University of Huddersfield

Analysis and Design of Non-Orthogonal Multiple Access (NOMA) Techniques for Next Generation Wireless Communication Systems

Zeyad Elsaraf

PhD Supervisor: Dr. Faheem A. Khan

October 31, 2023

A report submitted in partial fulfilment of the requirements for the PhD degree in Electrical and Electronic Engineering

Copyright Statement

Acknowledgements

First and foremost, I wish to express my deep gratitude to my supervisor Dr. Faheem Khan for his guidance, support and encouragement throughout the PhD period. His friendly approach and ability to provide quick solutions to any problem I faced was a big help to me. Without his support, this thesis would not have been possible. I would also like to acknowledge my co-supervisor Dr. Qasim Zeeshan Ahmed for his valuable help and suggestions during this period. I am also grateful to Dr. Abbas Ahmed for his kind support in my research.

I am grateful to the University of Huddersfield for supporting my PhD studies through Vice-chancellor studentship award. I also thank the school of computing and engineering for providing me all facilities for research.

Last but not least, I want to acknowledge the support of my family. I am deeply grateful to my parents who instilled in me a love of learning. Without their support, I would not have been able to continue my research. They always encouraged me and cheered me up whenever I was struggling with my mental health. Their support, both financial and emotional, has been invaluable to me and I am forever grateful.

Abstract

The current surge in wireless connectivity, anticipated to amplify significantly in future wireless technologies, brings a new wave of users. Given the impracticality of an endlessly expanding bandwidth, there's a pressing need for communication techniques that efficiently serve this burgeoning user base with limited resources.

Multiple Access (MA) techniques, notably Orthogonal Multiple Access (OMA), have long addressed bandwidth constraints. However, with escalating user numbers, OMA's orthogonality becomes limiting for emerging wireless technologies. Non-Orthogonal Multiple Access (NOMA), employing superposition coding, serves more users within the same bandwidth as OMA by allocating different power levels to users whose signals can then be detected using the gap between them, thus offering superior spectral efficiency and massive connectivity.

This thesis examines the integration of NOMA techniques with cooperative relaying, EXtrinsic Information Transfer (EXIT) chart analysis, and deep learning for enhancing 6G and beyond communication systems. The adopted methodology aims to optimize the systems' performance, spanning from bit-error rate (BER) versus signal to noise ratio (SNR) to overall system efficiency and data rates.

The primary focus of this thesis is the investigation of the integration of NOMA with cooperative relaying, EXIT chart analysis, and deep learning techniques. In the cooperative relaying context, NOMA notably improved diversity gains, thereby proving the superiority of combining NOMA with cooperative relaying over just NOMA. With EXIT chart analysis, NOMA achieved low BER at mid-range SNR as well as achieved optimal user fairness in the power allocation stage. Additionally, employing a trained neural network enhanced signal detection for NOMA in the deep learning scenario, thereby producing a simpler signal detection for NOMA which addresses NOMAs' complex receiver problem.

Contents

Copyright Statement **i**

Acknowledgements **ii**

Abstract **iii**

List of Figures **vi**

List of Abbreviations **x**

List of Symbols **xiv**

1 Introduction **1**

1.1 Introduction 1

1.2 Motivation for Research 3

1.3 Aims and Objectives 5

1.4 Novel Contributions 5

1.5 Thesis Structure 7

1.6 Summary 7

2 Background and Related Works **9**

2.1 Introduction 9

2.2 Requirements of the Next Generation of Wireless Communication Networks . . . 9

2.3 Motivation for Multiple Access (MA) 10

2.4 MA techniques: Past and Present 11

2.5 NOMA: Principle, Model, and Performance 14

2.6 Cooperative NOMA 18

2.7 EXIT Chart Analysis . . . 28
2.7.1 The Turbo Principle . . . 30
2.7.2 Mutual Information (MI) and EXIT Charts . . . 32
2.7.3 The EXIT Chart . . . 33
2.7.4 EXIT Chart Applications . . . 34
2.8 Deep Learning . . . 37
2.8.1 Background . . . 37
2.8.2 Deep learning in NOMA Systems . . . 45
2.8.3 Application of Deep Learning in Power Allocation . . . 47
2.8.4 User Clustering Using Deep Learning . . . 49
2.8.5 Signal Detection Using Deep Learning . . . 50
2.9 Methodology for Research . . . 62
2.9.1 NOMA in Cooperative Relaying Wireless Communication Systems . . . 62
2.9.2 NOMA with EXIT Chart Analysis . . . 63
2.9.3 NOMA with Deep Learning Techniques . . . 63
2.10 Summary . . . 65

3 Cooperative Non-Orthogonal Multiple Access for 5G Networks and Beyond 66
3.0.1 Introduction . . . 66
3.0.2 System Model . . . 67
3.0.3 Performance Analysis . . . 73
3.0.4 Simulation Results . . . 77
3.1 Summary . . . 83

4 EXIT Chart Analysis in Cooperative NOMA 85
4.1 Introduction . . . 85
4.2 System Model . . . 86
4.3 Performance Analysis . . . 86
4.3.1 Achievable Rate Analysis . . . 87
4.3.2 System Throughput Analysis . . . 87
4.4 Simulation Results . . . 88
4.5 Summary . . . 93

5 Enhancing the Performance of NOMA Systems Using Deep Learning Techniques 95
5.1 Introduction . . . 95
5.2 Signal Detection in a Downlink NOMA system using Deep Learning . . . 96
5.2.1 Introduction . . . 96
5.2.2 System Model . . . 96
5.2.3 Simulation Results . . . 118
5.3 Summary . . . 122

6 Conclusion and Future Works 124
6.1 Answers to Research Questions . . . 124
6.2 How the Aims and Objectives were Met . . . 125
6.3 Contribution of research . . . 126
6.3.1 Cooperative Relaying with NOMA . . . 126
6.3.2 EXIT Chart Analysis in Cooperative NOMA . . . 127
6.3.3 Deep Learning for Signal Detection in NOMA . . . 128
6.4 Future Works . . . 129
6.4.1 NOMA with Cooperative Relaying . . . 129
6.4.2 NOMA with EXIT Chart Analysis . . . 129
6.4.3 NOMA with Deep Learning Techniques . . . 130

A Appendix 144

List of Figures

2.1 Time Division Multiple Access [6] 11
2.2 Frequency Division Multiple Access [6] 13
2.3 TDMA/FDMA hybrid System 13
2.4 Code Division Multiple Access Signal Spreading [6] 14
2.5 Basic Downlink NOMA 15
2.6 Direct Phase [28] 19
2.7 Co-operation Phase [28] 19
2.8 System model for relay selection [30] 20
2.9 Successive Cooperative NOMA Relaying System Model [29] 22
2.10 Coordinated direct and relay transmission system model [30] 23
2.11 Cooperative NOMA with SWIPT 23
2.12 Cooperative NOMA for load balancing and user scheduling 24
2.13 Cooperative NOMA for a mobile user 25
2.14 Cooperative NOMA for MIMO scenario 27
2.15 EXIT chart analysis block diagram example 28
2.16 Example of an EXIT chart 29
2.17 Serial concatenated system with iterative detection/decoding [43] 31
2.18 Turbo decoder including inner and outer decoders 31
2.19 Coded system with convolutional codes as outer code and QAM transmission as inner code with BICM [43] 35
2.20 Multipath transmission as inner code as well as 7 convolutional codes behaving as outer codes EXIT chart [43] 36
2.21 Irregular code construction [43] 36

2.22 A network with many convolutional layers. Filters are applied to each training image at different resolutions, and the output of each convolved image serves as the input to the next layer. [89] 39
2.23 Rudimentary example of a neural network as it is used in deep learning systems [62] 41
2.24 Simple block diagram for RL [77] 44
2.25 Illustrating Multi-cell NOMA uplink resource allocation by using optimisation algorithm to efficiently cluster users for each resource block at the base-station side[102] 50
2.26 System model for user clustering in a MIMO-NOMA system.[38] 53

3.1 Proposed cooperative NOMA system 67
3.2 The cooperative relaying NOMA system comprises two distinct phases, namely Phase-I and Phase-II. 70
3.3 Total transmission time for the proposed cooperative relaying NOMA system for a range of SNR values 78
3.4 BER Vs SNR for the proposed system comprising of the performance of the NUs and the FU with and without cooperative relaying 79
3.5 Allocation of power within the NOMA framework and its correlation with decoded data speeds in both cooperative and traditional NOMA setups 80
3.6 The association between power distribution and decoded data speeds when both the NU and FU receive equivalent power allocations 81
3.7 The NOMA system's EXIT chart, the outer decoder employs an RSC code with a SF of 2, while the inner decoder is adapted for multiple SNR values within the MUD framework. 82
3.8 EXIT chart analysis using IRCC codes for the proposed NOMA system 83
3.9 BER Vs SNR for the MUD system, where various number of MUD DES/DEC iterations are utilised when SF=2 84

4.1 Comparison of BER performance between near and far users 89
4.2 Near and far users' data rates 90
4.3 System throughput for proposed NOMA system 91

4.4 EXIT diagram for a NOMA system with the outer decoder utilizing RSC code at an SF=2, while the inner decoder functions across diverse SNR values within the MUD framework 92
4.5 EXIT chart for the proposed NOMA system 93
4.6 Normalised throughput analysis for MUD model 94

5.1 General diagram for the proposed deep learning model 97
5.2 BER Vs SNR graph for 2 users using different signal detection techniques 119
5.3 Training neural network in terms of system accuracy and data loss 120
5.4 Training neural network in terms of system accuracy and data loss for 30 epochs 121

A.1 Flowchart for the power allocation algorithm (allocatePower) for the proposed system 144
A.2 Flowchart for the CE algorithm for the proposed system 145
A.3 Flowchart for the data transmission and reception algorithm for the proposed system 146
A.4 Flowchart for the signal detection algorithm for the proposed system 147
A.5 Flowchart for acquiring the feature and label algorithm for the proposed system 148
A.6 Flowchart for decoding transmitted symbols for the proposed system 149
A.7 Flowchart for decoding transmitted symbols using SIC for the proposed system . 150
A.8 Flowchart for Testing the transmitted data for the proposed system 151
A.9 Flowchart for training the transmitted data of the proposed system 152
A.10 Flowchart for training the neural network of the proposed system 153

List of Abbreviations

AUD **A**ctive **U**ser **D**etection

AWGN **A**dditive **W**hite **G**aussian **N**oise

AI **A**rtificial **I**ntelligence

BS **B**ase **S**tation

BN **B**atch-**N**ormalisation

BP **B**elief **P**ropagation

BER **B**it **E**rror **R**ate

BOMP **B**lock **O**rthogonal **M**atching **P**ursuit

CE **C**hannel **E**stimation

CSI **C**hannel **S**tate **I**nformation

CDMA **C**ode **D**ivision **M**ultiple **A**ccess

CNN **C**onvolutional **N**eural **N**etwork

CRS **C**ooperative **R**elaying **S**ystem

CF **C**ost **F**unction

CP **C**yclic **P**refix

DEC **D**ecoding

DL **D**eep **L**earning

DNN **D**eep **N**eural **N**etwork

DRL **D**eep **R**einforcement **L**earning

DES **D**espreading

EXIT **EX**trinsic **I**nformation **T**ransfer

FU **F**ar **U**ser

5G **F**ifth **G**eneration

4G **F**ourth **G**eneration

FDMA **F**requency **D**ivision **M**ultiple **A**ccess

HD **H**alf **D**uplex

i.i.d **i**ndependant and **i**dentically **d**istributed

IoT **I**nternet **o**f **T**hings

IRCC **IR**regular **C**onvolution **C**oding

LoS **L**ine **o**f **S**ight

LS **L**east **S**quare

LLR **L**og **L**ikelihood **R**atio

LSTM **L**ong **S**hort **T**erm **M**emory

LDS **L**ow **D**ensity **S**preading

MDP **M**arkov **D**ecision **P**rocess

MRC **M**aximal **R**atio **C**ombining

MAP **M**aximum **A** **P**osteriori

MaxMSE **M**aximum **M**ean **S**quare **E**rror

ML **M**aximum **L**ikelihood

MMSE **M**inimum **M**ean **S**quare **E**rror

MSE **M**ean **S**quare **E**rror

mmWave **m**illi**m**eter **W**ave

MA **M**ultiple **A**ccess

MAS **M**ultiple **A**cess **S**ignature

MIMO **M**ultiple **I**nput **M**ultiple **O**utput

MUD **M**ulti-**U**ser **D**etection

MUSA **M**ulti-**U**ser **S**hared **A**ccess

NU **N**ear **U**ser

NN **N**eural **N**etwork

NOMA **N**on-**O**rthogonal **M**ultiple **A**ccess

OFDMA **O**rthogonal **F**requency **D**ivision **M**ultiple **A**ccess

OFDM **O**rthogonal **F**requency **D**ivision **M**ultiplexing

OMA **O**rthogonal **M**ultiple **A**ccess

PCA **P**ilot **C**ontamination **A**ttack

PPP **P**oisson **P**oint **P**rocess

PD **P**ower **D**omain

QAM **Q**uadrature **A**mplitude **M**odulation

QPSK **Q**uadrature **P**hase **S**hift **K**eying

QoS **Q**uality **o**f **S**ervice

ReLU **Re**ctified **L**inear **U**nit

RNN **R**ecurrent **N**eural **N**etwork

RSC **R**ecursive **S**ystematic **C**onvolution

RL **R**einforcement **L**earning

SNR **S**ignal to **N**oise **R**atio

SWIPT **S**imultaneous **W**ireless **I**nformation and **P**ower **T**ransfer

SISO **S**ingle **I**nput **S**ingle **O**utput

6G **S**ixth **G**eneration

SCMA **S**parse **C**ode **M**ultiple **A**ccess

SF **S**preading **F**actor

SIC **S**uccessive **I**nterference **C**ancellation

SC **S**uperposition **C**oding

SER **S**ymbol **E**rror **R**ate

3D **T**hree **D**imensional

TDMA **T**ime **D**ivision **M**ultiple **A**ccess

UAV **U**nmanned **A**erial **V**ehicle

UE **U**ser **E**quipment

VR **V**irtual **R**eality

VLC **V**isible **L**ight **C**ommunications

List of Symbols

E_s/N_0	Symbol Energy-to-Noise Ratio
E_b/N_0	Energy per Bit-to-Noise Ratio
R_t	Reward (Reinforcement Learning)
S_t	State (Reinforcement Learning)
P_{FU}	Power Allocated to the Far User
$h_{n,1}$	Channel Coefficient for Rayleigh Fading Channel during Phase-I
$h_{n,2}$	Channel Coefficient for Rayleigh Fading Channel during Phase-II
$h_{FU,1}$	Channel Coefficient for Rayleigh Fading Channel for Far User
$w_{n,1}$	Added White Gaussian Noise for the First Phase of Transmission
x_{FU}	Far User Signal
x	Superposed Signal
x_N	Signal of n^{th} User
S	System Throughput
R_n	Data Rate of the n^{th} User
$L_{ex}(i^n)$	Extrinsic Log-Likelihood Ratio (LLR) of the n^{th} User's Multi-User Detector (MUD) Output
$L_{apr}(i^n)$	Apriori LLR of the n^{th} User's MUD Output
$P(r)$	Apriori Symbol
r	Codeword for Extrinsic Information Transfer (EXIT)
$P(v \mid r)$	Channel Probability with Respect to the Apriori Codeword
P	Total Allocated power
H	Channel Statuses of all Network Users
f_r	Cost Function of the MUD
η_{DG}	Diversity Gain
η_{IG}	System Diversity Gain
B	Number of Bits per MUD
R	Coding Rate
C	System Capacity
W	Noise Matrices
N^{MAP}	Cost Function of the Maximum APoseteriori (MAP)
SF	Spreading Factor

b^n	Number of Bits per Codeword in MUD for the n^{th} User
$O\left(\sqrt{N^{MAP}_{(CFEs/bits)}}\right)$	Hard Input Hard Output for the MUD
α_1	Power Splitting Ratio
$L^d_e(c)$	A Priori Information When QAM Mapping is Selected
$L^d_a(c)$	Extrinsic Information When QAM Mapping is Selected
C	Interleaved Encoded Data
C_k	Data After Passing Through The Serial-to-Parallel Converter
s_k	Modulated Information
r_k	Information After Passing Through the Channel
σ	Signal Variance
E_s	Energy Per Symbol
T	Symbol Duration
$E_s(1/T)$	Signal Power
N_0	Noise Spectral Density
$f(u)$ and $f(v)$	Marginal Probability Density Functions
$f(u,v)$	Joint Probability Density Function
P_u	Transmit Power
P_n	Noise Power
$I(U;V)$	Shannon Mutual Information
$H(V)$	Entropy of V
$H\left(V \mid U\right)$	Conditional Entropy of V Given U
R_k	Coding Rate

Chapter 1

Introduction

1.1 Introduction

With the ongoing deployment of Internet of Things (IoT) systems and ever-increasing use of wireless connectivity by billions of user around the world, the next generation of wireless communication systems will not be able to cope up with the demands for much higher data rates and ultra-reliable connectivity with extremely-low latency [1,2]. To overcome these issues, innovative solutions are essential. Non-Orthogonal Multiple Access (NOMA) has recently been proposed as a promising technology for the next generation of wireless communication systems due to its superior spectral efficiency in comparison to conventional Multiple Access (MA) techniques [1]-[3].

The NOMA technique facilitates the coexistence of multiple users within the same code, frequencies, and time domains through the allocation of distinct power levels to each user. The resulting gap is subsequently utilised for signal detection [4]. Furthermore, it has been observed that users who experience superior channel conditions are assigned a reduced amount of transmission power. This approach serves to preserve transmission power for users located at the cell-edge or in close proximity to it. Conversely, users who encounter weaker channel conditions are allocated a greater amount of transmission power [5]

The key features of NOMA include:

Superposition Coding: Unlike traditional schemes where users occupy orthogonal resources, in NOMA, multiple users can share the same frequency/time resource. Their signals are superimposed in the power domain, with different power levels assigned based on their respective channel conditions.

Spectral Efficiency: Due to superposition coding, NOMA can serve multiple users in the same resource block, leading to a higher spectral efficiency compared to Orthogonal Multiple Access (OMA) techniques.

Massive Connectivity: NOMA is suitable for scenarios with a large number of devices, such as IoT, because it can accommodate more users simultaneously.

Complex Decoding at the Receiver: With the use of Successive Interference Cancellation (SIC) at the receiver end, NOMA allows users with better channel conditions to subtract out the signals intended for other users, thereby decoding their own intended signal.

Flexible User Pairing: In the downlink, users can be paired based on their channel conditions to optimize the system's performance.

Enhanced Throughput: By accommodating more users in a given resource block, NOMA can potentially increase the overall system throughput.

