbb{E}\{|s_k|^2\} = p_k$, the $(u0, u1)$, are polynomial coefficients, which are complex-valued in the case of quasi-memoryless and real-valued in the case of the memoryless polynomial. $\varsigma_k \in \mathbb{C}$ is the actual symbol with power $\mathbb{E}\{|\varsigma_k|^2\} = 1$ and zero mean normal distribution.

2.3.6.2 BS distorting polynomial

The same quasi-memoryless polynomial will be used here to model the impact of the BS hardware distortions on the received signal in Eq.(2.60) [23] as follows:

$$z_m = \tilde{b}_{0m} r_m + \tilde{b}_{1m} |r_m|^2 r_m, \tag{2.63}$$

for $m = 1, 2, \ldots, M$, such that z_m, is the impaired signal that is received on the m^{th} antenna element. The coefficients (b_{0m}, b_{1m}) are the BS's distorting polynomial coefficients, which are complex. However, after adding receiver noise at the m^{th} BS antenna, the complex signal will be given as $y_m = z_m + n_m$. In a vector notation that considers all M antennas at the BS, the baseband signal will be

$$\mathbf{y} = \mathbf{z} + \mathbf{n}, \tag{2.64}$$

such that $\mathbf{y} = [y_1, \ldots, y_M]^T \in \mathbb{C}^M$, $\mathbf{z} = [z_1, \ldots, z_M]^T \in \mathbb{C}^M$, and $\mathbf{n} \sim \mathbb{C}_{\mathcal{N}}(\mathbf{0}_M, \sigma^2 \mathbf{I}_M)$.

2.3.6.3 Effective Uplink channels and hardware distortion correlation

Consider the general block system in Fig. 2.4 between the transmitted sym-

bol and baseband received signal vectors; the Bussgang decomposition that models the signal from end-to-end of the system, i.e., from ς to $\mathbf{y}$ is as follows [33]:

$$\mathbf{y} = \mathbf{C}_{\mathrm{y}\varsigma}\mathbf{C}_{\varsigma\varsigma}^{-1}\varsigma + \mu, \tag{2.65}$$

such that $\mathbf{C}_{\mathrm{y}\varsigma} \in \mathbb{C}^{M\times K}$, represents the covariance matrix between the transmit

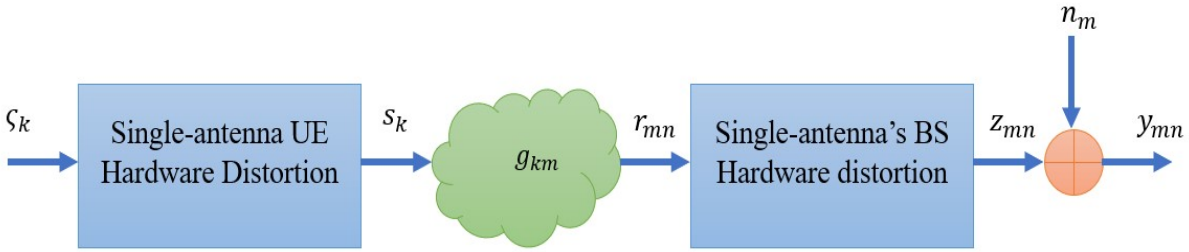

Figure 2.4: General block diagram of the uplink transmission system

vector ς and the received vector $\mathbf{y}$, it is given by $\mathbf{C}_{\mathrm{y}\varsigma} = \mathbb{E}_{|\mathbf{G}.}\{\mathbf{y}\,\varsigma^H\}$, the matrix $\mathbf{C}_{\varsigma\varsigma} \in \mathbb{C}^{K\times K}$, denotes the correlation matrix of the transmitted symbol vector, that is $\mathbf{C}_{\varsigma\varsigma} = \mathbb{E}_{|\mathbf{G}.}\{\varsigma\,\varsigma^H\}$. If the transmitted symbols have zero mean and unity variance, the symbol correlation matrix will be $\mathbf{C}_{\varsigma\varsigma} = \mathbf{I}_K$. Hence, the signal term in Eq.(2.65) becomes $\mathbf{C}_{\mathrm{y}\varsigma}\varsigma$; that is why, $\mathbf{C}_{\mathrm{y}\varsigma}$ is denoted by the effective channel covariance matrix, which is like the case of no distortion. The term μ is the additive term that characterises the impact of the hardware distortion on the UEs and BS. The $(m,k)^{th}$ elements of the effective channel covariance matrix are derived in [33] and given as follows:

$$\begin{aligned}\left[\mathbf{C}_{\mathrm{y}\varsigma}\right]_{mk} =& \tilde{b}_{0m}\tilde{g}_{km}\left(\tilde{u}_0 + \zeta_4\tilde{u}_1\right) + \tilde{b}_{1m}\left|\tilde{g}_{km}\right|^2\tilde{g}_{km} \\ &\times\left(\zeta_{10}C_{1,1,1}^{\mathrm{UE}} + 2\zeta_8 C_{1,1,0}^{\mathrm{UE}} + \zeta_8 C_{1,0,1}^{\mathrm{UE}} + 2\zeta_6 C_{0,0,1}^{\mathrm{UE}} + \zeta_6 C_{0,1,0}^{\mathrm{UE}} + \zeta_4 C_{0,0,0}^{\mathrm{UE}}\right) + \\ &2\tilde{b}_{1m}\tilde{g}_{km}\left(\tilde{u}_0 + \zeta_4\tilde{u}_1\right) \times \left(\zeta_6 C_{1,1}^{\mathrm{UE}} + \zeta_4 C_{1,0}^{\mathrm{UE}} + \zeta_4 C_{0,1}^{\mathrm{UE}} + C_{0,0}^{\mathrm{UE}}\right) \times \sum_{l=1,l\neq k}^{K}\left|\tilde{g}_{lm}\right|^2.\end{aligned} \tag{2.66}$$

where $\tilde{g}_{km} = g_{km}\sqrt{\eta_k}$ is the scaled channel gain, $C_{r_1,r_2}^{\mathrm{UE}} = \tilde{u}_{r_1}\tilde{u}_{r_2}^*, C_{r_1,r_2,,r_3}^{\mathrm{UE}} = \tilde{u}_{r_1}\tilde{u}_{r_2}^*\tilde{u}_{r_3}$ are simplified terms of the UE coefficients multiplication, $\zeta_l = \mathbb{E}\{|\varsigma_k|^l\}$, denotes the even expectation of the actual symbol ς_k.

Now, to derive the distortion covariance matrix, the following tractable way can be followed: from Eqs. (2.64) and (2.65), $\mu = \mathbf{y} - \mathbf{C}_{\mathrm{y}\varsigma}\mathbf{C}_{\varsigma\varsigma}^{-1}\varsigma$, and $\mathbf{y} = \mathbf{z}+\mathbf{n}$, then $\mu = \mathbf{z}+\mathbf{n}-\mathbf{C}_{\mathrm{y}\varsigma}\mathbf{C}_{\varsigma\varsigma}^{-1}\varsigma$, now taking $\mathbb{E}_{|\mathbf{G}}\{(.)^2\}$ for both sides and substituting

$\mathbf{C}_{\varsigma\varsigma} = \mathbb{E}_{|\mathbf{G}}\{\varsigma\ \varsigma^{\mathrm{H}}\} = \mathbf{I}_K$, then:

$$\mathbf{C}_{\mu\mu} = \mathbb{E}_{|\mathbf{G}}\{\mu\ \mu^{\mathrm{H}}\} = \mathbf{C}_{zz} + \sigma^2\mathbf{I}_M - \mathbf{C}_{\mathrm{y}\varsigma}\mathbf{C}_{\mathrm{y}\varsigma}^{\mathrm{H}}, \tag{2.67}$$

where $\mathbf{C}_{zz} = \mathbb{E}_{|\mathbf{G}}\{zz^{\mathrm{H}}\}, \mathbf{C}_{nn} = \mathbb{E}_{|\mathbf{G}}\{nn^{\mathrm{H}}\}$, and $\mathbf{C}_{\mathbf{C}_{\mathrm{y}\varsigma}\mathbf{C}_{\mathrm{y}\varsigma}} = \mathbb{E}_{|\mathbf{G}}\{\mathbf{C}_{\mathrm{y}\varsigma}\mathbf{C}_{\mathrm{y}\varsigma}^{\mathrm{H}}\}$. However, $\mathbf{C}_{\mu\mu}$ will be calculated using the Monte Carlo simulation since there is very hard and messy calculations will be arises when determining the covariance matrices $\mathbf{C}_{zz}$ and $\mathbf{C}_{\mathrm{y}\varsigma}$, according to ref.[33].

2.3.7 Bayesian channel estimates with nonlinear distortion

During the pilot transmission phase, the effective uplink channels will be estimated using the linear minimum mean squared error (LMMSE) Bayesian estimators. Let $\phi_k \in \mathbb{C}^{\tau_p}$ denotes the transmitted pilot sequence of any arbitrary k^{th} UE with τ_p length. Let $\tilde{\varphi}_{kn}$ be the distorted version of the original pilot symbol as follows:

$$\tilde{\varphi}_{kn} = \tilde{u}_0\varphi_{kn} + \tilde{u}_1\left|\varphi_{kn}\right|^2\varphi_{kn}, \tag{2.68}$$

for $n = 1,\ldots,\tau_p$, and would be transmitted with controlled power as $\sqrt{\tilde{\eta}_k}\tilde{\phi}_{kn}$, such that $\tilde{\eta}_k = \frac{\tau_p p_k}{\sum_{n=1}^{\tau_p}\left|\tilde{\phi}_{kn}\right|^2}$. The received baseband pilot signal before adding noise can be given in general formula of the third order quasi-memoryless polynomial as follows [23]:

$$z_{mn}^{\mathrm{p}} = \sum_{l=0}^{1}\tilde{b}_{lm}\left(\sum_{k=1}^{K} g_{km}\sqrt{\tilde{\eta}_k}\tilde{\phi}_{kn}\right) \times \left|\sum_{k=1}^{K} g_{km}\sqrt{\tilde{\eta}_k}\tilde{\phi}_{kn}\right|^{2l}, \tag{2.69}$$

for $n = 1,\ldots,\tau_p$ and after adding noise will be

$$\mathbf{y}_m^p = \mathbf{z}_m^p + \mathbf{n}_m^p \tag{2.70}$$

such that $\mathbf{y}_m^p, \mathbf{z}_m^p, \mathbf{n}_m^p$, and $\mathbf{n}_m^p \sim \mathbb{C}_{\mathcal{N}}\left(\mathbf{0}_{\tau_p}, \sigma^2\mathbf{I}_{\tau_p}\right)$, are the vectors at the m^{th} antenna index and are time-dependent with the pilot elements of length τ_p. The next subsections will discuss two Bayesian estimates that would be aware and unaware of the hardware distortions on the UE and the BS.

2.3.7.1 Hardware Distortion-Aware LMMSE estimate (DA-LMMSE)

In this subsection, the estimates of the covariance elements in Eq.(2.66) will be presented here using the distortion-aware LMMSE estimate [32]:

$$[\hat{\mathbf{C}}_{\mathrm{y}\varsigma}]_{mk} = [\bar{\mathbf{C}}_{\mathrm{y}\varsigma}]_{mk} + \mathbf{C}_{[C_{\mathrm{y}\varsigma}]\mathrm{y}_m^{\mathrm{p}}} \mathbf{C}^{\mathrm{p}}_{\mathrm{y}_m^p \mathrm{y}_m^{-1}} \left(\mathbf{y}_m^{\mathrm{p}} - \bar{\mathbf{y}}_m^{\mathrm{p}}\right) \tag{2.71}$$

for $k = 1,\ldots,K,\ m = 1,\ldots,M$, where $\bar{\mathbf{y}}_m^p = \mathbb{E}\{\mathbf{y}_m^p\} \in \mathbb{C}^{\tau_p}, \bar{\mathbf{C}}_{\mathrm{y}\varsigma} = \mathbb{E}\{\mathbf{C}_{\mathrm{y}\varsigma}\} \in \mathbb{C}^{M\times K}$, describes the LOS part effect, which is included in the Rician fading case, and $\mathbf{C}_{[C_{\mathrm{y}\varsigma}]_{mk}\mathrm{y}_m^p} = \mathbb{E}\{\left(\left[\mathbf{C}_{\mathrm{y}\varsigma}\right]_{mk} - \left[\bar{\mathbf{C}}_{\mathrm{y}\varsigma}\right]_{mk}\right)\left(\mathbf{y}_m^p - \bar{\mathbf{y}}_m^p\right)^{\mathrm{H}}\} \in \mathbb{C}^{1\times\tau_p}$, and $\mathbf{C}_{\mathrm{y}_m^p\mathrm{y}_m^p} = \mathbb{E}\{(\mathbf{y}_m^p - \bar{\mathbf{y}}_m^p)(\mathbf{y}_m^p - \bar{\mathbf{y}}_m^p)^{\mathrm{H}}\} \in \mathbb{C}^{\tau_p\times\tau_p}$, are the covariance matrices of the received signal-effective channel and received-received signal, respectively.

2.3.7.2 Hardware Distortion-unAware LMMSE estimate (DuA-LMMSE)

In this subsection, the LMMSE estimate will be equivalent to the LMMSE estimate under ideal hardware conditions, which means zero polynomial coefficients will be in Eqs. (2.62) and (2.63). Hence, the DuA-LMMSE estimate can take the formula [32]:

$$\hat{g}_{km} = \bar{g}_{km} + \frac{\sqrt{p_k}\beta_k}{\tau_p p_k \beta_k + \sigma^2} \times \left(\phi_k^H \mathbf{y}_m^p - \sqrt{p_k}\tau_p \bar{g}_{km}\right) \tag{2.72}$$

Chapter 3

Proposed Models

3.1 Introduction

This chapter presents the proposed models in this thesis that tends to estimate the channels under the transceiver's hardware consideration. The chapter will be subdivided into two parts according to the transceivers' hardware quality of the BS and UEs: The first part considers the uplink channel estimation model when ideal hardware transceivers are assumed on both the BS and UEs, see Fig.3.1, both transceivers are assumed as a linear filter model. The second part of this chapter estimates the channels under non-ideal hardware transceivers; see Fig.3.2 where non-linear filters are assumed at both transceivers of the BS and UEs.

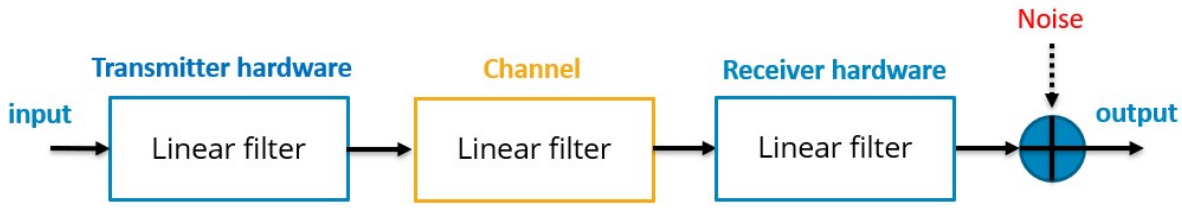

Figure 3.1: Block diagram of the ideal hardware system

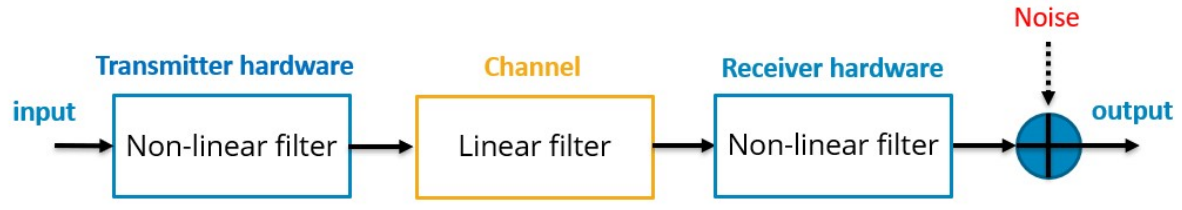

Figure 3.2: Block diagram of the non-ideal hardware system

3.2 Part I: Channel Estimation with Ideal Transceivers

This part presents the channel estimation model when the transceiver's hardware is assumed ideal on both the BS and UEs. It uses the Bayesian MMSE estimator in Eq.(2.21) to estimate uplink channels and aims to reduce The MMSE computational complexity at the BS. Two types of channel models are used here to be estimated: the local scattering channel model with completely known statistical information at the BS and the practical channel model when the BS does not have any knowledge about the channel statistical information. A single-cell single-active UE scenario is used here, and spatial correlation characteristics' impact on the estimated channels is considered to reduce the computational complexity for both channel models.

3.3 Proposed Model-1

The channel covariance matrix describes the spatial characteristics of the channel at the BS, and its knowledge plays a crucial role in investigating an optimal channel estimate. However, the covariance matrix of the local scattering model can be calculated in Eq.(2.15) depending on several parameters like the azimuth angle, the standard deviation of the nominal angle, the distance between antennas on the array, etc. But, in practice, the BS has incomplete knowledge about these statistics of the covariance matrices of the channels and needs to be estimated. The common approach to capturing such matrices is by evaluating sample covariance matrices, in which the BS needs to see many pilot signal observations from each UE in the network. However, the estimated channels produced by this method are full-rank covariance matrices with large dimensions. Since these large matrices must be manipulated every millisecond, their computational complexity will be a bottleneck problem. Hence, promising approaches have been suggested in [17, 16, 14, 60] to estimate such practical large-dimensional matrices. On the other hand, the Bayesian MMSE estimator comes with a large computational complexity compared to other conventional estimators.

Motivated by these facts, our proposed model-1 in Fig. 3.3 uses the reduced

version eigenstructure of the covariance matrices of the full rank channel models to estimate the uplink channels in the single cell single active user scenario. It leverages the channel's spatial correlation characteristics and the possibilities of linear algebra to decompose the channel covariance matrices into their eigenstructure and then truncate their eigenvalues with the help of SVD truncation rules.

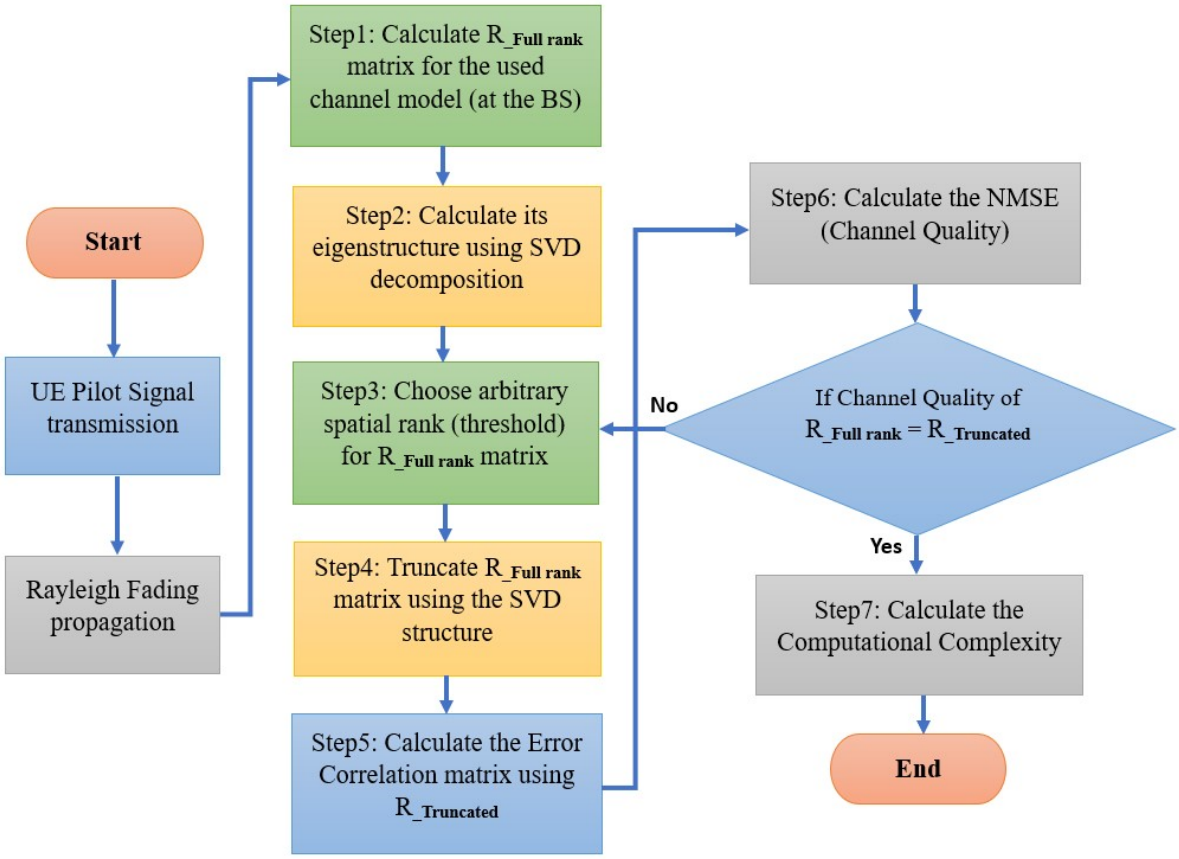

Figure 3.3: Proposed model-1 of channel estimation

The proposed model-1 in Fig.3.3 will be used for both channel models as follows:

3.3.1 Using local scattering channel model

The following algorithm steps in Fig.3.3 can be applied to the local scattering channel model as follows:

- **Step 1:** Calculate the channel covariance matrix of the local scattering model given in Eq.(2.15).

- **Step 2:** Decompose the full-rank covariance matrix calculated in Step 1 into its eigenstructure with three subspaces using the following SVD decomposition method:

  Reduced Eigenstructure of R: According to linear algebra rules, if $\mathbf{R}$ is a positive semi-definite square matrix, then its eigenvalue decomposition can be given by the SVD structure as follows: $\mathbf{R} = \mathbf{U}\mathbf{D}\mathbf{U}^{\mathrm{H}}$. Since $\mathbf{R}$ is a large matrix with M by M dimensions and full-rank $r = M$, then it can be reduced the eigenstructure of $\mathbf{R}$ into an arbitrary choosable rank $\mathbf{r} \leq M$ as follows:

$$\begin{aligned}\mathbf{R} &= \mathbf{U}\mathbf{D}\mathbf{U}^{\mathrm{H}} \\ &= [\mathbf{U}_r \mid \mathbf{U}_{r+1\to M}] \begin{bmatrix} \mathbf{D}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} [\mathbf{U}_r \mid \mathbf{U}_{r+1\to M}]^{\mathrm{H}} \\ &= [\mathbf{U}_r \mid \mathbf{U}_{r+1\to M}] \begin{bmatrix} \mathbf{D}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{U}_{\mathbf{r}}^{\mathrm{H}} \\ \mathbf{U}_{r+1\to M}^{\mathrm{H}} \end{bmatrix} = [\mathbf{U}_r][\mathbf{D}_r] \begin{bmatrix} \mathbf{U}_{\mathbf{r}}^{\mathrm{H}} \end{bmatrix}\end{aligned} \tag{3.1}$$

  where $\mathbf{U}_r$ contains the first r columns of $\mathbf{U}$ and $\mathbf{D}_r$ is the first $r \times r$ block of $\mathbf{D}$. The last term is the reduced eigenstructure of $\mathbf{R}$ and is sometimes denoted by $\tilde{\mathbf{R}} \approx \tilde{\mathbf{U}}\tilde{\mathbf{D}}\tilde{\mathbf{U}}^{\mathrm{H}}$.

- **Step 3:** Choose successive arbitrary rank values to produce different eigenspaces matrices of the eigenstructure calculated in Step 2.

- **Step 4:** Use the eigenspaces matrices evaluated in Step 3 with successive spatial ranks to reduce the full-rank matrix in Step 1 by applying the following SVD instruction $\{\mathbf{R}_{\text{truncated}} = \mathbf{U}(1 : \text{spatial rank}, 1 : \text{spatial rank}) * \mathbf{D}(1 : \text{spatial rank}, 1 : \text{spatial rank}) * \mathbf{U}^{H}(1 : \text{spatial rank}, 1 : \text{spatial rank}\}$' in Matlab to produce more spatial channels with arbitrary eigenvalue subspaces.

- **Step 5:** Calculate the error correlation matrix given in Eq.(2.35) in terms of the resultant spatially reduced matrices in Step 4 to show the impact of the spatial correlation characteristics on the MMSE computational complexity.

- **Step 6:** Check the quality of the estimated channels by evaluating the NMSE versus the effective SNR and ASD parameters to show how the

spatially reduced channels converged to the full-rank channel.

- **Step 7:** Compute the computational complexity of the MMSE estimator in Section 3.3.3 in terms of the spatially reduced covariance matrices and compare the results with the full-rank computational complexity.

3.3.2 Using practical channel model

Motivated by the fact that the practical channels estimated by the sample covariance matrix are large dimensional matrices; hence, the algorithm steps of the proposed model-1 shown in Fig. 3.3 are applied here to reduce the eigenstructure of such large matrices as well as the computational complexity at the Bayesian MMSE estimator as follows:

- **Step 1:** Calculate the Practical channel covariance matrix using the sample covariance matrix in Eq.(2.39)
- **Step 2:** Decompose it into its eigenstructure using the SVD structure $\mathbf{R} = \mathbf{U}\mathbf{D}\mathbf{U}^{\mathrm{H}}$.
- **Step 3:** Choose the spatial rank using the following well-known SVD thresholding rules:

  SVD thresholding rules: Let $\mathbf{R}$ be a semi-definite matrix with large M by M dimensions; two well-known thresholding rules can be used to reduce its eigenstructure as follows: The Truncated SVD (TSVD) rule [61], which is given by the formula:

$$\mathbf{R}_{\mathrm{TSVD}} = \sum_{i=1}^{r} \lambda_i u_i v_i^{'}, \tag{3.2}$$

  where r refers to the rank of the matrix $\mathbf{R}$, λ_i denotes the individual eigenvalue, $u_i v_i^{'}$ are the left and right columns in the eigenvector matrices, respectively. Since the expression in Eq.(3.2) uses the rank of the matrix to find its approximation, we suggest using arbitrary Successive-Empirical Thresholds (SETSVD) used for truncation to get more spatial characteristics for the estimated channels.

The second one is the Singular Value Hard Threshold (SVHT) rule [61], which has the formula:

$$\mathbf{R}_{\text{SVHT}} = \sum_{i=1}^{M} \Gamma_{\mathscr{H}}(\lambda_i, \mathscr{T})\, u_i v_i^{'} \tag{3.3}$$

where $\Gamma_{\mathscr{H}}(\lambda_i, \mathscr{T}) = \{\lambda_i : \lambda_i \geq \mathscr{T}\}$ denotes the hard thresholding rule. The optimal SVHT threshold for any arbitrary $\mathbf{R}$ is given by: $\mathscr{T} = 2.858\, y_{med}$, where $y_{med} =$ is the median of the eigenvalues. We apply both thresholding rules to the full-rank sample covariance matrices to show which rule provides better spatial characteristics over another.

- **Step 4:** Truncate the eigenstructure of the resultant eigenspace matrices in Step 3 to produce more spatial channels.
- **Step 5:** Calculate the error correlation matrix given in Eq.(2.35) in terms of the truncated matrices in Step 4 to show the impact of the spatial correlation characteristics of the reduced channels.
- **Step 6:** Evaluate the NMSE in Eq.(2.42) versus the number of samples (observations) at the BS to show how the spatially reduced channels converged to the full-rank channel.
- **Step 7:** Compute the computational complexity of the resultant channels and compare it with the full-rank channels using the following MMSE computations:

3.3.3 Computational complexity

The main drawbacks of having several antennas at the BS lie in the fact that the digital processor of the baseband signal will handle a large number of incoming observation signals. On the other hand, since the BS needs to estimate the uplink channels over different coherence blocks frequently; hence, the computational complexity of the MMSE estimate at the BS will be very large and intensive in terms of Floating Points Operations (FLOPs) number.

3.3.3.1 Standard MMSE computational complexity

The standard computational complexity of the MMSE estimator ($\hat{\mathbf{h}}^j = \mathbf{R}\Psi\mathbf{Y}_j^p\Phi^*$) at the BS j that has M_j antennas is given in [11] as follows:

- The correlating of $\mathbf{Y}_j^p$ with an arbitrary pilot sequence at the BS, i.e., ($\mathbf{Y}_j^p\Phi^* = \mathbf{y}$) costs $M_j\tau_p$ FLOPs.
- The multiplication of $\mathbf{y}$ with $\mathbf{R}$ and Ψ matrices costs $(4M_j^3 - M_j)/3$ FLOPs. This multiplication is precomputed once at the BS for all UEs using the same pilot sequences.
- The additional UE computations for each new user in the network will cost M_j^2.

Hence, the total cost can be given by: $(4M_j^3 - M_j)/3 + M_j^2 + M_j\tau_p$ FLOPs.

3.3.3.2 Proposed MMSE computational complexity

In this subsection, the Factor-Solve method will be proposed here to reduce the total cost of computations of the MMSE Bayesian estimator. First, the definition of the FLOPs number is discussed, and then the factoring procedure is described as follows:

FLOPs computations definition: Counting the Floating-Point Operations (FLOPs) of any arithmetic operations determines how the system is complex. Each flop point is defined as one addition, multiplication, subtraction, or division of two floating-point numbers and can be expressed as a polynomial function [62]. Also, the FLOPs can be simplified by ignoring the lower terms and keeping the leading terms (the terms with the largest exponents). For example, to evaluate the number of FLOPs of the inner product $\mathbf{x}^{\mathrm{T}}\mathbf{y}$ of the two vectors $\mathbf{x}$, $\mathbf{y} \in \mathbb{R}^M$. This process requires M multiplication and $M-1$ addition, i.e., $2M-1$ FLOPs are needed, which can be simplified to $2M$ FLOPs by keeping only the leading term. Hence, we can say that the inner product costs $2M$ FLOPs only. However, for matrix-vector multiplication with (M by M) and (M by 1), respectively, this requires $2M^2$ FLOPs, which result from the inner product of the M components of $\mathbf{y}$ with each row of the matrix.

Factor-Solve method: This method contains two essential steps: matrix factorization and backward solving. We used the matrix eigenstructure decomposition as the first factorization step and the FLOPs computation method as the second backwards-solving step. The computational complexity analysis of the Bayesian MMSE estimator in Eq.(2.34) under the full and reduced rank matrices will be shown as follows:

- For factorization: it will use the eigenstructure of $\mathbf{R} = \mathbf{UDU}^{\mathrm{H}}$ with $\mathrm{rank}(\mathbf{R}) = M$; hence, the MMSE will be given by:

$$\hat{\mathbf{h}}_{\mathrm{MMSE}} = \sqrt{p}\tau_p \mathbf{UDU}^{\mathrm{H}} \left(p\tau_p \mathbf{UDU}^{\mathrm{H}} + \sigma^2_{\mathrm{UL}} \mathbf{I}_M \right)^{-1} \mathbf{y} \tag{3.4}$$

- For backward solving: we start with inverse matrix computations:
  - Solving $\mathbf{DU}^{\mathrm{H}} = (M \text{ by } M) \times (M \text{ by } M) = (M \text{ by } M)$ matrix, which costs $(2M^2)$ FLOPs since D is a diagonal matrix.
  - Solving $\mathbf{U}(\mathbf{DU})^{\mathrm{H}} = (M \text{ by } M) \times (M \text{ by } M) = (M \text{ by } M)$ matrix, which costs $(2M^3)$ FLOPs.
  - Multiplication $(M \text{ by } M)$ matrix by $p\tau_p$ costs (M^2) FLOPs.
  - Addition $(M \text{ by } M)$ matrix with $\sigma^2_{\mathrm{UL}} \mathbf{I}_M$ costs (M^2) FLOPs.
  - –
  - Let $\left(p\tau_p \mathbf{UDU}^{\mathrm{H}} + \sigma^2_{\mathrm{UL}} \mathbf{I}_M \right) = \Psi$; then, let the quantity $\Psi^{-1}\mathbf{y}$ in the MMSE estimator to be equal to a dummy vector like $\mathbf{w}$, i.e., $\Psi^{-1}\mathbf{y} = \mathbf{w} \to \Psi\mathbf{w} = \mathbf{y}$, this is a set of M linear equations of M variables, which have cost of $(2/3)M^3$ FLOPs [62].
  - Now, the computations of the multiplication $\Psi^{-1}\mathbf{y}$ by $\sqrt{p}\tau_p \mathbf{UDU}^{\mathrm{H}}$ will cost $4M^2 + M$. Hence, the total number of FLOPs of the Bayesian MMSE estimator will be $(2/3)M^3 + 4M^2 + M$. Furthermore. Keeping only the leading terms, the total cost of computations will be $(2/3)M^3$ FLOPs, which has $\mathcal{O}(M^3)$ order of complexity.

A similar procedure will be followed for measuring the complexity of MMSE with reduced-rank matrices; the difference here is only by applying the reduced-rank covariance matrices. Thus, the total number of FLOPs will be $(2/3)r^3$,

which has $\mathcal{O}(r^3)$ order of complexity. Hence, the MMSE computations are now rank-dependent instead of M-dependent in terms of FLOPs complexity, which makes faster computations by an $(M-r)$ faster than before.

3.4 Part II: Channel Estimation with non-Ideal Transceivers

This part proposes channel estimation models that consider the non-ideal hardware on the transceivers of the BS and UEs. The first proposed model (proposed model-2 in this Chapter) estimates the practical channels under residual impairments (see Section 2.3.1) of the hardware issues using the sample covariance matrix, the Bussgang decomposition model, and a convex optimization procedure. The second proposed procedure (proposed model-3 in this Chapter) uses deep learning methods to learn the non-linearity distortion issue of the hardware components (see Section 2.3.5) and estimate the uplink effective channels from K UEs in the network. Our results in the first proposed model are validated and compared with the findings in [11, 21], while the results in the second part are more refined based on the findings in [33].

3.5 Proposed Model-2

In motivation with the fact that the BS can estimate the channel $\mathbf{h}_{\ell i}^{j}$ only when it knows the correlation matrices $\mathbf{R}_{\ell i}^{j}$ and $\mathbf{\Psi}_{\ell i}^{j}$ (see Section 2.2.7) which are in practice unknown at the BS. In this section, we have proposed a channel estimation model that takes into account the effects of this imperfection knowledge of the covariance matrices of the UEs at the BS while considering the joint impact of the residual impairments in both the BS and UEs. A massive MIMO system with the Wyner model (see Appendix B.3) of a single user per cell scenario is considered here.

However, due to the fact that the predicted channels with such impairments have high error floor, especially at high SNR [11],[20], we suggest first estimating the channel of the k^{th} UE practically at the BS based on the signal observations and the Law of the large numbers. Then, a convex-like optimization procedure is used to regularize the estimated channel and investigate a robust

channel estimate with minimum error. On the other hand, the impact of the residual distortion in the case of the local scattering model is derived here to validate the results of the proposed procedure, which will be used with the ideal hardware case as a quality checker for the estimated practical channels under residual hardware impairments.

3.5.1 Algorithm steps

The next algorithm steps will describe the proposed model-2 in this Chapter as follows:

- **Step 1:** Calculate the Standard Hardware Distortions impact using the error correlation matrix in Eq.(2.58) in terms of the Local scattering covariance matrix model with ASD $= 10^o$ and a nominal angle ranging from 0 to 360^o. Then calculate the NMSE $= \mathrm{tr}(\mathbf{C})/\mathrm{tr}(\mathbf{R})$ versus the effective SNR at $\tau_p = 10$ length and distinct hardware constants ($\rho = 0.95, 0.99,$ and 1) to demonstrate how the hardware imperfection impacts the estimated channel.

- **Step 2:** Estimate the practical channel using the sample covariance matrix method based on the law of large numbers of many observations at the BS as follows:

1. $(\hat{\Psi}^j_{\ell i})^{\text{sample}}$ estimation: in this part, the BS needs to see many observations from $\mathbf{y}^p_{j\ell i}$ as $\mathbf{y}^p_{j\ell i}[1], ..., \mathbf{y}^p_{j\ell i}[N_\Psi]$, where N_Ψ refers to the total number of signal observations, assumed the coherence time is large enough to include many channel realizations, then the sample matrix can then be given by:

$$(\hat{\Psi}^j_{\ell i})^{\text{sample}} = \frac{1}{N_\Psi} \sum_{n=1}^{N_\Psi} \mathbf{y}_{j\ell i}[n] (\mathbf{y}_{j\ell i}[n])^{\mathrm{H}}. \tag{3.5}$$

and according to the large numbers law, every element in the sample covariance matrix will converge to the corresponding element in the desired matrix when $N_\Psi \to \infty$, then

$$\left[\frac{1}{N_\Psi} \sum_{n=1}^{N_\Psi} \mathbf{y}_{j\ell i}[n] (\mathbf{y}_{j\ell i}[n])^{\mathrm{H}} \right]_{m,n} \to \left[\Psi^j_{\ell i} \right]. \tag{3.6}$$

2. $(\hat{\mathbf{R}}_{\ell i}^{j})^{\text{sample}}$ estimation: in this part, the same approach above can be followed with two stages of a procedure. In the first stage, the BS first estimate the $(\hat{\Psi}_{\ell i}^{j})^{\text{sample}}$ that is determined for both UEs in the network, including the desired UE. In the second stage, the BS gives extra pilot samples to the interfering user and uses its observations to estimate $(\hat{\mathbf{R}}_{\ell i}^{j})^{\text{sample}}$ as follows:

$$(\hat{\mathbf{R}}_{\ell i}^{j})^{\text{sample}} = (\hat{\Psi}_{\ell i}^{j})^{\text{sample}} - (\hat{\Psi}_{\ell i,-k}^{j})^{\text{sample}} \tag{3.7}$$

where $(\hat{\Psi}_{\ell i}^{j})^{\text{sample}}$ is the already computed matrix and $(\hat{\Psi}_{\ell i,-k}^{j})^{\text{sample}}$ is the sample covariance matrix of the interfering UE with extra pilot samples.

- **Step 3:** Apply the convex optimization to the estimated channels by the sample covariance matrices as follows: Due to the variance error with all M^2 entries of the practical covariance matrices, a convex optimization method will be used here to obtain a better estimate, where the statistical sample covariance matrices are regularized with their diagonal matrices to reduce the error as follows:

$$\hat{\Psi}_{\ell i}^{j}(c) = c(\hat{\Psi}_{\ell i}^{j})^{\text{sample}} + (1-c)(\hat{\Psi}_{\ell i}^{j})^{\text{diagonal}} \tag{3.8}$$

$$\hat{\mathbf{R}}_{\ell i}^{j}(\alpha) = \alpha(\hat{\mathbf{R}}_{\ell i}^{j})^{\text{sample}} + (1-\alpha)(\hat{\mathbf{R}}_{\ell i}^{j})^{\text{diagonal}} \tag{3.9}$$

where c and α denote the regularization factors. The diagonal elements of the $\hat{\mathbf{R}}_{\ell i}^{j}(\alpha)$ and $\hat{\Psi}_{\ell i}^{j}(c)$ are the same as in $(\hat{\mathbf{R}}_{\ell i}^{j})^{\text{sample}}$ and $(\hat{\Psi}_{\ell i}^{j})^{\text{sample}}$ respectively, while the elements in the off-diagonal locations are regularized corresponding to c and $\alpha \in [0,1]$.