Robustness in Varied Environments: NOMA can work in diverse scenarios, from broadband multimedia services to narrow-band IoT applications.

Integration with Other Technologies: NOMA can be integrated with other technologies such as MIMO (Multiple Input Multiple Output) and mmWave to further boost its performance.

Dynamic Power Allocation: Depending on the requirements and channel conditions, the power allocated to different users in NOMA can be dynamically adjusted to optimize performance.

While NOMA offers several advantages, it's worth noting that its practical implementation requires sophisticated signal processing techniques and might pose challenges, especially in a highly dynamic environment. However, its potential benefits make it an attractive option for next-generation wireless networks.

The NOMA approach is fundamentally grounded in the notion of non-orthogonal resource allocation, which aims to facilitate network user support. However, this approach is not without its drawbacks, as it can lead to heightened receiver complexity that scales in direct proportion to the number of users employing SIC within the system, as well as increased inter-user interference. The latest NOMA solutions can be classified into two distinct categories, namely power-domain NOMA and code-domain NOMA. Prominent NOMA techniques comprise Sparse Code Multiple Access (SCMA) [8, 9], Low Density Spreading (LDS) [10], MultiUser Shared Access (MUSA) [11], Spatial Division Multiple Access [12], and Successive Interference Cancellation

Amenable Multiple Access [13].

Literature such as [13], [14], and [15] provide detailed insights into the efficacy of NOMA as opposed to conventional MA techniques such as Orthogonal Frequency Division Multiple Access (OFDMA). The utilisation of a NOMA system model, despite its high spectral efficiency, may lead to a reduction in the performance of individual users due to the prioritisation of users with weaker channel conditions over those with stronger channel conditions [15].

Cooperative relaying, with its natural increase in diversity gains, EXtrinsic Information Transfer (EXIT) chart analysis, with its enhanced user fairness during allocation scheme, and deep learning models that aid in signal detection can be used to minimise these concerns. All of these factors result in improved user fairness and receiver dependability for users with better channel settings [14], utilising NOMA's propensity to be paired with other transmission systems.

1.2 Motivation for Research

The future of wireless communications (fifth generation (5G) and beyond) needs three critical attributes: low latency, ultra fast speeds, and massive connectivity. NOMA manages to not only satisfy these three demands but also massively improve the overall spectral efficiency of the network. This makes NOMA a highly sought after technology for the future of wireless communications.

According to the requirements of both 5G and the coming sixth generation (6G), the motivation behind adopting NOMA can be critically understood.

The present deployment of 5G mobile networks exhibits a speed that is 100 times faster than fourth-generation (4G) networks. The implementation of 5G networks is expected to offer data transmission rates reaching 10 Gbps, reduced latency measured in milliseconds, and enhanced reliability. A high-definition film can be downloaded within a matter of seconds. This technology has the capability to provide support for a wide range of IoT devices as well as intelligent automobiles. In order to address the continuous requirements of 5G, it is imperative to possess a wireless access technology that is efficient and capable of enhancing throughput without necessitating an increase in bandwidth.

Currently, 6G mobile networks are in the developmental phase and are expected to offer transmission speeds in the terabit level. The implementation of this technology would need the integration of an advanced antenna system, a substantial increase in the memory capacity of mobile devices, and the establishment of extensive optical networks. The forthcoming 6G networks are anticipated to be cell-free, hence enabling the integration of Artificial Intelligence (AI) within wireless networks. The frequency range to be utilised by 6G networks is currently subject to uncertainty. However, it is apparent that a substantially higher frequency band will be necessary in order to fulfil the data rate requirements of 6G networks.

The utilisation of frequencies exceeding 30 GHz and reaching up to 300 GHz, commonly referred to as millimetre waves, characterises 5G technology. Conversely, 6G is linked to substantially elevated frequencies within the terahertz (THz) bands, ranging from 300 GHz to 3 THz. It is anticipated that within the forthcoming 5-7 year period, the THz spectrum will be employed for the purpose of 6G communications. In the realm of 6G network technology, potential applications encompass interconnected robotic and autonomous systems, wireless brain-computer interface mechanisms, advancements in blockchain methodologies, immersive multisensory extended realities, interstellar expeditions, deep-oceanic exploration ventures, tactile internet innovations, and the industrial internet paradigm.

The challenges facing the sixth generation of wireless communication technologies are very demanding. NOMA is suggested as a prime candidate to address some of these demands. Especially demands such as spectral efficiency and connection density can be met by adopting NOMA. Taking advantage of NOMA's proclivity to being combined with other techniques, such as cooperative relaying, EXIT charts, and deep learning, the rest of these demands can be met quite vehemently.

The following are some **key gaps in NOMA**:

- While the NOMA principle is designed to serve cell edge users or users with high QoS requirements, there is room for improvement in regards to BER performance and massive connectivity. This improvement can be achieved by combining NOMA with other communication techniques such as cooperative relaying.

- NOMA is designed to prioritize users with poor channel conditions and, as such, there can exist a lack of user fairness in the power allocation stage. EXIT charts can be used to evaluate and improve user fairness in a NOMA system.

- NOMA provides higher BER performance and connection density when compared to the conventional OMA approach. This improvement in performance comes at the price of higher receiver complexity since SIC is required to receive the signals of user with lower power levels. Deep learning can applied here in order to simplify and streamline the signal detection process.

1.3 Aims and Objectives

The main aim of this research is to analyse NOMA techniques for next generation communications systems by applying cooperative relaying, EXIT chart analysis and deep learning to achieve improved performance. The key research objectives are listed as follows:

- Maximising diversity gains for a proposed cooperative NOMA system through exploitation of key NOMA features. The aim is to prove that combining NOMA with cooperative relaying yields better BER performance versus deploying NOMA by itself.
- Investigating the effects of employing EXIT chart analysis on a cooperative NOMA system. The aim is to prove the efficacy of applying EXIT chart analysis on determining the user fairness of the proposed NOMA system as well producing the exact SNR value that yields an infinitesimally small BER.
- Applying deep learning techniques in order to enhance the signal detection of a NOMA system. The aim is to provide a deep learning-aided alternative to normal signal detection in NOMA that is streamlined and simpler to implement.

1.4 Novel Contributions

The main contributions of this thesis are discussed as follows:

1. First contribution (Chapter 3): NOMA techniques were applied to a cooperative relaying system and the results show a significant improvement in overall system performance in regards to BER vs SNR. This shows that it is better to combine NOMA with cooperative relaying than deploying NOMA by itself in terms of BER performance

   This contribution has led to the following publications:

A. Ahmed, Z. Elsaraf, F. A. Khan and Q. Z. Ahmed, "Cooperative Non-Orthogonal Multiple Access for Beyond 5G Networks," in IEEE Open Journal of the Communications Society, vol. 2, pp. 990-999, 2021.

2. Second contribution (Chapter 4): The effects of applying EXIT chart analysis on a cooperative NOMA system were investigated theoretically and practically. The results show a remarkable improvement in user fairness in power allocation as a result of using EXIT chart analysis.

   This contribution has led to the following publications:

   Elsaraf, Z., Ahmed, A., Khan, F.A., Q. Z. Ahmed, "Cooperative Non- Orthogonal Multiple Access for Wireless Communication Networks by Exploiting the EXIT Chart Analysis," J Wireless Com Network 2021, 79 (2021).

   Z. Elsaraf, A. Ahmed, F. A. Khan and Q. Z. Ahmed, "EXIT Chart Analysis of Cooperative Non-Orthogonal Multiple Access for Next Generation Wireless Communication Systems," 2020 European Conference on Networks and Communications (EuCNC), Dubrovnik, Croatia, pp. 281-285,2020.

3. Third contribution (Chapter 5): The application of deep learning in communication systems was investigated as well as its deployment in a NOMA system in regards to signal detection. The results show an improved performance in signal detection with less complexity than standard SIC in NOMA.

   This contribution has led to the following publications:

   Z. Elsaraf, F. A. Khan and Q. Z. Ahmed, "Deep Learning Based Power Allocation Schemes in NOMA Systems: A Review," 2021 26th International Conference on Automation and Computing (ICAC), Portsmouth, United Kingdom, pp. 1-6, 2021.

1.5 Thesis Structure

Following the main aims and objectives, the thesis is structured as follows:

- Chapter 1 introduces the research as well as outlining the motivation for researching NOMA. The research objectives are listed here as well.
- Chapter 2 presents the literature review for cooperative NOMA, EXIT chart analysis, and deep learning in NOMA. The background of each area is vividly detailed and the challenges facing each area are defined.
- Chapter 3 focuses on cooperative NOMA and the new research that has been carried out in this area.
- Chapter 4 describes EXIT chart analysis and its use with NOMA as well as the research that has been carried out in this area.
- Chapter 5 is entirely concerned with deep learning in NOMA. It begins with a brief background about deep learning in general before hyper-focusing on its application in NOMA and the research that was undertaken in this area.
- Chapter 6 concludes the thesis by first summarising the research and objectives met as well as the methodology adopted to meet said objectives. Then the overall contributions the research has made are outlined. The future works are then outlined in concise yet vivid detail.

1.6 Summary

The primary objective of this thesis is to address some of the challenges facing the next generation of wireless communications by utilising NOMA and its features, such as SIC, power allocation, and signal detection, to serve cell edge users and users with a higher quality of service.

NOMA has been shown to outperform other conventional MA techniques, such as the widely adopted OMA, due to its higher spectral efficiency as well as its ability to achieve massive connectivity while effectively serving users with limited resources. As a result of having such features, NOMA has been poised as a reliable approach for future wireless communication systems to adopt in order to meet their stringent demands.

However, NOMA has some flaws, including a greater receiver complexity for SIC, a high IUI, and poor user fairness for users with a higher CSI. Incorporating additional approaches, such as cooperative relaying, EXIT chart analysis, and deep learning for signal detection into NOMA can address these issues.

Chapter 2

Background and Related Works

2.1 Introduction

The trajectory of our society seems to be steering towards one dominated by fully automated and remotely operated systems. The swift progression of nascent technologies like AI, Virtual Reality (VR), 3D media, and the Internet of Things (IoT) has led to a significant surge in data traffic [1]. Autonomous systems are gaining traction across various sectors - from industry to health, transportation, maritime, and even space exploration. The anticipation is that urban areas, vehicles, residences, industrial sites, food items, toys, and more will be embedded with millions of sensors to enable an intelligent, automated way of life. To bring these advanced applications to fruition, there's a pressing need for greater spectral efficiency coupled with ultra-low latency and energy efficiency. Studies have indicated that NOMA can deliver these performance metrics without additional transmission resource demands. Consequently, NOMA emerges as a promising solution for the forthcoming wave of wireless communication innovations.

2.2 Requirements of the Next Generation of Wireless Communication Networks

To accommodate the requirements of these advanced applications, both a high data transfer rate and reliable connectivity are essential. 5G networks, however, fall short in delivering a fully automated, intelligent system and a comprehensive immersive experience for users utilizing VR technologies [3]. Even though 5G communication infrastructures present substantial improvements over their antecedents, they might not be adequate to support the anticipated intelligent and automation systems a decade hence [4]. This is in spite of 5G networks boasting heightened capabilities and an elevated Quality of Service (QoS) when juxtaposed with 4G technologies [5]-[8]. The innovations within 5G encompass several novel strategies, which include the introduction of new frequency bands like millimetre wave (mmWave) and optical spectra, refined spectrum allocation and oversight, and the amalgamation of both licensed and unlicensed frequency bands [4].

However, the swift proliferation of data-driven and automated technologies could potentially eclipse the potential of 5G wireless infrastructures. Certain devices, such as VR systems, necessitate transitioning beyond 5G due to their requirement for data rates not less than 10 Gbps [1]. Consequently, as the boundaries of 5G are projected to be reached by 2030, research pursuits are already exploring the architectural objectives for the subsequent version. To surmount the constraints of 5G and address emerging challenges, the formulation of a 6G wireless architecture endowed with groundbreaking characteristics becomes imperative. The foundational pillars of 6G will encompass an amalgamation of attributes observed in previous iterations, such as heightened network density, substantial throughput, elevated reliability, energy efficiency, and extensive interconnectivity. Moreover, 6G is anticipated to perpetuate the trends of its forerunners by ushering in innovative services and technologies. These novel services will span AI, smart wearable gadgets, biomedical implants, self-driving vehicles, computational reality tools, sensory mechanisms, and 3D spatial modeling [9]. Among the paramount prerequisites for 6G wireless infrastructures is their ability to handle vast data quantities and exceedingly high device-specific data speeds [1].

2.3 Motivation for Multiple Access (MA)

Multiple access was proposed to enhance the 5G capabilities of Spectrum efficiency, Connectivity density, and Peak data rate. In essence, MA provides concurrent access to the same communication resources by several users. To achieve multiple access, numerous dimensions, including time, frequency, and code, were examined. Hence, Time Division Multiple Access (TDMA), Frequency Division Multiple Access (FDMA), TDMA/FDMA hybrid, and Code Division Multiple Access (CDMA) were developed, in that order [5].

Conventional MA techniques such as OMA experience a limit on the number of users they can support due to the orthogonality of their approach. This means that OMA systems cannot support a higher number of users without requiring more bandwidth.

NOMA, on the other hand, due to the non-orthogonality of its approach, is able to support double the number of users while utilising the same allocated bandwidth. This very desirable trait makes NOMA a highly sought after technique for efficient MA.

2.4 MA techniques: Past and Present

Time Division Multiple Access

TDMA, as illustrated in Fig.2.1, makes use of the time domain to facilitate multiple access. Users in a TDMA system are allotted distinct time slots or time intervals during which they may transmit signals. Other network users are expected to wait until their corresponding transmission time slot before transmitting. This method permits the division of a single channel into many time slots, allowing a greater number of users to transmit per channel. However, the reception of signals in a TDMA system is time-based and, as a result, is extremely error-sensitive; a minor delay at one user could result in the propagation of error to every subsequent user [6].

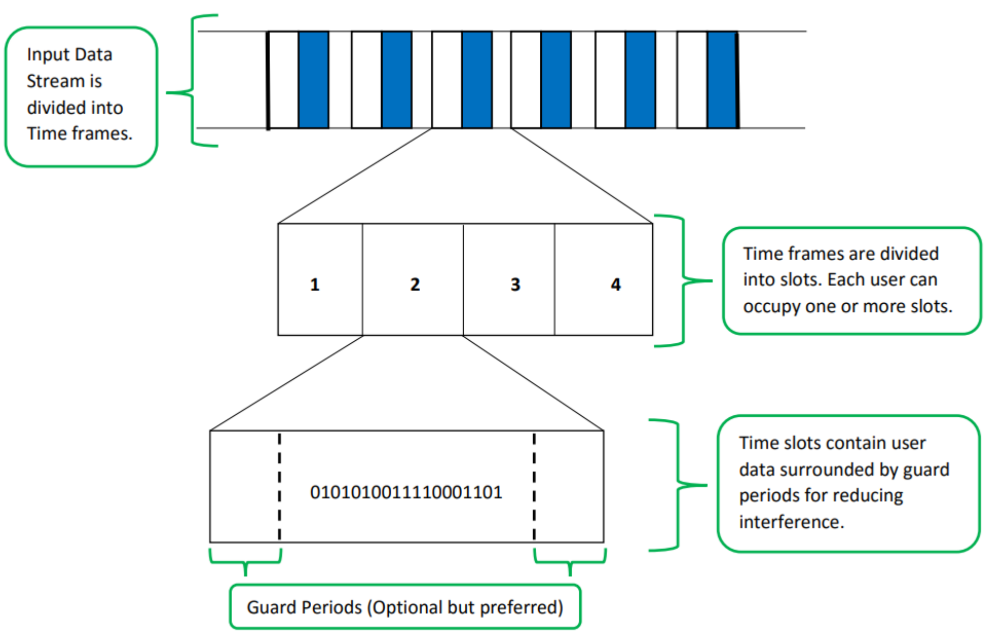


Figure 2.1: Time Division Multiple Access [6]

Frequency Division Multiple Access

As shown in Fig.2.2, using the frequency domain, FDMA achieves multiple access by allocating each user a distinct transmission frequency band. By assigning each user a specific frequency band, each user's signal becomes distinct enough to be easily identified by a receiver with low complexity. Whereas FDMA would struggle to handle numerous users in a frequency-limited system, the multiple access strategy is able to provide reliable and easy communications since only one user can use a frequency band at any given moment.

Nonetheless, FDMA is extremely sensitive to phase shift delays, which can cause adjacent channel interference, as shown in Fig.2.2. Adjacent channel interference occurs when a change in the phase shift of a signal, even a small one, causes two or more signals from adjacent bands to overlap, which makes it very difficult to extract the information from the signal using a low complexity receiver. Adjacent channel interference can occur as a result of propagation delay or even simple noise interference. Propagation delay refers to the time it takes for a signal to travel from the sender to the receiver over a wireless medium and noise interference refers to unwanted random signals or disturbances that interfere with the desired signal in a communication system.

A strategy for combating this issue is to leave a portion of the available bandwidth empty so that there is no information to overlap in the event of a phase shift. The unoccupied frequency bands are known as guard bands. This strategy, however, reduces spectral efficiency because a significant portion of the available bandwidth is underused.

TDMA/FDMA Hybrid

As Fig.2.3 illustrates, combining the mechanics of FDMA and TDMA, a system was constructed using both approaches. This multiple access technique, termed "Hybrid FDMA/TDMA," utilises both the time and frequency domains to enable multiple access to numerous users. The method is based on the premise that users within the same frequency range might occupy distinct time slots. This permits the allocation of the same carrier frequency to several users, provided that these users transmit at different times to prevent interference.

This hybrid scheme allows for a network to accommodate a much larger number of users when compared to FDMA and TDMA selectively but with the added cost of requiring a more complex receiver to function.