3.6 Proposed Model-3

In motivation with the facts of the non-linear distortion in Section 2.3.5, this proposed model-3 aims to use the deep learning approach to estimate the effective uplink channels from K UEs under the joint impact of non-linear hardware distortions of the BS and UEs in a single-cell scenario and Rician fading environment. Two Deep Neural Networks (DNNs) are proposed here based on the state-of-the-art literature [33], which are more refined and trained with different optimization algorithms, loss functions, types of modulation symbols, and distorting polynomials. The performance of this approach is validated with the state-of-the-art aware and unaware Bayesian estimators in Section 2.3.7. However, the algorithm steps of the proposed model-3 in this Chapter can be concluded as follows: we have

- **Step 1:** Develop two Deep Neural Networks (DNNs) with fully connected layers to estimate the effective uplink channels and the distortion correlation of K = 10, 20, and 30 UEs in a single-cell scenario.
- **Step 2:** Generated the deep learning dataset using the general order derivations of the effective channels in [33].
- **Step 3:** Refined the learning process of the DNNs with different optimization algorithms and loss functions.
- **Step 4:** Suggested used complex Gaussian symbols with 90^o circular shift symmetry instead of QPSK symbols used in [33] to improve estimation quality.
- **Step 5:** Suggested used real coefficients memoryless polynomial to model the non-linearity effect in addition to the complex coefficients quasi-memoryless polynomial used in [33] to simplify learning computations.

3.6.1 Deep learning based-estimation model

The two feed-forward Deep Neural Networks (DNNs) in Figs.3.4 and 3.5 are proposed here to estimate the effective uplink channels and the distortion

correlations of the K UEs, respectively. The following subsections will describe the DNNs parameters as follows:

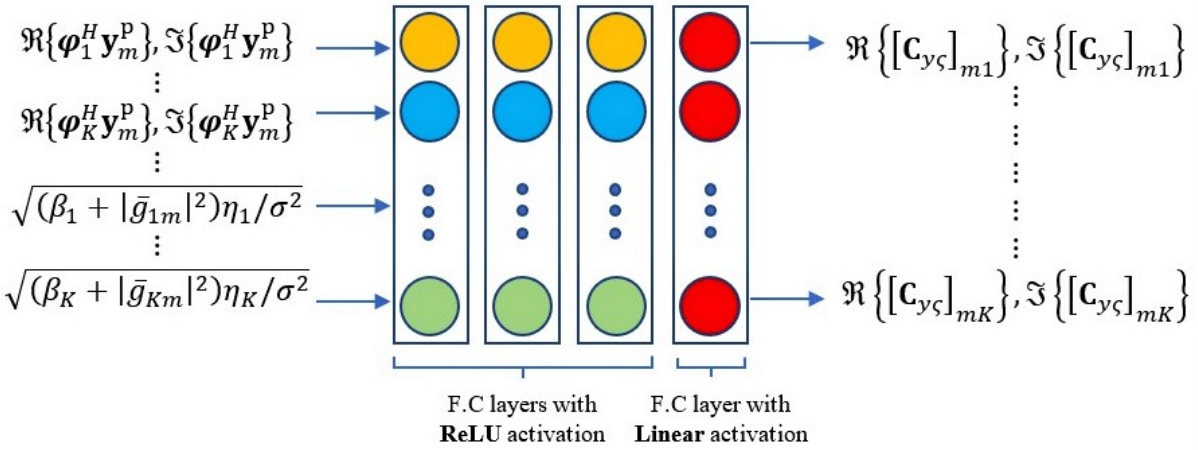

Figure 3.4: The deep neural network for estimating the effective channel elements.

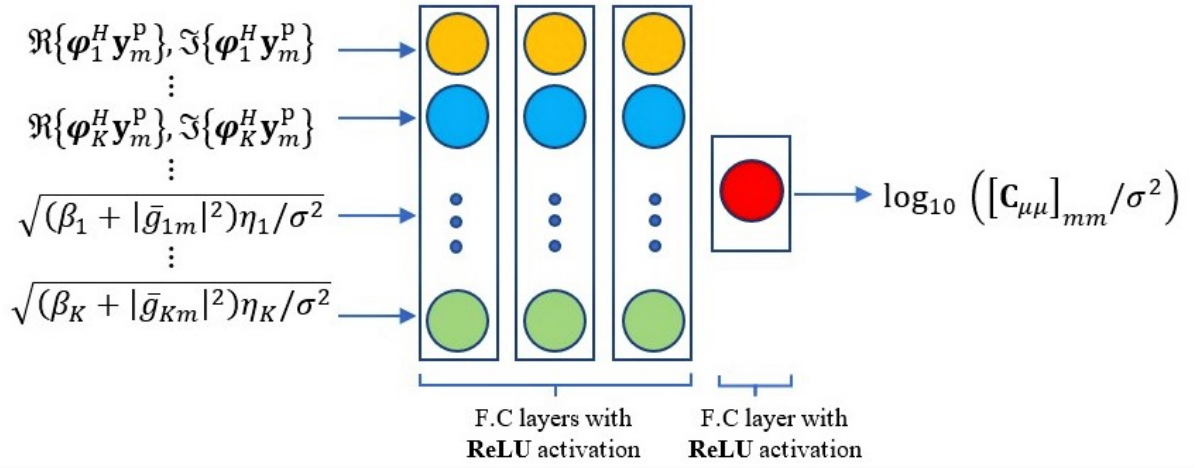

Figure 3.5: The deep neural network for estimating the distortion correlation elements

3.6.1.1 DNNs' architecture

The architecture of the proposed DNNs has fully connected layers and consists of one input layer, three hidden layers, and one output layer. The hidden layers have $30K$ neurons, while the input layer has $2K$ neurons. The output layers in Fig.3.4 and Fig.3.5 contain $2K$ and 1 neurons, respectively.

3.6.1.2 DNNs' activation function

For a DNN with **L** layers, the activation output $\mathbf{a}_l \in \mathbb{R}^{N_{\mathbf{L}}}$ of the l^{th} layer can

be written as:

$$\mathbf{a}_l = \Gamma_l\left(\mathbf{W}_l\mathbf{a}_{l-1} + \mathbf{b}_l\right), \;\; l = 1, \ldots, \mathbf{L} \tag{3.10}$$

such that $\mathbf{W}_l \in \mathbb{R}^{N_\mathbf{L} \times N_\mathbf{L}}$ and $\mathbf{b}_l \in \mathbb{R}^{N_\mathbf{L}}$, are the weight matrix and the bias vector at l^{th} layer, respectively. $\Gamma_l(.)$ denotes the activation function of the l^{th} layer.

The Rectified Linear Unit (ReLU) activation function is used in the input and the hidden layers of the DNNs in Figs 3.4 and 3.5. On the other hand, a Linear activation function is used in the output layer of Fig 3.4, while the ReLU function is utilized in the output of Fig.3.5.

3.6.1.3 DNNs' cost function and optimization algorithms

Generally, the mean squared error (MSE) [63] is utilized as the overall cost function of deep neural networks shown in Figs 3.4 and 3.5, which is given by

$$\mathbb{L}\left(\{\mathbf{W}_l, \mathbf{b}_l\}_{l=1}^{\mathbf{L}}\right) = \frac{1}{T}\sum_{t=1}^{T} l\left(\tilde{\mathbf{a}}_l^t, \mathbf{a}_o{}^t\right) \tag{3.11}$$

where T is the size of training samples, and $l\left(\tilde{\mathbf{a}}_l^t, \mathbf{a}_o{}^t\right)$, denotes any arbitrary loss function between the actual and desired outputs of t^{th} training sample applied to the neural networks. However, different optimization algorithms with gradient descent underlining math are applied here to refine the performance of the deep learning approach.

Hence, the DNNs are trained with different optimization algorithms and loss functions. The mean squared error and Huber loss functions are tested with various optimization algorithms, including Adam, RMSprop, and AdaDelta. A learning rate of 0.001 is used, and the epochs are set to 50, with a batch size of 1000. A callback function is used to stop training after 5 validation loss iterations that come with no improvements.

3.6.1.4 DNNs' training dataset

The general order derivation of the effective channels in [33] will be used here to provide the input-output sample pairs of the training dataset for the supervised learning in Figs 3.4 and 3.5. The input samples of the DNNs in Figs 3.4 and 3.5 are the processed signal $\phi_k^H \mathbf{y}_m^p$ of all UEs, i.e., for $k = 1, 2, \ldots, K$

plus the other input from the normalized channel gain over noise used to ordered the training inputs in descending manner to improve learning. The output in Fig.3.4 is the imaginary and real terms of the effective channel elements of all K UEs, while in Fig.3.5, the output is the logarithmic quantity of the normalized distortion correlation elements.

The input training dataset size is 3×10^6, while the validation set size is 2×10^5. However, an outlier elimination strategy is applied to the training dataset to remove the large deviations in the DNN's input. Also, the dataset normalization methods are performed using the Standard-Scaler for the first $2K$ inputs and the Min-Max scaler for the last K inputs.

To show how the deep learning based-estimator is performed compared to the aware and unaware Bayesian estimators in Section 2.3.7, the proposed model-3 uses the Rician channel environment with large-scale fading parameters from the 3GPP model used in [105]. The complex coefficients $\tilde{b}_m = \tilde{b}_l$ are assumed equal across all antennas at the base station (BS). The pilot book matrix is created using the Discrete Fourier Transform (DFT) with independent and orthogonal columns. The UE locations are randomly distributed in a cell area of $250m \times 250m$ at each training sample. However, Table 1 concludes both parameters of the single-cell scenario and the parameters of the DNNs shown in Figs. 3.4 and 3.5, as follows:

Table 3.1: DNNs and single-cell scenario parameters

THE PARAMETER	VALUES
Polynomial coefficients of the BS	$\tilde{b}_m = \tilde{b}_l$, for $m = 1,\ldots,M$
Pilot book matrix	Discrete Fourier Transform (DFT)
Pilot length	$\tau_p = K$
Fading channel	Rician fading channel
Environment	Urban Microcell from 3GPP
Cell area	$250m \times 250m$
Carrier frequency	2GHz
Bandwidth	200 MHz
Number of neurons in each hidden layer	$30K$
Number of neurons in each input layer	$2K$
Learning rate	0.001
Training dataset length	3e6
The validation set length	2e5
Epochs and batch sizes	50 and 1000
Callback patience	5, for the same validation loss

Chapter 4

Results and Discussions

This chapter shows the simulation results and their discussions of the proposed models in Chapter 3 as follows:

4.1 Results and Analysis of Proposed Model-1

The results and their analysis of the proposed model-1 are discussed for both channel models (the local scattering and the practical channel model) and compared with the models that have full rank channels in [11, 13, 12] as follows:

4.1.1 Using local scattering channel model

To evaluate the performance of the proposed model-1 in Section 3.3.1 on the estimated channels, the normalized mean squared error (NMSE) is simulated as a function of the effective SNR^p and ASD parameters for both the full-rank and spatial(reduced)-rank channels of the local scattering model with different numbers of antennas and different ASD distributions, as follows:

4.1.1.1 NMSE as a function of effective SNR^p

The figures in this subsection simulate the NMSE of the Bayesian MMSE estimator versus a range of effective SNR^p for both the full-rank and reduced-rank channels taking into account the following parameters: ASD$= 10^o$ with Gaussian distribution, the number of antennas at the BS are $M = 1, 10, 64$, or 100. The simulation results are evaluated in Matlab R2020a and calculated across a range of nominal arrival angles ranging from 0 to 2π.

Fig.4.1 reveals a special case of the simulation results: the highly spatially correlated channel (i.e., all multi-paths of the channel will reach the BS as a single path only) compared to the full-rank (multi-paths) channel obtained by the local scattering model.

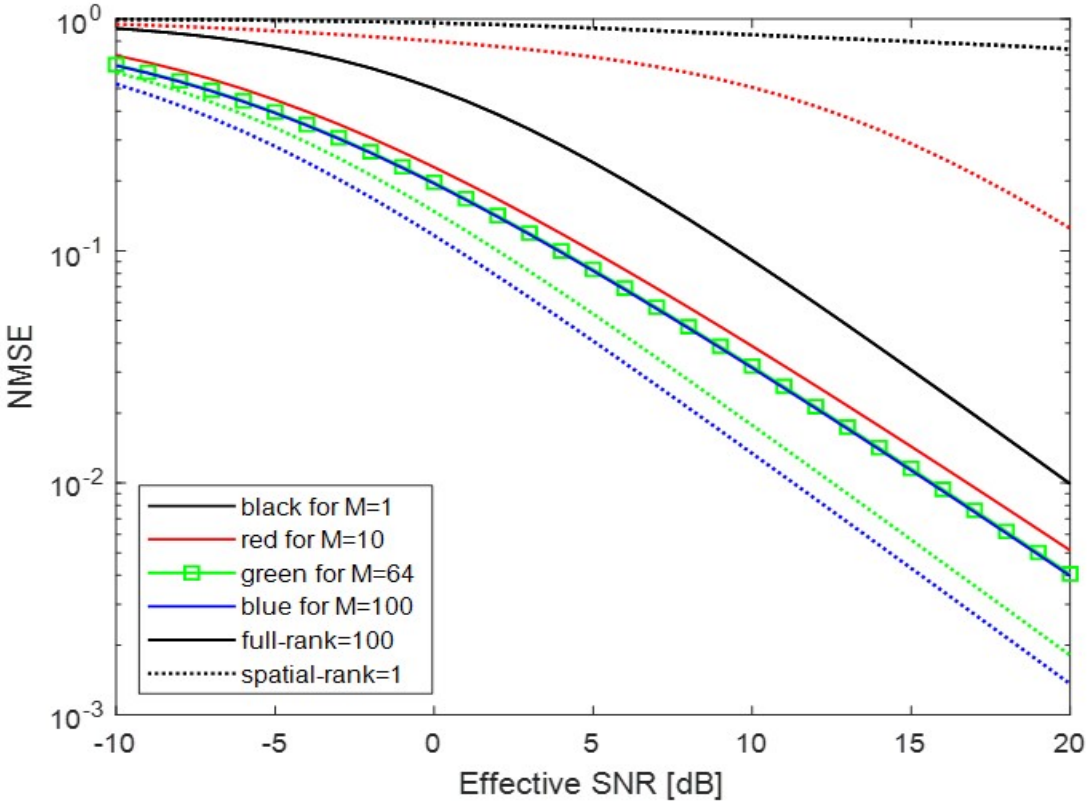

Figure 4.1: NMSE as a function of effective SNR of spatial-rank=1 and full-rank=100 channels with Gaussian ASD=10^o and M=1, 10, 64, and 100

Generally, the NMSE is a decreasing function with the SNR. It is noticed that with $M = 1$, the estimation error is high (about ≈ 1, even at high SNR values) since the single antenna cannot collect the signal power from one strong path only. While in the case of $M = 100$, the estimation error is reduced to less than (10^{-2}) at high SNR values since the BS can benefit from the array of antennas to collect the power of the signal more easily. Hence, as the M increases, the error is reduced at high SNR, and this is expected according to the spatial correlation structure in the expression of Eq.(2.37), which is M-dependent, in contrast to the spatially uncorrelated channels that have a spatial structure which is independent of M since the NMSE $= 1/(\text{SNR}^p + 1)$.

The figures from 4.2 to 4.5 show the simulation results of the NMSE with different spatial ranks of (10, 20, 40 and 64). It is noticed that each time in-

creasing the rank, the results become asymptotically aligned with the full rank result. Also, it is noticed that the total error is increasing synchronously with the increase of the spatial rank in the case of M=100 and 64; this is true due to the NMSE formula in Eq.(2.32), which depends on the trace of the error correlation matrix, i.e., the sum of the diagonal variance error across all M antennas at the BS.

Fig.4.5 shows an important unique case of the spatial correlation characteristics as follows: at the spatial rank=64, the full-rank channels will be completely matched with the reduced rank channels; this gives us an important intuition that we can benefit from the spatial correlation characteristics to sufficiently use only 64 antennas (which is the lower bound of the number of antennas used in the massive MIMO system) instead of 100 antennas at the BS; this also has a great impact on reducing the cost of building such large MIMO system since it is reducing the number of radio frequency chain that is needed for each antenna at the BS.

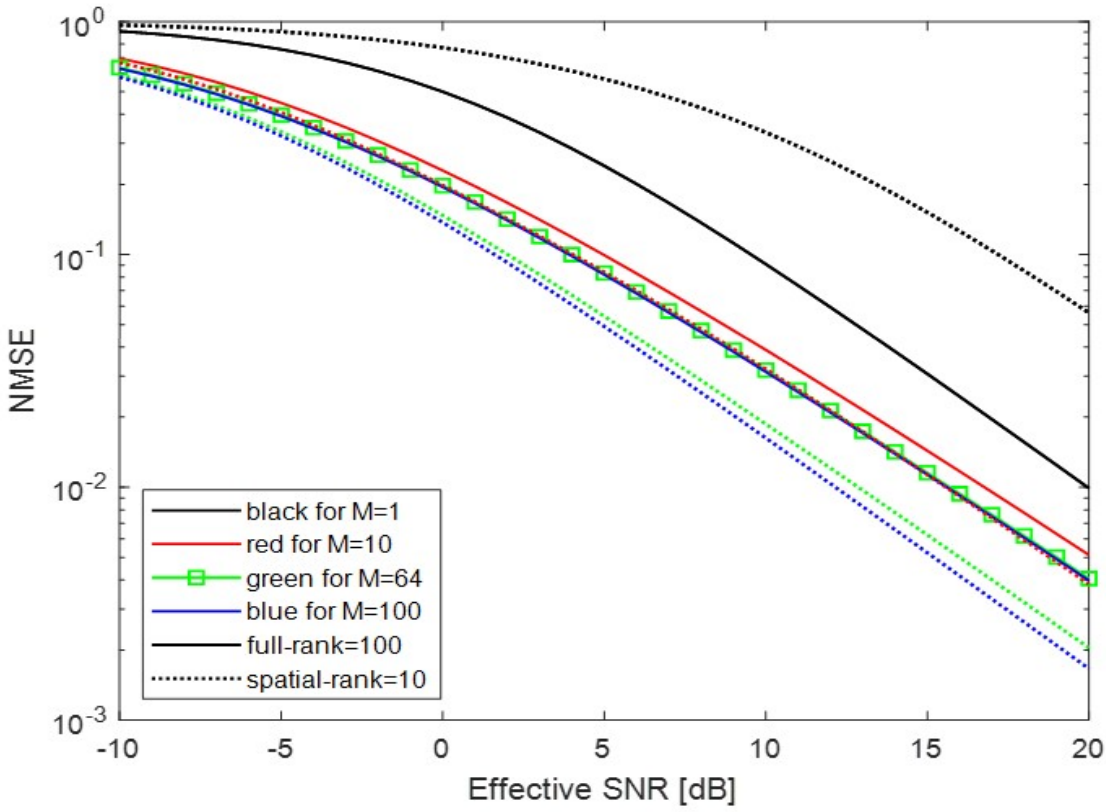

Figure 4.2: NMSE as a function of effective SNR of spatial-rank=10 and full-rank=100 channels with Gaussian ASD=10^o and M=1, 10, 64, and 100

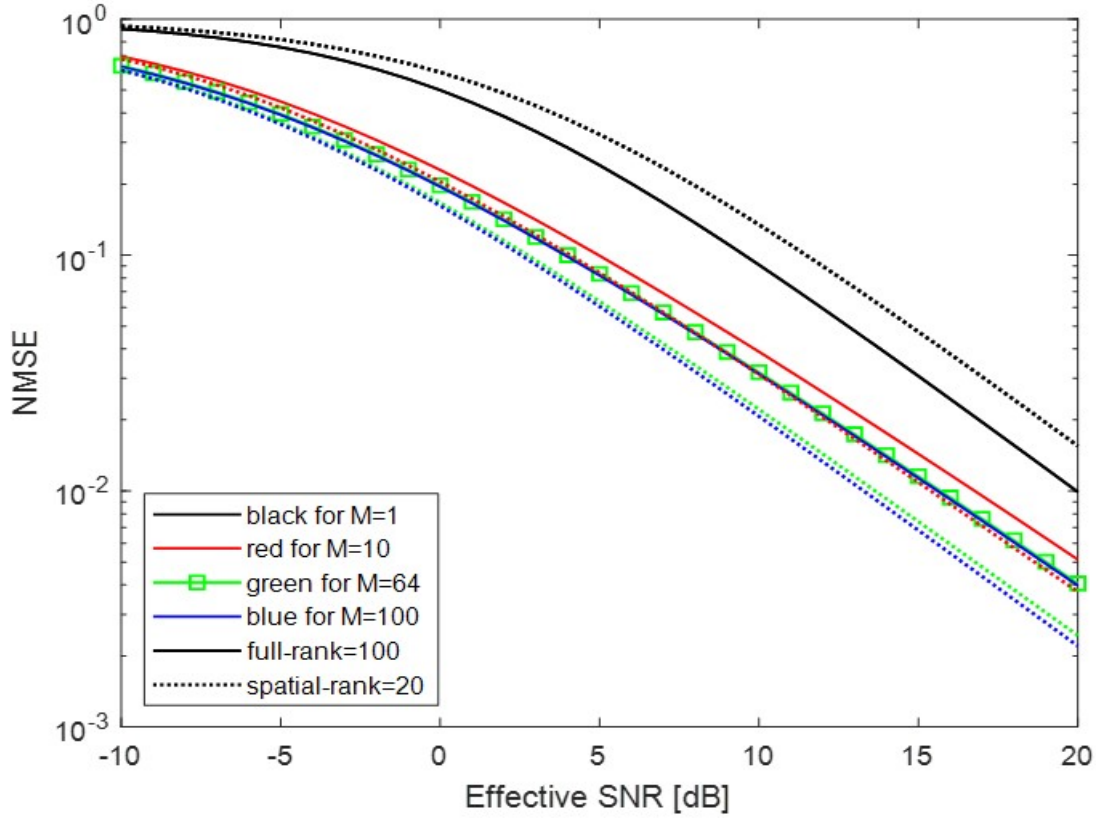

Figure 4.3: NMSE as a function of effective SNR of spatial-rank=20 and full-rank=100 channels with Gaussian ASD=10^o and M=1, 10, 64, and 100

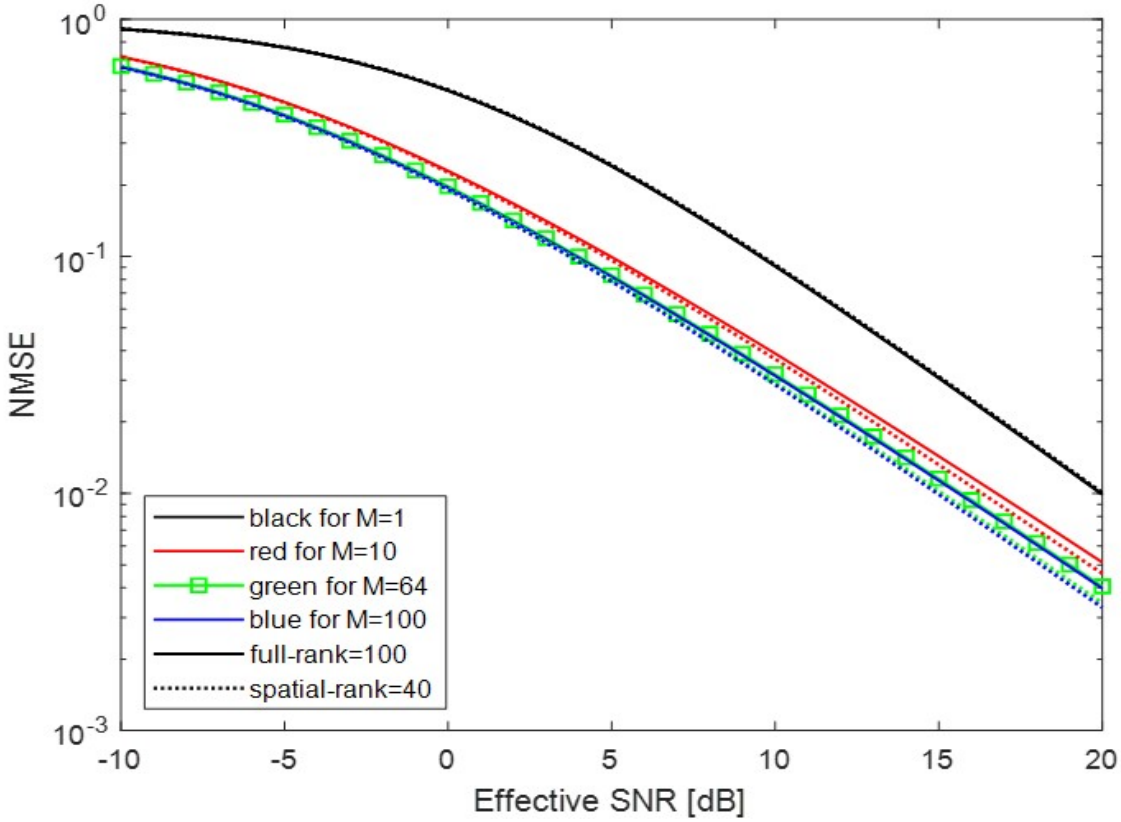

Figure 4.4: NMSE as a function of effective SNR of spatial-rank=40 and full-rank=100 channels with Gaussian ASD=10^o and M=1, 10, 64, and 100

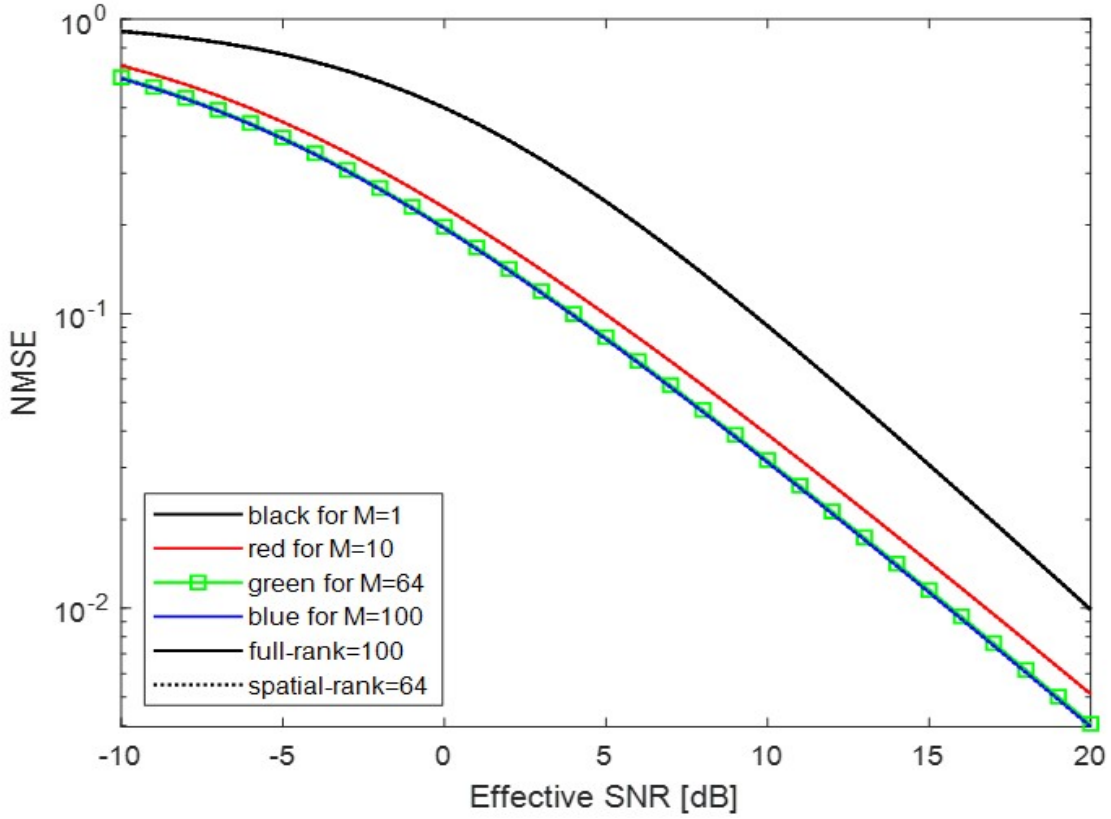

Figure 4.5: NMSE as a function of effective SNR of spatial-rank=64 and full-rank=100 channels with Gaussian ASD=10^o and M=1, 10, 64, and 100

4.1.1.2 NMSE as a function of effective ASD(σ_φ):

The influence of the spatial characteristics is more studied in Figures (4.6-4.11) such that the normalized MSE is simulated as a function of the angular standard deviation (ASD). The Gaussian, Uniform, and Laplacian distributions are considered with a 10dB effective signal-to-noise ratio and $M = 100$ array of antennas at the BS.

Referred to figures (4.6-4.11), it is seen that, at small values of ASD, the NMSE is approximately zero and is less (by two orders of magnitude) than the uncorrelated channel case (the upper limit of the estimation error), while it grows monotonically when the ASD increases.

In the case of Fig. 4.6, the estimation error is reduced by nearly one order of magnitude at the spatial rank=1 channel compared with the full-rank=100 channel, even at high values of ASD. Also, it is noticed that the uniform distribution of ASD has appeared as the lowest estimation error in the case of full-rank channels; that is because of the inherent spatial correlation characteristics of the ASD uniform distributions at the BS, which is equal to 20% with non-zero eigenvalue

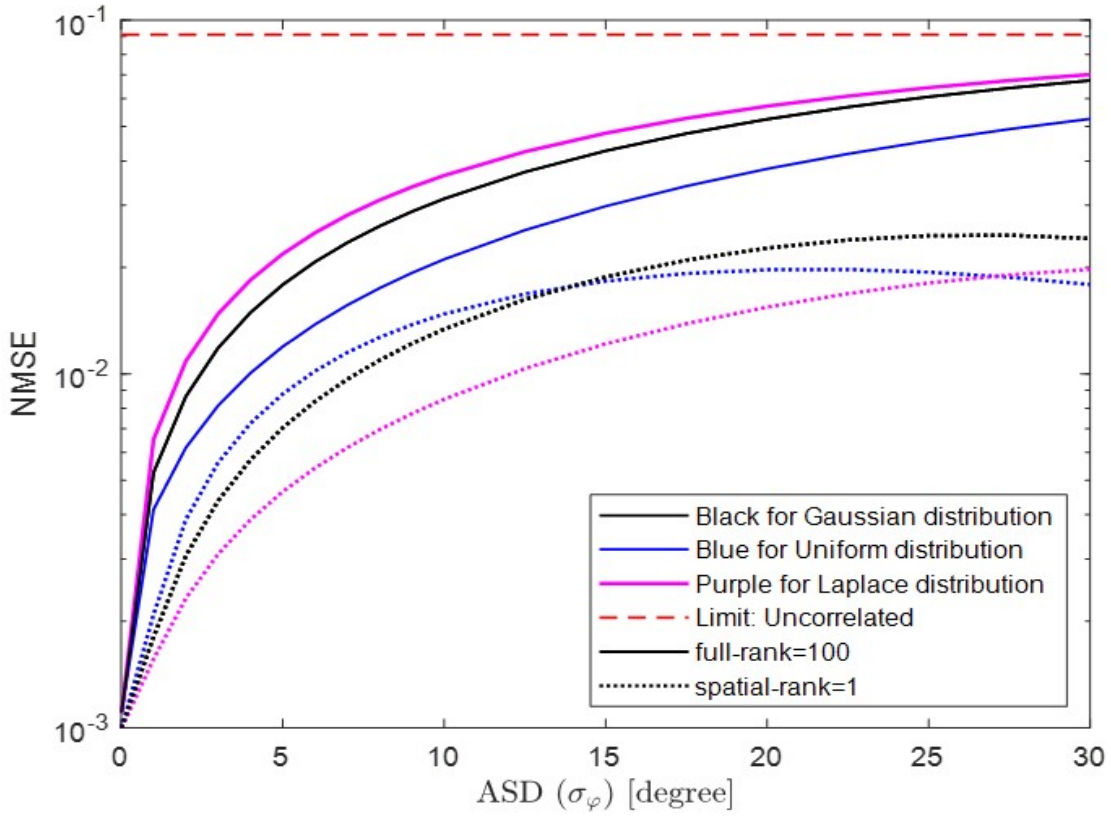

Figure 4.6: NMSE as a function of ASD, for full-rank=100 and spatial rank=1 using Gaussian, Uniform, and Laplace ASD distributions, $M = 100$ and SNR=10dB

subspace (i.e. it is more spatial than the Gaussian and Laplacian distributions that have 40% and 60% non-zero eigenvalue subspaces [11], respectively).

Note that, in all figures (4.7-4.11), the error variance increases when the rank increases; this is because NMSE $= \mathrm{tr}(\mathbf{C})/\mathrm{tr}(\mathbf{R})$. The case in Fig.4.10 asserts the simulation result obtained in Fig.4.5, which benefits from the spatial characteristics of using only 64 antennas instead of 100 at the BS. This result also has been proved in Fig.4.11 using two cases of BS array-antennas $M = 64$ and $M = 100$ with different ASD distributions, and it showed the same result in terms of channel quality but with less elapsed time for the spatial rank=64 as follows:

- For the full-rank channel, the elapsed time was 0.000603 seconds.
- For the reduced-rank channel, the elapsed time was 0.000305 seconds.

These simulation results are implemented in Matlab R2020a installed on a Laptop with a graphic card of NVIDIA RTX2070 MAx-Q, 16GB RAM, and Core i7-10750H processor.

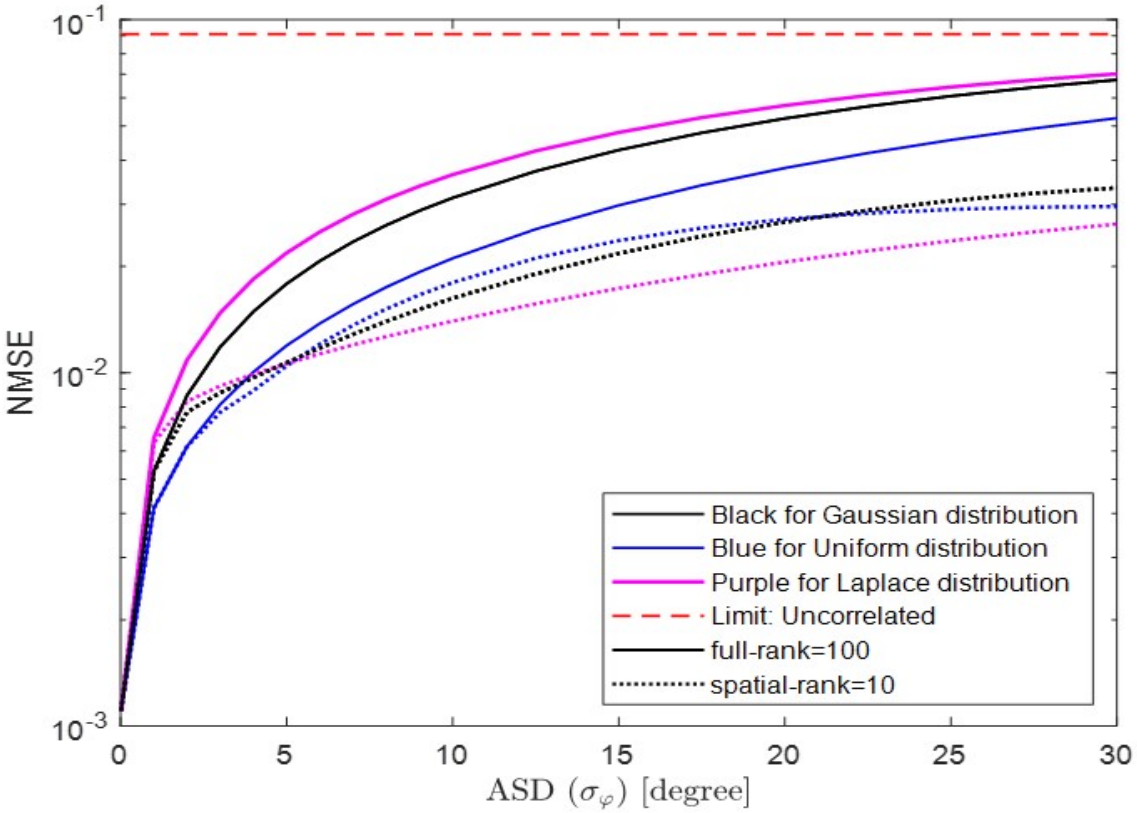

Figure 4.7: NMSE as a function of ASD, for full-rank=100 and spatial rank=10 using Gaussian, Uniform, and Laplace ASD distributions, $M = 100$ and SNR=10dB

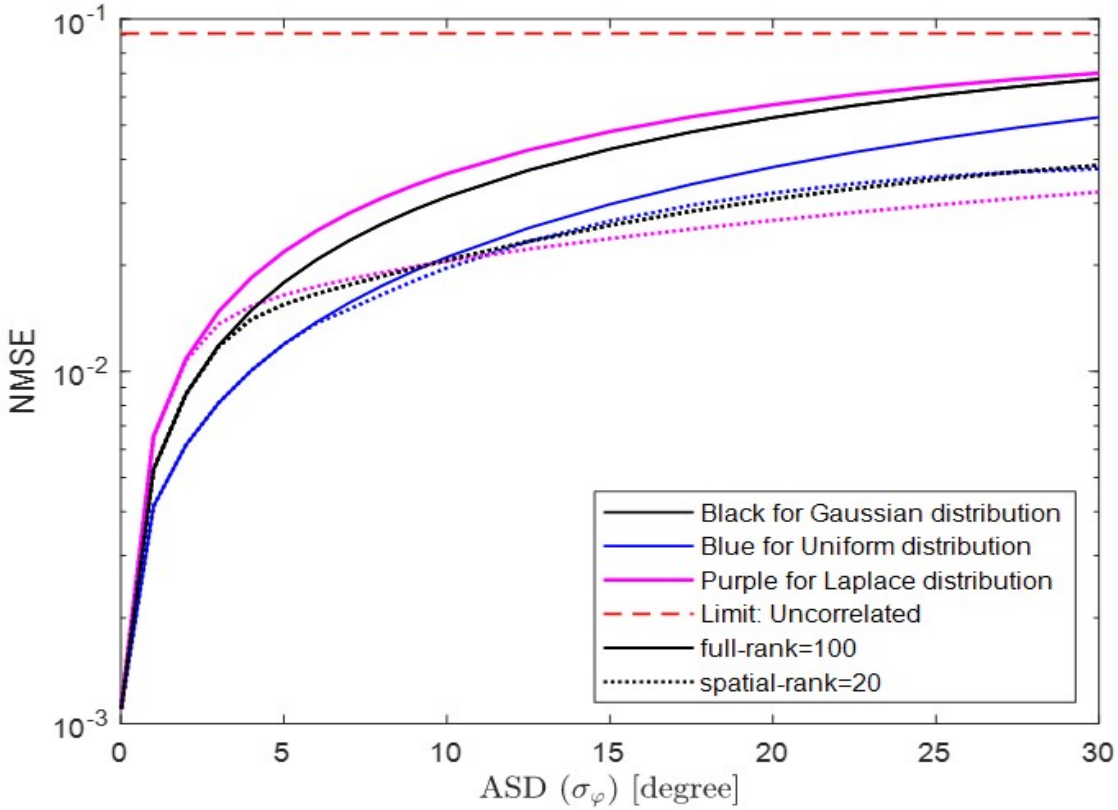

Figure 4.8: NMSE as a function of ASD, for full-rank=100 and spatial rank=20 using Gaussian, Uniform, and Laplace ASD distributions, $M = 100$ and SNR=10dB