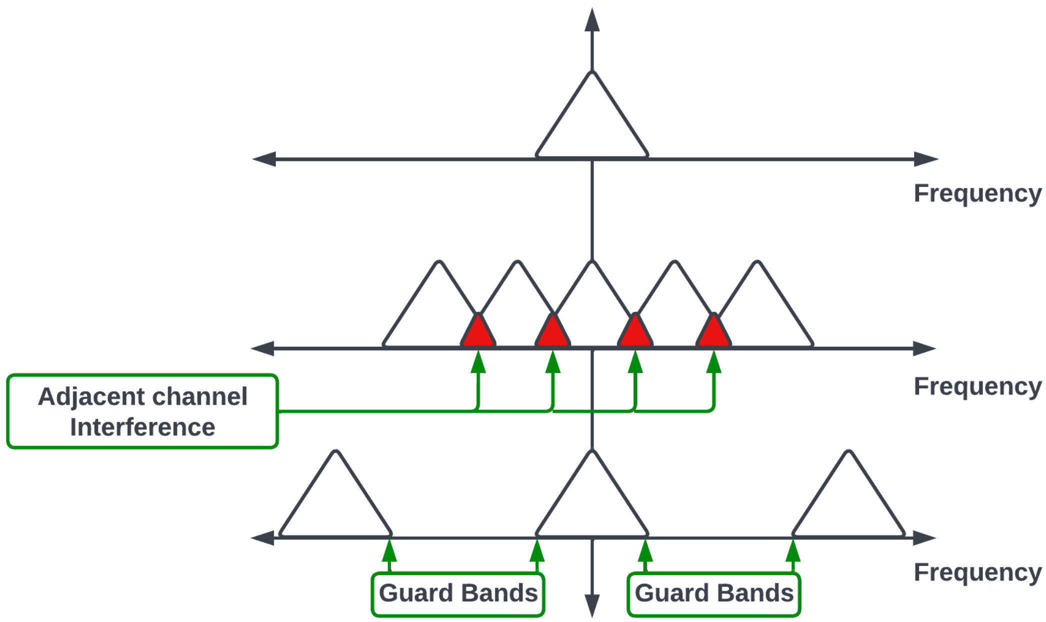


Figure 2.2: Frequency Division Multiple Access [6]

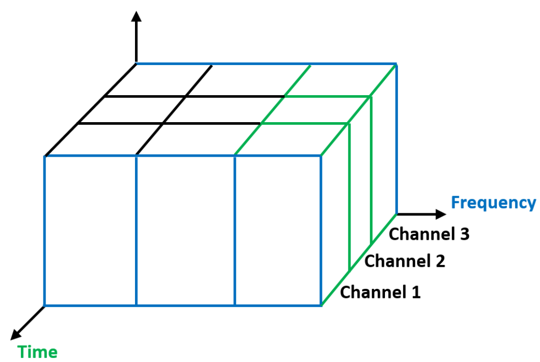


Figure 2.3: TDMA/FDMA hybrid System

Code Division Multiple Access

CDMA, as shown in Fig.2.4, is a technology for multiple access that leverages spread spectrum signalling. CDMA uses a set of specially created pseudo-random code words to spread the user signal, allowing it to consume more bandwidth than is required. CDMA permits the multiplexing of numerous users on the same physical channel by utilising the code domain, as opposed to FDMA and TDMA, which disperse their users throughout the frequency domain and the time domain, respectively. Signal identification in a CDMA-based system is simple since each codeword used to spread user signals is unique and can thus be used to de-spread those signals at the receiver. [9]-[14]

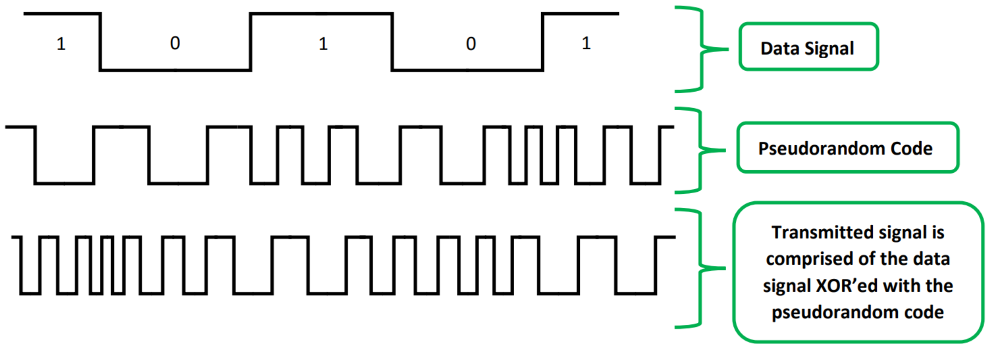


Figure 2.4: Code Division Multiple Access Signal Spreading [6]

In order to increase spectral efficiency, CDMA can also enable users to overload the system. A system is termed overloaded when multiple users are permitted to utilise the same spreading code to propagate their message. Theoretically, this strategy is capable of enhancing the spectrum efficiency of a network, but at the expense of an increase in inter-user interference, as user signals are given more opportunities to overlap. [6]-[15]

2.5 NOMA: Principle, Model, and Performance

NOMA, shown in Fig.2.5, is a type of superposition coding that permits several users to occupy a single sub-carrier by assigning them various power levels based on the channel conditions of each user. The majority of the available transmit power is allocated to users with worse channel conditions, i.e. cell-edge users, while the remainder is allocated to customers with better channel conditions. Due to its use of superposition, NOMA is characterised to as a type of multi-user superposition transmission in [16].

According to the model presented in [5], consider a downlink scenario with n users n_i, and channel gain, h_i, where $i = 1, 2$, and one base station. Assuming $h_1 <= h_2$ and $a_i = [1, 2]$ as $a_1^2 + a_2^2 = 1$ for power allocation coefficients. After performing all encoding and modulation, the Base Station (BS) would superpose both user signals and concurrently transmit the composite signal through each channel. As $a_1 > a_2$, User 1 can decode its own signal by considering the other signal as noise, for instance by employing Maximal Ratio Combining (MRC), which weighs each received signal copy according to its channel gain and then combines them. By doing so, it coherently adds the signals in phase, thereby maximizing the signal-to-noise ratio (SNR) for the weaker user's signal. User 2 must decode User 1's message before subtracting it from the received message to decode its own message. A procedure also referred to as Successive Interference Cancellation or SIC, which is explored in detail in [17] and [18].

NOMA can work with both single antenna and multiple antenna scenarios. In a basic Single-Input Single-Output (SISO) scenario, NOMA typically uses power domain separation to serve multiple users on the same frequency and time resource. SIC is applied at the receiver. The user with the stronger channel gain (typically the near user) is decoded first, and its signal is subtracted from the combined received signal, thereby allowing the weaker user's signal (typically the far user) to be decoded next.

MIMO-NOMA can leverage spatial domain along with the power domain or code domain for user multiplexing. The spatial degrees of freedom offered by multiple antennas can be used to serve more users simultaneously or to enhance the reliability and robustness of user detection. Techniques like beamforming can be integrated with NOMA to direct signals toward specific users, thereby increasing the efficiency of user multiplexing. MRC and other techniques can be employed to improve the decoding, especially in the presence of inter-user interference.

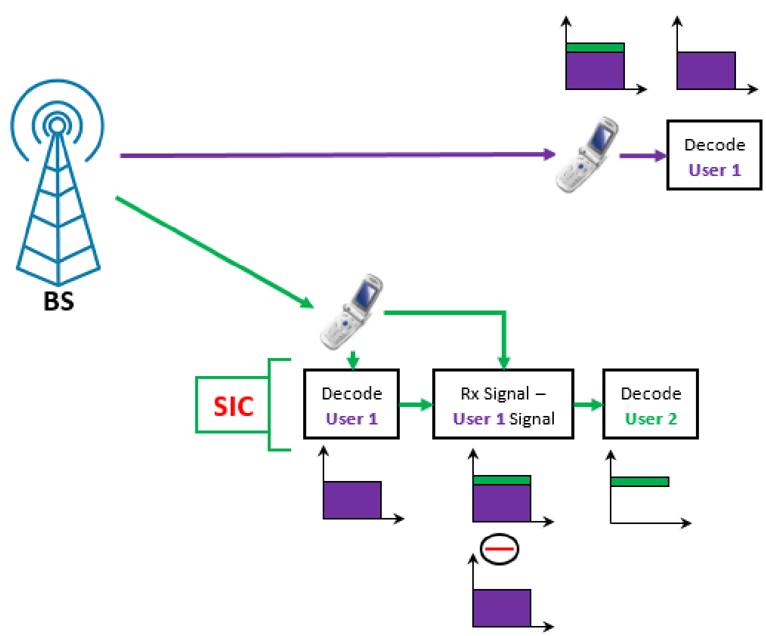


Figure 2.5: Basic Downlink NOMA

Studies such as [19]-[22] analyse and study the performance of NOMA in terms of its Outage, data rate, and bit-error performance. NOMA has improved data rates compared to the current industry standard OFDMA, as demonstrated in [23]. NOMA has much to offer to meet the requirements of beyond 5G technology as Orthogonal Frequency Division Multiplexing (OFDM) and its derivatives (Rel.15 CP-OFDM) are exceeded by NOMA in terms of spectral efficiency and data throughput.

[24] provides a clear mathematical demonstration why NOMA is superior to OMA. Outlining various facets of each technique and conducting a comprehensive assessment of basic Power Domain (PD) NOMA in comparison to Type I and Type II OMA in terms of the optimization difficulties associated with user fairness. Even with diverse user fairness case scenarios, [25] indicates that the minimum power required for NOMA is always less than that of OMA, and the total rate is always higher.

[25] examines NOMA from the perspective of user fairness. The research investigates NOMA under two conditions: Instantaneous and Averaged Channel State Information (CSI) at the transmitter, with maximum-minimum and minimum-maximum approaches, respectively. Utilizing conventional TDMA as a test standard, the results indicate that NOMA provides greater user fairness. The work alludes to the fact that such performance was obtained by careful power allocation.

The performance of NOMA with Internet of Things applications is investigated in [27], where the spectrum usage of NOMA is compared against the fixed spectrum allocation of OMA in a MIMO system optimised for tiny packet transmissions. Due to a lack of flexibility in spectrum resource distribution, the article concludes that NOMA provides superior performance benefits than OMA overall. NOMA also outperforms OMA in the low to mid SNR range. Adaptive resource allocation for OMA is investigated, and while it introduces dynamic changes to the features of orthogonal resource blocks, it may necessitate the use of time slots with extremely short durations, which may not be practical [28].

NOMA relies on superposition coding and SIC at the receiver to separate and decode the signals from different users. The central concept is around the allocation of power levels based on the strength of channel conditions. Users with less favourable channel conditions are assigned greater power levels, whereas users with more favourable channel conditions are assigned lower power levels. This allows for efficient spectrum utilization and improved system capacity.

NOMA can be used with different channel models, as long as the basic principles of superposition coding and SIC can be applied. Common channel models used in NOMA systems include:

1. **Additive White Gaussian Noise (AWGN) Channel:** In this channel model, noise is assumed to be Gaussian and additive to the received signal. NOMA can be applied in

AWGN channels by adjusting power levels and using appropriate codebooks for users.

2. **Rayleigh Fading Channel:** This channel model accounts for the effects of multipath propagation and can be modeled as a complex Gaussian random process. NOMA can also be applied in Rayleigh fading channels, but it may require more sophisticated techniques for power allocation and codebook design to account for the varying channel conditions.

3. **Frequency-Selective Fading Channel:** In this model, the channel response varies with frequency, leading to frequency-selective fading. NOMA can still be applied by considering the frequency-selective nature of the channel and adapting the superposition coding and SIC techniques accordingly.

4. **MIMO (Multiple-Input, Multiple-Output) Channel:** NOMA can be extended to MIMO channels, where multiple antennas are used at both the transmitter and receiver. In this case, NOMA can exploit spatial multiplexing to serve multiple users simultaneously.

NOMA does not rely on the use of the same codes for different users. Instead, it focuses on power and modulation level allocation to achieve multiple access in wireless communication systems. In NOMA, users are distinguished by their power levels and modulation schemes, and they typically use their own unique sets of codes, spreading sequences, or modulation schemes.

NOMA is designed to exploit the differences in channel conditions among users to maximize spectrum efficiency. Users who experience less favourable channel conditions are assigned greater power levels and potentially more resilient modulation schemes, whereas users who experience more favourable channel conditions are assigned lower power levels and less resilient modulation schemes.

In contrast to OFDMA or CDMA, where users are assigned orthogonal codes or subcarriers, NOMA allows multiple users to share the same time-frequency resources non-orthogonally by distinguishing them primarily through power and modulation. As a result, the codes themselves are not required to be orthogonal.

Codes in NOMA do not become non-orthogonal within the same user subset. Instead, the non-orthogonality arises from the simultaneous transmission of multiple users over the same time-frequency resources. The non-orthogonality is managed in the power domain, with different power levels assigned to each user based on their channel conditions, ensuring that their signals can coexist without significant interference.

In summary, NOMA does not use the same codes for different users and does not create non-orthogonality among codes within the same user subset. It achieves non-orthogonality by allocating different power levels and modulation schemes to users with different channel conditions, allowing them to share the same resources efficiently.

2.6 Cooperative NOMA

Cooperation in a NOMA system is examined further in [27], which intends to harness the SIC mechanics intrinsic to the NOMA principle to maximise the diversity gain for cell-edge users and users with superior channel conditions. In the suggested system paradigm, transmission is divided into two phases: A) Direct Transmission and B) Co-operation Phase.

The Direct transmission phase, shown in Fig.2.6, employs the fundamental NOMA principle, wherein the BS superposes all user signals using superposition coding and transmits the signal to all users. Each user, save the lead user, uses SIC to detect its message.

Throughout the Co-operation phase, shown in Fig.2.7, each Near User (NU) transmits both the cell-edge user signal and the superposed signal to the other users. This phase is divided into (K-1) time slots, with each time slot occupied by a single user transmitting a cooperative signal to other network users. For instance, user 2 would use the first time slot to broadcast a signal comprising the messages of (K-1) users to user 1, followed by user 3 transmitting the identical signal to user 1. This procedure would then be repeated for the remaining network users. At the conclusion of the cooperative phase, all users integrate the signals from both phases using MRC. Using the SIC mechanism, the goal of this method is to increase the diversity gain of all network users. This method improves efficiency at the cost of increased communication overhead, as each user must wait until both phases are complete before decoding their own message. [28] suggests that the system should only be utilised for short-range communications so as to reduce the duration of time slots during the cooperative phase.

The system, however, has three significant flaws: The robustness of the system under realistic inter user interference, the operation of the system in a quasi-ideal environment with Mobile Users (MU), the energy efficiency and cost of the proposed system, and the feasibility of the system with limited communication overhead.

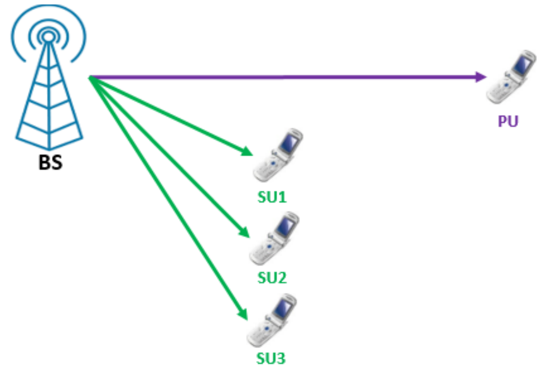
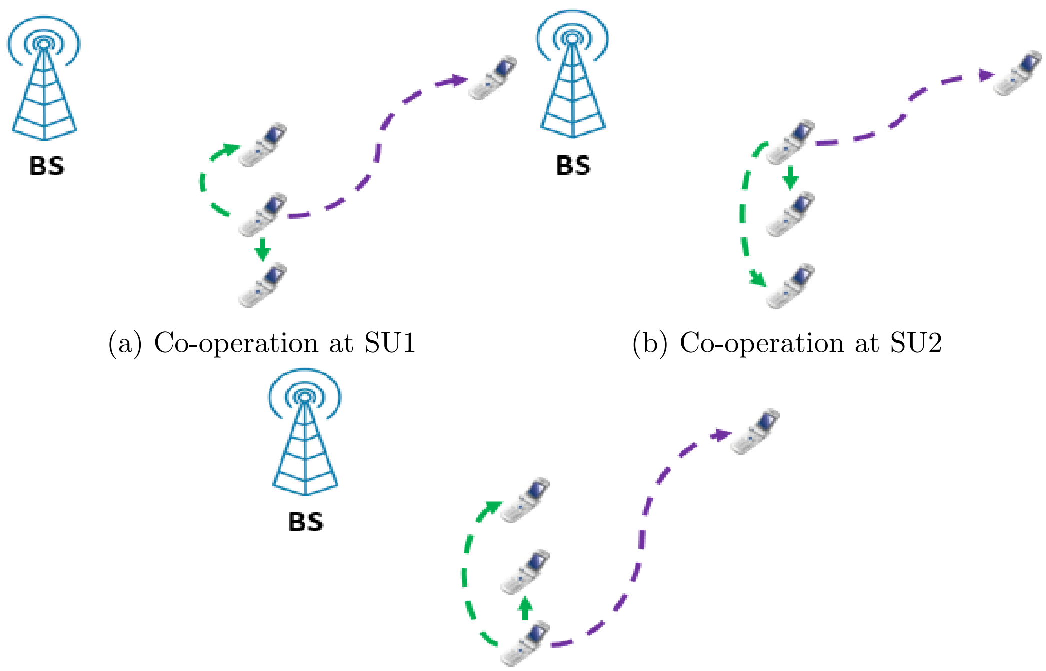


Figure 2.6: Direct Phase [28]



(a) Co-operation at SU1

(b) Co-operation at SU2

(c) Co-operation at SU3

Figure 2.7: Co-operation Phase [28]

[29] investigates energy efficiency in cooperative NOMA. Simultaneous Wireless Information Power Transfer (SWIPT) is used to a cooperatively relaying NOMA network. The objective was to examine the impact of the SWIPT energy harvesting protocol on network performance. Users are separated into two groups: nearby and distant. The suggested method intends to utilise users closest to the base station as energy collecting relays and then utilise the captured energy to make the close users relay the message of distant users without using additional energy. Users are distributed at random within discs using a Poisson Point Process (PPP). Experimental and theoretical results demonstrated that the diversity gain of the proposed system is equivalent to that of a standard cooperative NOMA system. With a little greater

outage probability reduction rate than the typical system.

[29] proposes a system with one base station, N relays, and two users. Assuming there are no direct links between the BS and the users, only relaying is used for communication. In accordance with CR principles, power is given to users not based on their channel conditions but on their individual QoS requirements such as data rates, latency, packet error rate, reliability, connection establishment time, mobility support, jitter (for voice and video applications), throughput, spectral efficiency, fairness, handover performance, energy efficiency, and security and/or privacy. One primary user (PU) has more stringent QoS needs than the other secondary user (SU).

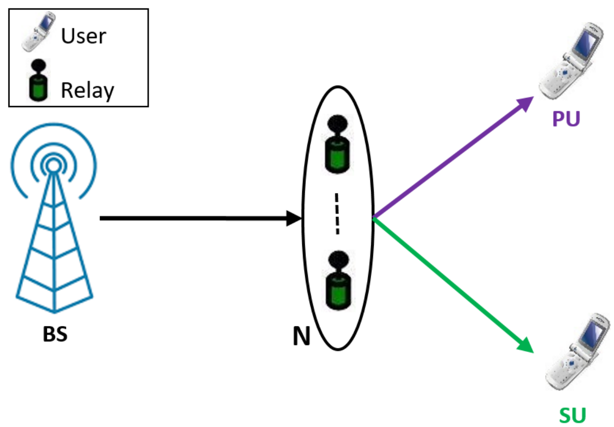


Figure 2.8: System model for relay selection [30]

[30] investigates two relay selection strategies. The standard max-min system, in which the objective is to select relays with the maximum minimal channel gains, and the suggested two-stage scheme, the objective of which is to maximise the data rate of the primary user in the first stage and the data rate of the secondary user in the second stage. In comparison to the conventional scheme, the results demonstrate a significant improvement in the outage performance of the system while achieving optimal diversity gain for both users. Notwithstanding the enhancements outlined in the study, the suggested system, shown in Fig.2.8, requires greater communication overhead to accomplish the desired outcomes.

The work in [31] also examines user relaying in cooperative NOMA. The suggested system model, shown in Fig.2.9, intends to address the issue of enhanced spectral efficiency in cooperative NOMA, where greater bandwidth is required for relaying, resulting in a decrease in system-wide spectral efficiency. The suggested model, employs consecutive relaying, with two Half-Duplex(HD) users decoding and forwarding the signal in an effort to enhance spectral efficiency. In an effort to increase user fairness, the research also provides an optimal power split method for complex power distribution.

The transmission occurs in three phases. At the initial phase, the BS transmits the overlaid signal to Near Users a and b. During the 2nd Phase, Near User a decodes and passes the far user message, while Near User b receives a second signal from the BS carrying the far user and user b messages. This introduces a self-interference element from user a to user b. Using the side information of the initial signal obtained in the first phase, it is claimed that this piece of paper is nullified. During the third phase, concurrent transmissions utilising user a and b as relays alternately occur in reverse.

In conjunction with optimal power split power allocation, the proposed successive relaying cooperative NOMA scheme provides superior outage performance compared to its OMA counterpart and traditional HD-NOMA counterpart. While the work's results are very encouraging in terms of developing a robust Full-Duplex NOMA system, the obtained results assume flawless SIC and disregard the error propagation issue. This system is also limited to the given user scenario, which involves two nearby users and one distant user. How the system would deal with the growing self intervention when more users are added has to be determined.

Coordinated direct and relay transmission, shown in Fig.2.10, is applied to NOMA in [32]. The suggested system concept contains a BS that connects directly with one user and relays with another user using a decode-and-forward method.

The objective is to increase diversity gains by leveraging the NOMA property, in which one user acquires the side information of another user to decode its own message. At the initial phase, the BS transmits the overlaid signal containing the messages of users 1 and 2. During the second phase, the relay transmits the decoded signal for user 2 while the base station simultaneously transmits a signal containing only the message for user 1. As a result of interference from the relay transmitting in the same time slot, user 1 uses the knowledge about user 2's signal to approximate the signal and, as a result, cancel the interference at the receiver. The provided results advise employing this system in small macro cells, where a BS communicates with a small number of users, as the interference generated by relaying would be too great to be properly cancelled in deployment scenarios with 3 or more users. [31] -[33]

[34] shows a cooperative NOMA system that utilises SWIPT energy harvesting to achieve cooperation by powering the cooperative relaying, as shown in Fig.2.11. The system model consists of one BS and two groups of users, one near the BS and the other near the cell edge.

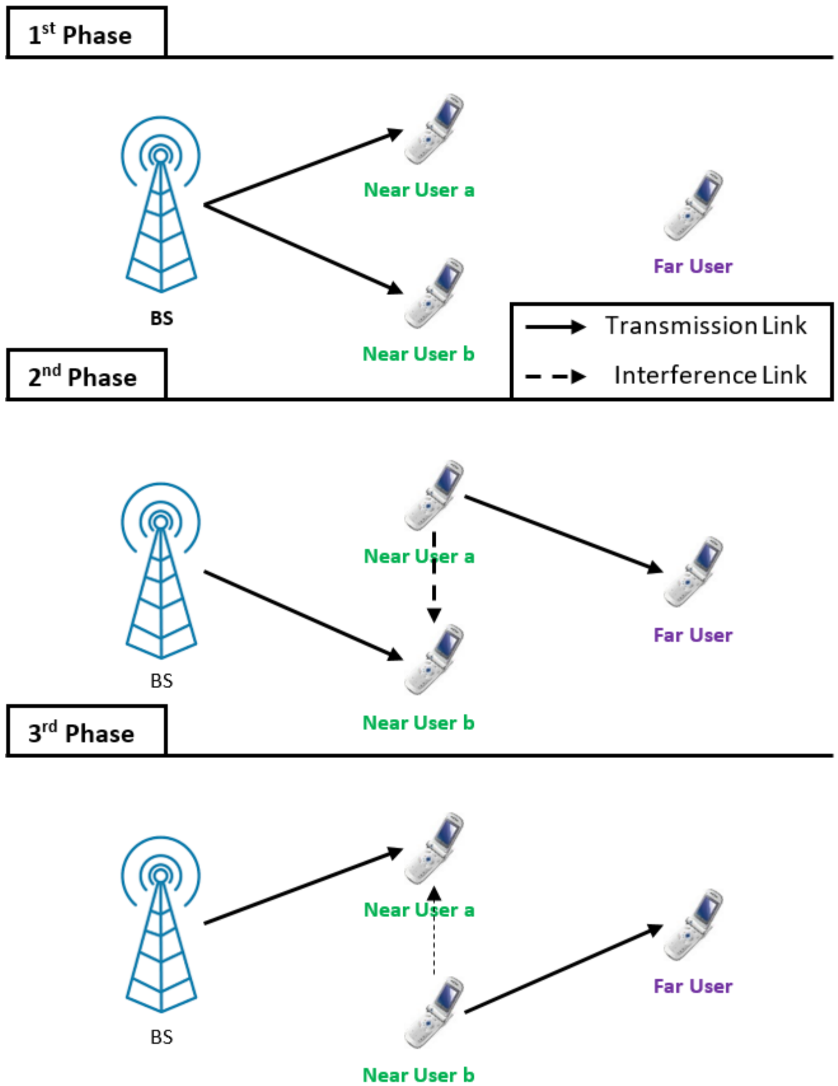


Figure 2.9: Successive Cooperative NOMA Relaying System Model [29]

The BS is placed within a disc and the two groups of users are scattered around the BS according to a pre-designed homogeneous PPP.

The users located near the BS are used as energy harvesting relays, used to both harvest energy using SWIPT and forward cell edge users messages. The transmission process is comprised of two phases: a direct transmission phase and a cooperative phase. During the direct transmission phase, the BS sends a signal of superposed messages according to the NOMA principle to both user groups Energy harvesting is done at the end of the direct transmission phase, which serves two purposes: firstly, the energy harvested is used to decode the messages of the near user(s). According to the NOMA principle, the decoding process relies heavily on SIC. This results in the near users gaining access to the data of the far users before decoding their own. Secondly, using the leftover energy, the near users prepare to function as cooperative

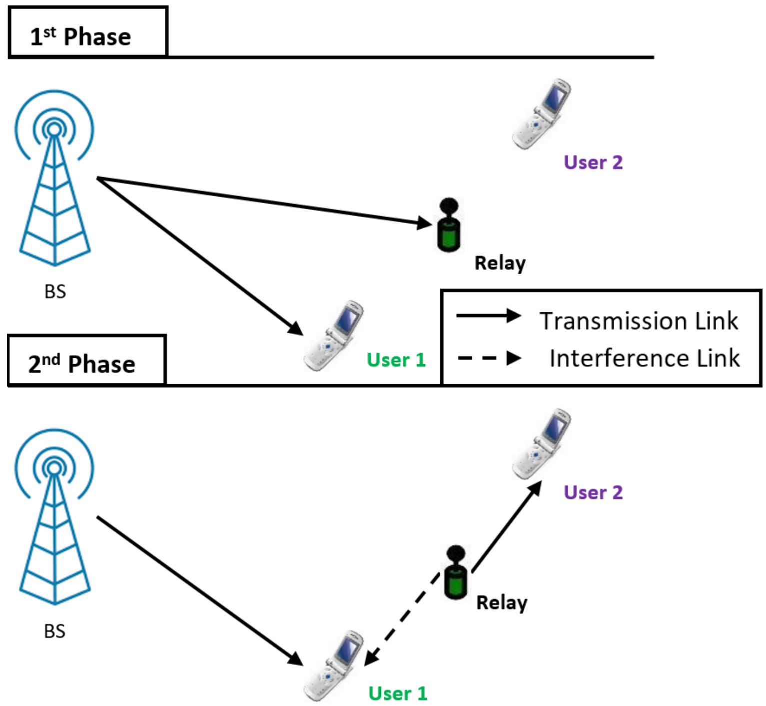


Figure 2.10: Coordinated direct and relay transmission system model [30]

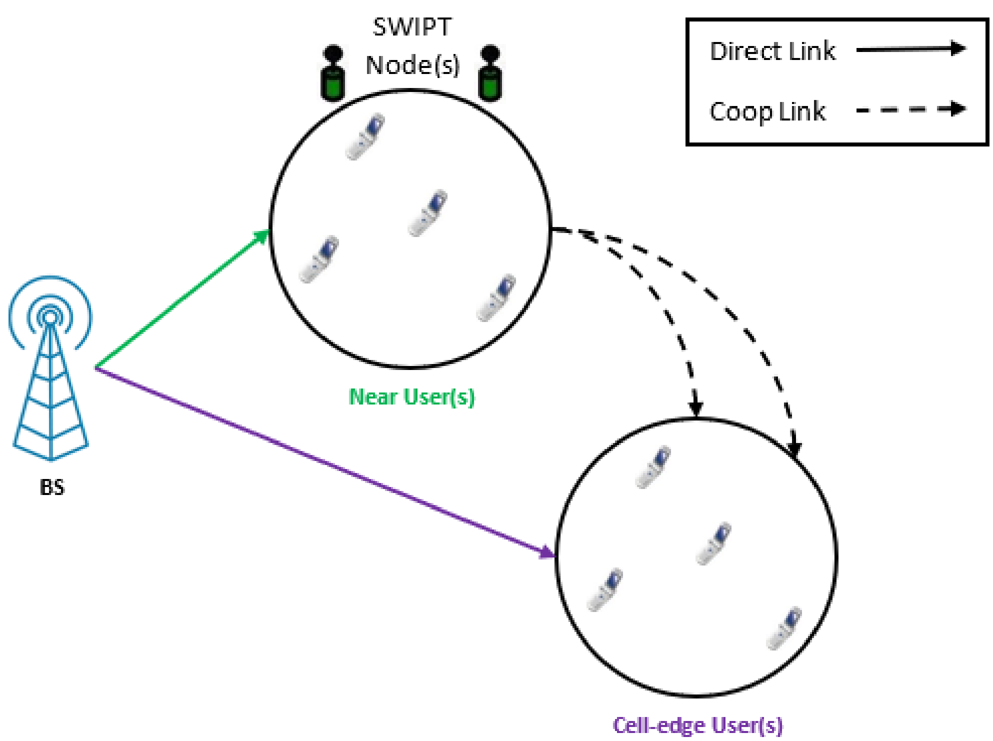


Figure 2.11: Cooperative NOMA with SWIPT

relays for the far cell edge users. During the cooperative transmission phase, the energy leftover from the previous phase is used to achieve cooperative relaying at the near user(s).

The analytical results cover the outage probability and diversity gains for both the near and far user groups. The outage probability is shown to be higher for the non-cooperative case and the diversity gains are the same across the cooperative NOMA and cooperative NOMA with SWIPT cases. This shows that, by utilising SWIPT, better cooperation between users can be achieved without sacrificing diversity gains.

[35] proposes a hybrid cooperative NOMA scheme, where NOMA and OMA are combined. Fig.2.12 shows the schemes are based on cooperative scheduling and load balancing among different groups of users based in cells.

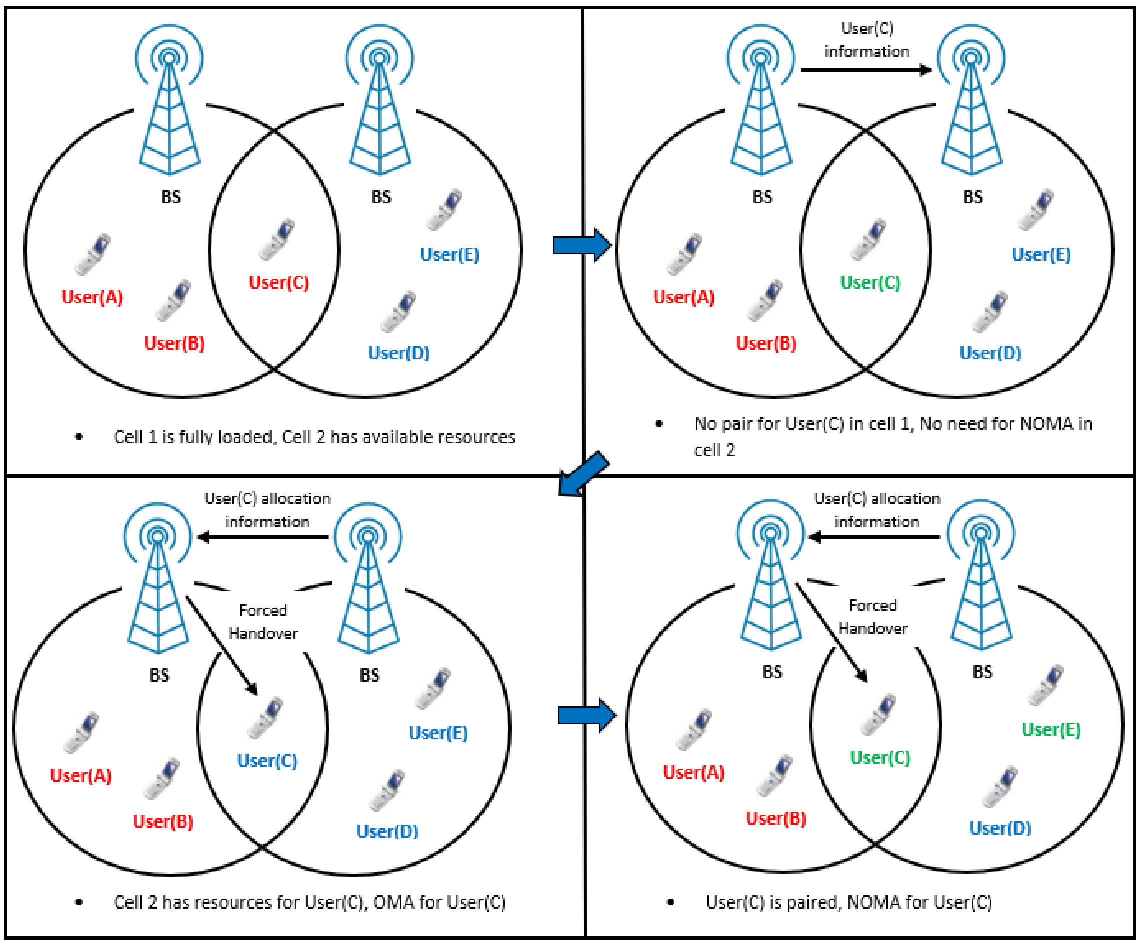


Figure 2.12: Cooperative NOMA for load balancing and user scheduling

The system model in this work consists of two neighbouring cells with two users each and one user on the cell edge of each cell. To achieve ideal performance, the aim is to pair as many users effectively together as possible according to the NOMA principle. This, however, is not always possible as one or both cells could have already paired users or just not enough resources to accommodate the cell edge user.

In this case, OMA is deployed. For example, if there is a user(A) that is the cell edge user for cells 1 and 2, two performance scenarios come into effect. One, cell 2 has enough resources to accommodate user(A), hence no NOMA pairing is required and OMA is used to serve user(A) or, two, user(A) needs more resources than there is available from cell 2 and there exists a suitable pair for it, hence, NOMA pairing is used to serve user(A).

After either option is selected, definitive resource allocation is carried out. A forced handover is performed where cell 2 signals to cell 1 that it can accommodate user(A), where it will only be served after other users in cell 2 have been served themselves.

The simulation results obtained for a multi-cell, multi-user system model have shown a 12-18% gain using cooperative NOMA over a non-cooperative NOMA system as well as an increase in probability of improving their system throughput for cooperative versus non-cooperative systems, 46% and 12% respectively.

By focusing on maximising the usage of NOMA over OMA, The cooperative NOMA system that has been suggested enhances the system capacity and is, therefore, the optimal choice in areas with unbalanced loads i.e hot spot areas.

[36] proposes a system model that aims to establish cooperative communications through cooperation between the BSs, as shown in Fig.2.13. The system model under consideration here consists of two BSs sharing one channel and serving two users, thereby exhibiting asymmetric interference, meaning one user suffers strong interference from one cell as it transmits the second user's signal at the same time as the first user.

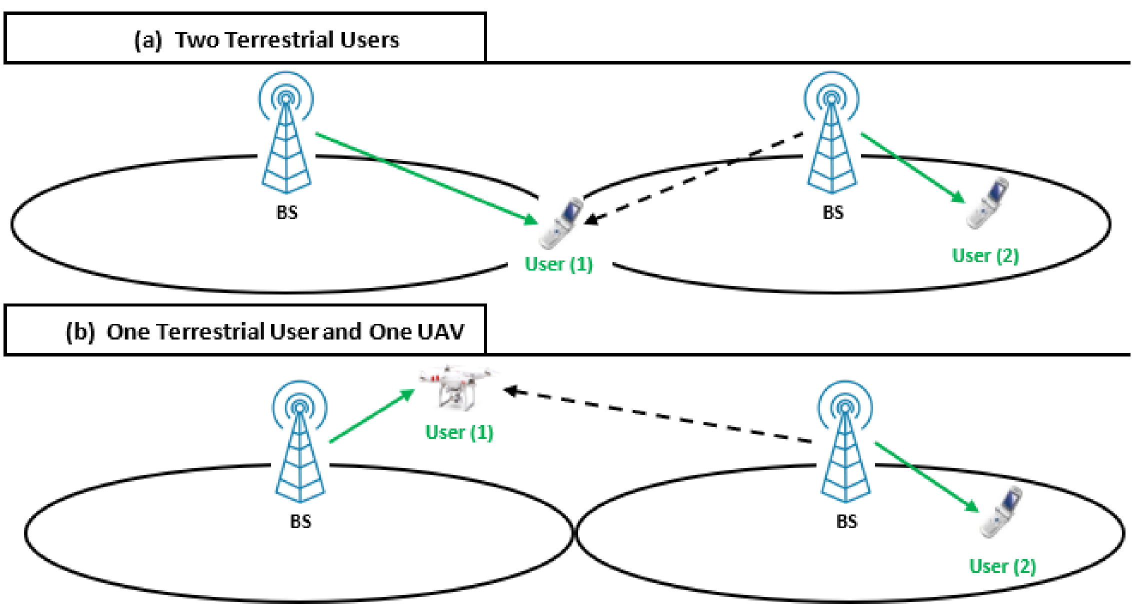


Figure 2.13: Cooperative NOMA for a mobile user

To manage this interference, a cooperative protocol is proposed where the second interfering BS shares its signal information with the first BS in order to aid in cancelling out the interference. The interfered user's BS sends out a superposed signal containing the messages of both users. The aim of such a scheme is for both signals, the original from the first BS and the interference signal from the second BS, to add constructively at the interfered user's receiver. The analytical results are derived for two cases: where both users are terrestrial and the other where one user is terrestrial and the other is an Unmanned Aerial Vehicle (UAV).