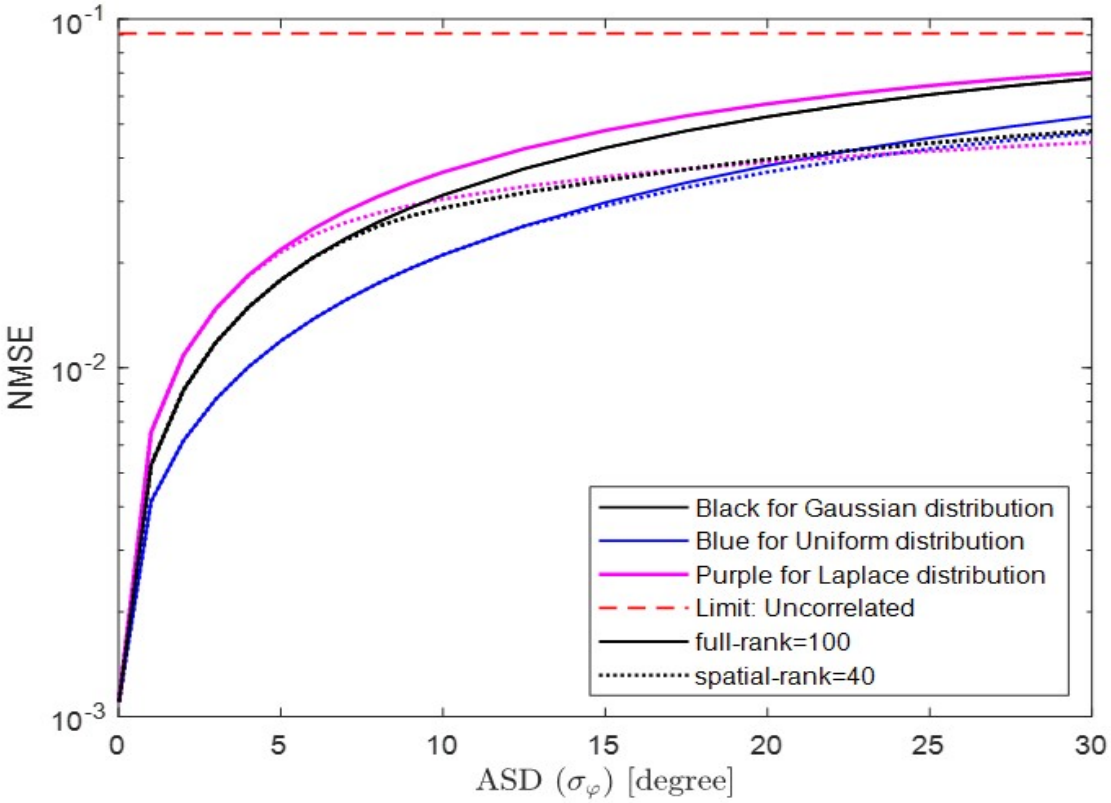

Figure 4.9: NMSE as a function of ASD, for full-rank=100 and spatial rank=40 using Gaussian, Uniform, and Laplace ASD distributions, $M = 100$ and SNR=10dB

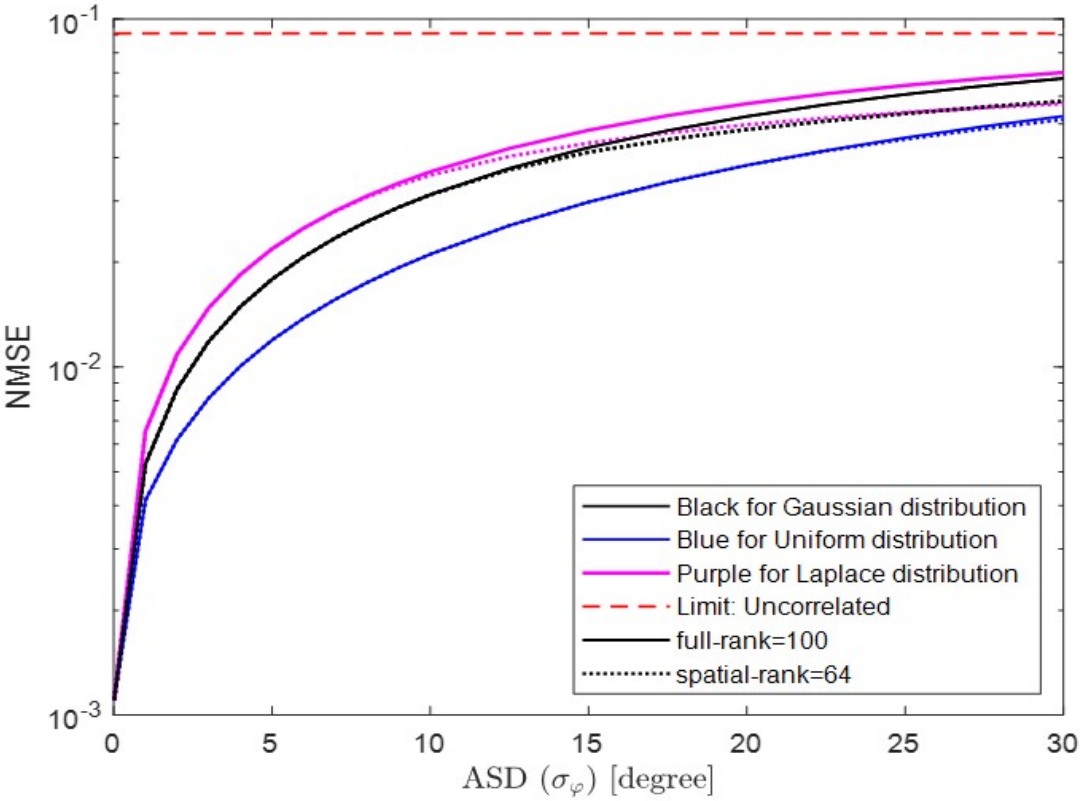

Figure 4.10: NMSE as a function of ASD, for full-rank=100 and spatial rank=64 using Gaussian, Uniform, and Laplace ASD distributions, $M = 100$ and SNR=10dB

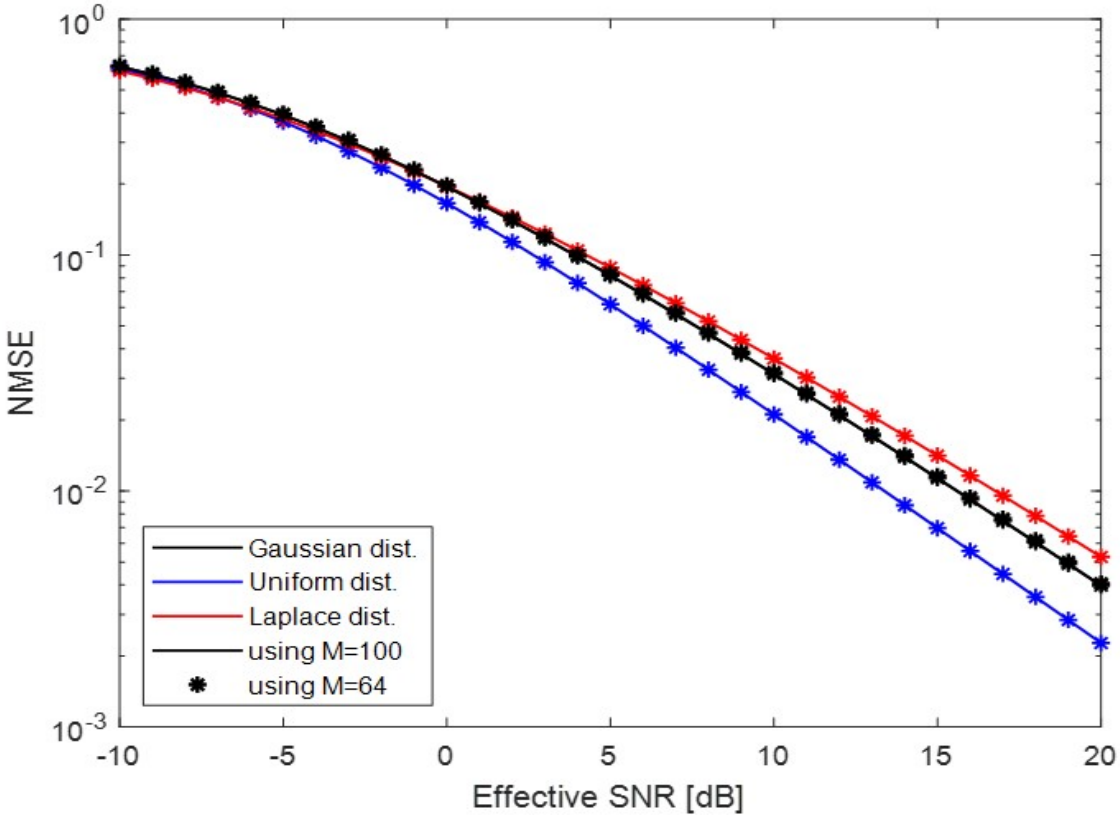

Figure 4.11: NMSE as a function of ASD, for full-rank=100 and spatial rank=64 using Gaussian, Uniform, and Laplace ASD distributions, $M = 100, 64$ and SNR=10dB

4.1.2 Using practical channel model

The effect of spatial correlation will be explained here by considering the normalized MSE as a relationship with samples (observations) number as well as with the UE's effective SNR, as follows:

4.1.2.1 NMSE versus samples number

To evaluate the proposed procedure in this part, the scenario of the local scattering model for a single-antenna UE is used here with the following parameters: $M = 100$ BS antennas, $d_H = 1/2$ wavelength antenna spacing, $\sigma_\phi = 10^o$ angular standard deviation with Gaussian distribution, and 10dB effective signal-to-noise ratio. The completely known covariance matrix $\mathbf{R}$ of the local scattering model and its diagonal version $\mathbf{R}_{\text{diagonal}}$ are leveraged here as the lower and upper bounds (as shown in Fig.4.12) to evaluate the asymptotic behaviour of the estimated channel in the practical case. Hence, a Monte Carlo simulation of the NMSE curve is used here in Fig.4.12, which averaged over a range of 500 realization samples and different nominal angles between (0 and 2π).

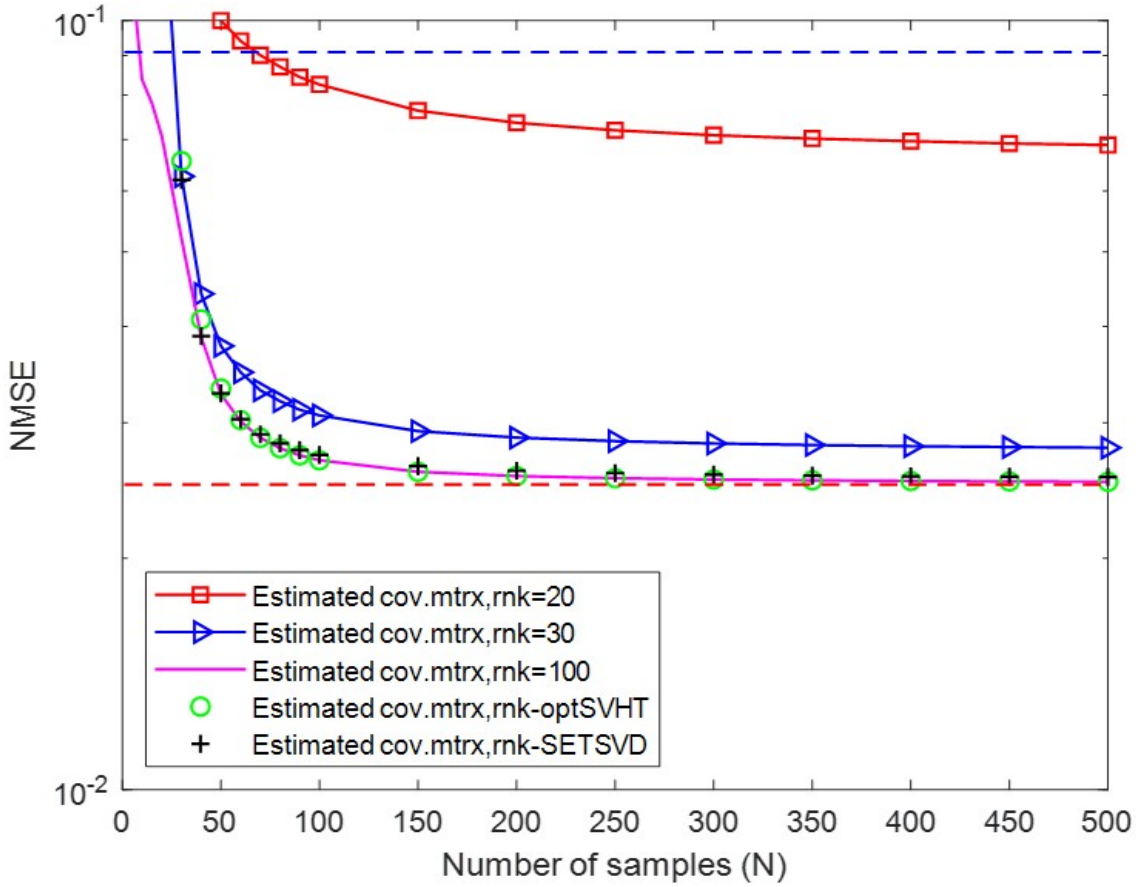

Figure 4.12: NMSE as a function of sample numbers of practical channels with full rank=100 and successive spatial ranks r=20, 30, SETSVD=35 and optimal-SVHT=48, $M = 100$, Gaussian ASD= 10^{o}, and effective SNR=10dB

It is noticed that the $\hat{\mathbf{R}}_{\text{diagonal}}$ matrix can be completely estimated from only a few samples/observations ($N = 10$) on the upper bound of Fig.4.12. At the same time, it is noticed that after 150 samples, the practical channel is asymptotically aligned with the lower bound of the figure; this means that the $\hat{\mathbf{R}}_{\text{sample}}$ matrix of the practical channel can be estimated with a number of samples of $1.5M$ which is comparable to the number of the BS antennas-array.

It is noticed that the NMSE curves of the truncated channels using thresholding rules in Section 3.3.1 are asymptotically aligned with the curve of the full-rank channels in Fig.4.12, i.e., they have the same quality as the full-rank channels. This is an important result from two perspectives regarding spatial correlation characteristics and computational complexity. For spatial characteristics, it is important to get on a channel that would be spatially correlated to reduce the interference between UEs and provide more degrees of freedom on the BS since only a small partial subspace (eigenvalues) will be provided for

each UE at the BS. On the other hand, the complexity will be reduced if we use antennas-array $M = 64$ rather than $M = 100$ with the same performance as will be seen in Subsection 4.1.2.2

The simulation results in Fig.4.12 reveal that the same quality is obtained when the proposed spatial ranks (SETSVD=35 and optimal-SVHT=48) are applied to the full rank channel. But these results can not be exploited to use M=35 or M=48 instead of M=100 antennas at the BS since the lower number of antennas at the massive MIMO BS is M=64 that keeps the favourable propagation regime true (see Appendix B.3.2.3).

However, undesirable estimated channels have been obtained when reducing the eigenstructure of the covariance matrices to the ranks $r = 20$ and $r = 30$, as illustrated by the blue and red curves in Fig.4.12.

4.1.2.2 NMSE versus the effective SNR

In this part, the spatial characteristics of the reduced channel can be further explained by measuring the quality of the estimated channels using the MMSE estimator as a function of the effective SNR. For this task, the scenario of a single UE per cell in the Wyner model in Fig.B.5 is considered, where the arbitrary BS per cell tends to estimate its own UE while another user in the network tends to interfere with the desired user at its serving BS. The effective SNR range is assumed to be (-10dB $\rightarrow$ 20dB) for the desired user, while a 10dB weaker SNR will be for the interfering UE along all SNR range values of the desired user. The NMSE of the local scattering channel model is considered here to represent the exact channel quality with ASD$= 10^o$ over a range of nominal angles from (0 to 2π). Hence, Fig.4.13 shows the quality of the channels used in this scenario using the NMSE relationship with the effective SNR. It reveals that the reduced rank channels (spatial channels with SETSVD rank) behave as the full rank channels with the same quality on the MMSE estimator.

However, the eigenvalues of the reduced and full-rank channels can be plotted in Figs.4.14 and 4.15 against their covariance matrices rank. The figures show keeping and eliminating eigenvalues of the covariance matrices and how it can ultimately reduce the eigenstructure of the covariance matrices using the proposed SETSVD spatial rank and the optimal SVHT rank.

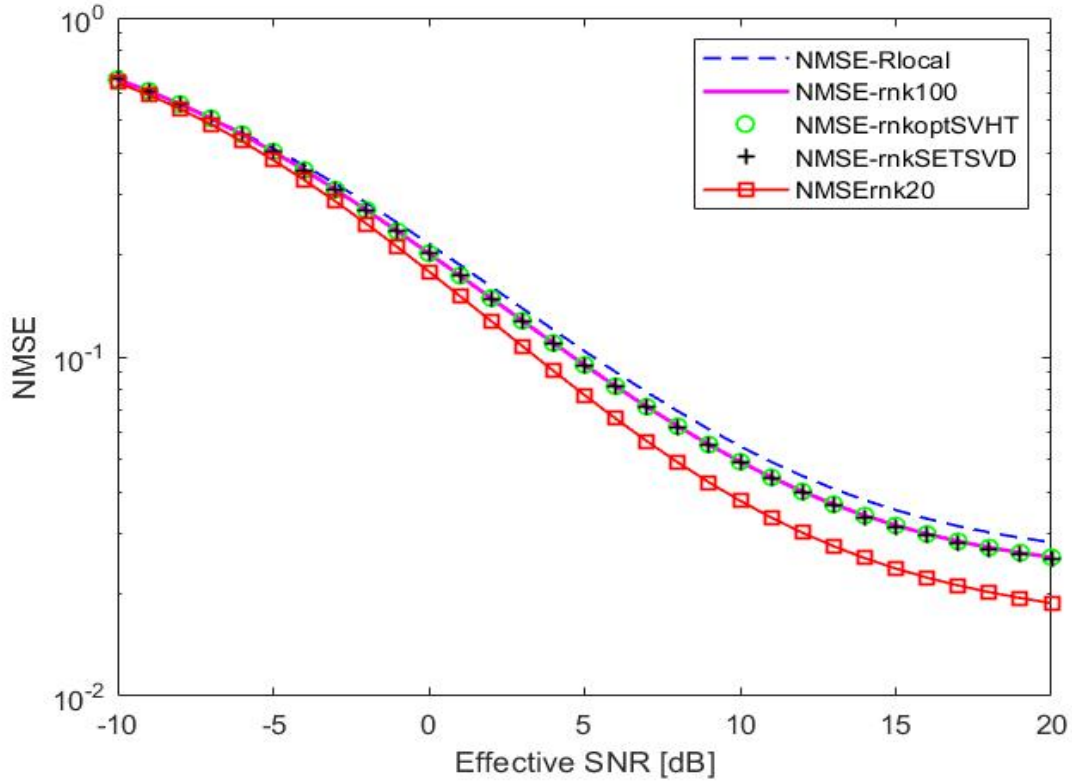

Figure 4.13: NMSE as a function of effective SNR of practical channels with full rank=100 and successive spatial ranks r=20, 30, SETSVD=35 and optimal-SVHT=48, $M = 100$, Gaussian ASD= 10^o, and effective SNR=10dB

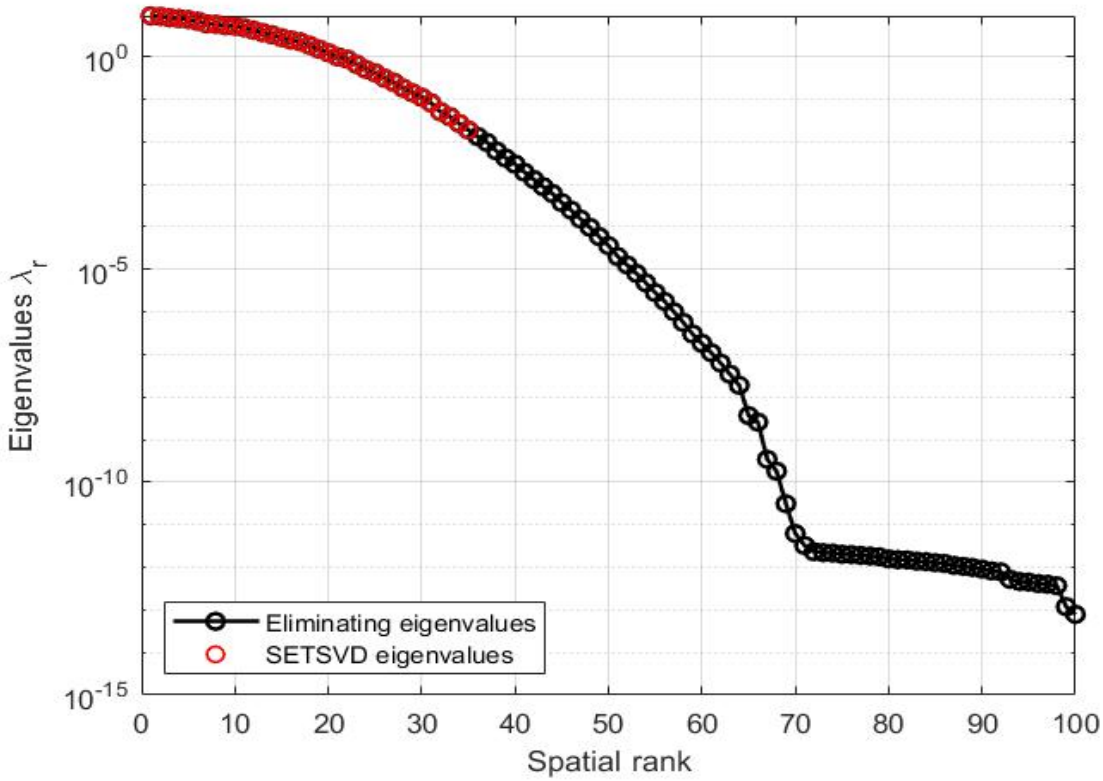

Figure 4.14: Eliminating and keeping eigenvalues using the proposed SETSVD spatial rank

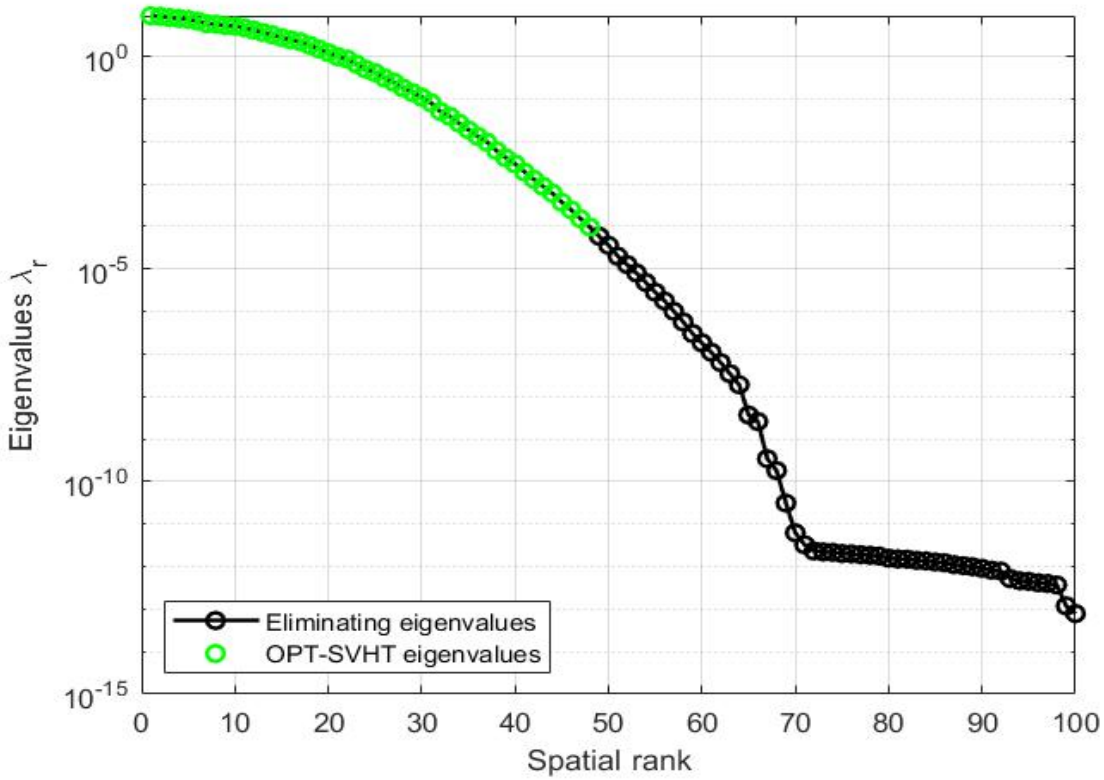

Figure 4.15: Eliminating and keeping eigenvalues using the optimal SVHT spatial rank

4.1.3 Unique results

The Figs 4.12 and 4.13 are reproduced here but with antennas-array $M = 64$ (minimum antennas-array required to have a massive MIMO system). The simulation results in Fig.4.16 show that the reduced rank channel has converged to the lower bound limit even faster (by about 50 samples) than the full rank channel in Fig.4.12. On the other hand, Fig.4.17 proves the same quality for both the full and reduced rank channels, especially when setting the SETSVD spatial rank to 64. The optimal SVHT rank also satisfies the same channel quality, but it is not desired here since it does not satisfy the number of massive antennas at the BS.

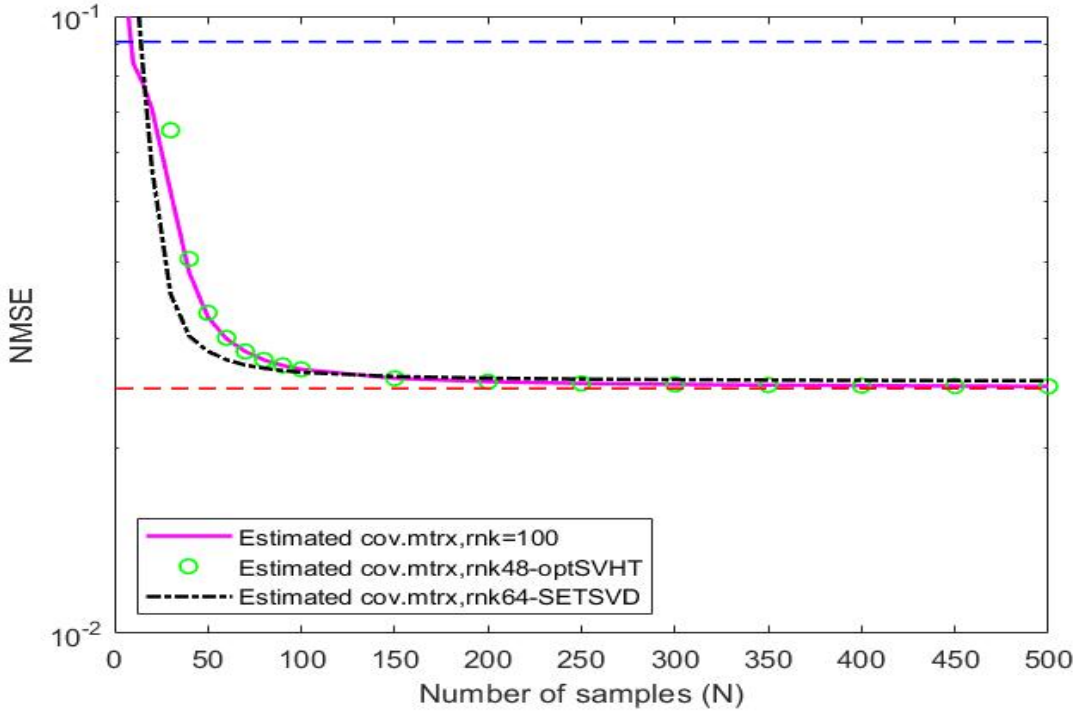

Figure 4.16: NMSE as a function of sample numbers of practical channels with full rank=100 and spatial ranks SETSVD=64 and optimal-SVHT=48, $M = 100$, Gaussian ASD= 10^o, and effective SNR=10dB

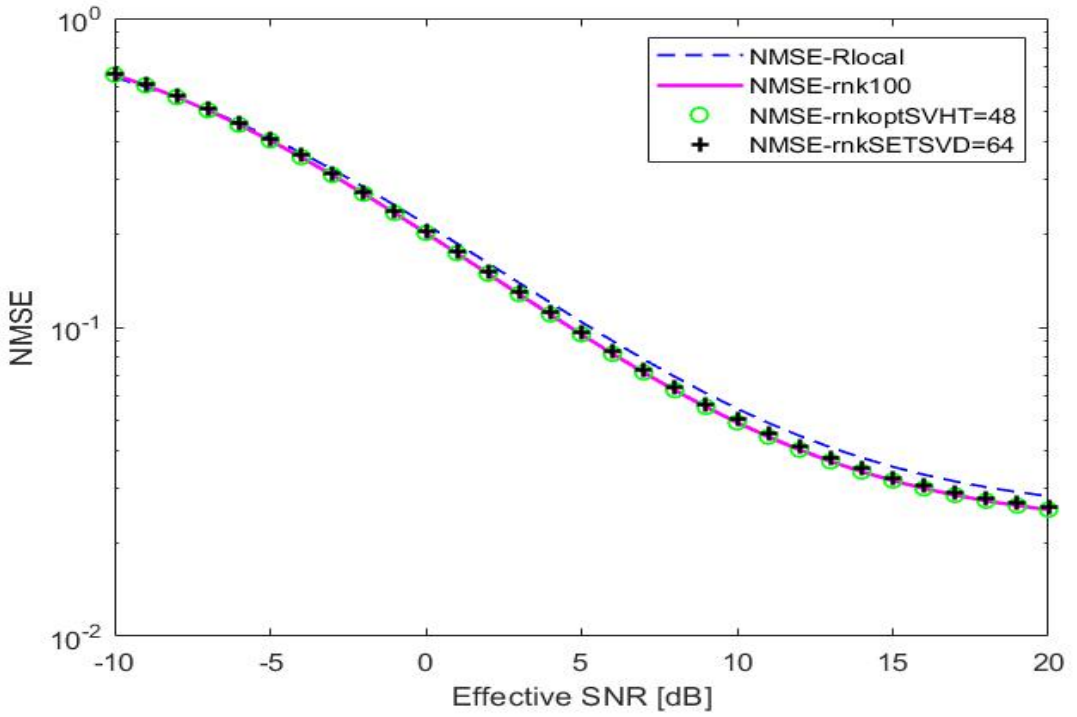

Figure 4.17: NMSE as a function of effective SNR of practical channels with full rank=100 and spatial ranks SETSVD=64 and optimal-SVHT=48, $M = 100$, Gaussian ASD= 10^o, and effective SNR=10dB

4.1.4 Computational complexity results

The complexity of the Bayesian MMSE estimator will be reduced in terms of the FLOPs number if the reduced-rank covariance matrices are applied instead of the full-rank matrices. As a comparison of the complexity between the standard and the proposed computational complexity in Section 3.3.3, Table 4.1 concludes the number of FLOPs required for both cases with different successive spatial ranks of the reduced channels as follows:

Table 4.1: Standard and proposed Computational Complexity (CC) comparison for the MMSE estimator with successive spatial ranks

Polynomial and Rank(r)	FLOPs of standard CC	FLOPs of proposed CC
Polynomial in terms of M	$(4/3)M^3 + M^2 + (\tau_p - 1/3)M$	$(2/3)M^3 + 4M^2 + M$
Polynomial in terms of r	$(4/3)r^3 + r^2 + (\tau_p - 1/3)r$	$(2/3)r^3 + 4r^2 + r$
r =20	11260	6953.333
r =35	58730	33518.333
r =48	150224	82992
r =64	354240	191210.666
r =100= M	1344300	706766.666

The Computational Complexity (CC) reduction can be computed using the percentage reduction ratio in terms of FLOPs number in Table 4.1, which can be determined by either computing the ratio of the proposed CC's FLOPS to the standard CC's FLOPS or by using the ratio between the FLOPs of the spatial ranks and the FLOPs of the full-rank used in Table 4.1. Hence, the number of FLOPs of the proposed CC is reduced to nearly 53% from the computations of the standard CC at spatial rank 64 in Table 4.1. Moreover, when computing the FLOPs in terms of the spatial and full ranks, the complexity will be reduced to 27% in computations at the spatial rank 64 compared to the full rank number of the proposed CC. However, the computation of the spatial rank 64 of the proposed CC will be significantly reduced to nearly 14% as compared to the full rank number of the standard CC.

4.2 Results and Analysis of Proposed Model-2

In this section, the impact of the hardware distortion on the practical estimated channels will be discussed here. The results of the proposed model-2 will be compared and validated with the local scattering channel model developed in [54, 11] in two cases with residual impairments (no-ideal hardware case) and without residual impairments (ideal hardware case) as follows:

The normalized mean squared error NMSE $= \mathrm{tr}(\mathbf{C})/\mathrm{tr}(\mathbf{R})$ shown in Fig.4.18 has been developed here with parameters ASD $= 10^o$ and a nominal angle ranging from 0 to 360^o to demonstrate how the hardware imperfection impacts the estimated channel. The NMSE curves are averaged throughout a range of effective SNR at pilot length $\tau_p = 10$ and distinct hardware constants.

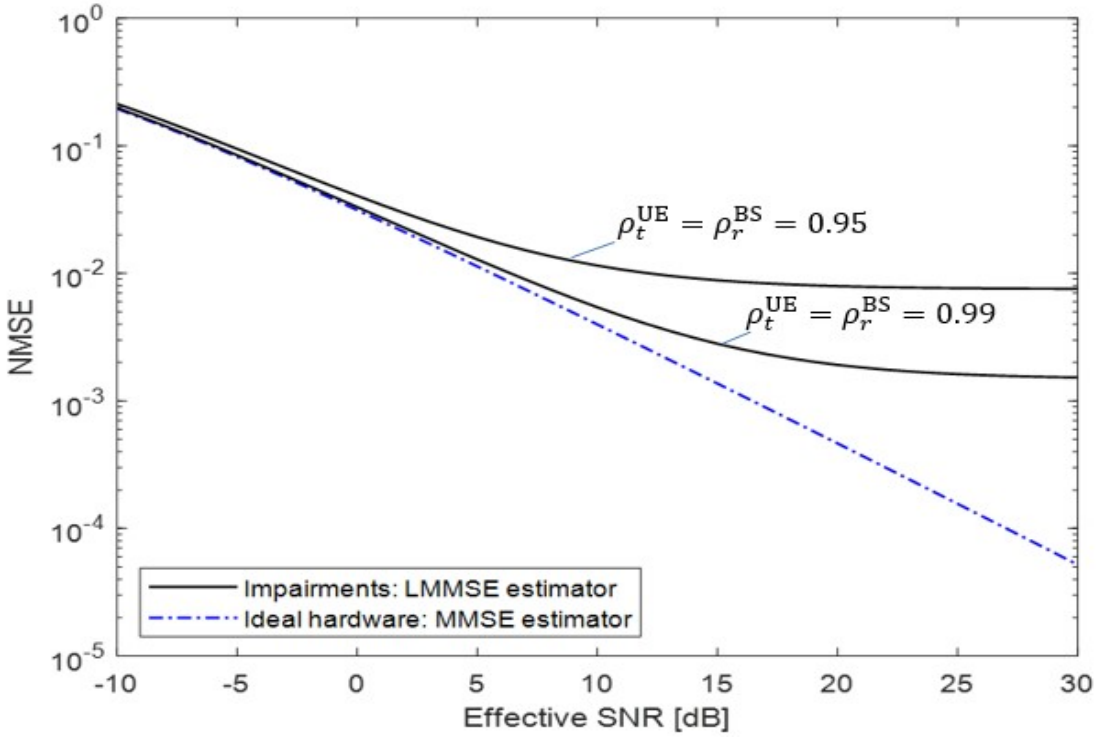

Figure 4.18: The normalized MSE of the channel estimation using the local scattering model with equal hardware quality at the transceiver ends.

4.2.1 Performance evaluation of practical channel estimation

For performance evaluation, the average NMSE curves in Fig.4.18 of the local scattering model will be considered here to validate our results in Figs. 4.19 and 4.20. The NMSE curves in Fig. 4.18 are averaged over different nominal angles of the UE in the cell between $(0^o - 360^o)$ with Gaussian distribution

and ASD $= 10^o$ using an array of $M = 100$ antennas at the BS. Fig.4.18 depicts how the NMSE curves are changing with the effective SNR $= (p/\sigma^2_{\text{UL}})$, using a $\tau_p = 10$ pilot sequence length and two cases of hardware constants as follows, $\rho_t^{\text{UE}} = \rho_r^{\text{BS}} = 0.99$ and $\rho_t^{\text{UE}} = \rho_r^{\text{BS}} = 0.95$ in case of hardware distortions, and $\rho_t^{\text{UE}} = \rho_r^{\text{BS}} = 1$ in case of ideal hardware. The figure emphasizes that there is an error floor in the estimated channels, particularly in the range of [20dB, 30dB] SNR, while it is small or negligible at lower SNR values. On the other hand, hardware impairments show a substantial impact on the estimated channels when the quality of hardware components decreases, particularly at $\rho_t^{\text{UE}} = \rho_r^{\text{BS}} = 0.95$.

Recall that the proposed model-3 involves using an optimized procedure to estimate the channel in practice first and then applying the convex optimization procedure to reduce the error floor in the NMSE curves at high SNR values.

The simulation results of the proposed model-2 are illustrated in Fig.4.19 and Fig.4.20. However, The NMSE curves of the practical channel in Fig.4.19 are obtained using the Monte Carlo simulation in Matlab with many sample realizations and averaged over the effective SNR range from (-10dB to 30dB). The simulation result in Fig.4.19 is compared to the local scattering channel in Fig.4.18 with two cases: under the joint impact of the transceiver's residual impairments case and the case of the ideal hardware transceivers at the BS and UEs. It shows that the practical channels are closely aligned with the NMSE curves of the developed results cases of the local scattering model.

On the other hand, Fig.4.20 shows the optimized NMSE curves when the convex optimization procedure is applied to the practical channels obtained by the sample covariance matrix method in the first step of the procedure. As a result, the error floor in Fig.4.20 is significantly reduced (nearly by one order of magnitude) when the hardware quality $\rho_t^{\text{UE}} = \rho_r^{\text{BS}} = 0.99$, while it is slightly better than before in the case of $\rho_t^{\text{UE}} = \rho_r^{\text{BS}} = 0.95$. On the other hand, it is noticed that the NMSE curve when $\rho_t^{\text{UE}} = \rho_r^{\text{BS}} = 0.99$ is closely aligned to the ideal hardware case of $\rho_t^{\text{UE}} = \rho_r^{\text{BS}} = 1$, which means that less quality hardware components can be used now at the BS with only a little performance loss in the channel estimation.

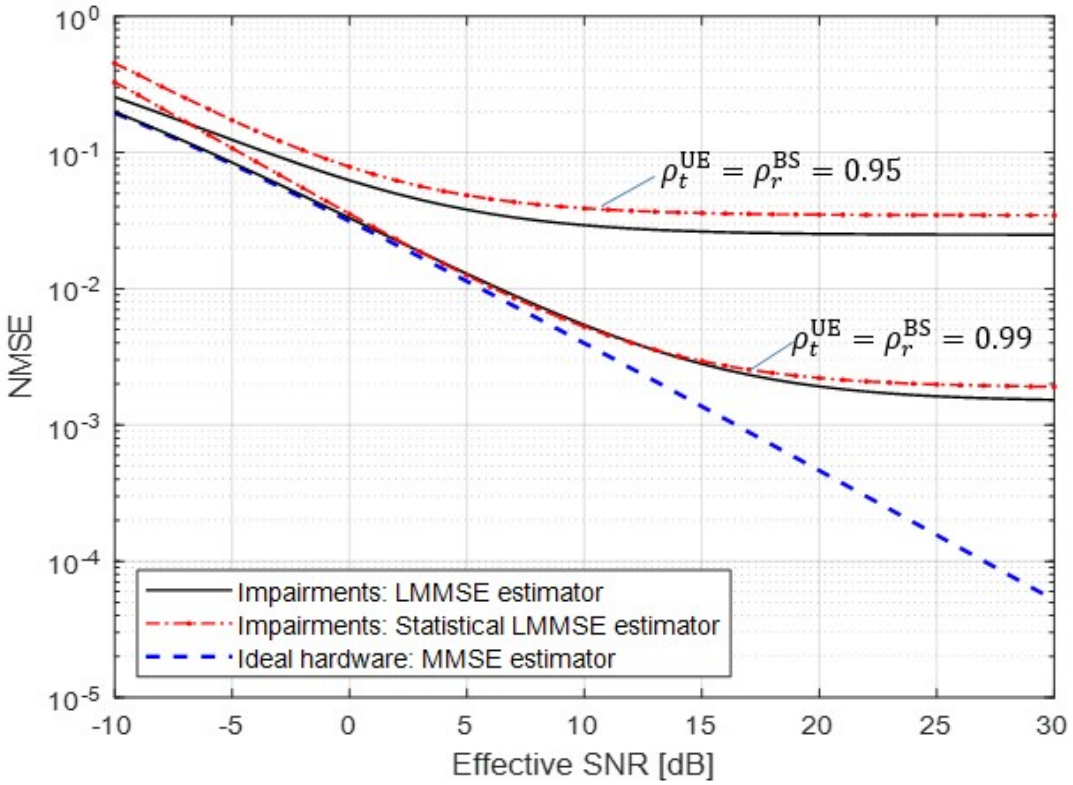

Figure 4.19: The normalized MSE of the practical channel estimation using the sample covariance with equal hardware quality at the transceiver ends.