The simulation results however are obtained for the second case only. The results presented are obtained for the case of the UAV having a fixed altitude and travelling in a horizontal line from BS1 to BS2, thereby having the interference caused by BS2 gradually increase as the UAV flies. The results show that the achievable data rates of the proposed cooperative scheme start to increase up to capacity as the UAV travels. The proposed scheme is also shown to outperform it non-cooperative counterparts even when the interference is sufficiently high. The proposed work is shown to work in both ideal and practical scenarios as the results show.

A MIMO case for cooperative NOMA is examined in [37] where it proposes a spatial modulation aided cooperative relaying system (CRS) NOMA. Spatial modulation is a relatively new MIMO technique that was develpoled for wireless communications. Unlike regular MIMO systems that leverage multiple antennas to transmit multiple data streams simultaneously, spatial modulation uses the indices of the transmit antennas themselves to convey additional information. This provides a unique combination of digital modulation and multiple antenna indexing. The communication for two users is transmitted via two information-carrying entities of SM. The suggested framework is examined in conjunction with multiple receiving antennas in the context of the MIMO scenario.

Figure 2.14 illustrates the system model, which encompasses a singular BS transmitting a composite signal based on the NOMA principle to two users: A and B. Here, user A is proximate, while user B is positioned farther away from the BS. Users A and B are respectively furnished with N_A and N_B receiving antennas, and the BS comes equipped with N_t transmitting antennas. The communication channels - from BS to A, BS to B, and A to B - are characterized as independent and identically distributed (i.i.d) complex Gaussian random entities. The entire transmission procedure is bifurcated into two stages. In the first stage, the BS transmits a layered signal to both users. Subsequently, user A employs Maximum Likelihood (ML) detection for message decoding, whereas user B retains the received signal, postponing its decoding to

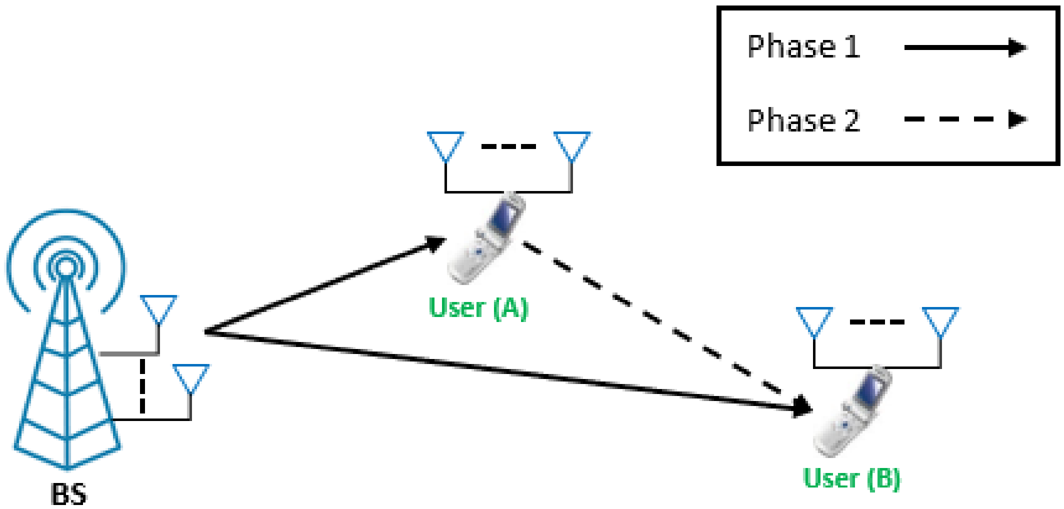


Figure 2.14: Cooperative NOMA for MIMO scenario

the conclusion of the second stage.

During phase 2, user A forwards the message of user B through a space shift keying signal. User B then combines the received message at the end of phase 2 with the one received at the end of phase 1 to finally decode its own message. Monte Carlo simulations are carried out in order to evaluate the proposed schemes against two benchmarks: CRS-NOMA, without SM, and SM-OMA. Two performance metrics are considered here, namely, BER and data rates. The results are obtained for a varying number of receive antennas where it is shown that with an increasing number of receive antennas comes an increase in diversity gains and a slight enhancement of BER performance in the low SNR region. CRS-NOMA utilising spatial modulation is shown to outperform its CRS-NOMA counterpart by gains of 5dBs and 2.5dBs at 10^{-5} BER. CRS-SM-NOMA is also shown to vastly outperform SM-OMA with about 10dBs gains. Moreover, since the BER curves almost overlap, CRS-SM-NOMA is shown to have better user fairness as well.

The achievable data rates for CRS-SM-NOMA also outperform the benchmark schemes. The sum rate is also shown to improve with N_t. The ergodic sum rate of CRS-SM-NOMA is shown to be higher than SM-OMA in the low SNR region and almost equal in the high SNR region. The proposed system model, through the simulations, is shown to be robust and capable in a MIMO scenario. However, the proposed model included only two two users in it's testing environment, thus, there exists the question whether this proposed system can achieve massive connectivity.

Challenges

According to the research conducted to date, one of the particular issues of cooperative NOMA is a problem with spectral efficiency, as more relaying will result in higher spectrum use. Full Duplex is being considered as a possible alternative. Energy efficiency problems also exist at each cooperative relay in case of large scale deployment.

2.7 EXIT Chart Analysis

EXIT charts have emerged as a fundamental technique for monitoring the convergence tendencies of systems employing iterative decoding [38], [39], [40]. They enable the pinpointing of the SNR threshold where a markedly minimal Bit Error Rate (BER) can be realized, circumventing the need for traditional Monte Carlo simulations [39], [40].

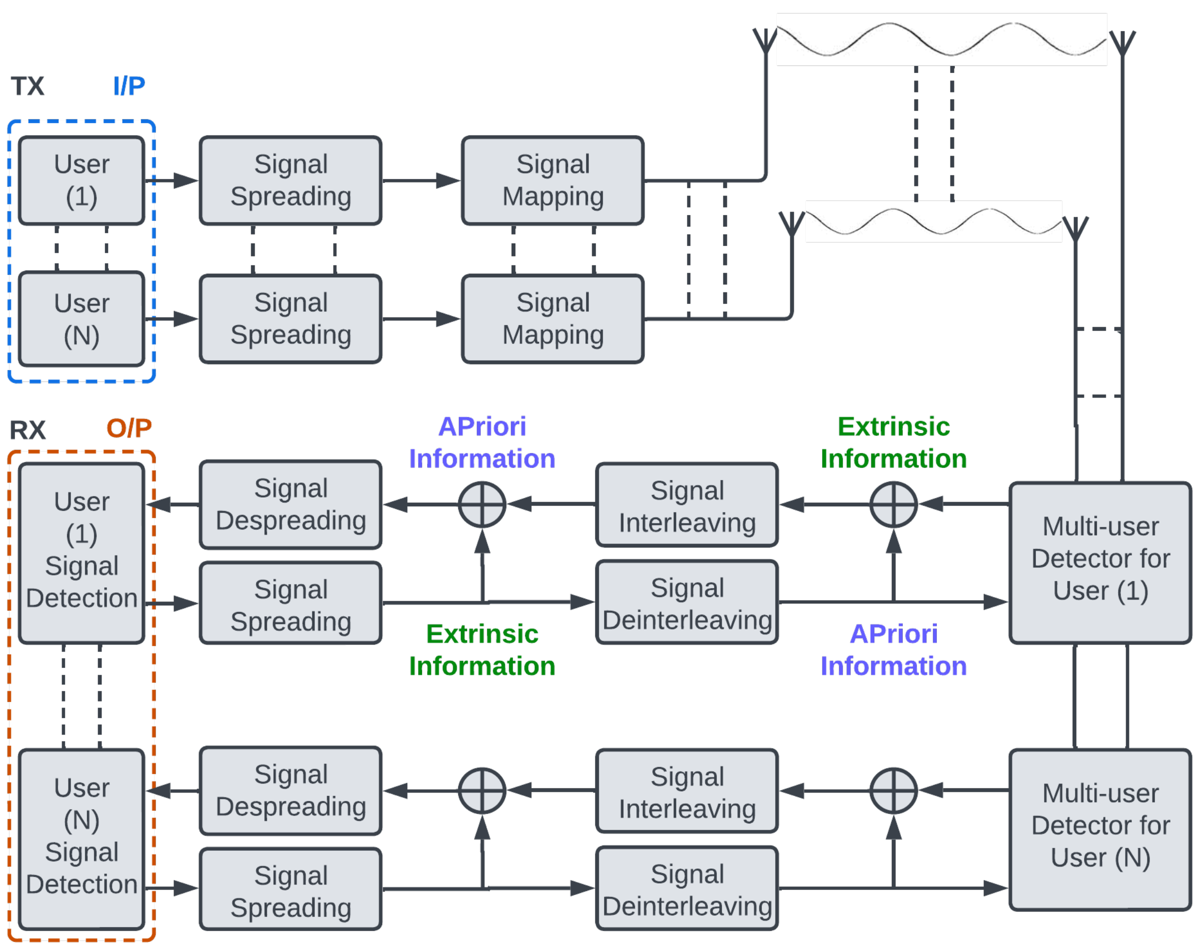


Figure 2.15: EXIT chart analysis block diagram example

Information is exchanged between the input and output module of the MUD by varying the channel conditions as shown in Fig.2.15. The proposed system model for the EXIT chart is comprised of a transmission and a reception stage. The transmission box is represented by the top-left corner of Fig.2.15 while the lower left-hand side represents the reception box. There are three parts to an EXIT chart: The outer curve, inner curve, and the stair shaped trajectory, as shown in Fig.2.16.

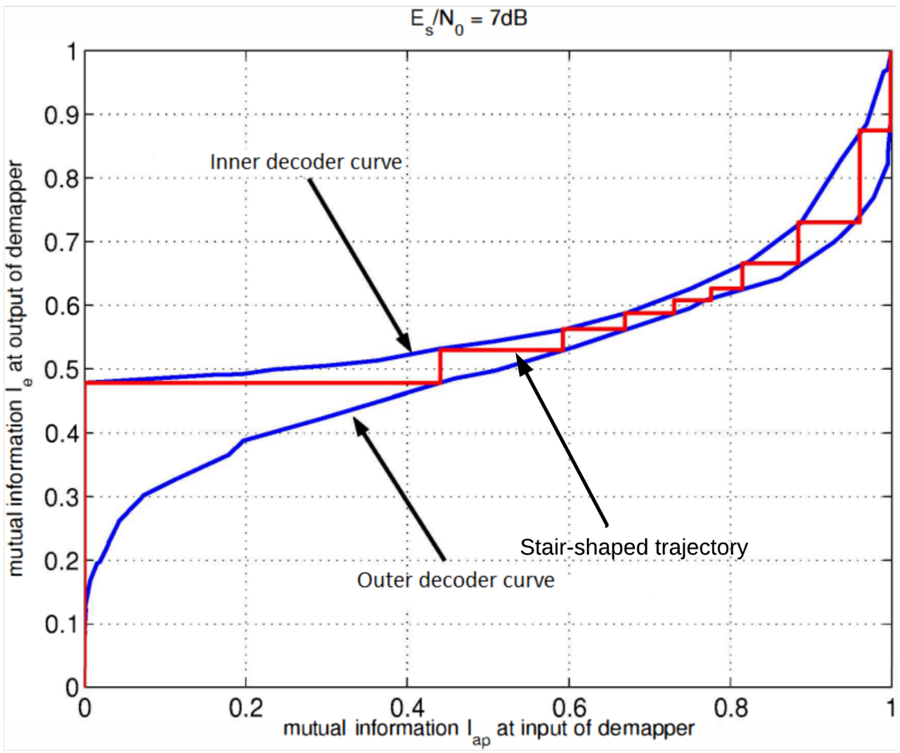


Figure 2.16: Example of an EXIT chart

The utilisation of interleavers allows for the simulation of both the inner and outer curves, enabling the prediction of values pertaining to the stair-shaped trajectory. An open tunnel is present when the trajectory, resembling a staircase, aligns with both the inner and outer curves. This signifies that the system has achieved its maximum convergence, resulting in unity gain for the system, provided that the inner and outer curves do not collide prior to this point. The relationship between the length of the interleavers and the alignment of the inner and outer curves with the stair-shaped trajectory is direct. Given that the length of the interleavers in our particular scenario is limited, it is possible that the trajectory will not align precisely with the inner and outer curves.

2.7.1 The Turbo Principle

The 'Turbo-Approach,' which was initially devised for deciphering concatenated codes [41], [42], is a generalized decoding and detection principle applicable to a wide range of detection and decoding challenges. Examples of such considerations include parallel or serial concatenation, normalization, encoded modulation, multi-user detection, multiple input/output (MIMO) detection, joint source and channel decoding, and low-density parity checks. Often, the system can be characterized by a serial concatenation, as depicted in Fig.2.17. In a basic mobile system, the multipath channel would act as the innermost layer's "encoder."

A feature of the receiver is that both detectors/decoders at the reception antenna are soft-in/soft-out decoders, capable of discerning and generating likelihoods or soft values. Importantly, the extrinsic segment of one decoder's soft-output is relayed to the subsequent decoder as a priori input, underscoring the receiver's principal attribute. Leveraging this approach, decoding of two-dimensional product-like codes [41] and other concatenated codes [42] has been proposed.

The program that Berrou utilises is sometimes referred to as a "turbo code." To be more specific, nothing in the codes can be interpreted as "turbo". The only component that uses turbo feedback is the decoder.

This is comparable to what happens in a mechanical turbo engine, as depicted in Fig.2.18. Fig.2.18 depicts the operation of a standard turbo coder. Initially, the data is fed to the turbo decoder, which processes and yields intrinsic information, such as the Log-Likelihood Ratios (LLR). This intrinsic data is subsequently relayed to the deinterleaver, resulting in the extraction of extrinsic information, such as a priori information. This extrinsic data is then conveyed to the outer decoder, which processes it further before forwarding it to the interleaver. Following this, the interleaver produces decoded intrinsic data, which is then passed to the inner decoder. Along with the received signal, this data is used to generate intrinsic details for the subsequent iteration. This cycle repeats until the error observed at the outer decoder becomes infinitesimally small. The trubo coding example shown in Fig.2.18 has the interleaver length set to 2048 bits as to provide operable error correction capability (longer interleavers provide better error correction but more complex turbo coders), convergence speed, latency, memory requirements, and performance in varying channels.

Turbo bit Interleaved Coded Modulation (BiCM), serial code concatenation, turbo equalisation, turbo Differential Phase Shift Keying (DPSK) modulation, turbo MIMO, turbo source channel, and Low Density Parity Check (LDPC) encoder/decoder are some of the established configurations for a turbo coding system. Most encoders and some decoders utilise Forward Error Correction (FEC) to carry out their tasks.

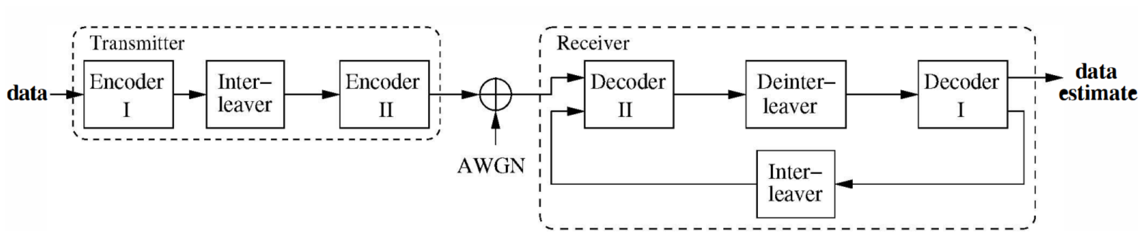


Figure 2.17: Serial concatenated system with iterative detection/decoding [43]

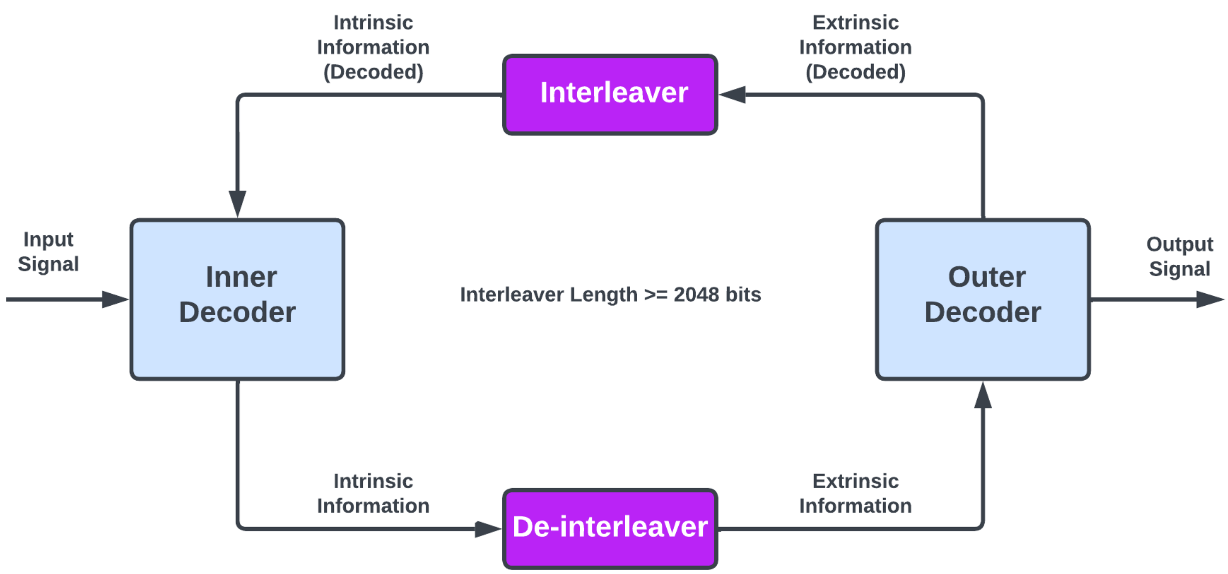


Figure 2.18: Turbo decoder including inner and outer decoders

Similar to how compressed air is taken back from the compressor to the main engine, the other decoder receives extrinsic information. This iterative approach that entails the exchange of information among the two decoders is challenging to examine and define. The EXIT chart created by Stephan ten Brink [43] for extrinsic transfer of information is an incredibly useful tool.