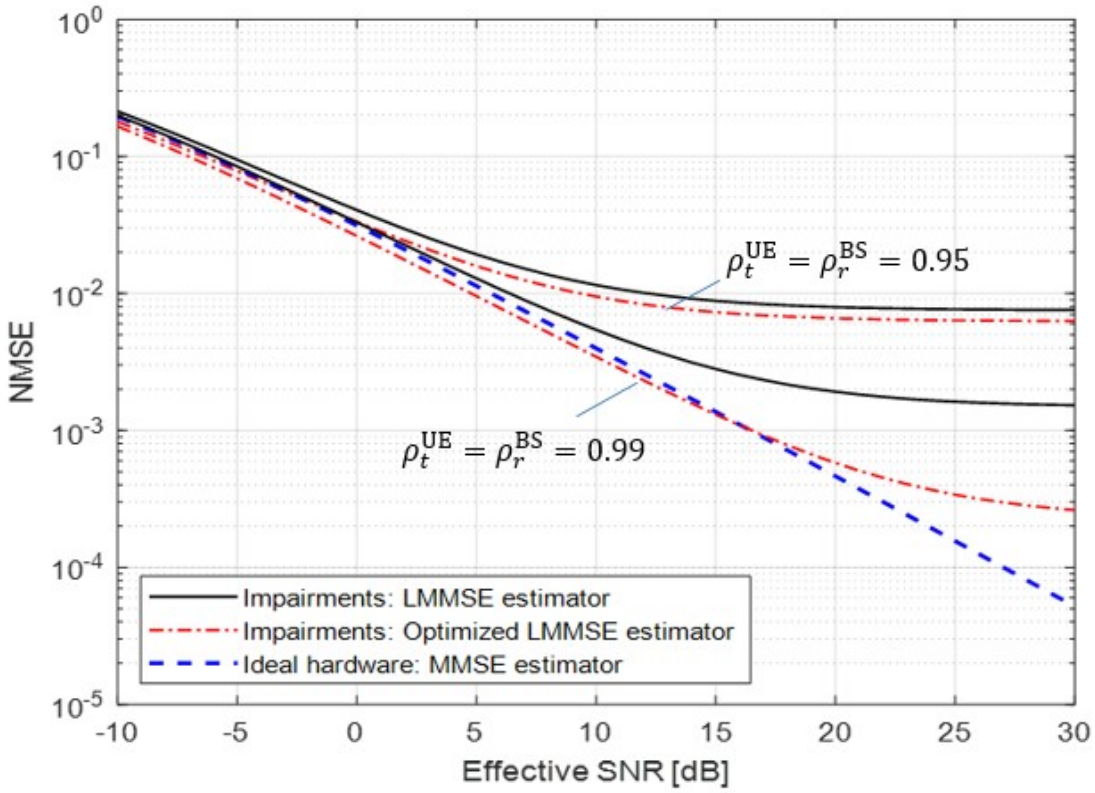

Figure 4.20: The normalized MSE of the optimized channel estimation using the sample covariance with equal hardware quality at the transceiver ends.

4.3 Results and Analysis of Proposed Model-3

In this section, the simulation results of the proposed model-3 will be discussed in terms of the refined parameters: The optimization algorithms, loss functions, the modulating type symbol, the non-linear polynomial types, and the number of UEs (K) in the cell as follows:

4.3.1 Optimization algorithm and loss function effect

The simulation results in Figs (4.21 to 4.32) demonstrate the deep learning-based estimation approach versus state-of-the-art estimators from Bayesian in terms of the normalized MSE for the effective channel and distortion correlation estimation. The simulation results in Figs (4.21 to 4.26) show the impact of applying different optimization algorithms and loss functions when estimating the effective channels using the DNN in Fig.3.4. On the other hand, Figs. (4.27 to 4.32) show the estimation quality in terms of the distortion correlation elements using the DNN in Fig.3.5.

The simulation results are averaged over 1000 different positions of each UE in the cell with 1000 channel realizations for each position. The QPSK constellations for $K = 10$ UEs with a third-order quasi-memoryless polynomial have been used for this case of simulation results. The DA-LMMSE estimate denotes the Distortion-Aware estimate given in Eq.(2.71) and is validated by the Monte Carlo simulation (DA-LMMSE Monte-Carlo). Also, the DuA-LMMSE refers to the Distortion-unAware estimate in Eq.(2.72), which is equivalent to the LMMSE estimate when there is ideal hardware on the BS and UEs.

The simulation results in Figs (4.21 to 4.26) show that the deep learning-based-Adam optimizer significantly outperforms the curves of the Bayesian estimators. It is noticed that the median of the NMSE has improved by about 3.2dB to 4.6dB when using the deep learning-based Adam optimizer compared to the distortion-aware and unaware estimators from Bayesian.

The simulation results of the distortion correlation matrix in Figs (4.27 to 4.32) are implemented using two schemes as follows:

1. A scheme of a Monte Carlo simulation for the LMMSE estimate of the normalized distortion correlation $\left[\mathbf{C}_{\mu\mu}\right]_{mm}/\sigma^2$, which is denoted by Linear

LMMSE estimate, and

2. A scheme of a Monte Carlo simulation for the LMMSE estimate of the logarithm of the distortion correlation in (1) and denoted by Logarithmic LMMSE estimate.

Although the Adam optimiser generally provides significant performance gain compared to the other deep learning optimizers like RMSprop and AdaDelta, the worst case in the simulation results arise in Figs 4.23, 4.26, 4.29 and 4.32 when using the AdaDelta optimizer. As a result, it can be deduced that the deep learning-based-estimating approach is not necessarily the best, but it may be the worst in some cases of optimizers when compared to the state-of-the-art Bayesian estimators and is optimiser-dependent.

The final convergence of the learning DNNs in Figs 3.4 and 3.5 regarding the training and validation loss are shown in Tables 4.2 and 4.3 for the effective channel and distortion correlation estimates in the case of $K = 10$ and QPSK constellation, respectively. Tables 4.2 and 4.3 show that the Huber loss function provides slightly better convergence than the MSE loss function.

Table 4.2: DNN convergence: effective channel estimate case, $K = 10$, QPSK

Optimizer and Loss function	Train loss	Train MSE	Val loss	Val MSE
Adam, mean squared error	0.0836	0.0836	0.0839	0.0839
Adam, Huber	0.0405	0.0822	0.0421	0.0852
RMSprop, mean squared error	0.1560	0.1560	0.1604	0.1604
RMSprop, Huber	0.0730	0.1488	0.0733	0.1494
AdaDelta, mean squared error	0.6689	0.6689	0.6597	0.6597
AdaDelta, Huber	0.3599	0.9979	0.3569	0.9855

Table 4.3: DNN convergence: : distortion correlation estimate case, $K = 10$, QPSK

Optimizer and Loss function	Train loss	Train MSE	Val loss	Val MSE
Adam, mean squared error	$2.3815e-04$	$2.3815e-04$	$2.9549e-04$	$2.9549e-04$
Adam, Huber loss	$1.0131e-04$	$2.0263e-04$	$1.1069e-04$	$2.2138e-04$
RMSprop, mean squared error	$5.9452e-04$	$5.9452e-04$	$6.6382e-04$	$6.6382e-04$
RMSprop, Huber loss	$4.0413e-04$	$8.0825e-04$	$6.3974e-04$	0.0013
AdaDelta, mean squared error	0.0071	0.0071	0.0072	0.0072
AdaDelta, Huber loss	0.0051	0.0103	0.0051	0.0102

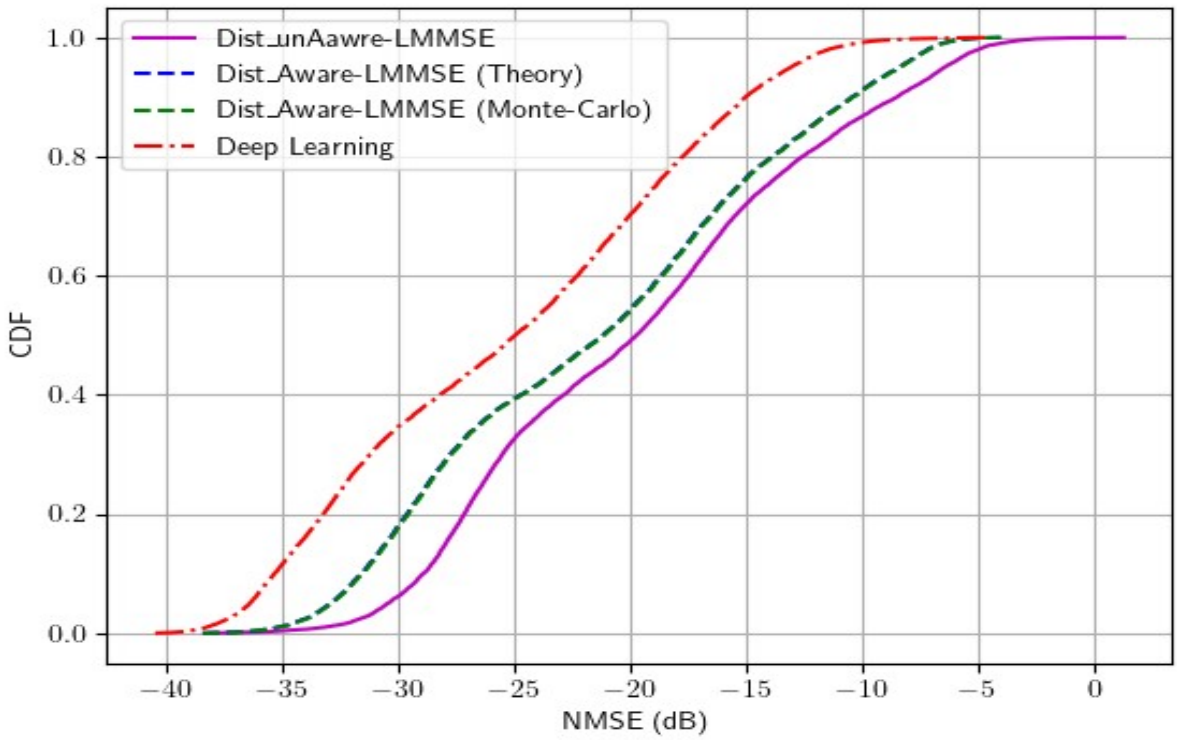

Figure 4.21: Effective channel estimates, loss: mean squared error, optimizer: Adam

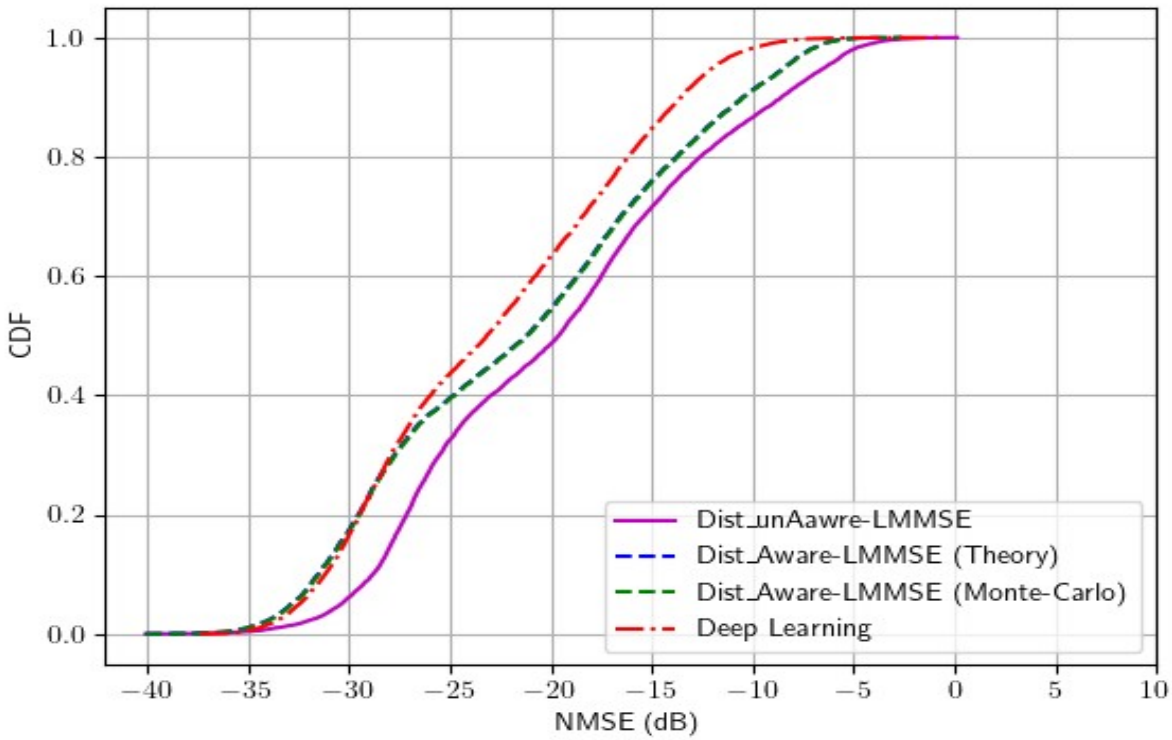

Figure 4.22: Effective channel estimates, loss: mean squared error, optimizer: RMSprop

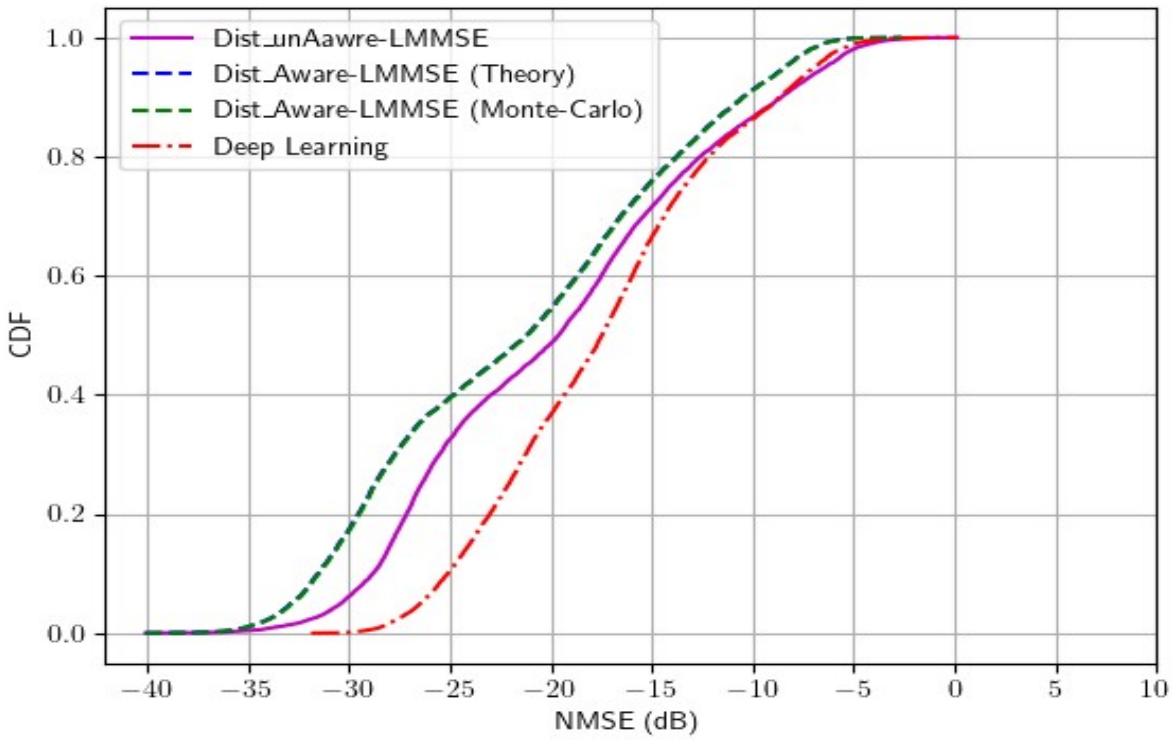

Figure 4.23: Effective channel estimates, loss: mean squared error, optimizer: AdaDelta

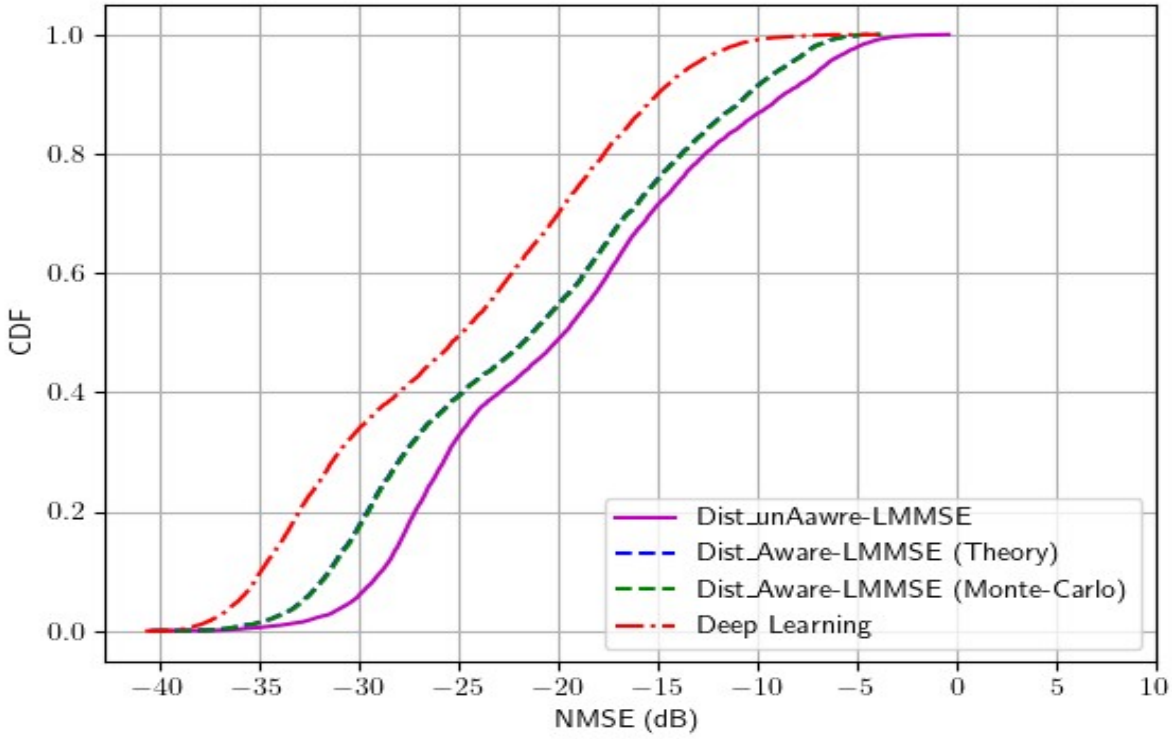

Figure 4.24: Effective channel estimates, loss: Huber, optimizer: Adam

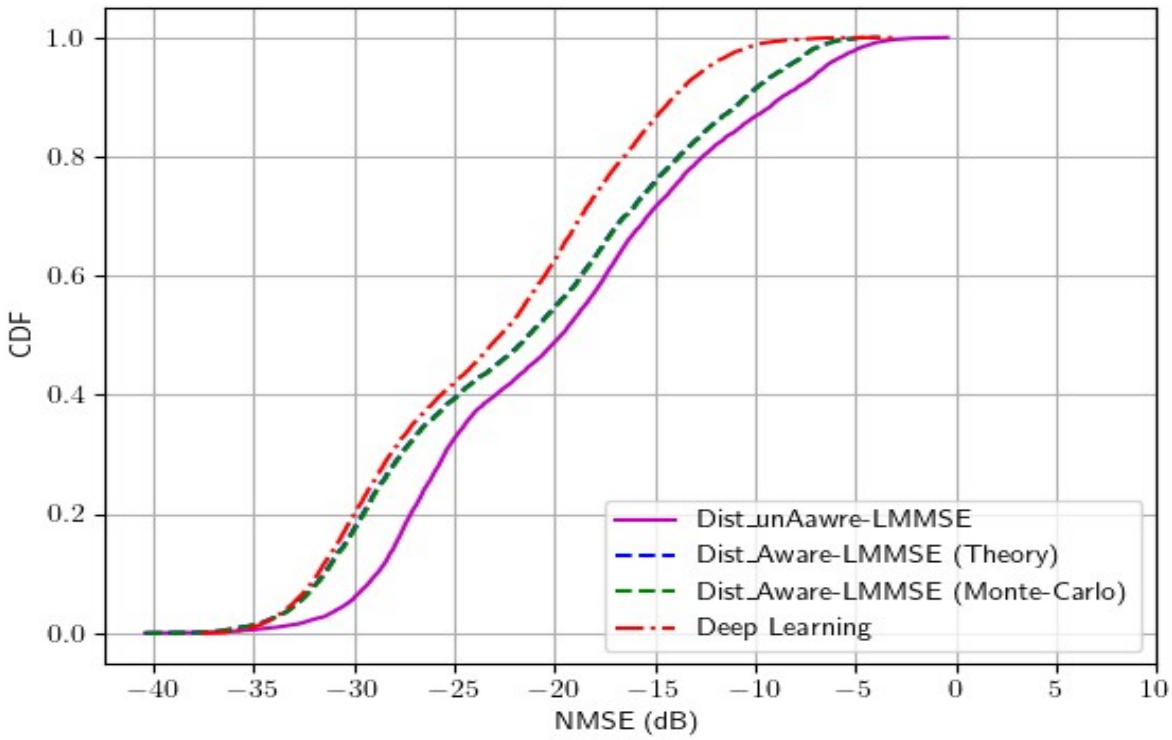

Figure 4.25: Effective channel estimates, loss: Huber, optimizer: RMSprop

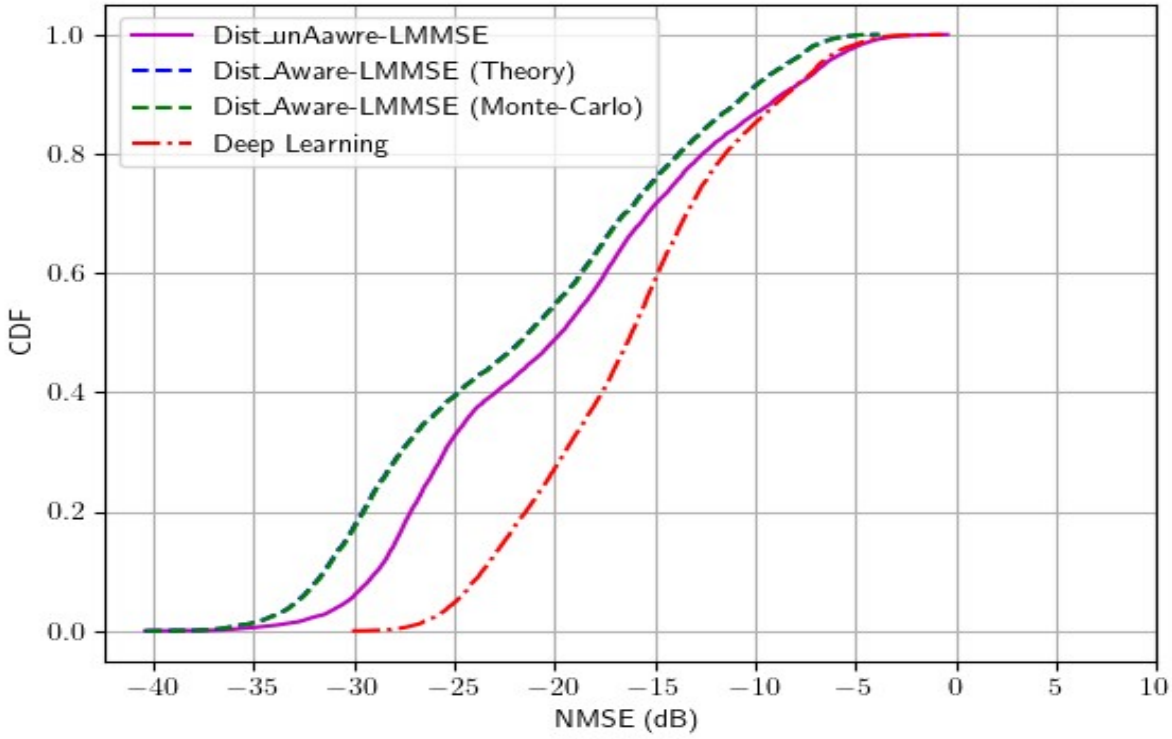

Figure 4.26: Effective channel estimates, loss: Huber, optimizer: AdaDelta

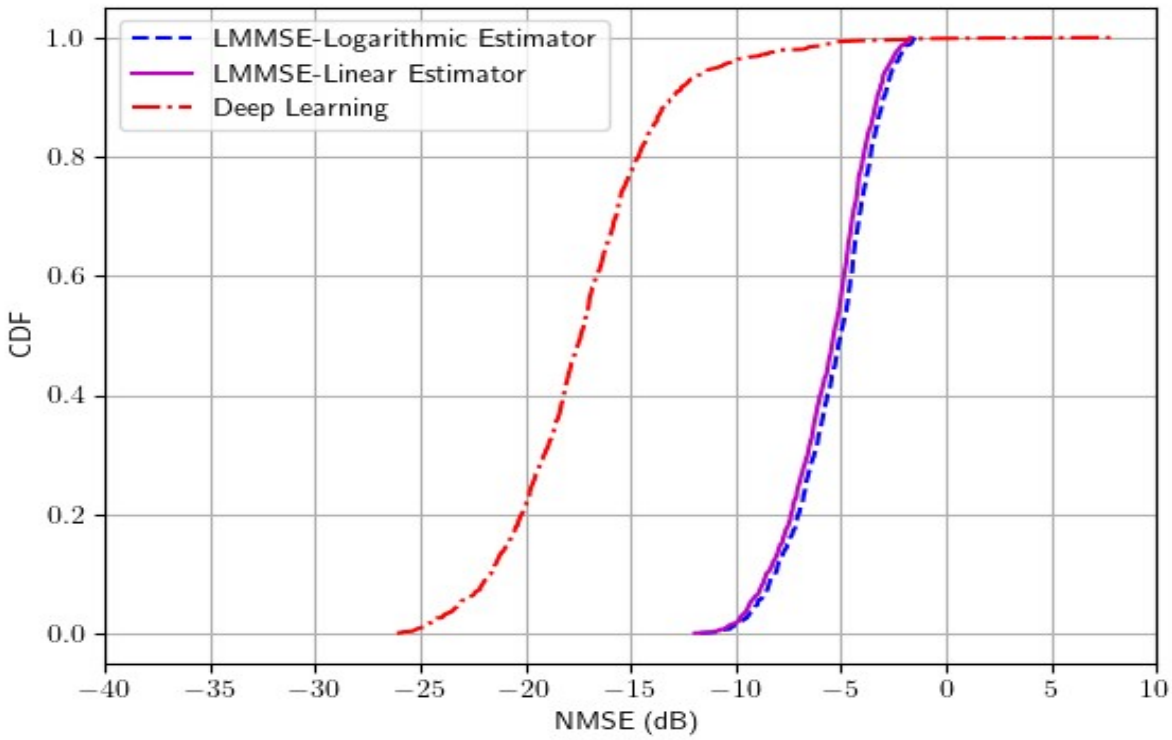

Figure 4.27: Distortion correlation estimates, loss: mean squared error, optimizer: Adam

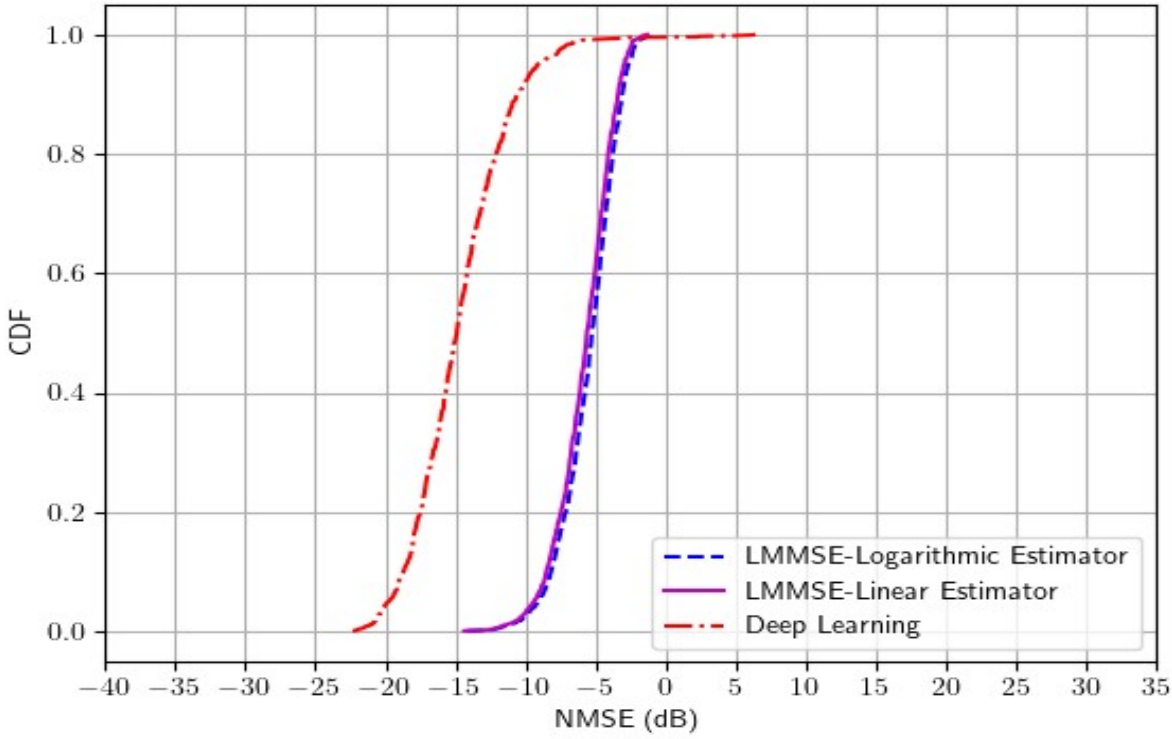

Figure 4.28: Distortion correlation estimates, loss: mean squared error, optimizer: RMSprop

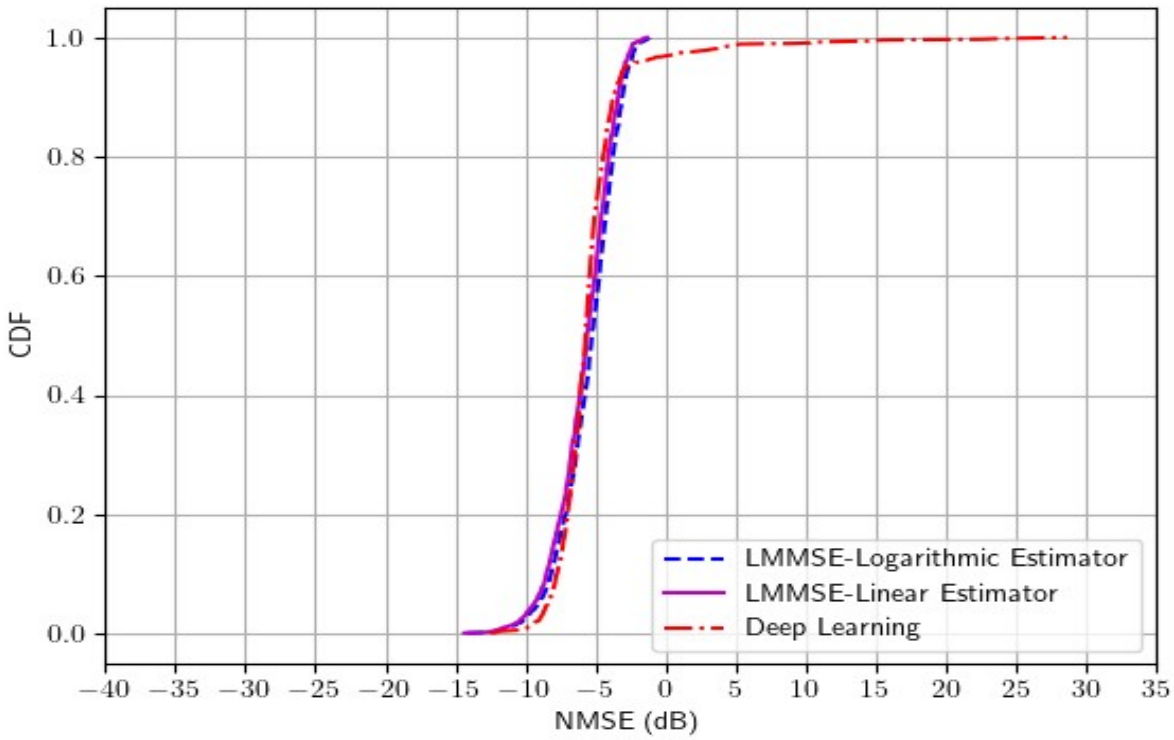

Figure 4.29: Distortion correlation estimates, loss: mean squared error, optimizer: AdaDelta

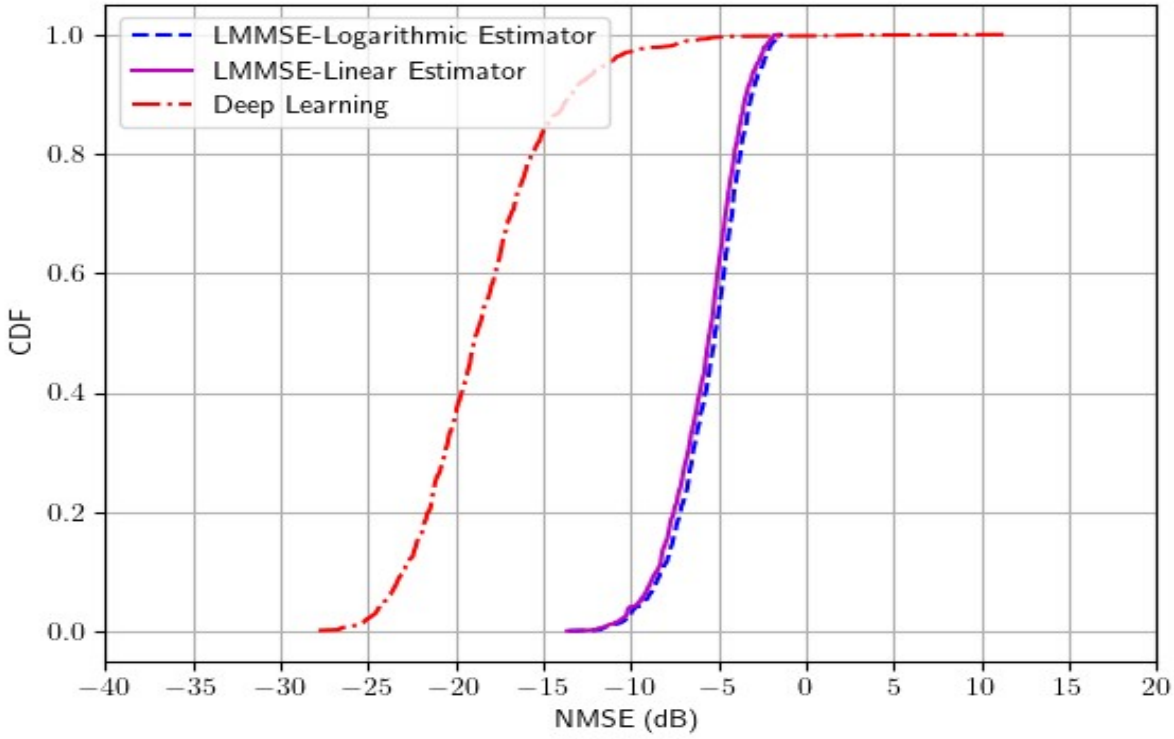

Figure 4.30: Distortion correlation estimates, loss: Huber, optimizer: Adam

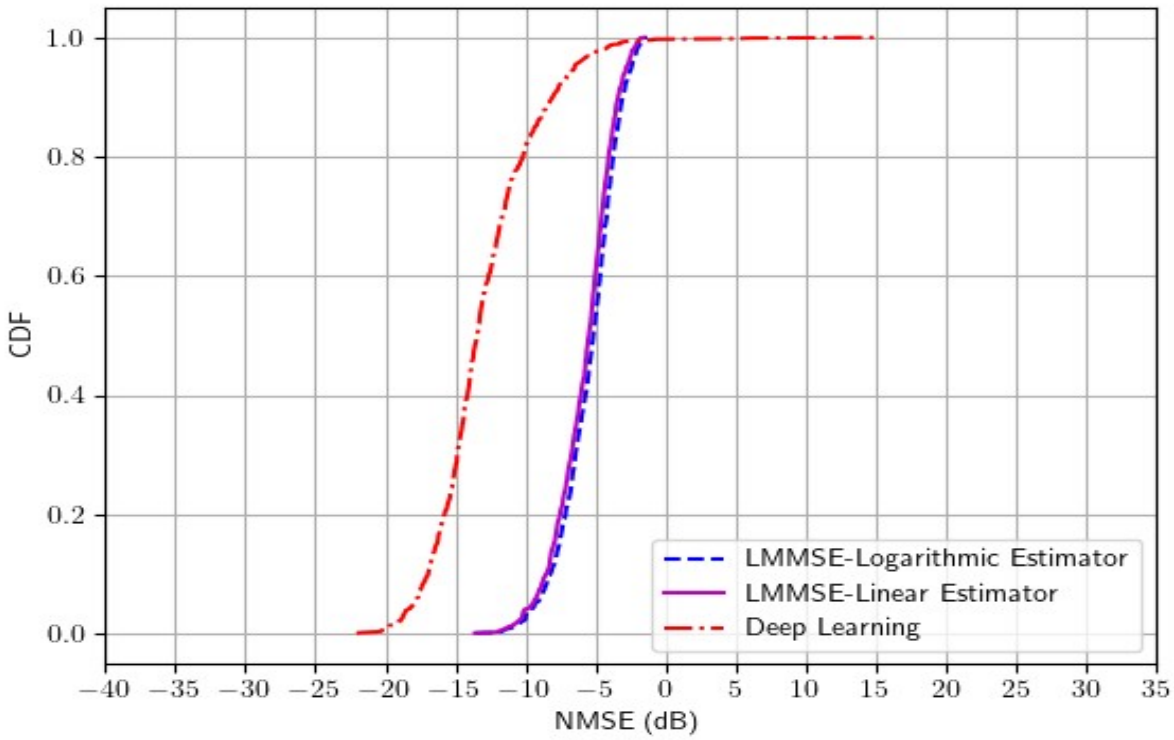

Figure 4.31: Distortion correlation estimates, loss: Huber, optimizer: RMSprop

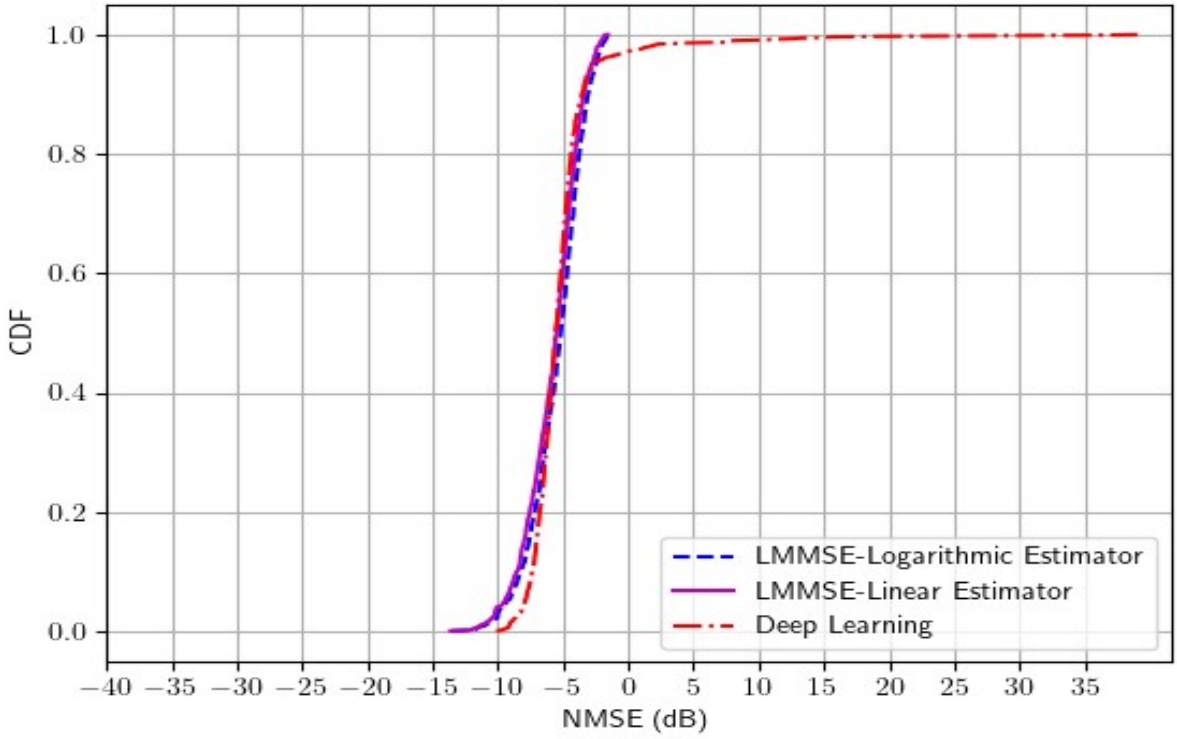

Figure 4.32: Distortion correlation estimates, loss: Huber, optimizer: AdaDelta

4.3.2 Modulating symbol and polynomial type effect

The simulation results in Figs. (4.33-4.40) show the performance of the deep learning approach compared to the Bayesian estimators using the third-order memoryless polynomial (with real coefficients) in addition to the complex coefficients quasi-memoryless polynomial and two types of symbols, the QPSK and complex Gaussian symbols.

The simulation results in Fig.4.33 use Gaussian symbols instead of QPSK and show a performance gain of about 4dB better than the deep learning-based effective channel estimate in Fig.4.34; this is because the Gaussian symbols increase the information entropy of the input signals and lead to enhance their estimation at the BS. On the other hand, the deep learning-based distortion correlation estimates in Fig.4.35 significantly improve the NMSE error compared to the deep learning performance in Fig.4.23 due to the same reason above.

More improvements of the deep learning approach (with a performance gain of about 6dB better than Bayesian) are gotten when using the suggested real coefficient memoryless polynomial with Gaussian symbols in Fig.4.37 compared to the results in Fig.4.38. This contributes to reduced computations by deep learning as an estimator when dealing with real coefficient polynomials instead of complex ones. On the other hand, when using Gaussian symbols, the entropy of the transmitted symbols will be increased, yielding to decrease in the probability of the estimation error at the BS.

The Other simulation results in Figs 4.39 and 4.40 show the distortion correlation estimation using Fig.3.5 when using Gaussian with memoryless polynomial and QPSK symbol modulation with memoryless polynomial, respectively.

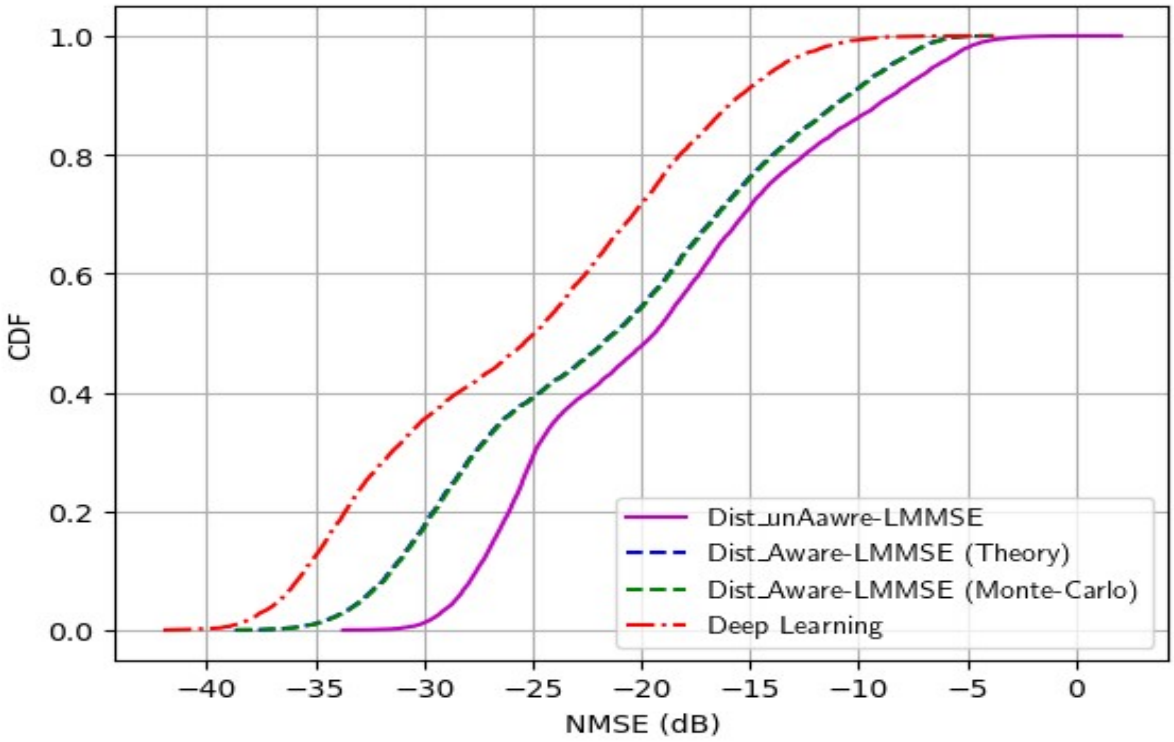

Figure 4.33: Effective channel estimates, Gaussian, quasi-memoryless

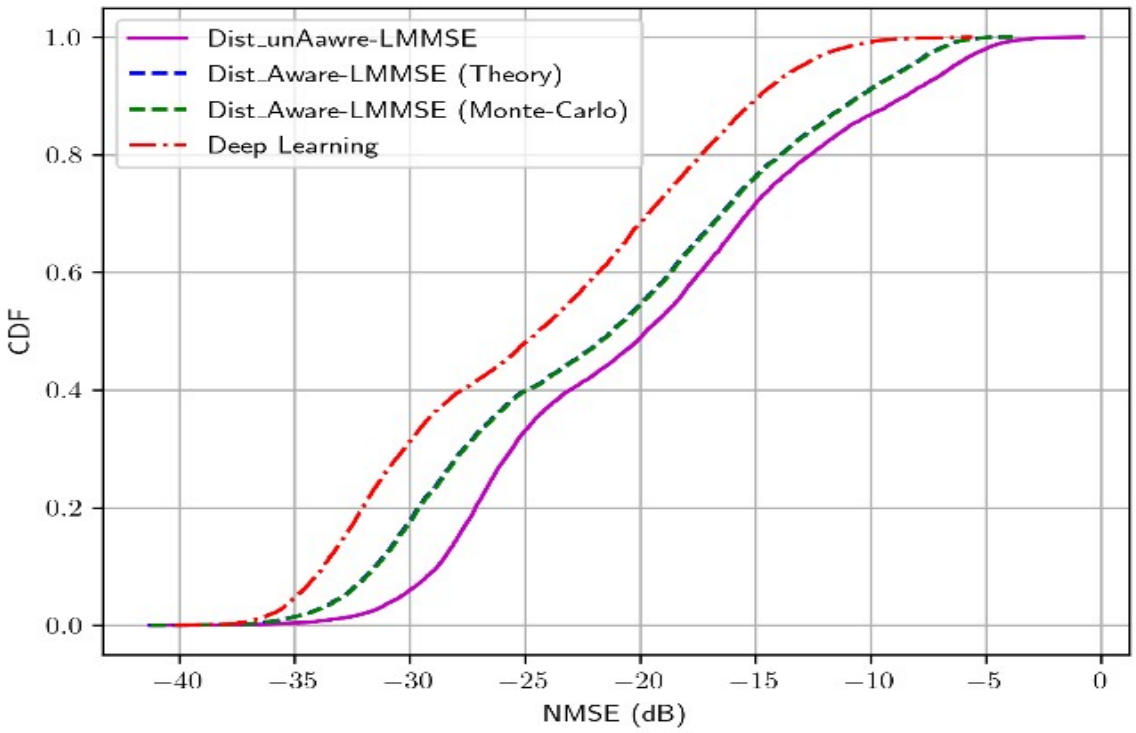

Figure 4.34: Effective channel estimates, QPSK, quasi-memoryless

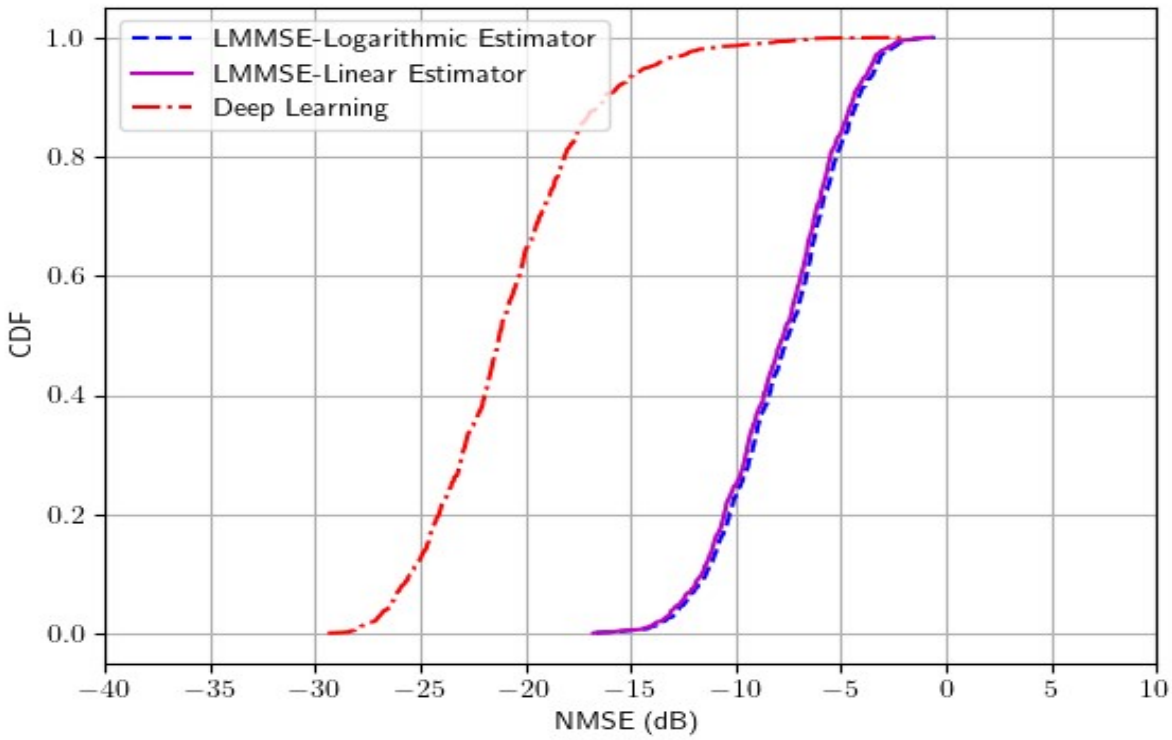

Figure 4.35: Distortion correlation estimates, Gaussian, quasi-memoryless

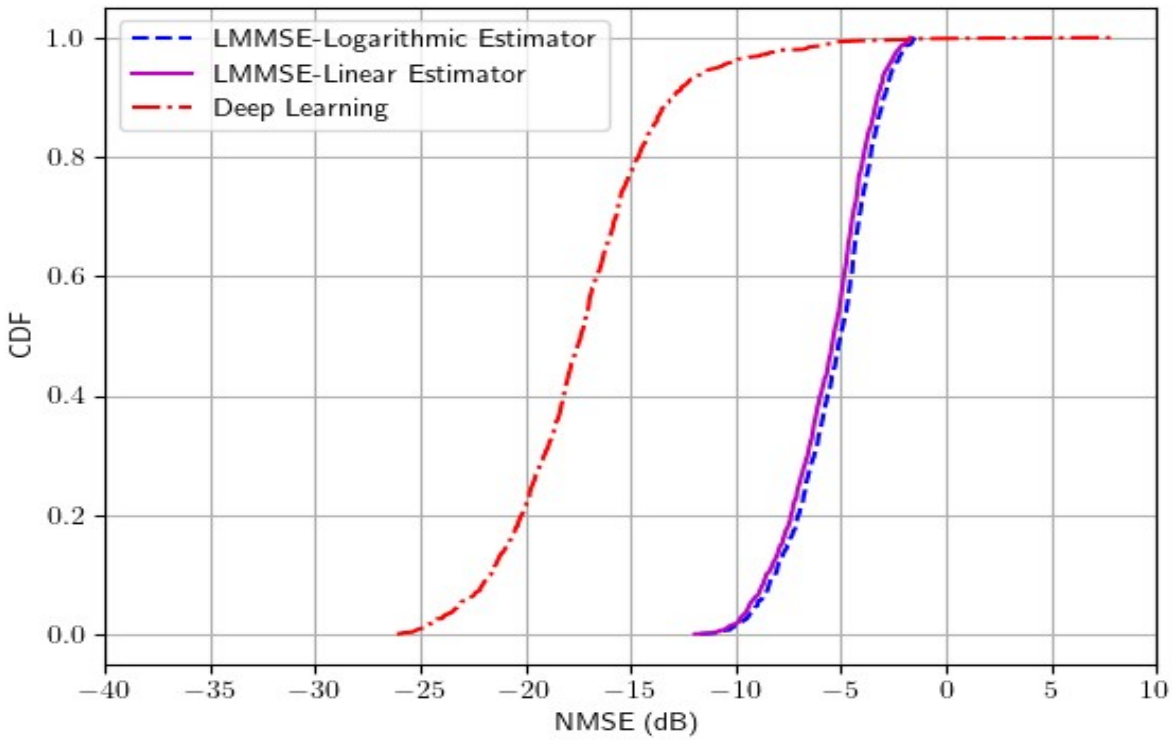

Figure 4.36: Distortion correlation estimates, QPSK, quasi-memoryless

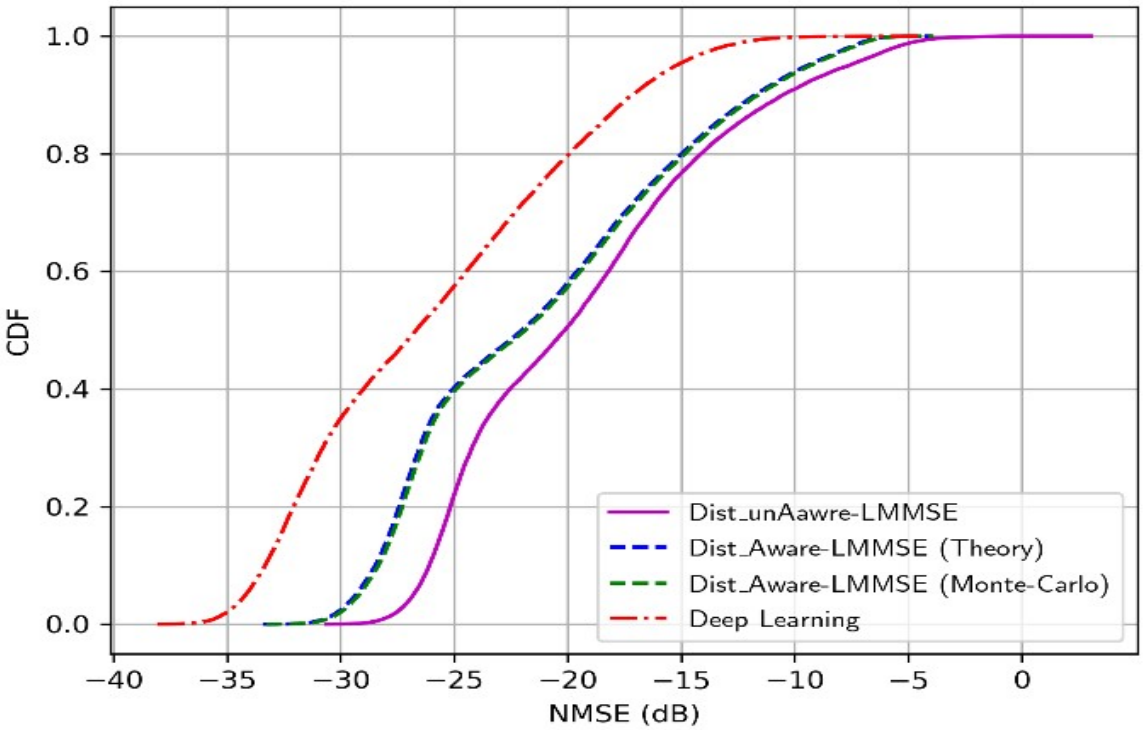

Figure 4.37: Effective channel estimates, Gaussian, memoryless

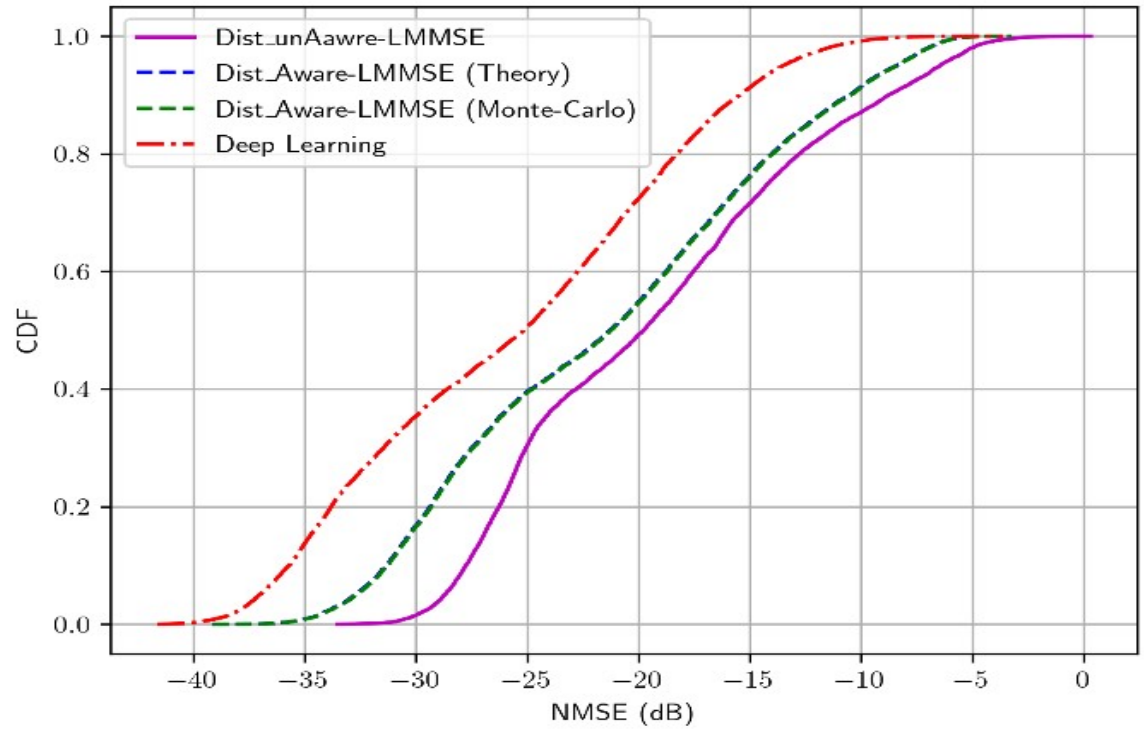

Figure 4.38: Effective channel estimates, QPSK, memoryless

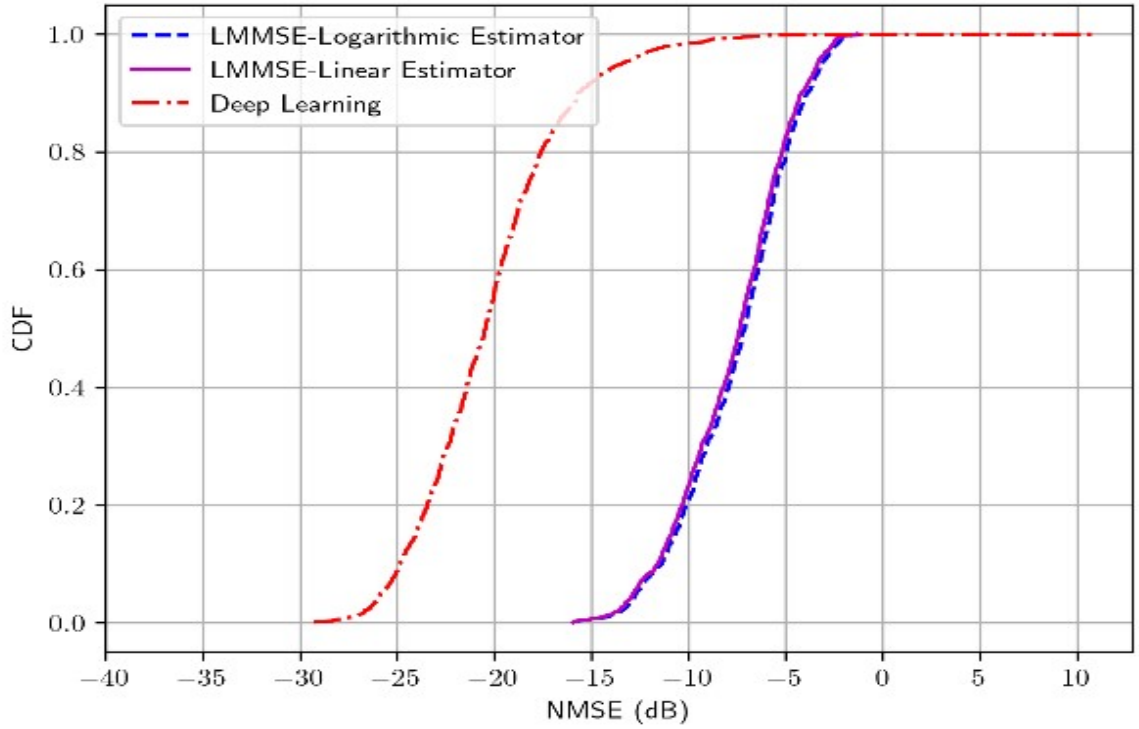

Figure 4.39: distortion correlation estimates, Gaussian, memoryless

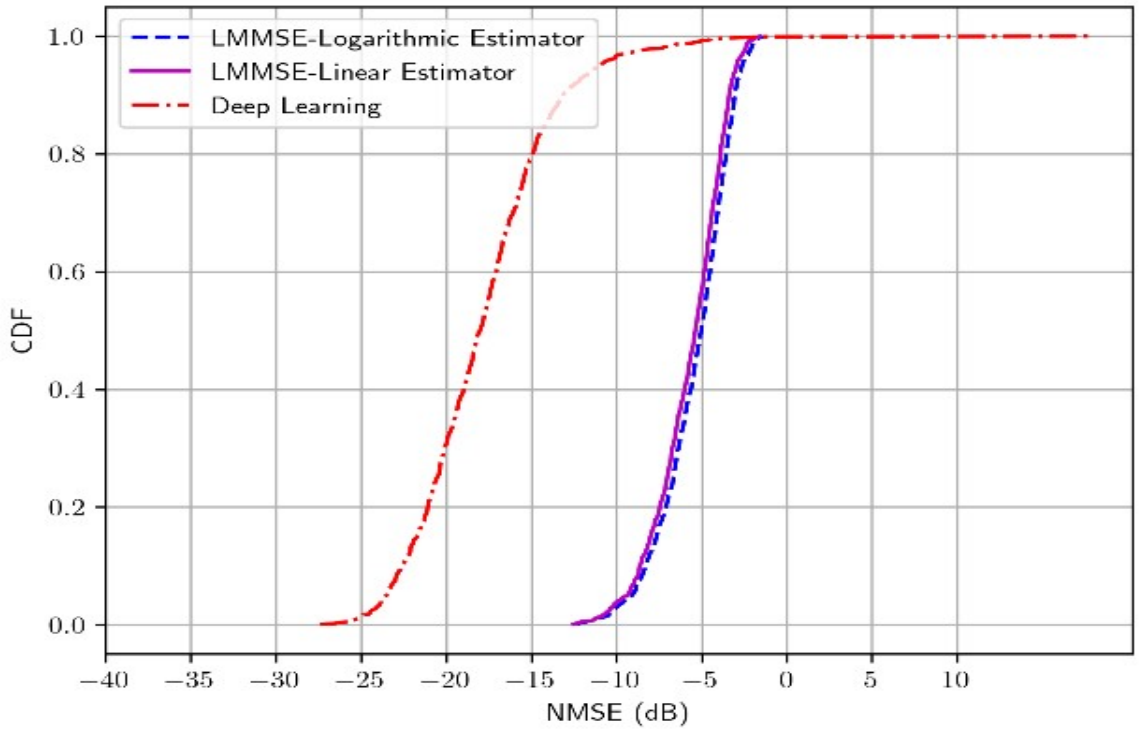

Figure 4.40: distortion correlation estimates, QPSK, memoryless

4.3.3 Number of users (K) effect

The effects of increasing the number of UEs on the quality of the estimated channels will be illustrated in Figs. (4.41 to 4.46) as follows:

The Figs. (4.41 to 4.44) show the simulation results of estimating the effective channel and the distortion correlation elements using the number of UEs (K=20) under the Adam optimizer tested with two loss functions (the mean squared error and Hubber loss functions). The simulation results in these figures show that the same performance with little improvement has gotten compared to the K=10 number of users.

The Figs. (4.45 and 4.46) show the simulation results of the effective channel estimates and the distortion correlation using the number of UEs (K=30) under the parameter: Adam optimizer, mean squared error loss function, Gaussian symbols, memoryless polynomial. Although these parameters were the best setting for better performance in the case of K=10, the simulation results with K=30 show a bad performance in the case of the distortion correlation elements while the same performance as the case of the effective channel estimation at k=10. The reason behind that is the increasing number of input features per training sample (distortions by K=30 UEs), which will cause the DNN in Fig.3.5 to be in a high bias problem, making the deep learning process worse than the Bayesian methods.

However, it is very important to mention that the enlarging of the architecture of the DNNs in Figs.3.4 and 3.5 by increasing the number of neurons or hidden layers has not provided any performance improvements, as shown in Figs. (4.47 and 4.48). Also, it is proved that using a larger dataset (5×10^6 instead of 3×10^6) did not give any improvements in the deep learning performance used in this thesis, as shown in Figs. (4.49 and 4.50).

It should be noted that all the simulation results in the proposed model-3 have been implemented using the TensorFlow environment in Python version 3.9 installed on a Laptop with a graphic card of NVIDIA RTX2070 MAx-Q, 16GB RAM and Core i7-10750H processor.

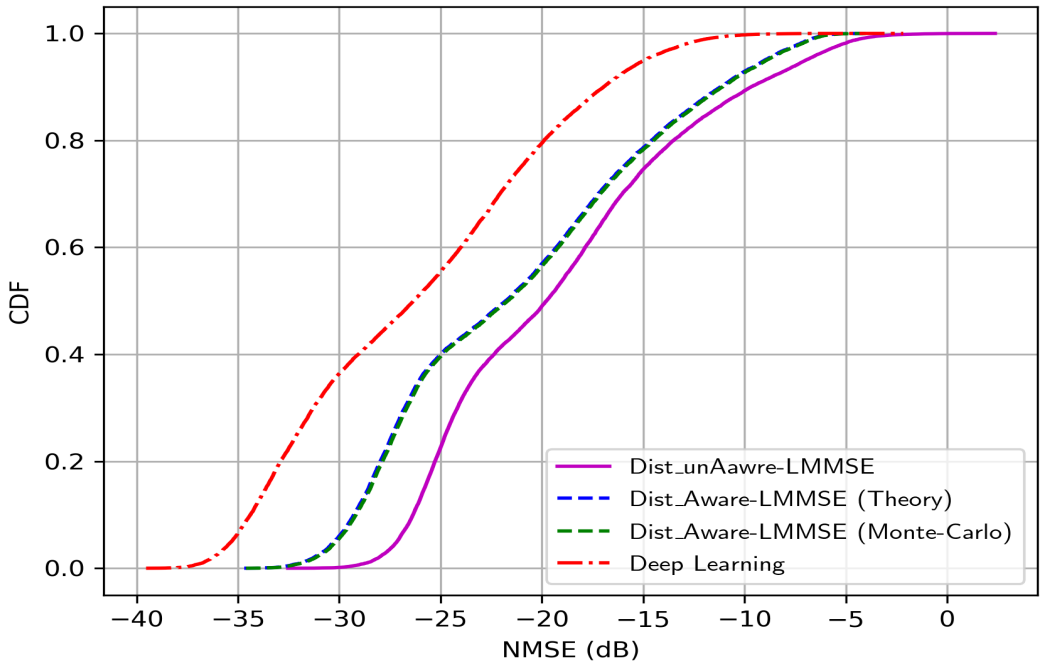

Figure 4.41: Effective channel estimates, Gaussian, quasi-memoryless, Adam optimizer, MSE loss function, K=20

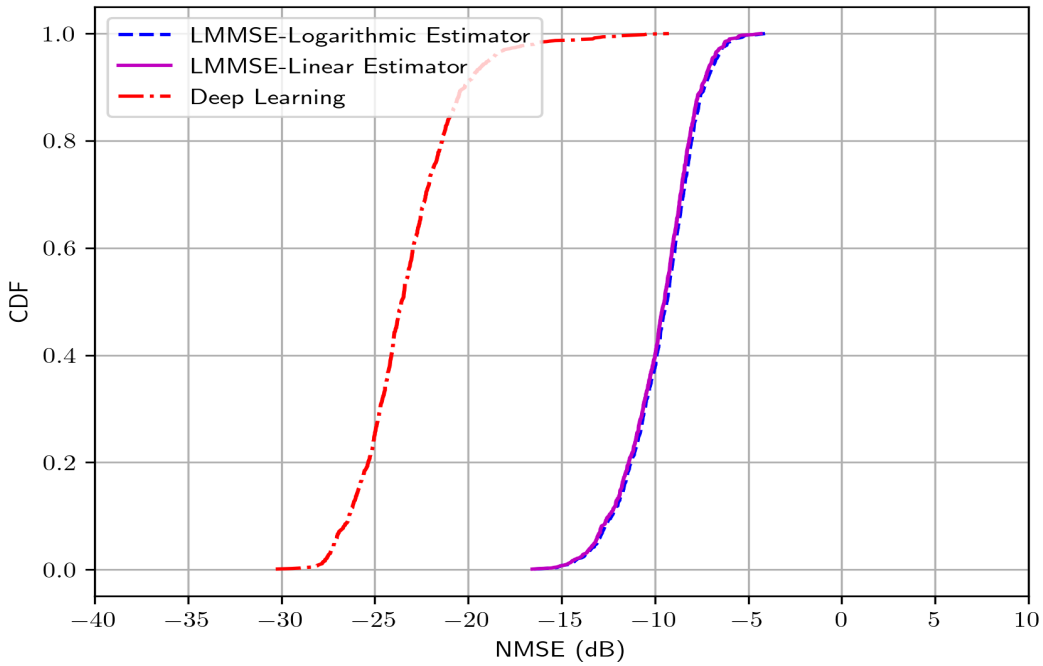

Figure 4.42: Distortion correlation estimates, Gaussian, quasi-memoryless, Adam optimizer, MSE loss function, K=20

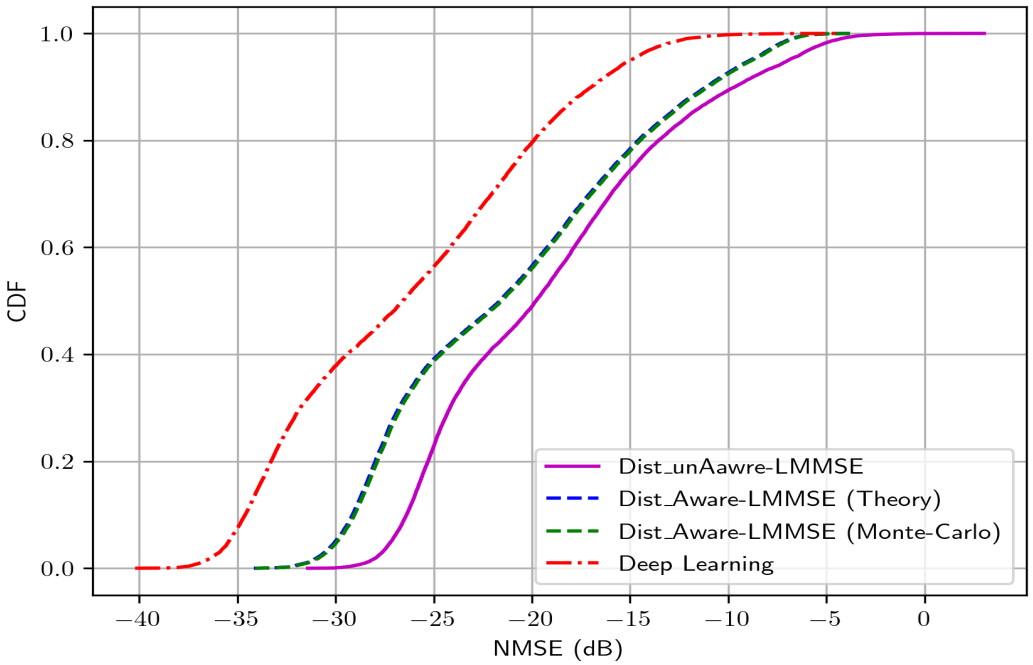

Figure 4.43: Effective channel estimates, Gaussian, quasi-memoryless, Adam optimizer, Huber loss function, K=20

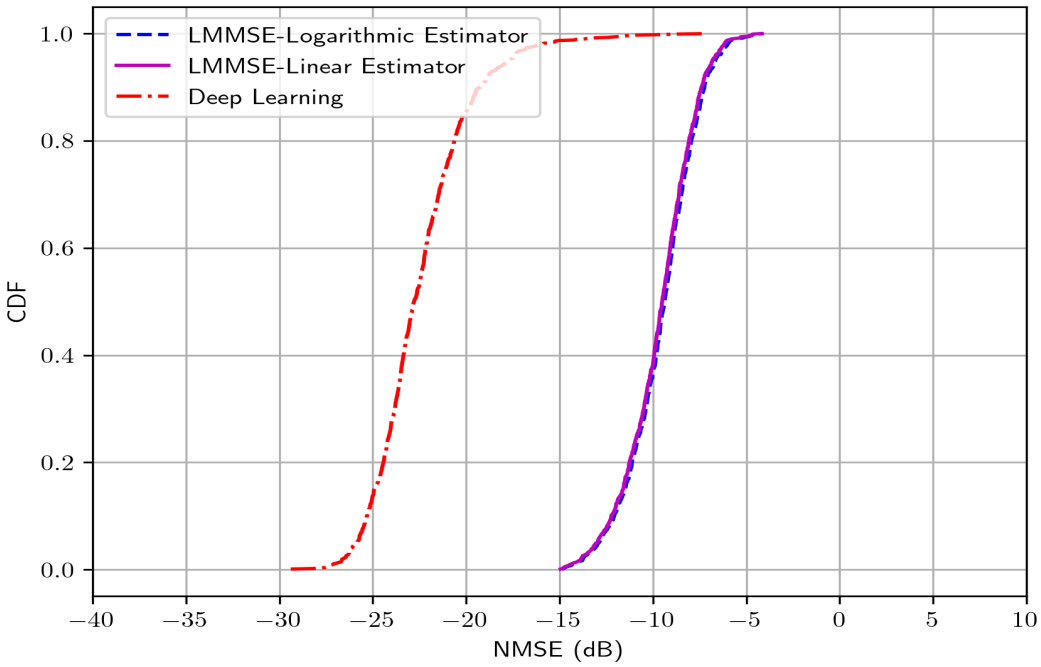

Figure 4.44: Distortion correlation estimates, Gaussian, quasi-memoryless, Adam optimizer, Huber loss function, K=20

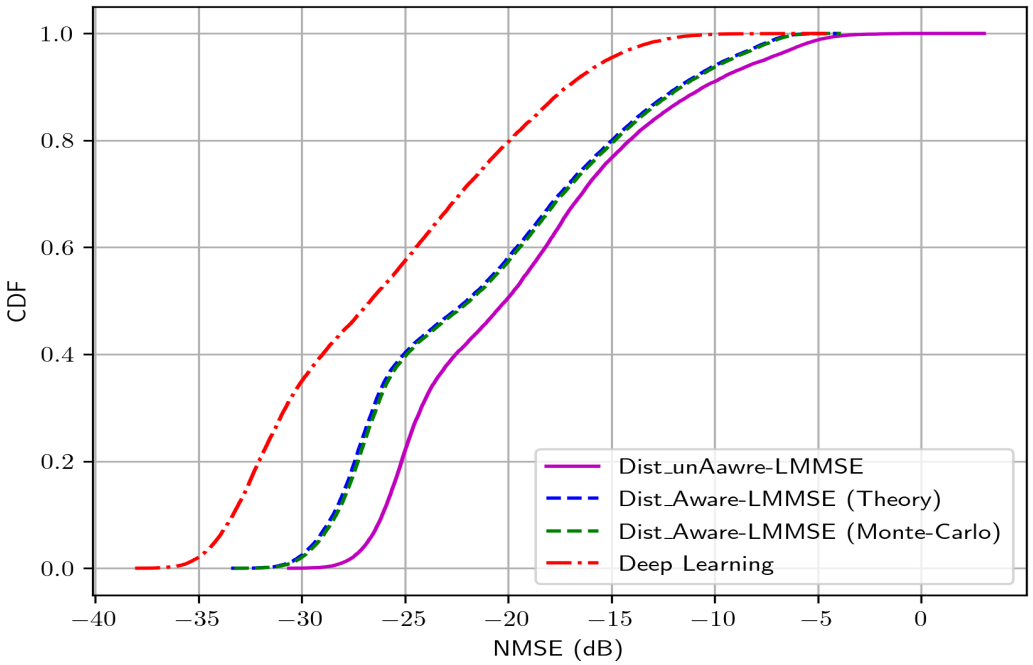

Figure 4.45: Effective channel estimates, Gaussian, memoryless, Adam optimizer, MSE loss function, K=30