2.7.2 Mutual Information (MI) and EXIT Charts

Shannon's MI

Consider U and V represent two random variables with real (as opposed to imaginary) values. The MI of Shannon is therefore defined as:

$$I(U;V) = \int\int f(u,v) \log \frac{f(u,v)}{f(u).f(v)} dudv \tag{2.1}$$

with

$$I(U;V) = H(V) - H(V|U) \tag{2.2}$$

where

$$H(V|U) = \int\int f(u,v) \log \frac{1}{f(v|u)} dudv \tag{2.3}$$

where $f(u,v)$ denotes the joint probability density function and $f(u)$ and $f(v)$ represent the marginal probability density functions, $H(V)$ represents the entropy of variable V, and $H(V \mid U)$ denotes the conditional entropy of V given U.

For an additive channel with $V = U + K$ with statistically independent K

$$H(V|U) = H(K) \tag{2.4}$$

with the transmit power $P_u = \sigma_u^2 = E_s(1/T)$, where σ represents the variance of the signal, T represents the symbol duration, E_s represents the energy per symbol, and $E_s(1/T)$ calculates the power of the signal, the noise power $P_n = \sigma_n^2 = \sigma_k^2 = (N_0/2) \times (1/T)$ and the receive power $P_v = \sigma_v^2 = \sigma_u^2 + \sigma_n^2 = P_v = P_u + P_n$ we have

$$I(U;V) = H(V) - H(K) \tag{2.5}$$

$$I(U;V) \leq \frac{1}{2} \log_2(1 + \frac{P_u}{P_n}) \tag{2.6}$$

$$I(U;V) \leq \frac{1}{2} \log_2(1 + \frac{2E_s}{N_0}) \tag{2.7}$$

The MI is consistent with the AWGN channel when there's a Gaussian input equality. Given both Gaussian input and noise, the output similarly assumes a Gaussian distribution, thereby achieving channel capacity.

$$C = \max_{f(U)} I(U;V) = \frac{1}{2}\log_2(1+\frac{2E_s}{N_0}) \tag{2.8}$$

If we restrict ourselves to binary inputs with $u \in \{+1,-1\}$ the MI becomes

$$I(U;V) = \sum_{u=+1,-1} \int_{-\infty}^{+\infty} f(v|u)P(u)\log_2\frac{f(v|u)}{f(v)}dv \tag{2.9}$$

and the maximal MI is achieved for equally likely inputs x as

$$I(U;V) = \frac{1}{2}\sum_{u=+1,-1} \int_{-\infty}^{+\infty} f(v|u)P(u)\log_2\frac{f(v|u)}{f(v)}dv. \tag{2.10}$$

with

$$f(v) = \frac{1}{2}(f(v|u=+1)+f(v|u=-1)), \tag{2.11}$$

and for $u = 1$:

$$f(v \mid u=1) = \frac{1}{\sqrt{2\pi}\sigma}\exp^{-\frac{(u-1)^2}{2\sigma^2}}, \tag{2.12}$$

and for $u = -1$:

$$f(v \mid u=-1) = \frac{1}{\sqrt{2\pi}\sigma}\exp^{-\frac{(u+1)^2}{2\sigma^2}} \tag{2.13}$$

where $\sigma^2 = N0/(2Es)$ The performance of the integral, which must be assessed numerically, is depicted in Fig.2.19. It indicates how much of a bit is known following transmission via a noisy channel.

2.7.3 The EXIT Chart

The EXIT Chart, which was formulated by Stephan ten Brink, graphically represents the transmission of information in turbo decoding and the corresponding decoding efficiency in the descending region. The analysis involves a comparison of the MI of the initial constituent decoder with that of the second constituent decoder, which is represented by the test (a priori) channel. To clarify, the quantified result of the inferior division establishes the numerical representation on the abscissa of the EXIT chart, while the quantified result of the superior division establishes the numerical representation on the ordinate. Both of the aforementioned values have a range that spans from 0 to 1. During the following (partial) iteration, the two decoders exchange their roles: The initial output of the primary decoder serves as the a priori input for the secondary decoder. This enables the independent analysis and optimisation of each individual code component. The output solely utilises extrinsic L-values denoted as L1o". This implies that the soft output L-value of the complete output is subtracted from the a priori input value. This hinders the spread of information that has already been established. The quantification of the information transfer function T is:

$$I(L_E; U) = T(I(L_A; U)) \tag{2.14}$$

with the following provisions:

- The model assumes a Gaussian distribution for the input L_A with parameter $\sigma_a^2 \leftrightarrow I(L_A; U)$, while also utilising a large interleaver to ensure statistical independence.
- The utilisation of inner decoders in a serial concatenated scheme and parallel concatenated schemes necessitates the incorporation of supplementary parameters, namely L_{CH} and σ_{CH}^2. The channel SNR or $I(L_{CH}; X)$ is featured as a parameter on a collection of EXIT curves.
- In the context of serial concatenation, the input for outer decoders is solely represented by $L_A^{(o)}$, which is derived from the interleaved serial $L_E^{(i)}$.

2.7.4 EXIT Chart Applications

Out of the many applications that could be used, we will only provide one example of an outer code and one example of an inner "code," and those examples will be Quadrature Amplitude Modulation (QAM) mapping and precoding.

QAM Mappings

Take, for instance, a QAM system with memory 2 and a rate 112 outer channel code, illustrated in Fig.2.19, which represents an application of turbo principles. This system is identified as bit-interleaved coded modulation (BICM). When the QAM mapping is appropriately selected, the a priori information of the other bits, denoted as $L_e^d(c)$, can enhance the detection of the prevailing bit, even in the absence of explicit memory. $L_a^d(c)$ represents the extrinsic information returning to the demodulator. C denotes the interleaved encoded data, C_k represents the data after passing through the serial to parallel converter, and s_k is the modulated information, r_k is the information after passing through the channel.

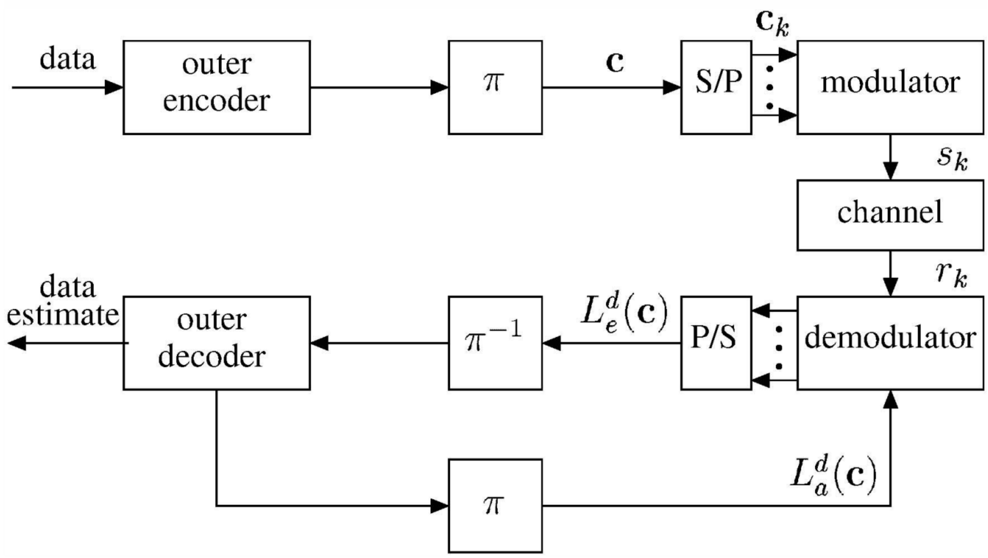


Figure 2.19: Coded system with convolutional codes as outer code and QAM transmission as inner code with BICM [43]

Irregular and Regular Convolutional Codes

Envision a serial concatenation that utilizes a convolutional outer code. For this external FEC coding mechanism, a suite of punctured convolutional codes with memory 4 and LR = 7 (4/1 2, 5/1 2, 6/1 2, 7/1 2, 8/1 2, 9/1 2, 10/1 2) rates is adopted. Their EXIT charts are represented by dashed trajectories in Fig. 2.20 and initiate at the coordinate (0,0) since the outer decoder solely intakes a singular input from the internal decoder. Postulate a specific EXIT chart for the internal system (termed decoder II; here, the A Posteriori Probability equalization of a multi-path channel [II]) at a defined channel SNR. This provides an initial value even when presented with a zero a priori input. The rate 1/2 code (visible as the fifth dashed trajectory from the apex) fails to achieve convergence through iterative processes due to the intersecting nature of the EXIT trajectories. Nevertheless, an irregular code maintaining an average rate of 1/2, as visualized in Fig. 2.21, can be formulated.

We have the following constraints on the α_k, which represents the fraction of edges in the

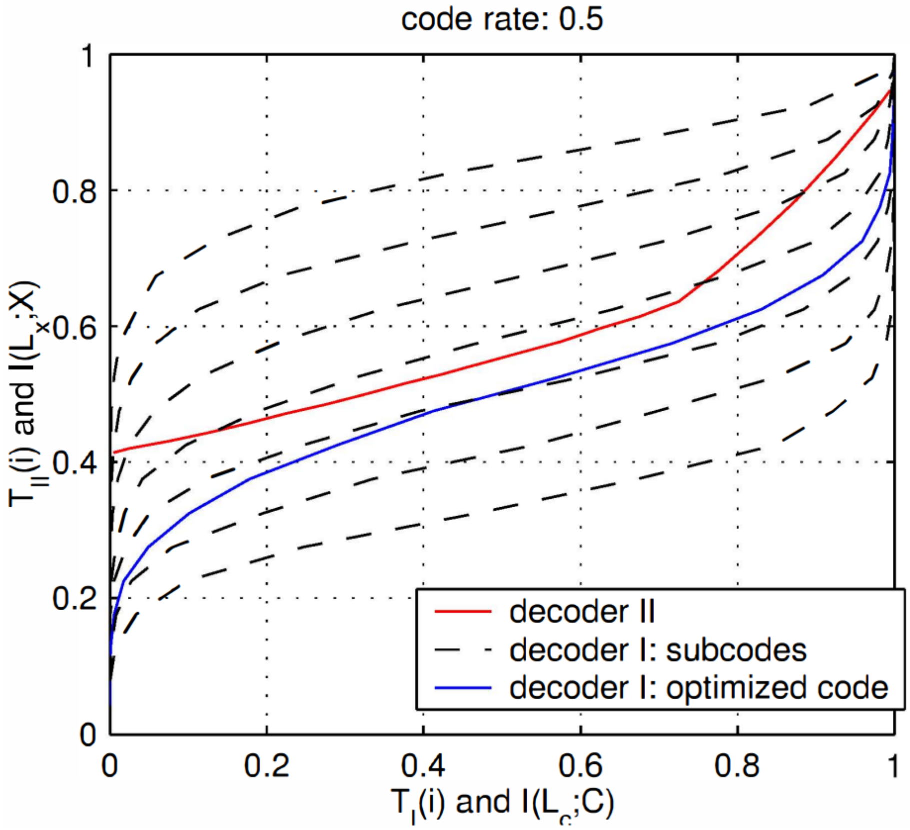


Figure 2.20: Multipath transmission as inner code as well as 7 convolutional codes behaving as outer codes EXIT chart [43]

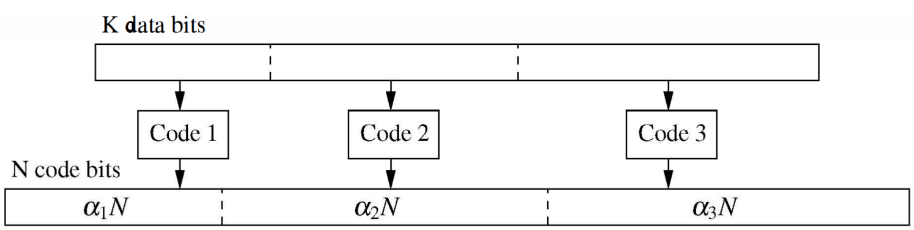


Figure 2.21: Irregular code construction [43]

tanner graph that are connected to variable nodes of degree k with L_R being the maximum degree of the variable nodes in the graph, and R_k being the coding rate, in this FEC code:

$$\sum_{k=1}^{L_R} \alpha_k = 1, \tag{2.15}$$

$$\sum_{k=1}^{L_R} \alpha_k R_k = 1/2, \tag{2.16}$$

and

$$\alpha_k \in [0, 1], k = 1, ..., L_R. \tag{2.17}$$

Given that the L-values across all subcode decoders exhibit symmetry and uniformity, the cumulative transfer function is the aggregate of the individual EXIT transfer curves.

$$T(i) = \sum_{k=1}^{L_R} \alpha_k T_k(i). \tag{2.18}$$

By fine-tuning the collection $\{\alpha_i\}$ to ensure that curves I and II closely align without intersecting, thus preserving a suitable open channel for the iterations [43], we derive the continuous curve denoted as I.

2.8 Deep Learning

2.8.1 Background

Types of AI

AI, machine learning, neural networks, and deep learning can be compared to Russian nesting dolls, which is perhaps the simplest method to conceptualise them. Each term is fundamentally a subset of the preceding.

That is, machine learning is a subfield of AI, as shown in Fig.5.1. Deep learning is a specialised area within the broader field of machine learning, wherein the fundamental building blocks of the algorithms are neural networks. The distinguishing factor between a neural network and a deep learning algorithm lies in the number of node layers, or depth, with the latter requiring more than three layers. [44]

Deep Learning Versus Machine Learning

Deep learning is distinguished from conventional machine learning by the data types and learning techniques it employs. Using structured, labelled data, machine learning algorithms make predictions. This means that certain features are extracted from the model's input data and grouped into tables. This does not necessarily imply that it does not employ unstructured data; if it does, it is typically pre-processed into a structured format [45].

Deep learning eliminates a portion of machine learning's standard data preprocessing requirements. These algorithms can process and analyse unstructured data such as text and images, as well as automate feature extraction, thereby reducing the need for human specialists. For instance, suppose we wished to organise a collection of photographs of various pets by "cat," "dog," "hamster," etc. Algorithms capable of deep learning can determine which characteristics (such as ears) distinguish one animal from another. This characteristic hierarchy is constructed manually by an expert in machine learning. Then, through gradient descent and backpropagation, the deep learning system modifies and optimises itself for accuracy, enabling it to make more precise predictions about a new image of an animal [46].

Machine learning and deep learning models are also capable of multiple forms of learning, which are typically categorised as supervised learning, unsupervised learning, and reinforcement learning (RL). Supervised learning employs labelled datasets to classify or generate predictions; this requires human intervention to label input data appropriately. Unsupervised learning, on the other hand, does not require labelled datasets; it discovers patterns in the data and clusters them based on any distinguishing criteria. RL is the process by which a model enhances its performance of an activity in a given environment in order to maximise the reward based on the feedback it receives.

The Functioning of Deep Learning

Deep learning neural networks, often termed artificial neural networks, endeavor to replicate the workings of the human brain utilizing data inputs, weights, and biases. In unison, these components accurately identify, categorize, and delineate data entities. In these networks, the predictive or classifying function is refined and enhanced through several tiers of interlinked nodes, a process termed forward propagation. Within such networks, the discernible layers comprise the input and output stages. The input layer introduces data for processing to the deep learning structure, while the output layer produces the concluding prediction or classification [47]-[50].

Back-propagation modifies the function's weights and biases by traversing backwards through the layers in an effort to train the model using techniques such as gradient descent to quantify errors in predictions. Forward and backward propagation allow neural networks to generate predictions and correct for errors. The precision of the algorithm develops continuously over time. The basic form of deep neural network has been described above. However, deep learning algorithms are extremely sophisticated, and there are a variety of neural network designs to address specific problems or datasets [51-53].

Convolutional neural networks (CNNs) are predominantly utilised in computer vision and image classification applications, enabling tasks such as object detection and recognition. In 2015, a CNN bested a human for the first time in an object identification competition. RNNs are frequently used in natural language and speech recognition applications because they leverage sequential or time series data [54]-[57].

CNNs obviate the requirement for manual feature extraction, thereby dispensing with the need to identify the features utilised for image classification. The CNN shown in Fig.2.22 for example, works by extracting features directly from images. The salient characteristics are not pre-existing; rather, they are acquired during the network's training process on a dataset of images. The utilisation of automated feature extraction renders deep learning models notably precise for computer vision applications, specifically object classification.

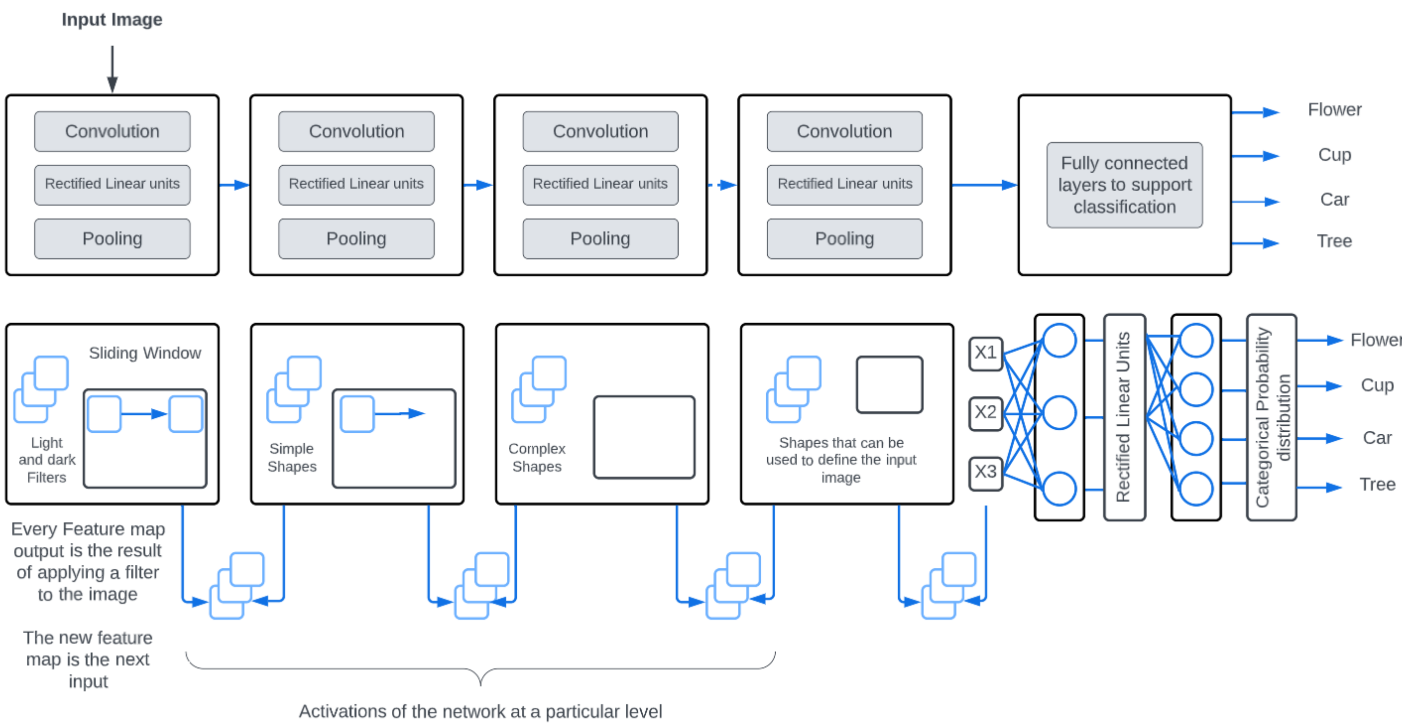


Figure 2.22: A network with many convolutional layers. Filters are applied to each training image at different resolutions, and the output of each convolved image serves as the input to the next layer. [89]

CNNs acquire the ability to identify distinct characteristics of an image through the utilisation of numerous hidden layers, ranging from tens to hundreds in number. The augmentation of hidden layers results in an elevation of the intricacy of the acquired image characteristics. As an illustration, the initial concealed layer may acquire the ability to identify edges, while the final layer acquires the ability to identify intricate patterns that are tailored to the contour of the object being identified.