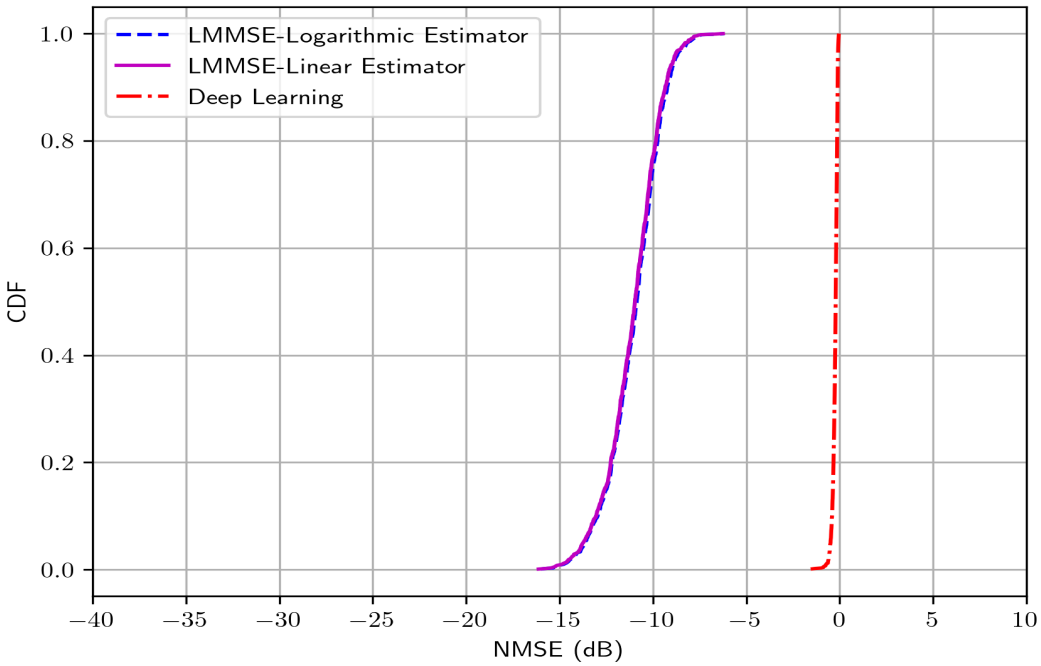

Figure 4.46: Distortion correlation estimates, Gaussian, memoryless, Adam optimizer, MSE loss function, K=30

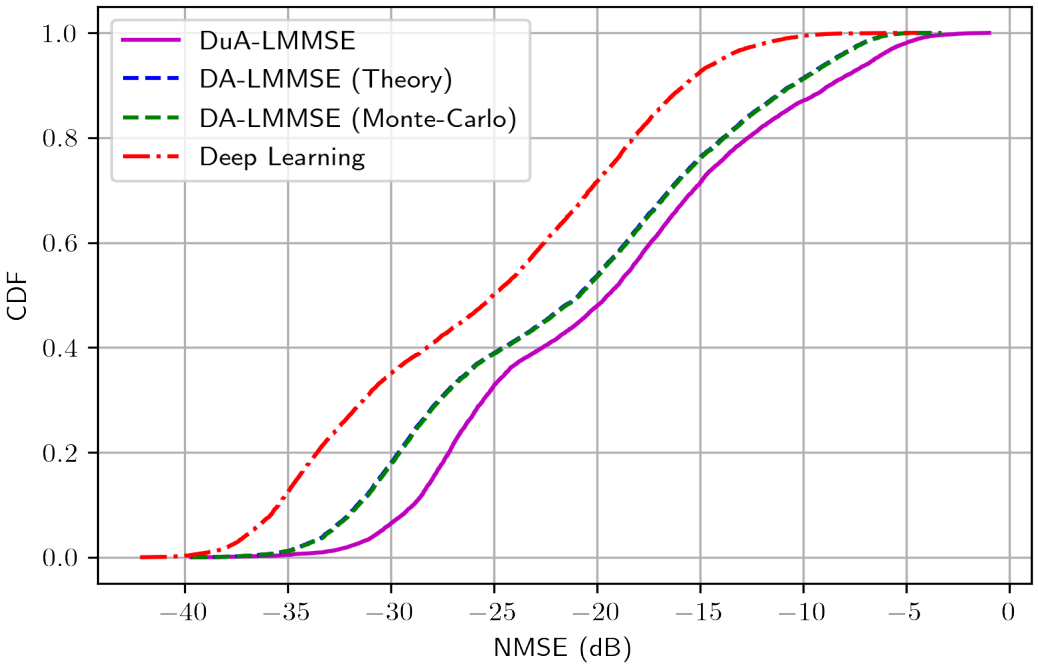

Figure 4.47: Effective channel estimates, QPSK, quasi-memoryless, Adam optimizer, MSE loss function, K=10, using Large DNN

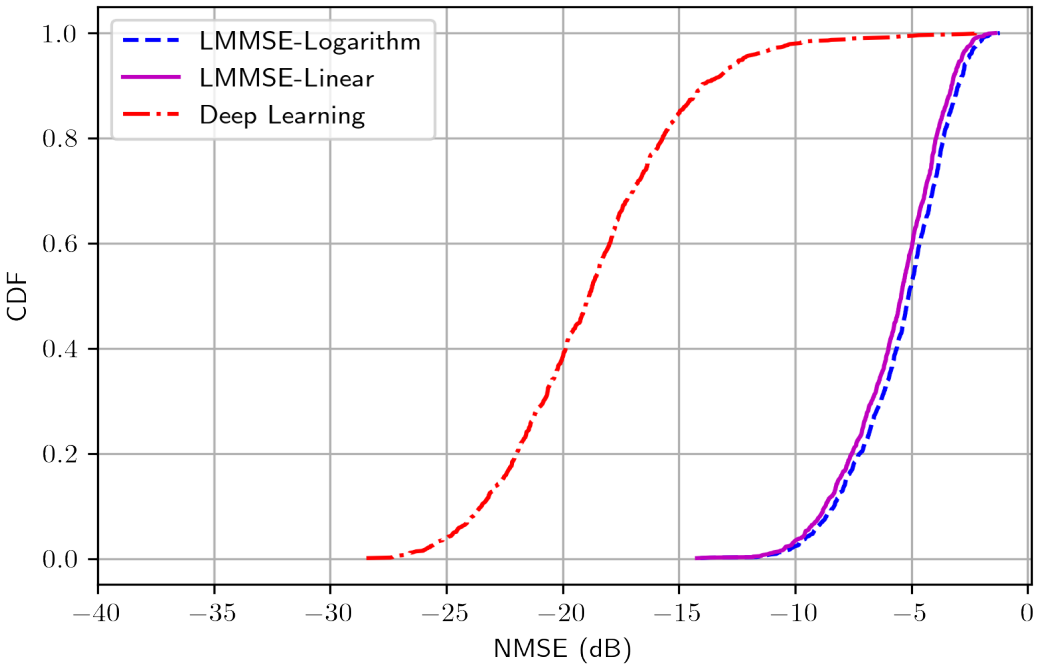

Figure 4.48: Distortion correlation estimates, QPSK, quasi-memoryless, Adam optimizer, MSE loss function, K=10, using Large DNN

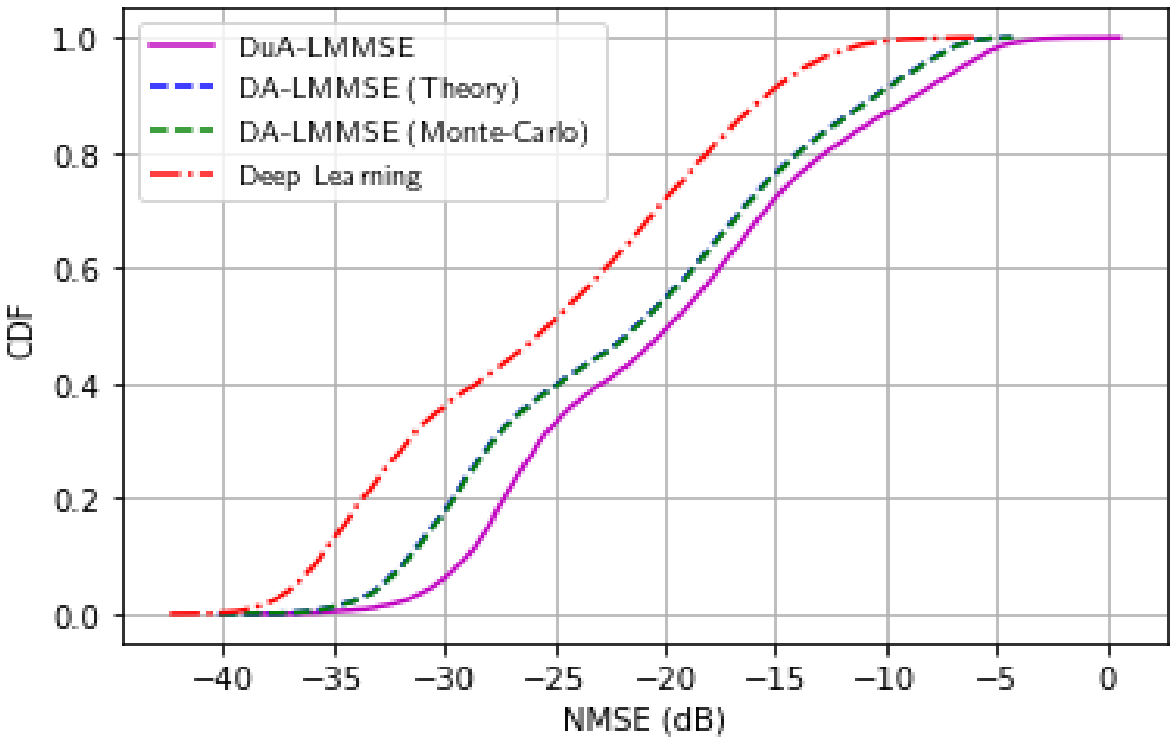

Figure 4.49: Effective channel estimates, QPSK, quasi-memoryless, Adam optimizer, MSE loss function, K=10, using Large dataset

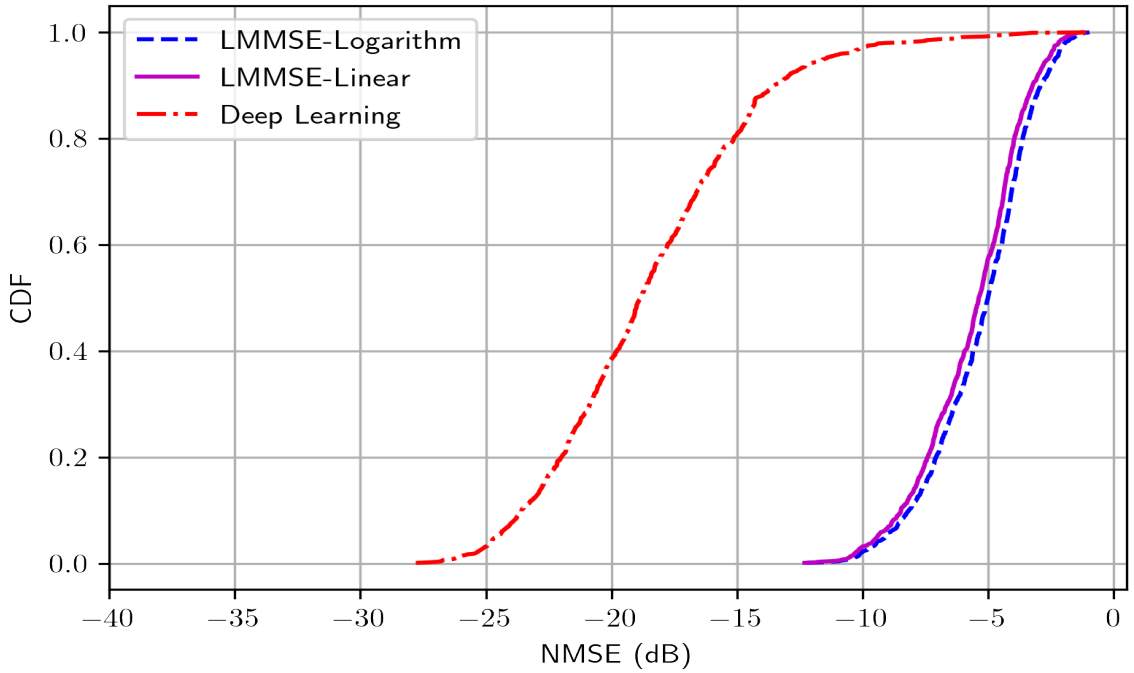

Figure 4.50: Distortion correlation estimates, QPSK, quasi-memoryless, Adam optimizer, MSE loss function, K=10, using Large dataset

4.3.4 Comparison works

The following Tables 4.4 and 4.5 compare the thesis work to some other state-of-the-art works used in this thesis as follows:

Table 4.4: Comparison with Proposed Model-1

Computational Complexity			
	Method	FLOPs Polynomial	reduction ratio/order
Thesis Works	Used the spatially correlated channels with Factor-Solve Algorithm	In terms of number of antennas: M^3+4M^2+M, and in terms of spatial rank: r^3+4r^2+r	To 30% of computations with $\mathcal{O}(r^3)$ where r is the spatial rank=64
[11]	Used the full rank channels	In terms of number of antennas: $(4/3)M^3+M^2+(\tau_p-1/3)M$, τ_p is the pilot length	Full computations with $\mathcal{O}(M^3)$ where M is the number of antennas
[16]	Used Polynomial ExpAnsion CHannel (PEACH) Algorithm	$k_c\left[(8L+4)M^2-(4L+2)\right]+k_s[2M^2-M]$ Where L is a polynomial degree, k_c is the precomputations per channel realization, and k_s is the precomputations per coherence time.	To $\mathcal{O}(M^2)$ of computations where M is the number of antennas at the BS
[19]	Used the Neumann series	Reduces the computational complexity of MIMO signal detection from $\mathcal{O}(BU^2+U^3)$	To $\mathcal{O}(BUN)$, where B, U, and N are the numbers of antennas at the base station, user equipment, and NS terms, respectively

Table 4.5: Comparison with Proposed Model-2

Error Floor Reduction			
	Channel model	Estimation Method	Error reduction
Thesis Works	Used the practical channel	Estimated the channel from the unknown covariance matrix using the sample covariance method with convex optimization	The error floor reduced to less than 10^{-2} when $\rho_t^{\mathrm{UE}}=\rho_r^{\mathrm{BS}}=0.95$ and to less than 10^{-3} in case of $\rho_t^{\mathrm{UE}}=\rho_r^{\mathrm{BS}}=0.99$
[11, 54]	Used the local scattering channel model	Estimated the channel from the known covariance matrix based on some parameters	The error floor reduced to 10^{-1} when $\rho_t^{\mathrm{UE}}=\rho_r^{\mathrm{BS}}=0.95$ and to less than 10^{-2} in case of $\rho_t^{\mathrm{UE}}=\rho_r^{\mathrm{BS}}=0.99$

Chapter 5

Conclusions and Future Works

5.1 Conclusions

This thesis presents three proposed channel estimation models in the massive MIMO system, taking into account the BS and UEs hardware qualities. Several conclusions from the works in this thesis can be listed as follows:

The conclusions from the first proposed channel estimation model can be listed as follows:

1. Converting the channel into spatially correlated channels with choosable rank; contributes to reducing the interference from other users since each user will have its own living subspace or degree of freedom at the BS.
2. Reducing the dimension of the covariance matrices according to the reduced rank will reduce the computational complexity of the Bayesian MMSE estimator.
3. The estimation error variance in some eigendirections will be reduced by SVD truncation, resulting in accurate channel estimation.
4. Exploiting the spatial characteristics to reduce the number of antennas at the BS (i.e., using 64 instead of 100); will also reduce the cost of building such a large MIMO system since it reduces the number of radio frequency chains that are needed for the BS.

The conclusions from the second proposed channel estimation model can be listed as follows:

1. The results show that our proposed technique beats the usual one in terms of lowering the error floor (almost by one order of magnitude).
2. Lowering the hardware quality at the BS while maintaining nearly the same quality of estimated channels will contribute to reducing the cost of building such a large system, which is the main advantage of this proposed procedure.

The conclusions from the third proposed channel estimation model can be listed as follows:

1. The simulation results have shown that the deep learning-based estimation is not always the best in performance, but it may be the worst with some optimizers like AdaDelta.
2. The Adam optimizer with the Gaussian sending symbols and memoryless distorting polynomial has shown a significant performance gain of about 11dB reducing the NMSE.
3. During the training process of DNNs, it was found that increasing the number of hidden layers and/or the number of neurons inside each layer did not enhance the learning process.

In Conclusion, this thesis contributes to the field of channel estimation in the massive MIMO system by suggesting accurate and efficient channel estimation models that take into account the hardware quality at the BS and UEs. The proposed methods have been evaluated using simulations in MATLAB and Python languages. The simulation results in all our proposed models have demonstrated their effectiveness in reducing the computational complexity and the error floor and improving the estimation quality.

5.2 Future Works

1. It can extend the first proposed model to multi-cell scenarios with more real channel models that consider the elevated angle in addition to the azimuth angle of the reaching paths to the BS. Also, it can use different low-complexity estimators with sub-optimal performance.

2. It can develop a novel uplink channel model for the second proposed channel estimation model using the nonlinear hardware distortion effect at both the BS and UEs in addition to the residual distortions.

3. It can use different DNN architectures like one-dimensional convolutional neural networks (CNN) to improve the learning of the distortion characteristics and enhance the estimation quality in the third channel estimation model.

References

[1] M. Cooper, “The myth of spectrum scarcity,” *DYNA llc, Tech. Rep., Mar*, 2010.

[2] A. Ericsson, “Ericsson mobility report. ericsson,” tech. rep., Sweden, Tech. Rep. EAB-17, 1. 1, Rapport technique, 2017.

[3] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, “Massive mimo for next generation wireless systems,” *IEEE communications magazine*, vol. 52, no. 2, pp. 186–195, 2014.

[4] T. L. Marzetta, “Massive mimo: an introduction,” *Bell Labs Technical Journal*, vol. 20, pp. 11–22, 2015.

[5] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, “An overview of massive mimo: Benefits and challenges,” *IEEE journal of selected topics in signal processing*, vol. 8, no. 5, pp. 742–758, 2014.

[6] F. A. P. de Figueiredo, “An overview of massive mimo for 5g and 6g,” *IEEE Latin America Transactions*, vol. 20, no. 6, pp. 931–940, 2022.

[7] J. Zheng, J. Zhang, E. Björnson, Z. Li, and B. Ai, “Cell-free massive mimo-ofdm for high-speed train communications,” *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 10, pp. 2823–2839, 2022.

[8] T. Schenk, *RF imperfections in high-rate wireless systems: impact and digital compensation*. Springer Science & Business Media, 2008.

[9] J. H. Kotecha and A. M. Sayeed, “Transmit signal design for optimal estimation of correlated mimo channels,” *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 546–557, 2004.

[10] D. Neumann, T. Wiese, and W. Utschick, "Learning the mmse channel estimator," *IEEE Transactions on Signal Processing*, vol. 66, no. 11, pp. 2905–2917, 2018.

[11] E. Björnson, J. Hoydis, L. Sanguinetti, *et al.*, "Massive mimo networks: Spectral, energy, and hardware efficiency," *Foundations and Trends® in Signal Processing*, vol. 11, no. 3-4, pp. 154–655, 2017.

[12] L. Sanguinetti, E. Björnson, and J. Hoydis, "Toward massive mimo 2.0: Understanding spatial correlation, interference suppression, and pilot contamination," *IEEE Transactions on Communications*, vol. 68, no. 1, pp. 232–257, 2020.

[13] V. Croisfelt Rodrigues, J. C. Marinello, and T. Abrão, "Exponential spatial correlation with large-scale fading variations in massive mimo channel estimation," *Transactions on Emerging Telecommunications Technologies*, vol. 30, no. 5, p. e3563, 2019.

[14] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *Journal of multivariate analysis*, vol. 88, no. 2, pp. 365–411, 2004.

[15] X. Mestre, "Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates," *IEEE Transactions on Information Theory*, vol. 54, no. 11, pp. 5113–5129, 2008.

[16] N. Shariati, E. Björnson, M. Bengtsson, and M. Debbah, "Low-complexity polynomial channel estimation in large-scale mimo with arbitrary statistics," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 815–830, 2014.

[17] E. Björnson, L. Sanguinetti, and M. Debbah, "Massive mimo with imperfect channel covariance information," in *2016 50th Asilomar Conference on Signals, Systems and Computers*, pp. 974–978, IEEE, 2016.

[18] X. Wei, W. Peng, D. Chen, R. Schober, and T. Jiang, "Uplink channel estimation in massive mimo systems using factor analysis," *IEEE Communications Letters*, vol. 22, no. 8, pp. 1620–1623, 2018.

[19] X. Zhang, H. Zeng, B. Ji, and G. Zhang, "Low-complexity implicit detection for massive mimo using neumann series," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 8, pp. 9044–9049, 2022.

[20] U. Gustavsson, C. Sanchéz-Perez, T. Eriksson, F. Athley, G. Durisi, P. Landin, K. Hausmair, C. Fager, and L. Svensson, "On the impact of hardware impairments on massive mimo," in *2014 IEEE Globecom Workshops (GC Wkshps)*, pp. 294–300, IEEE, 2014.

[21] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive mimo systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Transactions on information theory*, vol. 60, no. 11, pp. 7112–7139, 2014.

[22] S. Jacobsson, G. Durisi, M. Coldrey, U. Gustavsson, and C. Studer, "Throughput analysis of massive mimo uplink with low-resolution adcs," *IEEE Transactions on Wireless Communications*, vol. 16, no. 6, pp. 4038–4051, 2017.

[23] C. Mollén, U. Gustavsson, T. Eriksson, and E. G. Larsson, "Impact of spatial filtering on distortion from low-noise amplifiers in massive mimo base stations," *IEEE Transactions on Communications*, vol. 66, no. 12, pp. 6050–6067, 2018.

[24] Q. Zhang, T. Q. Quek, and S. Jin, "Scaling analysis for massive mimo systems with hardware impairments in rician fading," *IEEE Transactions on Wireless Communications*, vol. 17, no. 7, pp. 4536–4549, 2018.

[25] F. Athley, G. Durisi, and U. Gustavsson, "Analysis of massive mimo with hardware impairments and different channel models," in *2015 9th European Conference on Antennas and Propagation (EuCAP)*, pp. 1–5, IEEE, 2015.

[26] A. Papazafeiropoulos, B. Clerckx, and T. Ratnarajah, "Rate-splitting to mitigate residual transceiver hardware impairments in massive mimo systems," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 9, pp. 8196–8211, 2017.

[27] R. Raich and G. T. Zhou, “On the modeling of memory nonlinear effects of power amplifiers for communication applications,” in *Proceedings of 2002 IEEE 10th Digital Signal Processing Workshop, 2002 and the 2nd Signal Processing Education Workshop.*, pp. 7–10, IEEE, 2002.

[28] Y. Zou, O. Raeesi, L. Antilla, A. Hakkarainen, J. Vieira, F. Tufvesson, Q. Cui, and M. Valkama, “Impact of power amplifier nonlinearities in multi-user massive mimo downlink,” in *2015 IEEE Globecom Workshops (GC Wkshps)*, pp. 1–7, IEEE, 2015.

[29] S. Jacobsson, U. Gustavsson, G. Durisi, and C. Studer, “Massive mu-mimo-ofdm uplink with hardware impairments: Modeling and analysis,” in *2018 52nd Asilomar Conference on Signals, Systems, and Computers*, pp. 1829–1835, IEEE, 2018.

[30] D. Rönnow and P. Händel, “Nonlinear distortion noise and linear attenuation in mimo systems—theory and application to multiband transmitters,” *IEEE Transactions on Signal Processing*, vol. 67, no. 20, pp. 5203–5212, 2019.

[31] S. R. Aghdam, S. Jacobsson, and T. Eriksson, “Distortion-aware linear precoding for millimeter-wave multiuser miso downlink,” in *2019 IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 1–6, IEEE, 2019.

[32] M. Cherif and R. Bouallegue, “The effect of high power amplifier nonlinearity on mu-massive mimo system performance over rayleigh fading channel,” in *2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC)*, pp. 1426–1429, IEEE, 2019.

[33] Ö. Tugfe Demir and E. Björnson, “Channel estimation under hardware impairments: Bayesian methods versus deep learning,” *arXiv e-prints*, pp. arXiv–2208, 2022.

[34] T. L. Marzetta and A. Ashikhmin, “Mimo system having a plurality of service antennas for data transmission thereof,” Nov. 26 2013. US Patent 8,594,215.

[35] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE transactions on wireless communications*, vol. 9, no. 11, pp. 3590–3600, 2010.

[36] A. F. Molisch, *Wireless communications*, vol. 34. John Wiley & Sons, 2012.

[37] H. Yang and T. L. Marzetta, "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 172–179, 2013.

[38] A. Farhang, N. Marchetti, and B. Farhang-Boroujeny, "Filter bank multicarrier for massive mimo," *Signal Processing for 5G: Algorithms and Implementations*, pp. 67–89, 2016.

[39] O. Simeone, N. Levy, A. Sanderovich, O. Somekh, B. M. Zaidel, H. V. Poor, S. Shamai, *et al.*, "Cooperative wireless cellular systems: An information-theoretic view," *Foundations and Trends® in Communications and Information Theory*, vol. 8, no. 1-2, pp. 1–177, 2012.

[40] H. Wang, P. Wang, L. Ping, and X. Lin, "On the impact of antenna correlation in multi-user mimo systems with rate constraints," *IEEE communications letters*, vol. 13, no. 12, pp. 935–937, 2009.

[41] E. Björnson, M. Kountouris, M. Bengtsson, and B. Ottersten, "Receive combining vs. multi-stream multiplexing in downlink systems with multi-antenna users," *IEEE Transactions on Signal Processing*, vol. 61, no. 13, pp. 3431–3446, 2013.

[42] D. Gesbert, L. Pittman, and M. Kountouris, "Transmit correlation-aided scheduling in multiuser mimo networks," in *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, vol. 4, pp. IV–IV, IEEE, 2006.

[43] B. Clerckx, G. Kim, and S. Kim, "Correlated fading in broadcast mimo channels: Curse or blessing?," in *IEEE GLOBECOM 2008-2008 IEEE Global Telecommunications Conference*, pp. 1–5, IEEE, 2008.

[44] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE Journal on selected areas in communications*, vol. 31, no. 2, pp. 264–273, 2013.

[45] Z. Jiang, A. F. Molisch, G. Caire, and Z. Niu, "Achievable rates of fdd massive mimo systems with spatial channel correlation," *IEEE Transactions on Wireless Communications*, vol. 14, no. 5, pp. 2868–2882, 2015.

[46] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing—the large-scale array regime," *IEEE transactions on information theory*, vol. 59, no. 10, pp. 6441–6463, 2013.

[47] K. I. Pedersen, P. E. Mogensen, and B. H. Fleury, "Power azimuth spectrum in outdoor environments," *Electronics Letters*, vol. 33, no. 18, pp. 1583–1584, 1997.

[48] E. G. Larsson and E. Björnson, "Technique for assigning pilot signals to user equipments," Dec. 21 2021. US Patent 11,206,116.

[49] E. Björnson, E. G. Larsson, and M. Debbah, "Massive mimo for maximal spectral efficiency: How many users and pilots should be allocated?," *IEEE Transactions on Wireless Communications*, vol. 15, no. 2, pp. 1293–1308, 2015.

[50] K. J. Horadam, *Hadamard matrices and their applications*. Princeton university press, 2012.

[51] B. Gopalakrishnan and N. Jindal, "An analysis of pilot contamination on multi-user mimo cellular systems with many antennas," in *2011 IEEE 12th international workshop on signal processing advances in wireless communications*, pp. 381–385, IEEE, 2011.

[52] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell tdd systems," *IEEE Transactions on Wireless Communications*, vol. 10, no. 8, pp. 2640–2651, 2011.

[53] E. Bjornson and B. Ottersten, "A framework for training-based estimation in arbitrarily correlated rician mimo channels with rician disturbance,"

IEEE Transactions on Signal Processing, vol. 58, no. 3, pp. 1807–1820, 2009.

[54] E. Björnson, L. Sanguinetti, and J. Hoydis, "Hardware distortion correlation has negligible impact on ul massive mimo spectral efficiency," *IEEE Transactions on Communications*, vol. 67, no. 2, pp. 1085–1098, 2018.

[55] F. Horlin and A. Bourdoux, *Digital compensation for analog front-ends: a new approach to wireless transceiver design*. John Wiley & Sons, 2008.

[56] G. Hueber and R. B. Staszewski, *Multi-mode/Multi-band RF transceivers for wireless communications: Advanced techniques, Architectures, and Trends*. John Wiley & Sons, 2011.

[57] J. J. Bussgang, "Crosscorrelation functions of amplitude-distorted gaussian signals," 1952.

[58] P. Händel and D. Rönnow, "Dirty mimo transmitters: Does it matter?," *IEEE Transactions on Wireless Communications*, vol. 17, no. 8, pp. 5425–5436, 2018.

[59] E. U. T. R. Access, "Further advancements for e-utra physical layer aspects (release 9)," *European Telecommunications Standards Institute*, 2010.

[60] A. K. Kocharlakota, K. Upadhya, and S. A. Vorobyov, "On achievable rates for massive mimo system with imperfect channel covariance information," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4504–4508, IEEE, 2019.

[61] D. L. Donoho and M. Gavish, "The optimal hard threshold for singular values is $4/\sqrt{3}$," 2013.

[62] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[63] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[64] T. L. Marzetta and H. Yang, *Fundamentals of massive MIMO*. Cambridge University Press, 2016.

[65] C. V. N. Index, "Global mobile data traffic forecast update, 2015–2020," *White Paper, February*, vol. 1, 2016.

[66] C. E. Shannon, "A mathematical theory of communication," *The Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.

[67] T. M. Cover and J. A. Thomas, "Elements of information theory, 1991 john wiley & sons," *Inc. Print ISBN 0-471-06259-6 Online ISBN 0-471-20061-1*, 1991.

[68] E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive mimo: Ten myths and one critical question," *IEEE Communications Magazine*, vol. 54, no. 2, pp. 114–123, 2016.

[69] M. Medard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Transactions on Information theory*, vol. 46, no. 3, pp. 933–946, 2000.

[70] X. Shang and H. V. Poor, "Noisy-interference sum-rate capacity for vector gaussian interference channels," *IEEE transactions on information theory*, vol. 59, no. 1, pp. 132–153, 2012.

[71] A. D. Wyner, "Shannon-theoretic approach to a gaussian cellular multiple-access channel," *IEEE Transactions on Information Theory*, vol. 40, no. 6, pp. 1713–1727, 1994.

[72] J. Mietzner, R. Schober, L. Lampe, W. H. Gerstacker, and P. A. Hoeher, "Multiple-antenna techniques for wireless communications-a comprehensive literature survey," *IEEE communications surveys & tutorials*, vol. 11, no. 2, pp. 87–105, 2009.

[73] Y. S. Poberezhskiy, "Diversity schemes and coherent combining in digital receivers," in *2022 IEEE Aerospace Conference (AERO)*, pp. 1–20, IEEE, 2022.

[74] X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, "Linear pre-coding performance in measured very-large mimo channels," in *2011 IEEE vehicular technology conference (VTC Fall)*, pp. 1–5, IEEE, 2011.

[75] J. Hoydis, C. Hoek, T. Wild, and S. ten Brink, "Channel measurements for large antenna arrays," in *2012 International Symposium on Wireless Communication Systems (ISWCS)*, pp. 811–815, IEEE, 2012.

[76] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Aspects of favorable propagation in massive mimo," in *2014 22nd European Signal Processing Conference (EUSIPCO)*, pp. 76–80, IEEE, 2014.

[77] B. M. Hochwald, T. L. Marzetta, and V. Tarokh, "Multiple-antenna channel hardening and its implications for rate feedback and scheduling," *IEEE transactions on Information Theory*, vol. 50, no. 9, pp. 1893–1909, 2004.

[78] S. Venkatesan, A. Lozano, and R. Valenzuela, "Network mimo: Overcoming intercell interference in indoor wireless systems," in *2007 Conference Record of the Forty-First Asilomar Conference on Signals, Systems and Computers*, pp. 83–87, IEEE, 2007.

[79] W. Yu, "Uplink-downlink duality via minimax duality," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 361–374, 2006.

[80] M. Bengtsson and B. Ottersten, "Optimum and suboptimum transmit beamforming," in *Handbook of antennas in wireless communications*, pp. 18–1, CRC press, 2018.

[81] E. Björnson, E. Jorswieck, *et al.*, "Optimal resource allocation in coordinated multi-cell systems," *Foundations and Trends® in Communications and Information Theory*, vol. 9, no. 2–3, pp. 113–381, 2013.

[82] D. Gesbert, S. Hanly, H. Huang, S. S. Shitz, O. Simeone, and W. Yu, "Multi-cell mimo cooperative networks: A new look at interference," *IEEE journal on selected areas in communications*, vol. 28, no. 9, pp. 1380–1408, 2010.

[83] Y. Liu, T. F. Wong, and W. W. Hager, "Training signal design for estimation of correlated mimo channels with colored interference," *IEEE Transactions on Signal Processing*, vol. 55, no. 4, pp. 1486–1497, 2007.

[84] U. Spagnolini, *Statistical Signal Processing in Engineering*. John Wiley & Sons, 2018.

Appendix A

Basic Fundamentals and Definitions

A.1 Signal-to-Noise Ratio

Let's consider a single-cell single-user scenario of the cellular network with a single isotropic antenna at the Base station (BS) and at the user Equipment (UE). The transmitted power from such BS antenna will spread equally on the cell's surface area with spherical pattern radiation. The UE that resides in the BS zone will receive a fraction of the transmit power from the BS. Hence, the Signal-to-Noise Ratio (SNR) will be defined as follows:

$$\text{SNR} = \frac{\text{Received signal power}}{\text{Noise power}} = \frac{\text{Transmit power} \times \text{Channel gain}}{\text{Noise power}} \quad \text{(A.1)}$$

A.1.1 Transmit power

Generally, the transmitted power from the UE to the BS in the uplink direction is about 100mW. In contrast, the BS transmits about 40W in the downlink direction[1]. Since there is a huge difference in the transmit and receive power in Watts and milliWatts, it is usually using the dBm scale to measure the power as follows:

$$\text{Signal power in dBm} = 10\log_{10}\left(\frac{\text{Signal power}}{1\text{mW}}\right) \quad \text{(A.2)}$$

A.1.2 Channel gain

As mentioned in A.1, the transmitted power from the single isotropic antenna at the BS will take a spherical pattern, and the UE will receive only a small

[1] See the link: `https://github.com/emilbjornson/tsks14_multiple_antenna_communications`

fraction of the transmitted power. Hence, the channel gain at the UE located inside the BS zone pattern can be defined as follows:

$$\text{Channel gain} = \frac{\text{Surface area of UE's antenna}}{\text{Surface area of BS's spherical pattern}} \tag{A.3}$$

The UE's antenna surface area is determined by $\lambda^2/4\pi$, where λ refers to the wavelength of the carrier frequency. The BS's spherical pattern surface area is calculated using the area of sphere $4\pi d^2$, where d refers to the distance between the BS and the UE device. However, the channel gain in dB can be given by:

$$\text{channel gain in dB} = 20\log_{10}\left(\frac{\lambda}{4\pi}\right) - 20\log_{10}(d) \tag{A.4}$$

That is, if d=10m (one decade of distance), then the channel gain will be rewritten as:

$$\text{channel gain in dB} = 20\log_{10}\left(\frac{\lambda}{4\pi}\right) - 20, \tag{A.5}$$

which means it will be losing 20dB from the transmit power for every decade away from the BS. However, at high frequencies, the antenna size will be smaller, leading to lower channel gain at the UE device, i.e., only a tiny fraction of the transmitted power will be received. But, the communication system is still working because what matter is how much received power is compared to the noise power.

A.1.3 Noise power

The noise power or the variance of the noise is measured in Watts and given by:

$$\text{Noise power} = \mathrm{N_o B}, \tag{A.6}$$

where $\mathrm{N_o} = 10^{-14.4}$ Watt/Hz [2] refers to the noise power spectral density, and B, is the bandwidth in Hz. The noise power in dB can be given by:

$$\text{Noise power in dB} = -174 + 10\log_{10}\,(\mathrm{B}) \tag{A.7}$$

In summary, if p refers to the transmit power and β is the average channel

[2] See the link: `https://github.com/emilbjornson/tsks14_multiple_antenna_communications`

gain, the SNR will take the formula:

$$\mathrm{SNR} = \frac{p \times \beta}{\mathrm{N_o} * \mathrm{B}} = \frac{q \times \beta}{\mathrm{N_o}} \tag{A.8}$$

where q refers to the power spectral density of the signal and represents the signal's energy in Watt/Hz. In practice, the SNR ranges from -10dB to 40dB in the downlink direction.

A.2 Complex Random Vector

A.2.1 Circular-symmetry Gaussian complex vector

Since the baseband signals are inherently complex-valued random quantities, their representation will be characterized by the joint probability density function of their real and imaginary parts. That is, if $\mathbf{s}^{'}$ and $\mathbf{s}^{"}$ are $M \times 1$ real-valued random vectors with Gaussian distributions, then their joint probability will characterize the complex random vector [64]

$$\mathbf{s} = \mathbf{s}^{'} + j\mathbf{s}^{"}, \tag{A.9}$$

and the probability density function (PDF) of the complex vector $\mathbf{s}$ is a complex Gaussian distribution with a circular-symmetry property. The last property can be defined as follows:

The $\mathbf{s}$ is a circular-symmetry Gaussian complex vector if $e^{j\varphi}\mathbf{s}$ has identical Probability Density Function (PDF) as $\mathbf{s}$ for all real value of φ [64], i.e.,

$$\mathbb{E}\{e^{j\varphi}\mathbf{s}\} = \mathbb{E}\{\mathbf{s}\} \tag{A.10}$$

where the mean of $\mathbf{s}$ must be equal to zero, i.e., $\mathbb{E}\{\mathbf{s}\} = \mu_{\mathbf{s}} = 0$, this satisfies that the circular-symmetry condition will be strictly around the zero-mean. Hence, the circular-symmetry complex Gaussian vector [64] can be modelled as

$$\mathbf{s} \sim \mathbb{C}_{\mathcal{N}}(0, \mathbf{R}) \tag{A.11}$$

where $\mathbf{R} = \mathbb{E}\{\mathbf{s}\,\mathbf{s}^{\mathrm{H}}\}$ is the correlation matrix of $\mathbf{s}$, some times it is called by the covariance matrix.

A.2.2 PDF of circular-symmetry Gaussian complex vector

As an extension to the above case, let the circular-symmetry Gaussian complex vector with N-length has a non-zero mean and is modelled as:

$$\mathbf{s} \sim \mathbb{C}_{\mathscr{N}}(\mu_{\mathbf{s}}, \mathbf{R}) \tag{A.12}$$

where $\mu_{\mathbf{s}} \in \mathbb{C}^N$ is the mean vector, and the covariance matrix $\mathbf{R} \in \mathbb{C}^{N \times N}$, then the circularly symmetric PDF will be given by [11]

$$f(\mathrm{s}) = \frac{e^{-(\mathbf{s}-\mu_{\mathbf{s}})^{\mathrm{H}}\mathbf{R}^{-1}(\mathbf{s}-\mu_{\mathbf{s}})}}{\pi^N \det(\mathbf{R})} \tag{A.13}$$

In this case, the circular-symmetry property will be assumed to be satisfied around the mean $\mu_{\mathbf{s}}$ as follows: $\mathbf{s} - \mu_{\mathbf{s}} = e^{j\varphi}(\mathbf{s} - \mu_{\mathbf{s}})$ for any given φ, and this implies that both distributions are identical.

Now, if the rank of the covariance matrix $\mathbf{R}$ is less than N, then the complex random vector $\mathbf{s}$ with $\mu_{\mathbf{s}}$ mean can be written in terms of its mean and the eigenstructure of its covariance matrix as follows [11]:

$$\mathbf{s} = \mu_{\mathbf{s}} + \mathbf{U}\mathbf{D}^{\frac{1}{2}} \begin{bmatrix} \mathbf{g} \\ \mathbf{0}_{N-r} \end{bmatrix} \tag{A.14}$$

where $r = rank(\mathbf{R})$. The r-dimensional vector $\mathbf{g} \sim \mathbb{C}_{\mathscr{N}}(0_r, \mathbf{I_r})$ is a random vector with PDF given by $\frac{e-||\mathbf{g}||^2}{\pi^r}$. The decomposition of $\mathbf{R} = \mathbf{U}\mathbf{D}\mathbf{U}^{\mathrm{H}}$ is the eigenstructure of the covariance matrix with $\mathbf{U}$, $\mathbf{U}^{\mathrm{H}}$ are the left and right matrices, respectively, they contain orthogonal eigenvectors, $\mathbf{D}$ is the diagonal matrix with decreasing order of eigenvalues.

Appendix B

Principles of Cellular Network

B.1 Cellular Network Definition

The cellular network is defined as a group of adjacent cells, and each cell has a distinct geographical area with a fixed-location BS that serves some of the UEs inside its coverage area. Also, the BS may serve other users from other cells, as shown in Fig. B.1 [11]. Thus, each UE can communicate with more than one BS in the network, the own cell BS and the BSs from adjacent cells. The transmission direction from the UE to the BS is referred to as the Uplink (UL) transmission, while the direction from the BS to the UE is called the Downlink transmission (DL).

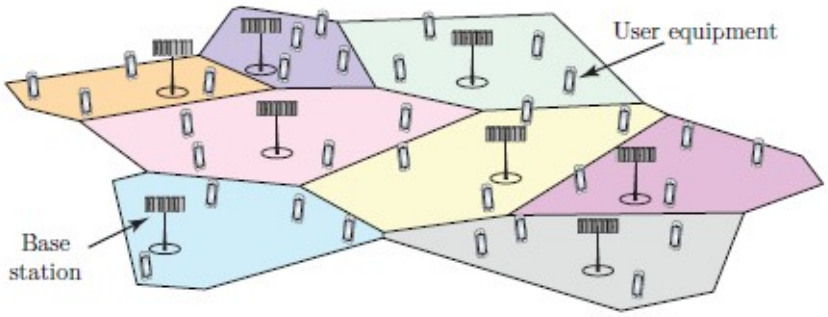

Figure B.1: Cellular network topology

The coverage of the network can be subdivided into two coverage areas, which are the coverage-tier area and the hotspot-tier area [11]. The former consists of outdoor environment BSs that serve a wide coverage area, while the latter consists of the indoor environment in which the BSs can serve a high-quality throughput within a small area.

Although the cellular network is mainly designed to provide voice communication [2], nowadays, it is used to serve large data transmission like the on-demand video transmission, which increases the traffic data in the network [65]. Thus, the throughput of the area in bit/sec/km^2 [11] is the high-performance metric of the cellular network that is needed and given as follows:

$$\text{Area throughput} = \mathrm{B} \cdot \mathrm{D} \cdot \mathrm{SE} \tag{B.1}$$

such that B in (Hz) is the bandwidth and D in (cells /km^2) denotes the average density of the cellular network. The SE in (bit /sec/Hz/ cell) stands for the overall spectral efficiency per cell, and it refers to the amount of user information that may be conveyed per second over a frequency range of one hertz from the available bandwidth. However, one can treat the three parameters of the throughput equation as independent components. Consequently, improving the total area throughput is done by either allocating more bandwidth, densifying the network, or improving the spectral efficiency of the cell. But each has its own potential on the global cellular networks of the sub-6GHz bandwidth standard. For example, if one wants to increase the throughput by a factor ($\times 1000$) by increasing the bandwidth and maintaining other factors as constant, more than 1THz of the bandwidth will be needed in the future. This is impractical from a physical point of view since the spectrum frequency is a global resource shared and organized between different applications.

Another procedure for increasing the throughput of the area is to densify the cellular network (add more cells to the network), which increases the number of BSs in the network by a factor of ($\times 1000$). However, the current deployment of the BSs in urban areas is established with a few hundred inter-distances and at elevated locations to prevent the shadowing problem. Hence, densifying the network causes the BSs to be moved closer to the UEs, and deepening the shadowing problem.

Moreover, densifying the network and allocating more bandwidth have already dominated the coverage of wireless services; they have reached saturation points, and further enhancements will be very complicated and expensive. Hence, more possible and practical improvements will be made by increasing the SE of the current cellular networks.

B.2 Spectral Efficiency (SE) Definition

The spectral efficiency (SE) is defined as the maximum amount of information (bits) per each baseband symbol that can be reliably transferred without error throughout the transmission channel. It is a deterministic value measured by bit/sec/Hz. However, the total uplink spectral efficiency of the channels of each user's equipment to the BS will be considered here. Hence, the term sum SE will be used to describe the total spectral efficiency from all UEs in the cell at the BS and is given by bit/sec/Hz/cell. The channel capacity by Shannon [66] is the main underline concept in determining the SE and can take different forms depending on what type of communication channel it is. The achievable SE can be determined based on the following theorem.

Theorem B.2.1 (Channel Capacity:) Consider Fig.B.2, for any input random variable x and any output random variable y, the channel capacity of the discrete memoryless channel [67] can be given by:

$$C = \sup_{g(x)} (\mathscr{H}(y) - \mathscr{H}(y \mid x)) \tag{B.2}$$

where (sup) denotes the supremum function taken for all possible distributions of the input $g(x)$, the $\mathscr{H}(y)$ refers to the output uncertainty (entropy), and $\mathscr{H}(y \mid x)$ stands for the conditional uncertainty of y given x. Hence, the SE is said to be achievable with low arbitrarily bit error probability if it is equal to or smaller than the channel capacity in Eq.(B.2)

Figure B.2: A general memoryless discrete channel

In wireless communication, the Additive White Gaussian Noise (AWGN) channel is the common channel of interest to many researchers because it is convenient and tractable. Thus, using the AWGN channel, it can calculate the channel capacity given in theorem B.2.1 in a closed form for the case of the single-input single-output (SISO) system in the following corollaries.

Corollary B.2.1 (Channel Capacity of SISO system) Consider the SISO system in Fig.B.3, the channel with memoryless and discrete characteristics can be given by

$$y = hx + n \tag{B.3}$$

where $y \in \mathbb{C}$ is the output, $x \in \mathbb{C}$ is the input, and $n \sim \mathbb{C}_{\mathcal{N}}\left(0, \sigma^2\right)$ is the additive white noise at the receiver and is independent of the channel. The parameter $h \in \mathbb{C}$ denotes the discrete channel, which is assumed to be known at the receiver. However, the channel capacity is assumed to be achievable under the condition that the input is power-limited as $\mathbb{E}\left\{|x|^2\right\} \leq p$.

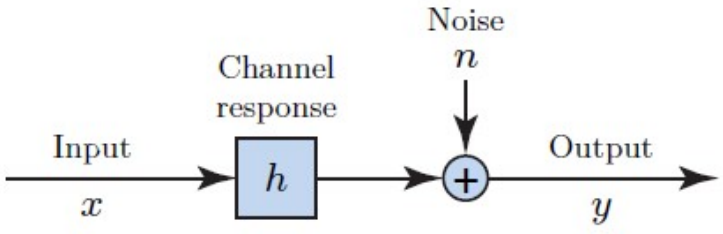

Figure B.3: Single-Input Single-Output (SISO) system

As mentioned earlier, the channel capacity depends on the type of channel considered. Hence, if the channel h in Eq.(B.3) is given as a *deterministic* channel, then the achievable channel capacity will be given by

$$C = \log_2\left(1 + \frac{p|h|^2}{\sigma^2}\right), \tag{B.4}$$

while for a *random* channel, the capacity will be called *ergodic capacity* and is given by

$$C = \mathbb{E}\left\{\log_2\left(1 + \frac{p|h|^2}{\sigma^2}\right)\right\} \tag{B.5}$$

where h, in this case, represents one realization from the channel's random variable $\mathbb{H}$. The expectation in front of the capacity refers to the fast-fading channel; that is, there is one channel realization for each new transmit symbol; hence, the averaged capacity overall channel realization will be denoted by the *ergodic capacity*.

The signal-to-noise ratio (SNR) in Eq(B.4) represents the *actual* SNR measured for the deterministic channel h, while the SNR in Eq.(B.5) denotes the *instantaneous* SNR measured for one arbitrary random channel realization h.

Hence, for many random realizations from the channel h, it is more convenient to consider the average of the SNRs values of the total channel realizations as follows: $p\mathbb{E}\left\{|h|^2\right\}/\sigma^2$ where $\mathbb{E}$ denotes the expectation that is carried out to evaluate the averaged value of the SNR over all channel realizations. However, the value $\mathbb{E}\left\{|h|^2\right\}$ denotes the average channel gain determined for all channel realizations.

The performance of the SISO system in Fig. B.3 can be measured using the SE as the good performance metric of the system [68] as follows: if the spectral efficiency of each complex-valued symbol is equal to or smaller than the channel capacity, then the symbol will be decodable at the receiver. A similar procedure can be followed for transmitting long information of block code with N baseband scalar symbols, which may be decodable at the receiver with an arbitrarily low bit error rate.

Since the communication system has inevitable interference problems behind the simultaneous transmissions, the following corollary will discuss the channel capacity of the interfered SISO system. It is a lower bound capacity since the exact capacity of such a system is generally unknown [69].

Corollary B.2.2 (Channel Capacity with interference) Consider the SISO system in Fig.B.4: the input signal $x \in \mathbb{C}$ is assumed to be power limited by p, the output signal $y \in \mathbb{C}$ is further assumed to be interrupted by an interference $v \in \mathbb{C}$ in addition to the noise $n \sim \mathbb{C}_{\mathcal{N}}\left(0, \sigma^2\right)$. Moreover, the channel $h \in \mathbb{C}$ is assumed to be known to the receiver with independent noise n.

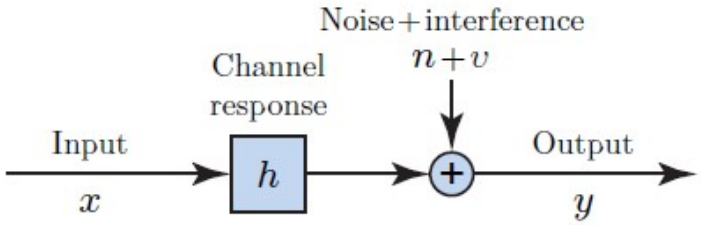

Figure B.4: Single-Input Single-Output (SISO) system with interference

Again, the channel capacity depends on what type of channel is going to be, as follows: If the channel h is given as a *deterministic* channel, the achievable

capacity of the channel will be lower bounded by

$$C \geq \log_2\left(1+\frac{p|h|^2}{p_v+\sigma^2}\right), \tag{B.6}$$

where $p_v \in \mathbb{R}$ stands for the variance of the interference, which is assumed to be uncorrelated with x, that is $(\mathbb{E}\{x*v\}=0)$ and is assumed known at the receiver.

Now, if the parameter h is given as one realization of the *random* variable $\mathbb{H}$, then the ergodic achievable capacity will be lower bounded [70] as follows:

$$C \geq \mathbb{E}\left\{\log_2\left(1+\frac{p|h|^2}{p_v(h,u)+\sigma^2}\right)\right\}, \tag{B.7}$$

where u is one realization from the interference random variable $\mathbb{U}$ and is associated with each realization h from $\mathbb{H}$. The interference v can be modelled by a conditional zero mean $\mathbb{E}\{v|h,u\}=0$ and a conditional variance $p_v(h,u)=\mathbb{E}\{|v|^2|h,u\}$. Moreover, v is assumed to be uncorrelated with the input signal x, which can be expressed as $\mathbb{E}\{x*v|h,u\}=0$. At the receiver, it is assumed that the values of h and u are known such that the additive noise n is conditionally independent of v given h and u. The expectation is taken to express the many realizations of the channel random variable, and the lower bound capacity is approximated with the complex Gaussian distribution of the input signal x.

Note that the ratio $\{p|h|^2\}/\{p_v(h,u)+\sigma^2\}$ inside the logarithmic expression in Eq.(B.7) is referred to as the signal-to-noise plus interference ratio (SNIR). It is deterministic when the channel h and interference v are deterministic values, while it is called instantaneous SNIR when the channel and the interference are random.

B.3 Improving Methods of Uplink SE

There are various methods for increasing the spectral efficiency in cellular networks, either by increasing the transmit power or deploying more antennas at the BS to serve more users in the cell. The Wyner model [71] shown in Fig.B.5 will be used here since it is the more tractable model used for analysing the characteristics of the cellular communication system. It contains only two cells, and the average channel gains between each user equipment and its served

BS are assumed identical. That is, β_0^0 refers to the average gain of the channel between the user in cell 0 to the BS in cell 0, and it represents the desired signal at the BS in cell 0, while β_1^0 refers to the average gain of the channel between the user in cell 1 to the BS in cell 0 and it represents the interference signal at the BS in cell 0. Note that the notations of the superscript denote the corresponding BS location, and the notations of the subscript denote the corresponding UE location in the cell.

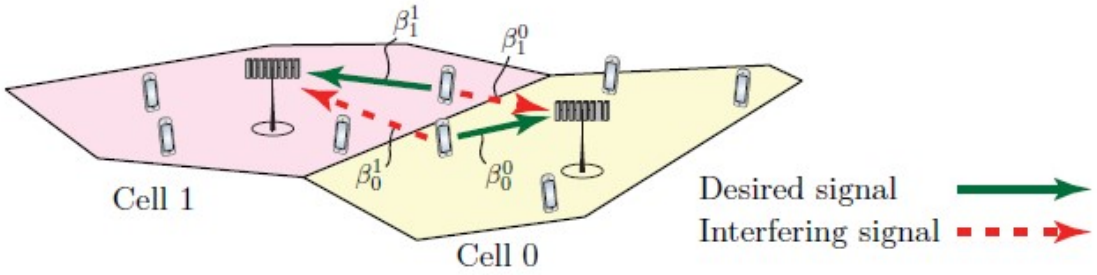

Figure B.5: Wyner model

The average channel gain is a very small positive quantity; this is because the power of the propagation signal decays faster as large as the propagation distance. It ranges from -70dB to -120dB inside the serving cell, while will be even smaller values for the interfering signals from other cells. More details about the channel gain are in Appendix A.1.2. In cellular communication, when evaluating the SE, what matter is the relative strength $\bar{\beta}$ [11] of the interference signal to the desired signal and not the desired channel gain only. It is expressed by the ratio of the inter-cell interference to the desired intra-cell gains of the channels, as follows:

$$\bar{\beta} = \frac{\beta_1^0}{\beta_0^0} = \frac{\beta_0^1}{\beta_0^0} = \frac{\beta_1^0}{\beta_1^1} = \frac{\beta_0^1}{\beta_1^1}. \tag{B.8}$$

In general, $0 \leq \bar{\beta} \leq 1$, where the two extremes $\bar{\beta} \approx 0$ denotes the weak inter-cell interference case, and $\bar{\beta} \approx 1$ refers to the case when the interference signal is strong as the desired signal, which may occur at the edge of the neighbouring cells. In the following subsections, we will discuss the different ways of improving the SE in the serving cell.

B.3.1 Increasing UEs' signal power

Before delving into what happens to the SE when increasing the power of the transmitted signal, the following analysis is very important to be considered first.

Let us consider the Wyner cellular model with a single active UE in each cell. Let

$$\mathrm{SNR}_0 = \frac{p\beta_0^0}{\sigma^2} \tag{B.9}$$

be the signal-to-noise ratio of the UE at the BS in cell 0, where p is the transmitted power of UE, and σ^2 denotes the noise power. Hence, the complex-valued received signal at the BS in cell 0 will be given by:

$$y_0 = \underbrace{h_0^0 s_0}_{\text{“Desired signal”}} + \underbrace{h_1^0 s_1}_{\text{“Interfering signal”}} + \underbrace{n_0}_{\text{“Noise”}} \tag{B.10}$$

such that $s_0, s_1 \sim \mathbb{C}_{\mathcal{N}}(0, p)$ are the transmitted information signals from the desired and interfering UEs, respectively. The scalars $h_0^0, h_1^0 \in \mathbb{C}$ are the channel responses of the desired UE_0 and interfering UE_1 respectively, and are assumed flat-fading. The noise $n \sim \mathbb{C}_{\mathcal{N}}(0, \sigma^2)$ is the additive noise at the BS antenna. To show the effect of these channel responses on the SE, two propagation channel cases, the LoS and NLoS, will be considered here.

B.3.1.1 LoS propagation channel model

In LoS propagation with a single antenna, the channel responses in Eq.(B.10) will be deterministic scalars as follows:

$$h_i^0 = \sqrt{\beta_i^0} \quad \text{for } i = 0, 1, \tag{B.11}$$

where β_i^0 is the i^{th} channel gain at the BS and also denotes the large-scale or path-loss fading effect. It is distance-dependent and may be changed when the transmitter and/or receiver move.

B.3.1.2 NLoS propagation channel model

In this case, the channel responses will be random variables and vary over

time. However, if there is a sufficient scattering of objects around the user equipment, the channel will be modelled as follows:

$$h_i^0 \sim \mathcal{N}_{\mathbb{C}}\left(0, \beta_i^0\right) \quad \text{for } i = 0, 1. \tag{B.12}$$

This model represents the Rayleigh fading channel model and describes the large-scale and small-scale fading effects [39]. The transmitted signal will reach the BS in multiple paths and different delays, and the composed received signals at the BS will either be a constructive signal (i.e., the coming multiple paths reinforce each other if reaching the BS at the same delay) or a deconstructive signal (i.e., the coming multiple paths cancel each other if reaching the BS at different delays). The average channel gain in both propagation models (LoS and NLoS) is given by

$$\beta_i^0 = \mathbb{E}\left\{\left|h_i^0\right|^2\right\}, \text{ for } i = 0, 1 \tag{B.13}$$

B.3.1.3 Uplink SE analysis

However, for known channel responses at the BS. The achievable SE in the uplink transmission of the desired UE [11] can be given as follows:

- In LoS case:

$$\mathrm{SE}_0^{\mathrm{LoS}} = \log_2\left(1 + \frac{p\left|h_0^0\right|^2}{p\left|h_1^0\right|^2 + \sigma^2}\right) = \log_2\left(1 + \frac{1}{\bar{\beta} + \frac{1}{\mathrm{SNR}_0}}\right) \tag{B.14}$$

- In NLoS case:

$$\begin{aligned} \mathrm{SE}_0^{\mathrm{NLoS}} &= \mathbb{E}\left\{\log_2\left(1 + \frac{p\left|h_0^0\right|^2}{p\left|h_1^0\right|^2 + \sigma^2}\right)\right\} \\ &= \frac{e^{\frac{1}{\mathrm{SNR}_0}} E_1\left(\frac{1}{\mathrm{SNR}_0}\right) - e^{\frac{1}{\mathrm{SNR}_0\bar{\beta}}} E_1\left(\frac{1}{\mathrm{SNR}_0\bar{\beta}}\right)}{\log_e(2)(1-\bar{\beta})} \end{aligned} \tag{B.15}$$

where $E_1(z) = \int_1^\infty \frac{e^{-zu}}{u} du$ is the exponential integral and $\log_e(.)$ is the natural logarithm. The closed forms of the SE in Eq.(B.14) and Eq.(B.15) and their proofs are referenced in [11]. It is clearly shown that the SE in the expressions

above is strongly dependent on the SNR_0 of the desired UE and on the relative gains $\bar{\beta}$. Hence, the following limits will illustrate the effect of growing the transmit power (i.e., increasing SNR) on the SE.

- In LoS case:

$$SE_0^{LoS} \rightarrow \log_2\left(1+\frac{1}{\bar{\beta}}\right) \quad \text{as } p \rightarrow \infty \tag{B.16}$$

- In NLoS case:

$$SE_0^{NLoS} \rightarrow \frac{1}{1-\bar{\beta}}\log_2\left(\frac{1}{\bar{\beta}}\right) \text{ as } p \rightarrow \infty \tag{B.17}$$

The behaviour of the SE in Eq.(B.16) and Eq.(B.17) is simulated in Fig.B.6 in [11] under two different relative strengths of $\bar{\beta} = -10dB$ and $\bar{\beta} = -30dB$.

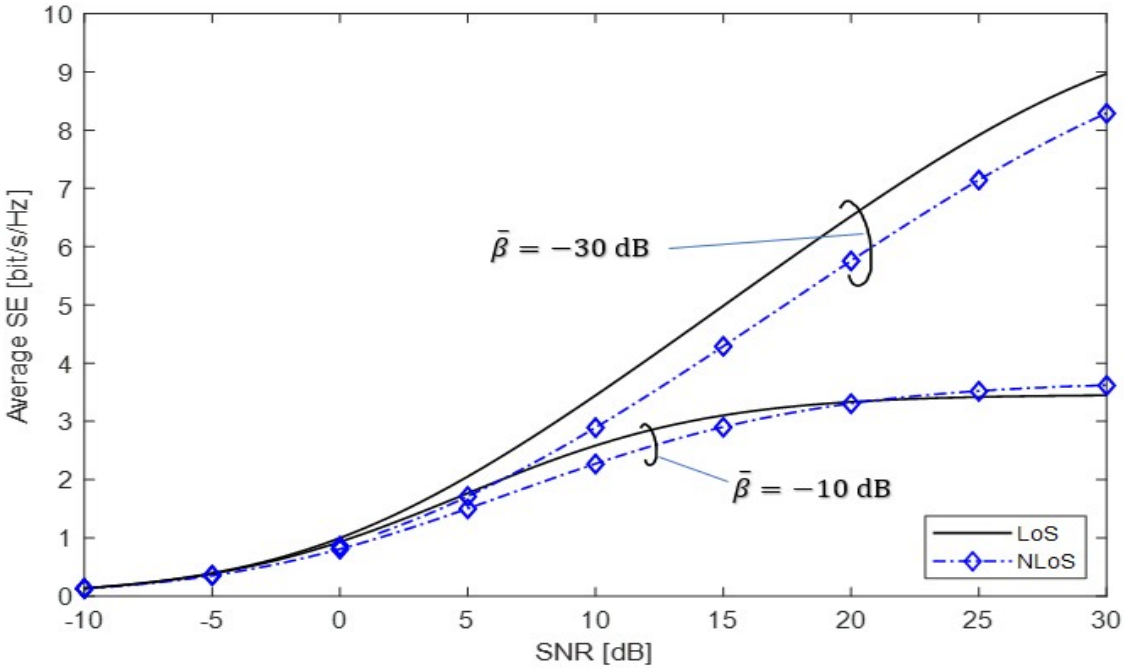

Figure B.6: Uplink SE as a function of SNR

It is noticed that in Fig.B.6 when the SNR goes from 10 dB to 30 dB, the SE only doubles, while 100 times more transmit power is needed. In conclusion, growing the SNR by transmitting more additional power enhances the spectral efficiency but quickly pushes the network into the interference-limited regime. The latter reduces the degree of freedom at the BS (i.e., the BS cannot split between the desired and interference signals) since they transmit the same strong signal. This makes each UE wait its turn since the BS can't listen to both UE

simultaneously. Hence, the simple power-increasing way did not present much to investigating higher SE in the network.

B.3.2 Using array of antennas at the BS

Rather than raising the transmitting uplink power, the BS can be equipped with multiple receive antennas to gather extra power from the reached electromagnetic waves. The use of multiple antennas at the BS has been studied in many seminal works either to increase the receive signal power or to achieve spatial diversity.

Now, if the BS in cell 0 is deployed with M antennas array, the channel responses of the desired and interfering UEs will be expressed by the channel vectors $h_0^0 \in \mathbb{C}^M$ and $h_1^0 \in \mathbb{C}^M$, respectively, where each m^{th} element will represent the observed channel response at m^{th} antenna of the BS. Hence, the vector of the complex received signal will be given by:

$$\mathbf{y}_0 = \underbrace{\mathbf{h}_0^0 s_0}_{\text{“Desired signal”}} + \underbrace{\mathbf{h}_1^0 s_1}_{\text{“Interfering signal”}} + \underbrace{\mathbf{n}_0}_{\text{“Noise”}} \tag{B.18}$$

such that $s_0, s_1 \sim \mathbb{C}_{\mathcal{N}}(0, p)$ are the transmitted information signals from the desired and interfering UEs, respectively, $n \sim \mathbb{C}_{\mathcal{N}}\left(0, \sigma^2 \mathrm{I}_M\right)$ is the additive noise on each antenna elements at the BS. Since there are M receive antennas at the BS and a single transmit antenna per UE, the propagation channels will be SIMO channels; and will be considered in two propagation cases as follows:

B.3.2.1 LoS propagation channel model

Again, the Wyner model is considered here assuming a horizontal Uniform Linear Array (ULA) at the BS with $d_{\mathrm{H}} \in [0, 0.5]$ antenna spacing and fixed far-field location for the UEs in each cell. Hence, the deterministic SIMO channel response [72] will be as follows:

$$\mathbf{h}_i^0 = \sqrt{\beta_i^0} \left[1 \; e^{2\pi \mathrm{j} d_{\mathrm{H}} \sin\left(\varphi_i^0\right)} \ldots e^{2\pi \mathrm{j} d_{\mathrm{H}} (M-1) \sin\left(\varphi_i^0\right)}\right]^{\mathrm{T}} \text{ for } i = 0, 1, \tag{B.19}$$

where β_i^0 represents the large-scale fading effect, and in this case of the LoS propagation channel, it describes the path loss gain. It is assumed the same

for all antenna elements since the propagation distance is much larger than the spacing between antennas. The angle $\varphi_i^0 \in [0, 2\pi]$ is the main nominal angle measured between the UE's planer array reached the BS antennas and the bore-sight direction of the ULA, as shown in Fig.B.7.

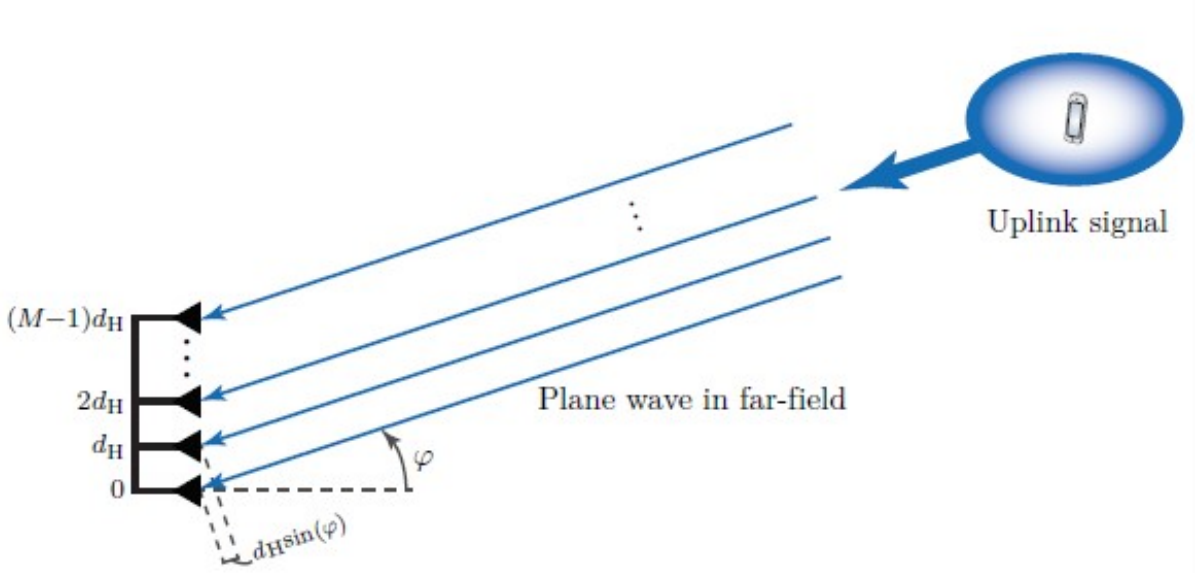

Figure B.7: LoS propagation-SIMO channel

B.3.2.2 NLoS propagation channel model

In this part, the uplink channels are assumed random and come to the BS from multiple paths. A rich scattering environment around the BS is assumed here, and the result is the spatially uncorrelated Rayleigh channel vector, which is given by:

$$\mathbf{h}_i^0 \sim \mathcal{N}_{\mathbb{C}} \left(0_M, \beta_i^0 \mathbf{I}_M \right), \quad \text{for } i = 0, 1, \tag{B.20}$$

where β_i^0 is the large-scale fading parameter that describes the path loss gain in addition to the shadowing effect in this case of the NLoS propagation. On the other hand, the complex Gaussian distribution of the channel describes the small-scale fading and is raised from the independent and identical distributed (i.i.d) multiple paths at the BS, as shown in Fig.B.8.

Corollary B.3.1 (Receive Combining Vector) The benefit behind using M antennas is only satisfied when the uplink channel vectors are known at the BS. This is because the BS will coherently combine the received signals from the

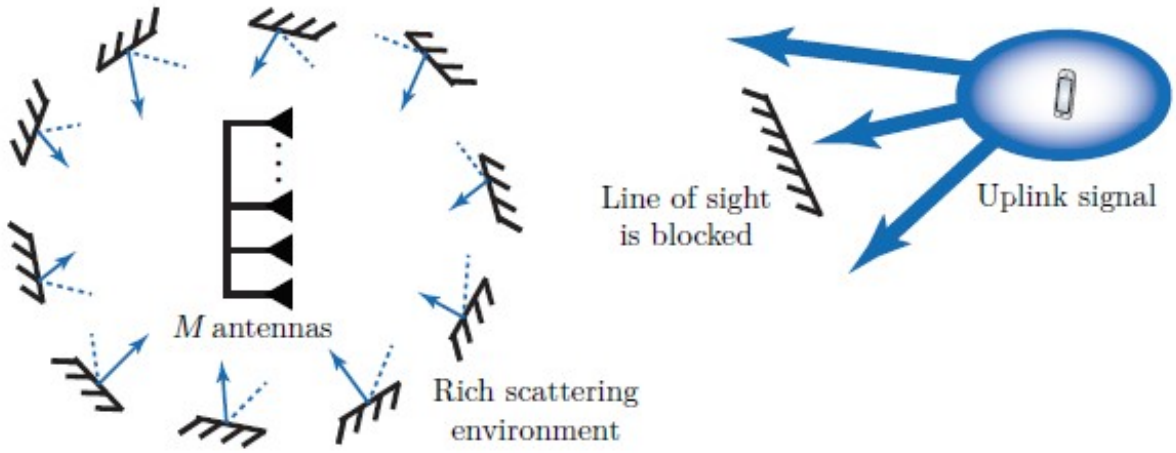

Figure B.8: NLoS propagation-SIMO channel

array of antennas to maximize the received SNR. On the other hand, this explains why the channel estimation subject performs an important key aspect of the MIMO system. Currently, the channel responses are assumed to be known at the BS. Hence, the receive combing vector $\mathbf{v}_0 \in \mathbb{C}$ can be applied at the BS as follows:

$$\mathbf{v}_0^{\mathrm{H}}\mathbf{y}_0 = \underbrace{\mathbf{v}_0^{\mathrm{H}}\mathbf{h}_0^0 s_0}_{\text{Desired signal}} + \underbrace{\mathbf{v}_0^{\mathrm{H}}\mathbf{h}_1^0 s_1}_{\text{Interfering signal}} + \underbrace{\mathbf{v}_0^{\mathrm{H}}\mathbf{n}_0}_{\text{Noise}}. \tag{B.21}$$

The receive combing vector is defined as the linear projection vector that tends to transform the SIMO channel vector into an effective SISO channel vector with a higher achievable SE than the single antenna case. The Maximum Ratio (MR) combining vector

$$\mathbf{v}_0 = \mathbf{h}_0^0, \tag{B.22}$$

is the simple combining scheme [73] that tends to maximize the power of the desired signal term in Eq.(B.21) as $\left|\mathbf{v}_0^{\mathrm{H}}\mathbf{h}_0^0\right|^2 / \|\mathbf{v}_0\|^2$.

B.3.2.3 Uplink SE analysis

After applying the MR combining vector to the received signal at the BS, the uplink achievable SE [11] will be given by:

- In LoS case:

$$\mathrm{SE}_0^{\mathrm{LoS}} = \log_2\left(1+\frac{p\left\|\mathbf{h}_0^0\right\|^2}{p\frac{\left|\left(\mathbf{h}_0^0\right)^{\mathrm{H}}\mathbf{h}_1^0\right|^2}{\left\|\mathbf{h}_0^0\right\|^2}+\sigma^2}\right) = \log_2\left(1+\frac{M}{\bar{\beta}g\left(\varphi_0^0,\varphi_1^0\right)+\frac{1}{\mathrm{SNR}_0}}\right) \tag{B.23}$$

where $g\left(\varphi_0^0,\varphi_1^0\right)$ is a function of the nominal angles of the desired and interfering UEs. Also, it depends on the number of antennas at the BS and can be given by:

$$g(\varphi,\psi) = \begin{cases} \frac{\sin^2(\pi d_{\mathrm{H}} M(\sin(\varphi)-\sin(\psi)))}{M\sin^2(\pi d_{\mathrm{H}}(\sin(\varphi)-\sin(\psi)))} & \text{if } (\varphi)\neq(\psi) \\ M & \text{if } (\varphi)=(\psi) \end{cases} \tag{B.24}$$

- In NLoS case:

In this case, the closed form of the SE will be very difficult and hard to calculate; hence, the SE will be lower-bounded as follows:

$$\mathrm{SE}_0^{\mathrm{NLoS}} = \mathbb{E}\left\{\log_2\left(1+\frac{p\left\|\mathbf{h}_0^0\right\|^2}{p\frac{\left|\left(\mathbf{h}_0^0\right)^{\mathrm{H}}\mathbf{h}_1^0\right|^2}{\left\|\mathbf{h}_0^0\right\|^2}+\sigma^2}\right)\right\} \geq \log_2\left(1+\frac{M-1}{\bar{\beta}+\frac{1}{\mathrm{SNR}_0}}\right) \tag{B.25}$$

As seen from expressions in Eq.(B.23) and Eq.(B.25), the SE is completely described by the signal-to-noise ratio of the desired UE, the relative gain $\bar{\beta}$, and the number of BS antennas. Note that the desired signal gain is linearly scaled with M antennas in the LoS case and with $M-1$ antennas in the NLoS case, which is referred to as the *linear array gain*. The term of the interference gain in the LoS case is $\bar{\beta}g\left(\varphi_0^0,\varphi_1^0\right)$, and is decaying as $1/M$ in Eq.(B.24), while the interference gain in the NLoS case is given by only $\bar{\beta}$ and is independent of M. The latter scaling behaviour explains why the LoS case provides more degrees of freedom at the BS than the NLoS case since the interference decreases as more antennas are placed at the BS. Hence, as M increases, the channel responses of the active users become increasingly orthogonal. Consequently, the base station (BS) can differentiate between the desired and interfering user

equipment (UE) signals [74, 75], that is:

$$\frac{\left(\mathbf{h}_i^0\right)^{\mathrm{H}}\mathbf{h}_k^0}{\sqrt{\mathbb{E}\left\{\left\|\mathbf{h}_i^0\right\|^2\right\}\mathbb{E}\left\{\left\|\mathbf{h}_k^0\right\|^2\right\}}} \to 0 \quad \text{as} \quad M \to \infty \tag{B.26}$$

In this case, the channel vectors $\mathbf{h}_i^0$ and $\mathbf{h}_k^0$ will provide what is called the *asymptotically favourable propagation* [76], which means that both channel directions will become asymptotically orthogonal as M increased.

Figure B.9 [11] describes the impact of adding more antennas on the SE in both cases, the LoS and NLoS.

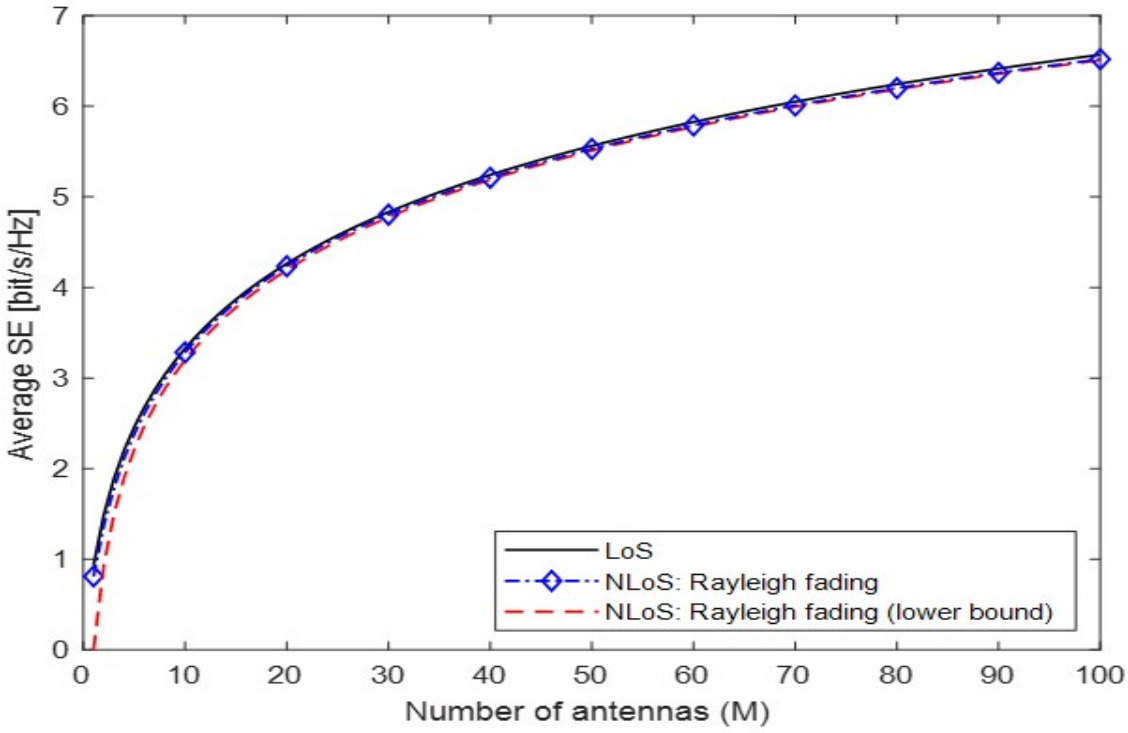

Figure B.9: Uplink SE as a function of the number of antennas

Figure B.9 is simulated under uncorrelated Rayleigh fading channels with a fixed 0dB SNR for the desired UE and -10dB lower SNR for the interfering UE. The simulation curves in Fig. B.9 are measured over different angles of each UE from $0 \to 2\pi$. Here, we can conclude several insights from the simulation results in Fig.B.9 as follows:

- Insight-1: Although we have weak conditions with poor SNR, the achievable SE(bit/sec/Hz) will go from $0.8 \to 3.3$ when M goes from $1 \to 10$. This is because of the offered array gain by the MR combining at the BS.

- Insight-2: The Uplink SE is increasing consistently with the number of placed antennas without limit in contrast to the power scaling case where it is saturated at high SNR. This is also because the MR combining coherently combines all energy from the desired UE and non-coherently combines the noise and interfering UE.