Neural Networks

A neural network is a functional unit of deep learning that operates by accepting input and producing an output. Artificial neural networks are utilised in the field of Deep Learning. Artificial neural networks emulate the cognitive processes of the human brain in order to address intricate data-related challenges. [58]-[60].

These technologies solve problems in image recognition, speech recognition, pattern recognition, and natural language processing, among others.

Before diving into neural networks, a number of key terms needed to be defined [61]-[62]:

- **Neuron:** A fundamental component of artificial neural networks. The system is accountable for receiving input data, executing computations, and generating output.
- **Input data:** Neurons receive information or data.
- **Artificial Neural Network:** The computational system is modelled after the neural networks found in the human brain, which are responsible for processing information.
- **Deep Neural Network:** A deep artificial neural network is characterised by the presence of multiple layers situated between the input and output layers.
- **Weights:** The synaptic efficacy refers to the degree of potency in the inter-neuronal connection. The impact of the input on the output is determined by the weights.
- **Bias:** A supplementary variable employed in conjunction with the summation of the multiplication of weights and inputs for the purpose of generating an outcome.
- **Activation Function:** Determines a neural network's output.

Overview

Figure 2.23 depicts a basic neural network comprised of synthetic neurons responsible for receiving and analysing input data. Information is transmitted sequentially through the input layer, the hidden layer, and the output layer. The initiation of a neural network process occurs

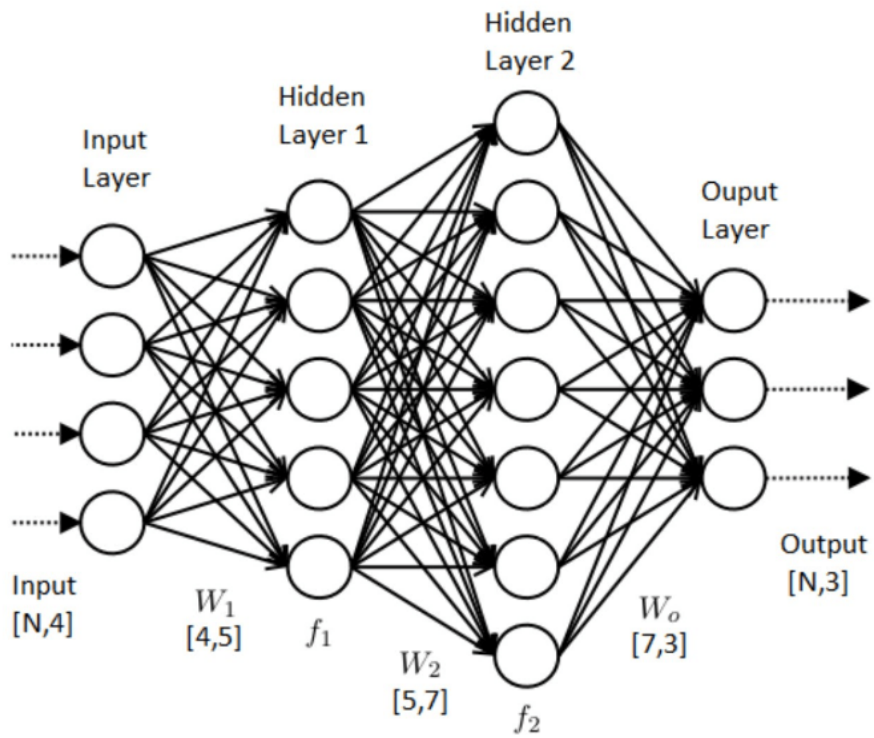


Figure 2.23: Rudimentary example of a neural network as it is used in deep learning systems [62]

upon the provision of input data to the system. The information undergoes a series of layers of processing in order to generate the intended output. A neural network is capable of acquiring knowledge from organised data and demonstrating the resulting output. The acquisition of knowledge that occurs within neural networks can be classified into three distinct categories:

- **Supervised Learning:** The algorithms are fed with inputs and outputs, aided by labelled data. Subsequently, the anticipated outcome is projected subsequent to receiving instruction on data interpretation.
- **Unsupervised Learning:** An artificial neural network is capable of autonomous learning without any human intervention. The absence of annotated information necessitates the utilisation of patterns discerned from the output data to determine the output.
- **RL:** learning takes place according to the information acquired during the feedback stage.

A neuron is considered to be the fundamental unit of a neural network. The application employs the supervised learning approach for data classification and acquisition.

Supervised Learning Algorithms

Supervised learning is a method utilised in the development of AI, whereby a computer algorithm is instructed using input data that has been categorised for a specific output. The model undergoes training until it attains the capability to identify the fundamental patterns and associations between the input data and the output labels, thereby facilitating the production of precise labelling outcomes when confronted with novel data [63].

Supervised learning has demonstrated efficacy in addressing classification and regression tasks, such as discerning the appropriate category for a news article or forecasting the sales volume for a forthcoming date. The objective of supervised learning is to interpret data in relation to a particular inquiry [64].

Unsupervised learning stands in contrast to supervised learning. This methodology involves providing unannotated data to the algorithm, which is programmed to autonomously identify patterns or similarities.

Unsupervised Learning Algorithms

The process of training models on unprocessed and unlabeled data is referred to as unsupervised deep learning. Frequently, it is utilised for the purpose of detecting patterns and trends within unprocessed data sets, or for grouping comparable data into a predetermined quantity of clusters. Frequently, it is a methodology employed during the initial exploratory stage to enhance comprehension of the data-sets [65].

Unsupervised deep learning is, as its name suggests, a more hands-off approach than supervised deep learning. The task of setting model hyper-parameters, such as determining the number of cluster points, is typically performed by a human operator. However, once these parameters are established, the model is capable of effectively processing vast arrays of data with minimal human oversight. Unsupervised deep learning is a suitable approach for addressing inquiries pertaining to latent patterns and correlations within data that have not been previously observed.[66]

A significant proportion of the data that is currently accessible is unannotated and unprocessed. Unsupervised learning is a potent technique employed to extract insights from data by categorising it based on shared characteristics or identifying underlying patterns within data sets. In contrast, the utilisation of supervised deep learning may require significant resources due to the requirement of annotated data [67]-[70].

Unsupervised deep learning is mainly used to:

- To cluster data-sets based on similarities between their features or to segment the data.
- Recognise the correlation among diverse data points, such as the automated suggestions for music.
- Conduct preliminary data analysis

Reinforcement Learning

RL, illustrated by Fig.2.24, pertains to the process of instructing deep learning models to execute a series of decisions. The agent acquires the ability to attain a specific objective in an environment that is characterised by uncertainty and may be intricate in nature. In the context of Reinforcement Learning, an AI agent is presented with a scenario that resembles a game. The problem is approached by the computer through the utilisation of a trial and error method to arrive at a solution. In order to achieve desired outcomes, programmers incentivize AI systems through a system of rewards and penalties based on their actions. The objective is to optimise the overall reward, as stated in [71].

Whilst the designer is responsible for establishing the reward policy and game rules, they do not provide any guidance or recommendations to the model for resolving the game. The model is responsible for determining the optimal approach to achieve the highest possible reward, commencing with random attempts and culminating in advanced strategies and exceptional abilities [72]-[75].

Through the utilisation of search algorithms and iterative experimentation, RL presently stands as the most effective means of stimulating artificial intelligence's creative capacity. Unlike humans, AI has the ability to accumulate experience from numerous simultaneous game-plays through the utilisation of a RL algorithm, provided that the computer infrastructure is powerful enough [76].

Historically, the utilisation of RL was constrained by inadequate computational infrastructure. Nonetheless, advancements were made. The recent advancements in computational technologies have brought about significant changes in the early progress, paving the way for novel and inspiring applications.

The utilisation of RL in the development of autonomous vehicles' control models is a noteworthy illustration of its potential application. Under optimal conditions, it is preferable for the computer to receive no directives pertaining to driving.

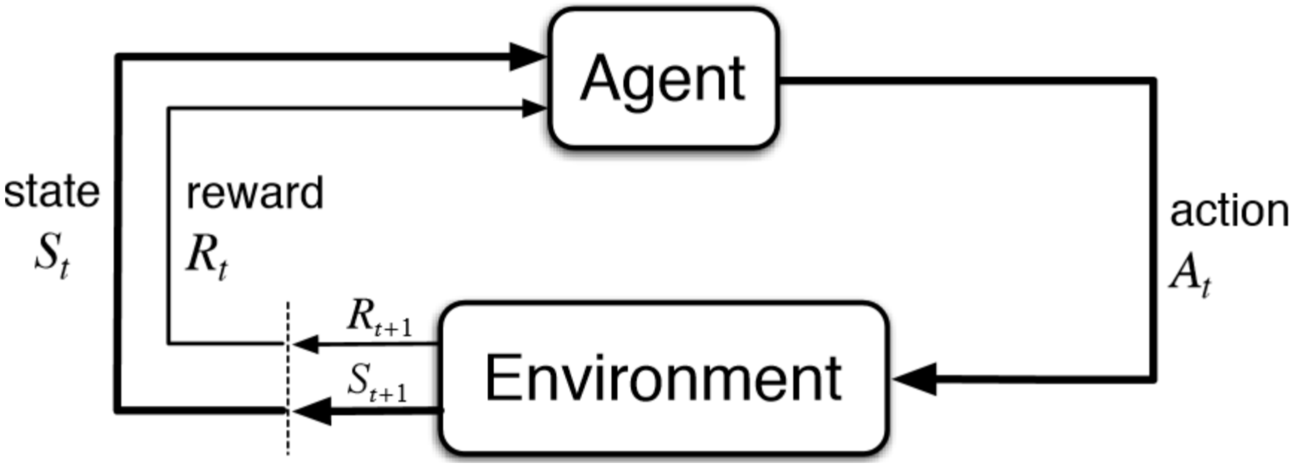


Figure 2.24: Simple block diagram for RL [77]

The software developer would refrain from implementing fixed solutions for any aspect related to the task at hand and instead enable the machine to acquire knowledge from its own mistakes. Ideally, the sole component that would be pre-programmed in an optimal scenario is the incentive mechanism.

Differences between RL and unsupervised learning

While both supervised learning and RL involve establishing a relationship between input and output, there is a fundamental difference in the way feedback is provided to the agent. In contrast to supervised learning, where the agent is given the correct set of actions to perform a task, RL employs rewards and punishments as signals to reinforce positive and negative behaviour.

In contrast to unsupervised learning, RL diverges in its objectives. The objective of unsupervised learning is to identify commonalities and discrepancies among data points, whereas in the context of reinforcement learning, the objective is to determine an appropriate action model that would optimise the overall cumulative reward of the agent. The diagram presented in Figure 2.24 depicts the feedback loop between action and reward in a typical reinforcement learning model.

RL problem formulation

Several fundamental terms that define the core components of an RL problem include:

- **Environment:** material environment where the agent functions.
- **State:** present state of the agent.
- **Reward/Penalty:** response from the surroundings.
- **Policy:** approach to associate the agent's state with particular actions.
- **Value:** anticipated reward an agent is expected to obtain by executing an action in a given state.

2.8.2 Deep learning in NOMA Systems

In the domain of wireless network communication, while the commercial implementation of deep learning devices is relatively infrequent [78], a considerable number of researchers are endeavouring to integrate deep learning techniques into communication systems primarily to enhance the performance of existing signal processing algorithms. In the process of decoding information transmitted across a channel, deep learning has the capability to acquire a decoding algorithm rather than a basic classifier. Several instances of effective outcomes have been documented in the literature. Instances such as the process of deciphering linear code [79] and the utilisation of polar code [80] serve as illustrative examples.

The study presented in [81] demonstrates the enhancement of the standard belief propagation (BP) decoder and its integration into a BP-CNN system. This integration leads to enhanced BER performance without significantly increasing the complexity of the receiver. According to deep learning theory [82], this method demonstrates superior performance compared to state-of-the-art compressed sensing techniques in terms of both signal recovery quality and computational efficiency. Furthermore, an area of research that has gained significant popularity is the field of modulation categorization and identification, which is mostly based on deep learning techniques [83].

The efficiency of the design schematic for the stacked auto-encoder is observed during the deployment of, protocol categorization, feature learning, and anomalous protocol detection [84]. Mobile traffic classifiers, based on Deep Learning (DL), have also been found to be able to operate under encrypted traffic while mirroring their complicated traffic patterns. Achievements made in blind detection for MIMO systems with DL are also continuously being discovered [85], [86]. Furthermore, a technique that implements DL into an OFDM system has been presented [87]; its simulation results divulge the promising performance of DNNs.

In [88], a comprehensive DL system was proposed, consisting of an encoding layer, noise layer, and decoding layer. This system incorporated a long-short-term memory (LSTM) network to facilitate uplink non-orthogonal multiple access (NOMA) transmission. Additionally, [89] extended this DL system to include MIMO scenarios. The auto-encoder's superb performance was show-cased by their simulation results for jointly learning transmit and receive functions. In most real-life environments however, it would be hard to artificially train and optimally utilise both send and receive functions of the DL system.

The research contributions of [90] have shown that deep learning allows for an effective solution to fast data clustering since various information features of higher dimensionality can be used in a flexible manner [91] as well as having unique optimisation behaviour. Adopting a deep learning approach also allows for the training of the input signals, which leads to better performance and can be used to design systems to be deployed in unknown channel environments. It also boasts low system complexity and allows for communication system to operate with minimal human intervention. Deep learning based communication systems experience low power consumption.

Moreover, with regards to the pursuit of knowledge discovery, it is essential to develop algorithms that are efficient through the utilisation of the foundational structures inherent in channel information. In order to reduce the computational strain associated with the complex user clustering problem, non-optimal methods such as matching theory-based user scheduling [92]-[93] were developed. These algorithms aim to provide a solution by reducing the need for brute-force calculation. The aforementioned research projects, however, mostly focus on algorithm development while neglecting the incorporation of the learning feature.

Deep learning algorithms supply a novel and effective solution for improving the system performance of NOMA systems by making expert use of adaptive learning properties. Deep learning algorithms are also able to discover the link between channel information and system performance [94], [95].

Motivated by the above-mentioned features and benefits from applying deep learning to NOMA systems, [96] reviews the techniques that combine NOMA with the aforementioned technologies while maintaining an emphasis on several fields or areas of research where deep learning in NOMA can be applied. Said fields of research are: optimal power allocation using deep learning, user clustering using deep learning, and signal detection using deep learning.

2.8.3 Application of Deep Learning in Power Allocation

In order to effectively leverage the benefits of the NOMA system, power allocation with restricted resources is a crucial issue that must be addressed. This problem of optimal power allocation has proven to be Non-deterministic Polynomial time (NP)-hard, meaning that in order to obtain an optimal solution, all channel assignment permutations must be studied, which is impractical if not extremely computationally difficult and costly.

Researchers have put forth a variety of approaches to address this challenge. The proposed solutions encompass power allocation techniques for a downlink single-input single-output non-orthogonal multiple access (SISO-NOMA) system with two users [97]. These solutions address power allocation for achieving optimal user fairness [98] as well as maximising energy efficiency [99]. Nevertheless, a considerable number of these proposed solutions have been proven to be less than ideal, thus requiring the implementation of deep learning methods. In the subsequent section, a comprehensive literature review will be provided, focusing on deep learning approaches to address the power allocation problem. The review will go into the subject matter with adequate depth and analysis.

The analysed research has attracted significant attention due to recent technological developments in power allocation, namely in the field of NOMA, which incorporates deep learning techniques. The aim of this study is to examine the power allocation mechanism within a NOMA system, with a particular focus on its integration with deep learning techniques. The utilisation of an Attention-based Neural Network (ANN) is employed for the purpose of conducting channel assignment. [100] presents a method based on Deep Reinforcement Learning (DRL) for efficiently distributing power to customers. The approach utilises an attention-based Neural Network for the purpose of channel assignment.

The system model in a downlink NOMA scenario comprises of N subscribers and a single BS. The RL algorithm based on deep learning operates by designating the BS as the agent and the users as the performance environment. Prior to allocating channels and resources to users, the BS elects a course of action (channel assignment) from a predetermined collection. Subsequently, a feedback signal is transmitted to the BS to facilitate the allocation of users for the next transmission, based on the environment's response (i.e., the users). The aforementioned procedure consists of three fundamental components, namely the status space, the action space, and the reward function. The channel information is allocated to a state space, which is denoted by the user channel pairs. The action space pertains to the selection of a transmission channel for a singular user by the agent, specifically the base station.

In order to comply with the user channel allocation specifications, the quantity of actions is limited to ensure that each user is associated with a distinct action. The allocation procedure terminates after N iterations. The reward function is the feedback mechanism that is conveyed to the agent upon the completion of each time slot, indicating the outcome of the transmission, whether it was successful or unsuccessful. The signal comprises the data rates experienced by individual users and transmitted by the base station. The objective of [101] is to enhance the incentive signal, thereby optimising the data rates that individual users experience. The collected results demonstrate a comparison of overall rates. [101] proposes a system model for achieving sum-rate maximisation through the application of DL methodology in joint resource allocation. The performance of the DL-based approach is compared to that of a non-DL-based approach, revealing a significant advantage for the former.