- Insight-3: The LoS and NLoS cases are asymptotically aligned. This is due to using more antennas at the BS, which leads to hardening the channel and overcoming fading variation problems. This spatial diversity attribute is denoted by the name of *asymptotic channel hardening* [77] and can be expressed as follows:

$$\frac{\left\|\mathbf{h}_i^0\right\|^2}{\mathbb{E}\left\{\left\|\mathbf{h}_i^0\right\|^2\right\}} \to 1 \tag{B.27}$$

- Insight-4: The LoS channels always provide channel-hardening issues since they are deterministic values and characterized like spatially uncorrelated channels (since the channel gains from all incoming planner arrays are the same and do not depend on the channel directions), while the NLoS case will provide the channel hardening characteristics only when

$$\frac{\left\|\mathbf{h}_i^0\right\|^2}{\mathbb{E}\left\{\left\|\mathbf{h}_i^0\right\|^2\right\}} = \frac{\left\|\mathbf{h}_i^0\right\|^2}{M\beta_i^0} \to 1 \quad \text{as} \quad M \to \infty. \tag{B.28}$$

  this means that the instantaneous normalized channel gains (in the NLoS case) will closely converge to the deterministic average channel gains (in the LoS case) when more antennas are added at the BS.

In fact, although the SE grows monotonically in a logarithmic way with the increasing number of antennas, this way does not provide us with sufficient improvements to investigate any order of magnitude in the SE since the scalability will be inside the logarithm function of the SE expression. Hence, a spatial multiplexing method will be discussed in the following subsection, which will significantly improve the throughput of the cellular network.

B.3.3 Increasing number of UEs per cell

As demonstrated in the last subsections, increasing the sending power by the UEs or employing several antennas at the BS can only bring minor improvements to the UL SE of the cell. This is due to the fact that these strategies improve only the SNIR that exists within the logarithmic expression of the SE, causing the SE to increase gradually. Identifying a method that makes improvements outside the logarithmic expression of the SE is the main interest of this part. The space division multiple access (SDMA) is the conceived spatial multiplexing method that has been used since the earlier time of using the MIMO system in (1980-1990) years. Potentially, spatial multiplexing is used when there are multiple UEs in the cell and served by a BS with M antenna array. Hence, a sum of SE from K UEs can be achieved within each cell. The only bottleneck from such a case is the multiuser interference. In 1990, the MIMO system was implemented by using ten antennas only. In the early 2000s, the information theory of the multiuser-single cell scenario was described [78, 79] where the Multiuser MIMO terminology was cemented. However, applying the terminology of multiuser MIMO in the cellular network has been surveyed in many papers and books like [80, 81, 82, 35].

In order to analyze the uplink signals using the SDMA, the following scenario of the Wyner cellular model will be considered with K active users in each cell instead of one UE per cell as follows:

Let $\mathbf{h}_{0k}^{0} \in \mathbb{C}^{\mathbb{M}}$ and $\mathbf{h}_{1i}^{0} \in \mathbb{C}^{\mathbb{M}}$, for $k,i = 1,2,...K$ denotes the uplink received channel vectors from the desired UEs in cell 0 and interfering UEs in cell 1 at the BS in cell 0, respectively. Then the multi-antenna received signal at the BS in cell 0 will be given as follows:

$$\mathbf{y}_0 = \underbrace{\sum_{k=1}^{K} \mathbf{h}_{0k}^{0} s_{0k}}_{\text{“Desired signals”}} + \underbrace{\sum_{i=1}^{K} \mathbf{h}_{1i}^{0} s_{1k}}_{\text{“Interfering signals”}} + \underbrace{\mathbf{n}_0}_{\text{“Noise”}}, \tag{B.29}$$

such that $\mathrm{s}_{lk} \sim \mathbb{C}_{\mathscr{N}}(0,p)$ is the transmitted signal by any k UE in cell l where $l = 0,1$, and $\mathrm{n}_0 \sim \mathbb{C}_{\mathscr{N}}\left(0_M, \sigma^2 \mathbf{I}_M\right)$ is the additive noise at the BS antennas. The channel responses in Eq.(B.29) can be modelled in two cases of LoS and NLoS propagation models as in Eq.(B.19) and Eq.(B.20), except the notation of the

channel vector, will be $\mathbf{h}_{lk}^{0}$ instead of $\mathbf{h}_{i}^{0}$, where the first subscript l denotes the cell number and the second subscript k denotes the UE that resides in that cell. Here, it can define two types of interference: intra-cell interference, which is raised inside the serving cell, and inter-cell interference, raised from the other adjacent cells. The interference from the own cell is much stronger than the interference from the adjacent cell and must be removed if (K-fold) growth in spectral efficiency is needed.

As mentioned earlier, the main objective of the receive combining scheme is to emphasize the relative power of the desired signal term and de-emphasize the interfering signal power in the received signal. The MR combining scheme defined in the corollary (B.3.1) is the popular sub-optimal vector that increases the desired signal's relative power. It is not the best option to choose when there are interfering signals. In addition to the sub-optimal vector, the combining vector in the following corollary is the optimal combining scheme that can be used in the cellular network to maximize the SE.

B.3.3.1 Multicell minimum mean squared error (M-MMSE) combining vector

It is given by:

$$\mathbf{v}_{0k} = p\left(p\sum_{i=1}^{K}\mathbf{h}_{0i}^{0}\left(\mathbf{h}_{0i}^{0}\right)^{\mathrm{H}} + p\sum_{i=1}^{K}\mathbf{h}_{1i}^{0}\left(\mathbf{h}_{1i}^{0}\right)^{\mathrm{H}} + \sigma^{2}\mathbf{I}_{M}\right)^{-1}\mathbf{h}_{0k}^{0}, \tag{B.30}$$

The intra-cell and inter-cell interfering signals are taken into account in Eq.(B.30) of the M-MMSE vector inside the inverse matrix term. Hence, the M-MMSE vector is the optimal option for the combining vector to be used since it makes a trade-off between amplifying the desired signal and/or weakening the interfering signals.

Applying the combing vector scheme at the BS means that the BS can direct its attention to the location of the k^{th} desired UE by projecting its received signal in the direction of the desired channel response. For now, if the MR combining

vector in Eq.(B.22) is applied at the BS in cell 0, then:

$$\mathbf{v}_{0k}^{\mathrm{H}}\mathbf{y}_0 = \underbrace{\mathbf{v}_{0k}^{\mathrm{H}}\mathbf{h}_{0k}^{0}s_{0k}}_{\text{Desired signal}} + \underbrace{\sum_{\substack{i=1\\ i\neq k}}^{K}\mathbf{v}_{0k}^{\mathrm{H}}\mathbf{h}_{0i}^{0}s_{0i}}_{\text{Intra-cell interference}} + \underbrace{\sum_{i=1}^{K}\mathbf{v}_{0k}^{\mathrm{H}}\mathbf{h}_{1i}^{0}s_{1i}}_{\text{Inter-cell interference}} + \underbrace{\mathbf{v}_{0k}^{\mathrm{H}}\mathbf{n}_0}_{\text{Noise}}. \quad \text{(B.31)}$$

where $\mathbf{v}_{0k} = \mathbf{h}_{0k}^{0}$; hence, the knowledge of the channel response $\mathbf{h}_{0k}^{0}$ is required at the BS. However, after combining the signal from multiple antennas using the MR combining vector, the achievable sum SE at the BS in cell 0 in the uplink transmission can be given in the following expression in bit/sec/Hz/cell [11], assuming the BS knows all UEs' channel vectors as follows:

- In LoS case: The closed form of the sum SE will be as follows:

$$\mathrm{SE}_0^{\mathrm{LoS}} = \sum_{k=1}^{K}\log_2\left(1+\frac{M}{\sum_{\substack{i=1\\ i\neq k}}^{K} g\left(\varphi_{0k}^{0},\varphi_{0i}^{0}\right)+\bar{\beta}\sum_{i=1}^{K} g\left(\varphi_{0k}^{0},\varphi_{1i}^{0}\right)+\frac{1}{\mathrm{SNR}_0}}\right) \quad \text{(B.32)}$$

  where $g(.,.)$ is defined in Eq.(B.24).

- In NLoS case: The achievable sum SE will be lower bounded as follows:

$$\begin{aligned}\mathrm{SE}_0^{\mathrm{NLoS}} &= \sum_{k=1}^{K}\mathbb{E}\left\{\log_2\left(1+\frac{p\left\|\mathbf{h}_{0k}^{0}\right\|^2}{\sum_{\substack{i=1\\ i\neq k}}^{K} p\frac{\left|\left(\mathbf{h}_{0k}^{0}\right)^{\mathrm{H}}\mathbf{h}_{0i}^{0}\right|^2}{\left\|\mathbf{h}_{0k}^{0}\right\|^2}+\sum_{i=1}^{K} p\frac{\left|\left(\mathbf{h}_{0k}^{0}\right)^{\mathrm{H}}\mathbf{h}_{1i}^{0}\right|^2}{\left\|\mathbf{h}_{0k}^{0}\right\|^2}+\sigma^2}\right)\right\} \\ &\geq K\log_2\left(1+\frac{M-1}{(K-1)+K\bar{\beta}+\frac{1}{\mathrm{SNR}_0}}\right)\end{aligned} \quad \text{(B.33)}$$

In addition to the array gain of the desired signal, the big advantage when using the SDMA is the multiplexing gain in front of the logarithmic expression of the SE that arises from multiplexing several K users in the individual cell. Although the big advantages of the array and multiplexing gains, a bottleneck problem will arise that results from the interfering signals of the other users in the network. The interior scaling behaviour in both expressions in Eq.(B.32) and Eq.(B.33) with MR combining can be further discussed as follows:

- In LoS case: The power of the desired signal increases by M, while the interfering signal power is proportionate to K/M. Hence, the SNIR will be M^2/K and is adequately scaled as $M/\sqrt{K}$. This ratio is named as the antenna-UE ratio and is a constant and preferable SNIR ratio in the multiuser scenario. It means that for every $\sqrt{K}$ additional UE number in the cell, we need M antennas at the BS to keep this antenna-UE ratio constant.
- In NLoS case: The power of the desired signal increases by M, while the interfering signal power linearly increases with K. That means the antenna-UE ratio will be M/K, which is not a preferable SNIR ratio compared to the LoS case since additional antennas will be required to suppress the interference in the NLoS case. We can serve roughly the same SNIR per each UE in the cell by rising M with K to preserve the antenna-UE ratio constant. The case with the M-MMSE is different and will be shown in Fig.B.10 in the next paragraph.

To reveal the behaviour of the two combining vectors, Figs.B.10 and B.11 show the achievable uplink sum SE as a function with the number of UEs in each cell for $M = 10$ and $M = 100$ cases. Assume fixed SNR at the desired UEs in cell 0 ($\text{SNR}_0 = 0$dB) and $\bar{\beta} = -10$dB relative strength.

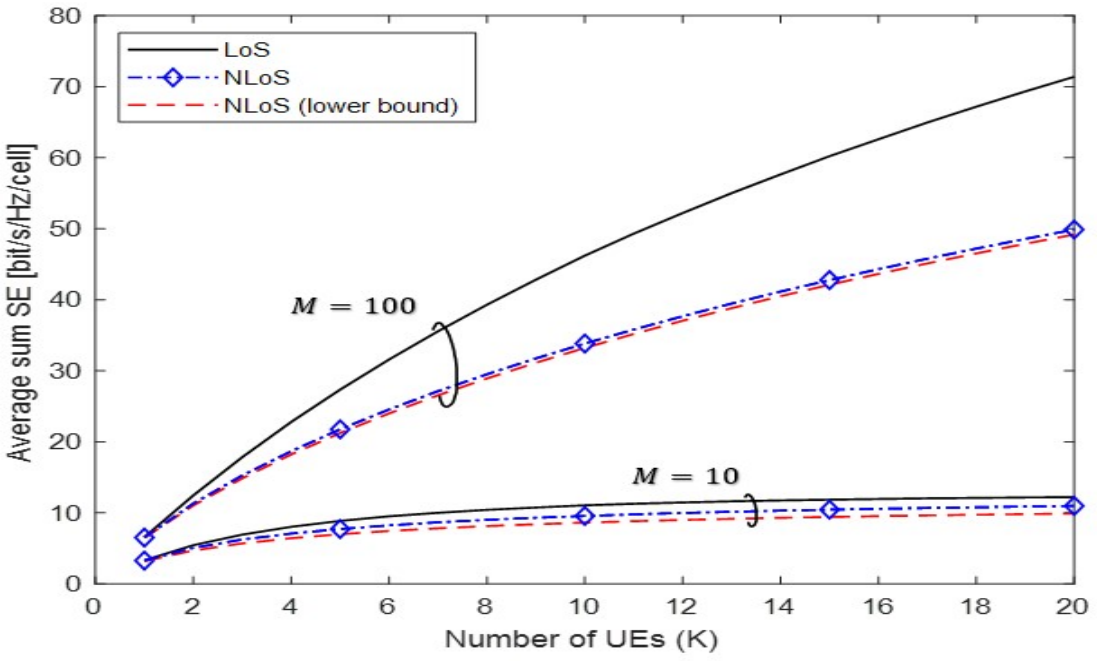

Figure B.10: Uplink sum SE as a function of the number of UEs using MR combining

The important insights from Fig.B.10 can be concluded as follows:

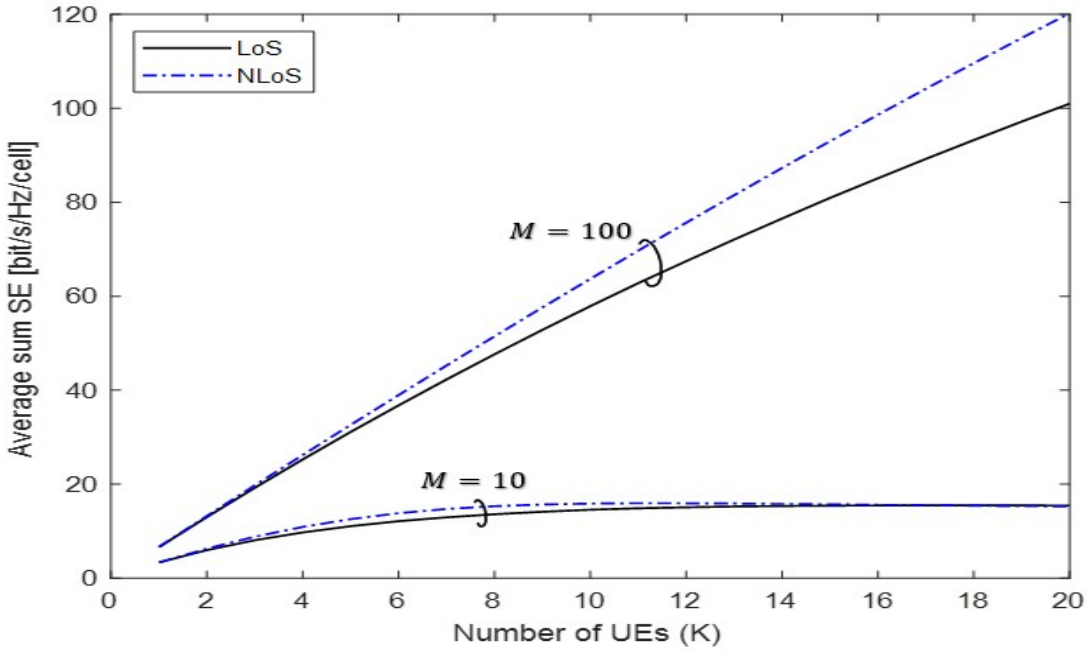

Figure B.11: Uplink sum SE as a function of the number of UEs M-MMSE combining

- Insight 1: The sum SE [bit/sec/Hz/cell] slowly increases at $M = 10$ antennas in both LoS and NLoS propagation cases when using the MR and M-MMSE combining vectors. This implies that the base station (BS) lacks sufficient degrees of freedom to distinguish between multiple user equipments at this number of antennas.

- Insight 2: The sum SE [bit/sec/Hz/cell] roughly linearly increases by (K-folds) at $M = 100$ antennas in both LoS and NLoS propagation cases when using the MR and M-MMSE combining vectors.

- Insight 3: When using MR combining, the average of the sum SE is better in the LoS than in the NLoS case. This is because the SNIR in the LoS case follow the preferable antenna-UE ratio ($M/\sqrt{K}$). On the other hand, in the LoS case, the strongest interference from a similar nominal angle is rarely happening in practice. However, when using M-MMSE combining vector, the average sum SE will be better in the case of the NLoS than that of the LoS case. This is due to its good structure that can suppress interference from other UEs. In other words, in the NLoS case, only a few channels from other UEs may parallel the desired UE's channel.

In conclusion, using the SDMA in the uplink transmission increases the average sum SE by more than one order of magnitude. This is accomplished by

increasing the number of antennas at the BS and serving multiple UEs in each cell of the network. That is, it can achieve more improvements in the SE when the antenna-UE ratio is increased to some large values. In the next sections, we will discuss important concepts related to the massive MIMO system and the channel estimation to be used later in this thesis.

B.4 Channel State Information (CSI)

Up to this point, it has been assumed that the channel responses are entirely familiar to the base station (BS), but in practice, the channel responses need to be estimated every time the channel changes. The reason behind that is the varying time-changing nature of the channels, where they are constant for a few milliseconds and only over a little hundred KHz bandwidths. Hence, the channels at the BS are commonly modelled by random variables, for example, Rayleigh fading channels. However, the current knowledge of these channel responses at the BS is called the channel state information (CSI).

B.4.1 Pilot signaling

The instantaneous CSI at the BS must be acquired simultaneously when the channel changes. Pilot signalling is the main method used to acquire CSI at the BS to estimate the uplink channel responses. Figure B.12 shows the pilot signalling method where any single-antenna UE in the network can send a predetermined pilot signal which is simultaneously received by the BS array antennas.

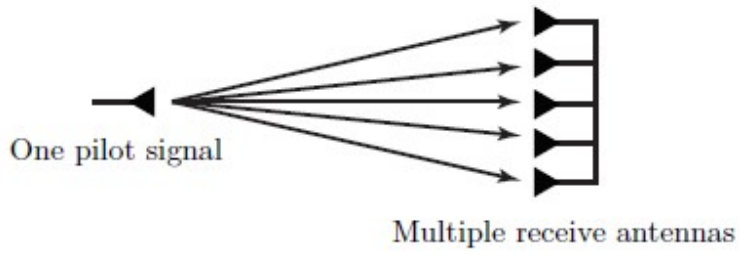

Figure B.12: Pilot signalling from single-antenna to multiple antennas

However, to determine the channel responses from two transmitting anten-

nas, it is necessary to utilize two distinct pilot signals that are orthogonal to each other [53, 83, 9]. The number of pilot signals is equivalent to the number of transmitting antennas. This means that if we have K transmitting antennas, then K orthogonal pilot signals are needed, while an unlimited number of receiving antennas can receive these pilots simultaneously.

B.4.2 Time Division Duplex (TDD) and reciprocity

Based on the concept of the pilot signalling above, K orthogonal pilot signals will be required in the uplink direction if K UEs with a single antenna reside in the cell. Similarly, when signalling in the downlink direction from M BS antennas to the UEs, M orthogonal pilot signals are needed. Hence, an overhead problem will be raised from the difference between the signalling process in the Uplink and downlink directions. However, the relationship between the number of BS antennas and the number of the single-antenna UEs follows the preferable antenna-UE ratio ($M/K \geq 4$), which is the preferred operating mode that provides many degrees of freedom at the BS [11]. To overcome the overhead problem, the time division duplex (TDD) in Fig. B.13 with reciprocal channels will be used to separate the uplink and downlink transmission. The TDD state that the estimated channel in the uplink directions (during the uplink signalling with K pilot signals) will be used in the downlink directions and doesn't require to be estimated again.

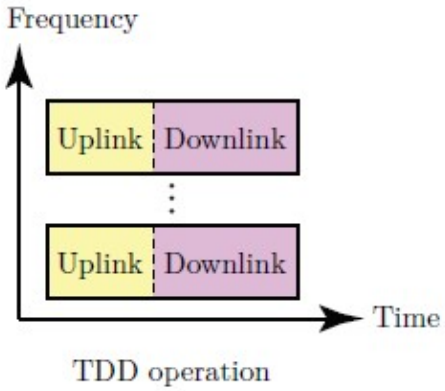

Figure B.13: Time Division Duplex (TDD) protocol

Hence, the TDD protocol depends only on the K pilot signals and is independent of the M BS antennas.

Appendix C

Principles of Estimation Theory

C.1 Parameter Estimation

The main goal of the estimation method is to get an approximate value of an unknown variable (the parameter needs to be estimated) based on some noisy observations. There are two parameter estimation methods that depend on whether the parameter variable is deterministic or random [84]. The maximum likelihood estimation (MLE) will be used if the unknown variable is deterministic. The second estimation method is the Bayesian estimator, applied when the unknown variables are random. In this section, the Bayesian estimator is one of our interests since our unknown variable is the channel response h, representing one realization of the random process $\mathbb{H}$. However, an important thing that the Bayesian estimator needs is the statistical distribution of the unknown random variable that may be known or partially known.

Definition C.1.1 (Bayesian estimation) Given a set of observations from the received signal $\mathbf{y} \in \mathbb{C}^T$ and wanted to estimate the random channel vector $\mathbf{h} \in \{\mathscr{H}\}$ based on these observations, the estimator of the channel vector $\mathbf{h}$ that minimizes the expectation:

$$\mathbb{E}\{\ell(\mathbf{h}, \hat{\mathbf{h}}(\mathbf{y}))\} \tag{C.1}$$

is called the Bayesian estimator and is denoted by $\hat{\mathbf{h}}(\mathbf{y})$. The function $\ell(.,.)$ denotes the loss function, and the Squared-Error ($\ell(\mathrm{h}, \hat{\mathrm{h}}(\mathrm{y})) = \|\mathrm{h} - \hat{\mathrm{h}}(\mathrm{y})\|^2$) is one of our interests in this thesis since its expectation gives us how the variance of the estimation error is large.

Definition C.1.2 (Minimum MSE (MMSE) estimator) It is defined by the conditional expected value

$$\hat{\mathbf{h}}_{\text{MMSE}}(\mathbf{y}) = \mathbb{E}\{\mathbf{h} \mid \mathbf{y}\} = \int_{\mathscr{H}} \mathbf{h} f(\mathbf{h} \mid \mathbf{y}) d\mathbf{h} \tag{C.2}$$

that minimizes the mean squared error

$$\mathbb{E}\left\{ \|\mathbf{h} - \hat{\mathbf{h}}(\mathbf{y})\|^2 \right\} \tag{C.3}$$

where $f(\mathbf{h} \mid \mathbf{y})$ is the conditional probability density function (PDF) of the unknown random vector $\mathbf{h}$ given the set of observations $\mathbf{y}$.

C.1.1 MMSE estimator of nonzero mean complex Gaussian vector

Given the observation $\mathbf{y} = \mathbf{A}\mathbf{h} + \mathbf{n} \in \mathbb{C}^L$, and need to estimate $\mathbf{h} \in \mathbb{C}^N$, where $\mathbf{h}$ is a complex Gaussian vector distributed as $\mathbf{h} \sim \mathbb{C}_{\mathscr{N}}(\mu_{\mathbf{h}}, \mathbf{R})$. The matrix $\mathbf{A} \in \mathbb{C}^{L \times N}$ is a defined matrix, and $\mathbf{n}$ is the independent additive receiver noise distributed as $\mathbf{n} \sim \mathbb{C}_{\mathscr{N}}(\mu_{\mathbf{n}}, \mathbf{S})$. The covariance matrices $\mathbf{R}$ and $\mathbf{S}$ are assumed positive semi-definite matrices. The MMSE of $\mathbf{h}$ depending on $\mathbf{y}$ will be take the formula:

$$\hat{\mathbf{h}}_{\text{MMSE}}(\mathbf{y}) = \mu_{\mathbf{h}} + \mathbf{R}\mathbf{A}^{\text{H}} \left(\mathbf{A}\mathbf{R}\mathbf{A}^{\text{H}} + \mathbf{S} \right)^{-1} (\mathbf{y} - \mathbf{A}\mu_{\mathbf{h}} - \mu_{\mathbf{n}}). \tag{C.4}$$

and the MSE, ($\mathbb{E}\left\{ \|\mathbf{h} - \hat{\mathbf{h}}(\mathbf{y})\|^2 \right\}$), will be given as

$$\text{MSE} = \text{tr}\left(\mathbf{R} - \mathbf{R}\mathbf{A}^{\text{H}} \left(\mathbf{A}\mathbf{R}\mathbf{A}^{\text{H}} + \mathbf{S} \right)^{-1} \mathbf{A}\mathbf{R} \right). \tag{C.5}$$

The error correlation matrix $\text{C}_{\text{MMSE}} = \mathbb{E}\left\{ \left(\mathbf{h} - \hat{\mathbf{h}}_{\text{MMSE}} \right) \left(\mathbf{h} - \hat{\mathbf{h}}_{\text{MMSE}} \right)^{\text{H}} \right\}$ will be given by:

$$\text{C}_{\text{MMSE}} = \mathbf{R} - \mathbf{R}\mathbf{A}^{\text{H}} \left(\mathbf{A}\mathbf{R}\mathbf{A}^{\text{H}} + \mathbf{S} \right)^{-1} \mathbf{A}\mathbf{R} \tag{C.6}$$

that is, $\text{MSE} = \text{tr}\left(\text{C}_{\text{MMSE}} \right)$.

C.1.2 MMSE estimator of zero mean complex Gaussian vector

Given the observation $\mathbf{y} = \mathbf{h}q + \mathbf{n} \in \mathbb{C}^N$, and need to estimate $\mathbf{h} \in \mathbb{C}^N$, where

$\mathbf{h}$ is a complex Gaussian vector distributed as $\mathbf{h} \sim \mathbb{C}_{\mathscr{N}}(0_N, \mathbf{R})$. The symbol $q \in \mathbb{C}$ denotes the pilot signal, and $\mathbf{n}$ is the independent additive receiver noise distributed as $\mathbf{n} \sim \mathbb{C}_{\mathscr{N}}(0_N, \mathbf{S})$. The covariance matrices $\mathbf{R}$ and $\mathbf{S}$ are assumed positive semi-definite matrices. The MMSE of $\mathbf{h}$ depending on $\mathbf{y}$ will be given by the formula:

$$\hat{\mathbf{h}}_{\text{MMSE}}(\mathbf{y}) = q^{\star}\mathbf{R}\left(|q|^2\mathbf{R}+\mathbf{S}\right)^{-1}\mathbf{y} \tag{C.7}$$

and the MSE is

$$\text{MSE} = \text{tr}\left(\mathbf{R} - |q|^2\mathbf{R}\left(|q|^2\mathbf{R}+\mathbf{S}\right)^{-1}\mathbf{R}\right). \tag{C.8}$$

The correlation matrix of the error in estimation is

$$\mathrm{C}_{\text{MMSE}} = \mathbf{R} - |q|^2\mathbf{R}\left(|q|^2\mathbf{R}+\mathbf{S}\right)^{-1}\mathbf{R} \tag{C.9}$$

Corollary C.1.1 (MMSE estimate of deficient matrices) The MMSE estimator is also used when the unknown variables (channels) have low-rank covariance matrices (spatially correlated channels), as follows: using the channel model of Karhunen-Loeve given in (2.3) by letting $\mathbf{D}e = \mathbf{g}$, hence, the channel model will be given as $\mathbf{h} = \mathbf{U}\mathbf{g}$, where $\mathbf{g} \sim \mathbb{C}_{\mathscr{N}}(0_r, \mathbf{D})$ and $\mathbf{h} \sim \mathbb{C}_{\mathscr{N}}(0_r, \mathbf{R})$, and $\mathbf{R} = \mathbf{U}\mathbf{D}\mathbf{U}^{\mathrm{H}}$, where $\mathbf{U} \in \mathbb{C}^{N\times r}$ and $\mathbf{D} \in \mathbb{C}^{r\times r}$.

Now, for the given observation $\mathbf{y} = \mathbf{h}q + \mathbf{n}$, we need to estimate the channel that given by $\mathbf{h} = \mathbf{U}\mathbf{g}$; use this notation, the observation yields

$$\mathbf{y} = \mathbf{U}\mathbf{g}q + \mathbf{n} \tag{C.10}$$

Since $\mathbf{U}q$ are known quantities like matrix $\mathbf{A}$, and $\mathbf{n}$ is the independent noise; hence, only the unknown variable is $\mathbf{g}$ and has full rank covariance matrix $\mathbf{D}$. To apply the MMSE estimator for the unknown variable $\mathbf{g}$, using the definition C.1.2 it yields:

$$\begin{aligned}
&\mathbb{E}\{\mathbf{h} \mid \mathbf{y}\} = \mathbf{U}\mathbb{E}\{\mathbf{g} \mid \mathbf{y}\} \\
&= q^{\star}\mathbf{U}\mathbf{D}\mathbf{U}^{\mathrm{H}}\left(|q|^2\mathbf{U}\mathbf{D}\mathbf{U}^{\mathrm{H}}+\mathbf{S}\right)^{-1}\mathbf{y} \\
&= q^{\star}\mathbf{R}\left(|q|^2\mathbf{R}+\mathbf{S}\right)^{-1}\mathbf{y}
\end{aligned} \tag{C.11}$$

and this is the same estimator in Eq.(C.7).

C.1.3 Linear MMSE estimator

In some cases, the unknown variable is non-Gaussian distributed; hence, applying the MMSE estimator of the Gaussian random variable becomes very difficult because either analytically impossible or because of the lack of obtaining the full statistical PDF. The linear Bayesian estimator is then valuable since just the first and second statistics of the unknown variable are required.

Definition C.1.3 (Linear MMSE (LMMSE)) The Linear estimator LMMSE is defined by the linear equation:

$$\hat{\mathbf{h}}_{\text{MMSE}}(\mathbf{y}) = \mathbf{A}\mathbf{y} + \mathbf{b} \tag{C.12}$$

where the linear equation parameters **A** and **b** are jointly selected to minimize the mean squared error

$$\mathbb{E}\left\{ \|\mathbf{h} - \hat{\mathbf{h}}(\mathbf{y})\|^2 \right\} \tag{C.13}$$

The closed formula of the Bayesian LMMSE estimator of the unknown variable **h** based on the observation **y** is given by:

$$\hat{\mathbf{h}}_{\text{LMMSE}}(\mathbf{y}) = \mathbb{E}\{\mathbf{h}\} + \mathbf{C}_{\mathbf{hy}}\mathbf{C}_{\mathbf{yy}}^{-1}(\mathbf{y} - \mathbb{E}\{\mathbf{y}\}) \tag{C.14}$$

The correlation matrix of the estimation error will the relation:

$$\mathrm{C}_{\text{LMMSE}} = \mathrm{C}_{\mathbf{hh}} - \mathrm{C}_{\mathbf{hy}}\mathrm{C}_{\mathbf{yy}}^{-1}\mathrm{C}_{\mathbf{hy}}^{\mathrm{H}}, \tag{C.15}$$

and can define the covariance matrices in Eq.(C.15) as follows:

$$\mathbf{C}_{\mathbf{hy}} = \mathbb{E}\left\{ (\mathbf{h} - \mathbb{E}\{\mathbf{h}\})(\mathbf{y} - \mathbb{E}\{\mathbf{y}\})^{\mathrm{H}} \right\} \tag{C.16}$$

$$\mathbf{C}_{\mathbf{hh}} = \mathbb{E}\left\{ (\mathbf{h} - \mathbb{E}\{\mathbf{h}\})(\mathbf{h} - \mathbb{E}\{\mathbf{h}\})^{\mathrm{H}} \right\} \tag{C.17}$$

$$\mathbf{C}_{\mathbf{yy}} = \mathbb{E}\left\{ (\mathbf{y} - \mathbb{E}\{\mathbf{y}\})(\mathbf{y} - \mathbb{E}\{\mathbf{y}\})^{\mathrm{H}} \right\} \tag{C.18}$$

The MSE can be calculated from the formula $\text{MSE} = \text{tr}\left(\text{C}_{\text{LMMSE}}\right)$. Note that the Linear Bayesian estimator relies only on the mean and covariance matrices of the unknown variables, and the full PDF distribution is not required. This pushes the LMMSE estimator to be quite suited for realistic implementations since extracting these moments is relatively simple.

C.1.4 Law of large numbers

Let h denote an arbitrary random variable, and the expectation $\mathbb{E}\{h\}$ is the mean of the random variable h. Suppose we have N observations from the random variable h denoted by $h_1, h_2, ..., h_N$, then the mean of these observations is called the (sample mean) and denoted by:

$$\begin{aligned} \bar{h_n} &= \frac{h_1 + h_2 +, ..., + h_N}{N} \\ &= \frac{1}{N} \sum_{n=1}^{N} h_n \end{aligned} \tag{C.19}$$

The law of large numbers says that: $\bar{h_n}$ converges to $\mathbb{E}\{h\}$ when $N \to \infty$.

اللاخطي الحاصل في أجهزة محطة القاعدة وأجهزة المستخدم. ثم يتم تعلم التشوه اللاخطي باستخدام طرق التعلم العميق واستغلالها لتقدير القنوات الفعالة. تم تدريب نماذج مختلفة للتعلم العميق (استنادًا إلى النماذج المطورة في الأدبيات الحديثة في هذه الأطروحة) باستخدام خوارزميات تحسين مختلفة ودوال فقدان ورموز تعديل وأنواع متعددات حدود معينة باستخدام بيئة TensorFlow في إصدار Python-9.3 مثبتة على كمبيوتر محمول يحتوي على بطاقة رسومات MAXQ RTX2070 NVIDIA وذاكرة وصول عشوائي بحجم 16 جيجابايت ومعالج i7-10750H. Core حيث تمت مقارنة النتائج التحليلية لإجراء التعلم العميق مع مقدرات البايزية الحديثة الواعية أو غير الواعية لتشوهات النظام. تظهر نتائج المحاكاة أن نهج التعلم العميق ليس دائمًا يحسن جودة التقدير للقناة، وتحت بعض خوارزميات التحسين، فإنه يكون الأسوأ. من جهة اخرى تم العثور على أن محسّن الأدم مع رموز الإرسال الغاوسية ومتعددات الحدود بدون الذاكرة أظهرت اكتسابًا ملحوظا في الاداء يبلغ حوالي 6 ديسيبل أقل في معدل الخطأ المربع المتوسط المعياري من المقدرات البايزية.

الخلاصة

تعد تقنية الاتصالات اللاسلكية الضخمة ذات المدخلات والمخرجات المتعددة واحدة من أحدث التقنيات لأجيال النظام اللاسلكي في الوقت الحالي والمستقبلي والسبب وراء ذلك هو العدد الكبير من الهوائيات M = 64 أو أكثر في محطة القاعدة (BS) والتي يمكن أن تزيد بشكل محتمل من معدل بيانات الشبكة من حيث الكفاءة الطيفية (SE) وبالتالي تتيح هذه التقنية التضاعف المكاني لعشرات من أجهزة المستخدم (UEs) في الشبكة، مما يجعل من الأصعب تقدير قنوات الارتباط الصعودي بدقة. ومن جهة اخرى، فإن تشويه الأجهزة الموجود على كل سلسلة تردد هوائي ستشكل ايضا تحديًا آخر في تقدير القناة. فهي تميل إلى تشويه الإشارات المستلمة وتؤدي إلى تقدير غير دقيق للقنوات. استنادا الى هذه الحقائق، تقدم هذه الأطروحة ثلاثة نماذج لتقدير القناة والتي تأخذ في نظر الاعتبار جودة الأجهزة في محطة القاعدة وأجهزة المستخدم.

يفرض النموذج الأول المقترح لتقدير القناة جودة مثالية لكل من محطة القاعدة وأجهزة المستخدم، ويستغل خصائص الترابط المكاني لقنوات الارتباط الصعودي لتقليل تعقيد الحسابات لمقدر الخطأ المتوسط المربع الأدنى (MMSE) البايزية عن طريق قص قنوات الرتبة الكاملة استنادًا إلى تحليل القيم الذاتية الخاصة بها. يتم استخدام قناتين: القناة المحلية للتشتت بمصفوفة تباين معروفة تمامًا في محطة القاعدة وسيناريو القناة العملي، حيث ليس لمحطة القاعدة أي معرفة بإحصاءات القناة. تظهر نتائج هذا النموذج المقترح جودة قناة قابلة للمقارنة في كلا الحالتين مع توفير تقليل بنسبة 30٪ في تعقيد الحسابات للتقدير البايزي.

يفرض النموذج الثاني المقترح جودة غير مثالية للأجهزة ويقدر القنوات في بيئات عملية باستغلال نموذج Bussgang لوصف التشوبه في اجهزة محطة القاعدة وأجهزة المستخدم، ثم يستخدم إجراء تحسين محدب لتنظيم القنوات المقدرة العملية. يتم التحقق من نتائج هذا الإجراء باستخدام القنوات المحلية للتشتت التي تم قياسها بواسطة معدل الخطأ المربع المتوسط المعياري (NMSE) مقابل نسبة الإشارة إلى الضوضاء الفعالة. تم الحصول على النتائج في برنامج MATLAB R2020a وتظهر تقليلًا كبيرًا (بواقع نحو ترتيب واحد في الحجم) في قاعدة الأخطاء مقارنةً بالقناة التقليدية، خاصةً في نطاق نسبة الإشارة إلى الضوضاء (SNR) العالية من 20 ديسيبل إلى 30 ديسيبل.

يحقق النموذج الثالث للتقدير تأثير التشويه اللاخطي في الأجهزة على القناة المقدرة للارتباط الصعودي في بيئة التلاشي الرايسي. يتم استخدام متعددات الحدود من الدرجة الثالثة لمحاكاة التأثير المشترك للتشوه

اقرار لجنة المناقشة

نحن اعضاء لجنة المناقشة ، نشهد باننا اطلعنا على رسالة الدكتوراه الموسومة (تحسين تقدير القناة لنظام الاتصالات اللاسلكية الضخمة ذات المدخلات والمخرجات المتعددة لأجهزة الارسال والاستقبال المثالية وغير المثالية)وقد ناقشنا الطالب في محتوياتها وفيما له علاقة بها ، نؤيد انها جديرة بالقبول لنيل درجة دكتوراه فلسفة في هندسة الالكترونيك والاتصالات

رئيس اللجنة
التوقيع
الاسم: ا.د. سمير جاسم محمد
التاريخ

عضو اللجنة
التوقيع
الاسم: ا.د. بيان مهدي صبار
التاريخ

عضو اللجنة
التوقيع
الاسم: ا.د. اسامة قاسم جمعة
التاريخ

عضو اللجنة
التوقيع
الاسم: ا.د. سعد سفاح حسون
التاريخ

عضو اللجنة
التوقيع
الاسم: ا.د. احمد عبد الكاظم حمد
التاريخ

عضو اللجنة (مشرفا)
التوقيع
الاسم: ا.د. ايهاب عبد الرزاق حسين
التاريخ

مصادقة رئيس القسم
التوقيع
الاسم: ا.د. قيس كريم عمران
التاريخ

مصادقة عميد الكلية
التوقيع
الاسم: ا.د. ليث علي عبد الرحيم
التاريخ

جمهورية العراق
وزارة التعليم العالي والبحث العلمي
جامعة بابل/كلية الهندسة
قسم الهندسة الكهربائية

تحسين تقدير القناة لنظام الاتصالات اللاسلكية الضخمة ذات المدخلات والمخرجات المتعددة لأجهزة الارسال والاستقبال المثالية وغير المثالية

رسالة

مقدمة إلى جامعة بابل - كلية الهندسة وهي جزء من متطلبات الحصول علي درجة دكتوراة فلسفة في الهندسة / الهندسة الكهربائية / الإلكترونيك والاتصالات

من قبل
احمد حسين شاطي حمزة

بإشراف
أ.د. أيهاب عبد الرزاق حسين

2023 م

1444هـ