The results encompass spectral efficiency, minimum data rates, and sum rates for different batch sizes and learning rates. Among them, the maximum performance is observed while using a batch size of 40 and a learning rate of 0.001. The results of all tests indicate that the DL-based method consistently achieves superior performance compared to its non-DL-based counterparts. In [101], a power allocation strategy is proposed that utilises DL approaches. The objective of this scheme is to optimise the system sum-rate in a downlink NOMA scenario where SIC is incomplete. The utilisation of an exhaustive search technique is employed in order to ascertain the most optimal distribution of power. A power allocation approach is given with the objective of maximising the total rate experienced by the system in the scenario of imperfect SIC. The proposed methodology employs DL techniques to anticipate the optimal power allocation variables by means of an exhaustive search approach. The system model consists of a singular BS that offers service to K users, each equipped with a single antenna. The BS is located at the central position within a cell, whereas the users are randomly dispersed across the cell.

The proposed inputs for the DNN include the Channel Response Normalised by Noise, the total transmission power, and the signal power percentage without SIC. The DNN incorporates all input data to effectively optimise the output, which consists of a set of factors related to power allocation. The obtained outcomes encompass a juxtaposition of the sum-rate in relation to the overall transmission power for two different signal power configurations in the pre-SIC scenario, involving two and three users. Additionally, the average Central Processing Unit (CPU) processing time is examined in relation to the total transmission power for the same two and three user scenarios.

The findings indicate that the proposed methodology yields performance that is close to optimal in a scenario involving two users. Furthermore, this level of performance is sustained even when there is a decline in performance within the range of low total transmission power. The processing time of the ideal non-DL system exhibits an exponential growth pattern as the total transmission power increases. In contrast, the proposed scheme demonstrates a low initial processing time that remains constant despite an increase in total transmission power. Furthermore, the simulation results illustrate that the proposed approach can get nearly ideal sum rate performance while greatly reducing computational complexity.

2.8.4 User Clustering Using Deep Learning

User clustering is a significant problem in NOMA systems. Users would be grouped together according to a clustering algorithm. Every individual user is then served by a NOMA beam. In a system like this, the problem of clustering the users in an optimal way presents itself. The aim being to cluster users in such a way that users in one group have

highly correlated channels (e.g: Indoor scenarios with dense users for an environment like a conference room or auditorium) while, simultaneously, having less correlation with the other users in the network. This will lead to an efficient use of the available resources since the system will experience much reduced interference while simultaneously optimising the throughput. Previous solutions to user clustering in NOMA include, dynamic user grouping in order to achieve better system throughput with better BER using a joint resource allocation algorithm [103], and user clustering for a downlink NOMA system that remarkably outperforms others in the literature [104]. The clustering problem, especially with a large number of users, is a combinatorial problem which is specifically why the application of deep learning is required to solve it optimally.

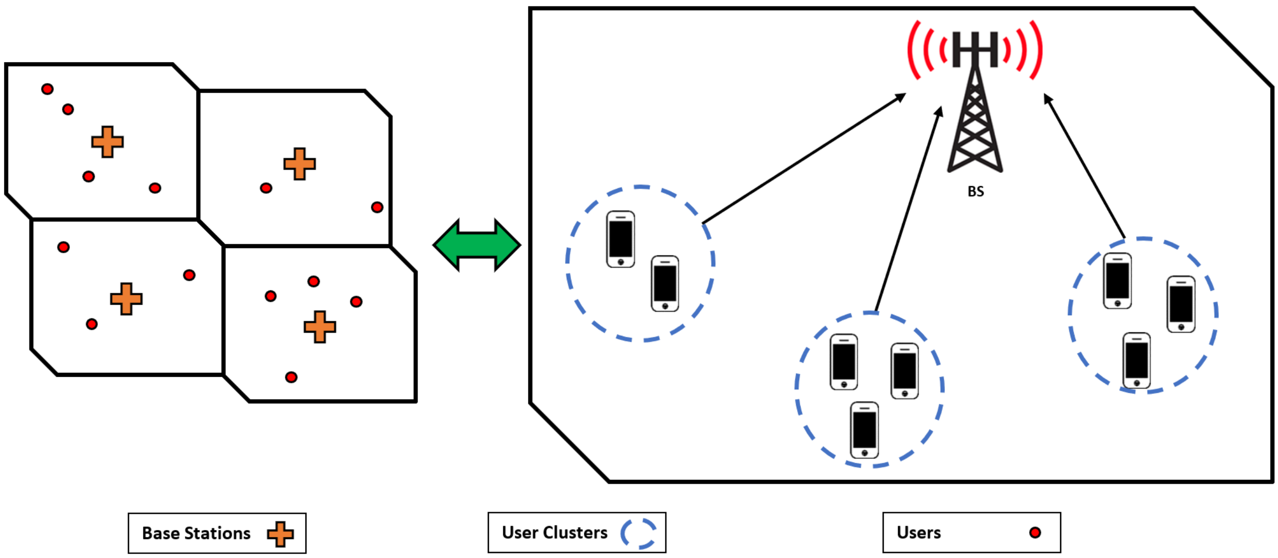


Figure 2.25: Illustrating Multi-cell NOMA uplink resource allocation by using optimisation algorithm to efficiently cluster users for each resource block at the base-station side[102]

[105] proposes a user clustering-based resource allocation with uplink NOMA techniques in a multi-cell system that performs user grouping based on network traffic using RL. The system model, illustrated by Fig.2.25, consists of N BS communicating with K IoT users via J orthogonal sub-channels. For each BS, the bandwidth is divided equally into orthogonal sub-channels. Users are grouped together following the NOMA principle, where more than one user can utilise one orthogonal sub-channel to transmit to the BS in an uplink scenario. The users start by transmitting the superposed signal containing the signals of both users occupying one sub-channel to the BS.

Then the BS applies SIC to sequentially decode each user's signals. The optimal user resource allocation and user grouping is considered as a Markov Decision Process (MDP) [106]. State-Action-Reward-State-Action Q-learning is used for light network traffic and DRL is used for heavy network traffic. The results obtained for this work show the proposed system is stable when deployed in different types of networks. The proposed algorithm is also shown to outperform its OMA counterpart.

2.8.5 Signal Detection Using Deep Learning

Originally, SIC has been widely employed as the prevailing approach for detecting NOMA signals. In the context of a downlink scenario, the generation of a transmission signal is achieved by means of Superposition Coding (SC), subject to a power constraint imposed at the BS. Subsequently, the impaired signal will employ the SIC technique at the receiver to decipher its communication, based on the User Equipment (UE) channel gain sequence within a specific user cluster. The SIC procedure shall be executed in accordance with the NOMA principle. SIC can be considered as an optimal detection scheme for MA from an information theoretic standpoint, as it enables the attainment of the multi-user capacity region in both the uplink and downlink. Several studies have employed SIC for signal detection. The aforementioned works encompass a collaborative approach that combines user activity and data detection using structured compressive sensing for NOMA [107]. Additionally, a scheme for detecting multiple users and signals in the terrestrial return channel with NOMA is proposed, which is based on Generalised Spatial Modulation [108]. Furthermore, a grant-free NOMA system is introduced, which incorporates joint user identification, Channel Estimation (CE), and signal detection [109].

Instead of SIC however, deep learning techniques are proposed as a viable substitute. Multi-layer NNs in particular are proposed to achieve signal detection in NOMA systems. As will become apparent through the rest of this section, there are many deep learning aided techniques proposed for signal detection in NOMA.

[110] proposes an online learning detection method for detecting users in large clusters that are utilising NOMA in a downlink scenario. The aim being the development of a sum space design that is robust against the variations of a changing wireless network environment. Such changes can deteriorate the performance of a non-linear adaptive filter. The results compare the proposed partially linear adaptive filter with non-linear adaptive filter (NLAF) and maximum mean square error (MaxMSE) SIC in terms of average Gray coded BER. The partially linear adaptive filter is shown to achieve better performance when compared to non-linear adaptive filters and MaxMSE-SIC.

The study in [111] presents a system model designed to blindly detect the modulation order of interference signals in a downlink NOMA system using a machine learning algorithm that is based on the Anderson-Darling test. The system model is comprised of one BS and N users in a downlink NOMA scenario. Based on the concept of PD-NOMA, the user signals are superposed, using a process called SC, and broadcast across the network.

The Machine Learning Algorithm based on Anderson-Darling test is divided into two phases, the training phase and the blind detection phase. The training phase consists of clustering, where all candidates of modulation order for interference user are used to cluster received constellation points, feature extraction, where accurate features of received constellation points are extracted, and classification training, where a model parameter of the logistic regression model is found. The blind detection phase consists of model parameter selection criteria, where one model for the classification model is chosen, and blind detection, where the output of the modulation order of interference user is presented.

The results obtained include blind detection rate comparisons and throughput comparison. The blind detection rate shows that the proposed scheme is more accurate than the conventional max-log algorithm on modulation order blind detection. The throughput comparison shows the proposed algorithm is close to the ideal curve and achieves larger gains compared to the max-log algorithm. The throughputs of NOMA and OMA are also addressed in the final result, due to the imperfect SIC, the throughput of NOMA maybe worse than OMA but the total throughput of NOMA is better than that of OMA in most SNR regions.

[112] proposes a learning technique that automatically evaluates the CSI of the network and detects the original transmit sequences. As opposed to traditional SIC methods, in which the search for the optimal order of channel gain must be done in order to remove the signal with the higher power allocation factor while detecting the signal with the lower power factor. The proposed DL algorithm combines the CE process while recovering the desired signal that is deteriorating from channel distortion and multiuser signal superposition. The system model, shown in Fig.2.26, consists of one BS and K number of users spread across N clusters. For brevity's sake, the work considers signal detection from users in a single cluster. The transmission process follows a conventional PD-NOMA, where the BS sends a superposed signal to users in the cluster. A DNN in applied in the signal detection process along with SIC to detect and decode each user's signal. Without any more signal processing at the receiver side, signals from the receive antennas are sent directly to the MIMO-NOMA-DL detector. As a whole, the MIMO-NOMA-DL system is comprised of three parts, a training block, a testing block, and a DNN detecting block.

The purpose of the training block is to generate the MIMO-NOMA signal and furnish the DNN with the corresponding NOMA signal labels. To obtain the MIMO-NOMA signal for the receive antennas, two sequences are generated for each antenna corresponding to user 1 and user 2. The technique of superposition coding is employed to modulate the signal, whereby distinct power factors are allocated to individual users. Subsequently, the signal is disseminated throughout the network. Upon experiencing the effects of both independent fading channel and the AWGN channel, the signal is subsequently obtained at the receiver. The labels, which are akin to the pilot sequence, are recognised by the recipient as the sequences in question. The utilisation of the testing block enables the emulation of real-time MIMO-NOMA transmission. Initially, the MIMO-NOMA signal is generated, wherein the labelling process is deemed unnecessary. The testing dataset is utilised to assess the efficacy of the deep neural network in detecting signals. In order to ensure optimal performance of the deep neural network (DNN) during both the training and testing phases, the channel models and generated data within the training and testing blocks are independently and identically distributed, thereby avoiding a perfect match. The training and testing process utilise different data sequences as to avoid overfitting.

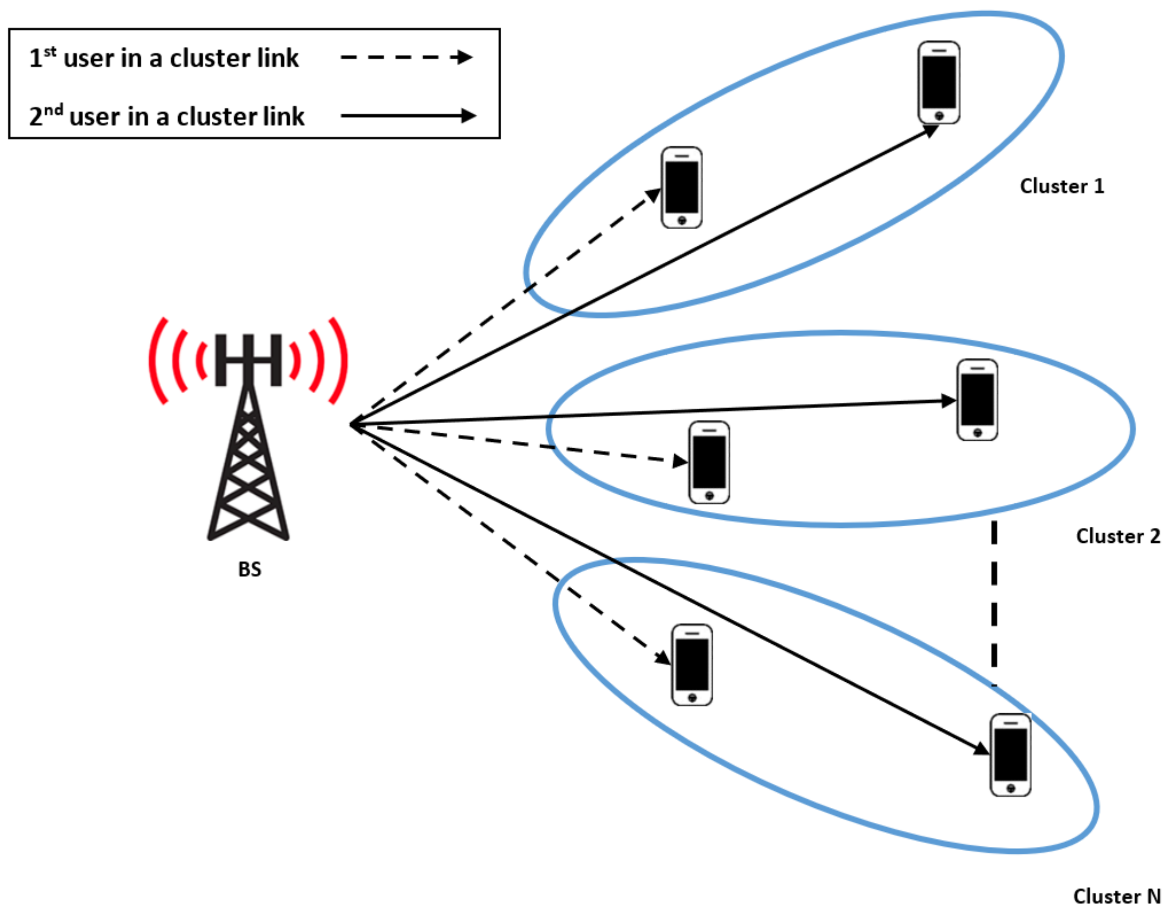


Figure 2.26: System model for user clustering in a MIMO-NOMA system.[38]

During the training phase, the SNR is stochastically generated, while the duration of the data time slot is varied within the desired range. The SNR within the testing block remains constant in order to evaluate the error of the DNN under certain SNR conditions. The primary component responsible for deciphering the received signal is the DNN block. The investigation focuses on the analysis of channel characteristics and the MIMO-NOMA decoding algorithm by effectively utilising the hyperparameters of the DNN in an optimal fashion. This block incorporates various levels, including the number of layers, the loss function, the optimisation criteria, and the iteration method. The DNN algorithm under consideration consists of a total of seven layers, specifically, one input layer, one output layer, and five hidden layers. The objective of the loss function employed in the proposed study is to quantify the disparity between the predicted values and the corresponding labels. The selection of the loss function and optimisation technique is of utmost importance for the MIMO-NOMA-DL network. The selected loss function is the usual MSE function. In this study, the cross-entropy function is being examined. The cross-entropy function exhibits rapid convergence and low computational complexity throughout the iterative optimisation procedure.

In addition, the optimisation algorithm employed for the self-adaptation of the learning rate and the enhancement of robustness is the Adam method [113]. Various parameters were utilised to obtain the results. In a particular transmission environment, a comparative analysis is conducted between the performance of the conventional SIC and the proposed method, MIMO-NOMA-DL, which utilises MIMO-NOMA-DL. The investigation also encompassed the impact of diverse categories of MIMO-NOMA modulations on the symbol error rate (SER). The simulation of the power allocation factor's impact was conducted and its influence on the system's performance was examined. Subsequently, an investigation was conducted into the scenario in which the approximated CSI diverged from the factual CSI. Various mini-batch sizes are utilised during the simulations to expedite the convergence of the MIMO-NOMA-DL algorithm. The findings demonstrate that the suggested algorithm is a feasible and efficient replacement for traditional SIC.

In [114], a demodulator utilizing a convolutional neural network (CNN) is proposed for Non-Orthogonal Multiple Access-Visible Light Communications (NOMA-VLC), facilitating simultaneous signal compensation and recovery. The objective of the proposed system is to mitigate the issue of error propagation, as well as linear and non-linear distortions caused by multipath propagation. Additionally, it aims to address the limitations imposed on the transmission performance of NOMA-VLC systems due to the nonlinearity of light emitting diodes and the constrained modulation bandwidth. Furthermore, the accuracy of CSI is crucial in the recovery of the NOMA signal, particularly in the context of a transportable VLC wireless system. However, obtaining precise CSI is a challenge in such a system mainly because of the dynamic environment, where any obstruction or change in environment can result in quick and dramatic variations in the channel, interference, since VLC is susceptible to interference from any other light sources, and multipath reflections, where indoor environments experience multipath reflection due to walls, floors, and ceilings which leads to a high degree of multipath dispersion. The system model comprises two components: a transmission side and a reception side. At the transmitting end, the data streams originating from the source data modules are further transformed into a 4-QAM format and then fed into the OFDM modulators. The resulting output is subsequently transmitted to the power allocation factor modules and then combined using superposition coding. Before being subjected to Intensity Modulation (IM) of the LED, the resulting NOMA signal is biased with Direct Current (DC). The optical NOMA signal is received in free space and subsequently detected by an optical receiver including a photo-diode and a trans-impedance amplifier.

At the recipient, subsequent to the process of frame synchronisation, the regenerated NOMA signal is introduced to the frame synchronisation module. Following this, the correlation operation is performed, resulting in the detection of the frame header. The NOMA signal is decoded through direct utilisation of a CNN based demodulator. In traditional receivers based on SIC, each user performs message decoding after the decoding of messages from users with higher power allocation factors, considering signals from other users with lower power allocation factors as noise. Also, if an error occurs where a signal in the SIC decoding chain is not decoded correctly, that error will propagate throughput the rest of the network. This phenomenon is referred to as an error propagation problem. In order to enhance performance, it is necessary to have precise CE in VLC systems. However, achieving accurate CE poses challenges due to user mobility and driving bias drift, which hinder the acquisition of reliable channel responses. In this work, the authors employ the MaxMSE based CE technique to compute the CSI in the SIC based receiver. The MaxMSE based CE approach is found to outperform other traditional CE methods. In the proposed CNN based demodulator, channel equalization is performed implicitly with signal demodulation, thus, it suffers less from the influence of the high correlation among user's channel responses.

Furthermore, no pilot overhead and CSI are required in the CNN based receiver. The results are obtained for the power-voltage characteristics for the LED as well as its frequency response where the simulated experimental results almost fit perfectly to the fitting curve. The average BER Vs SNR is obtained for the CNN, LSTM, and fully connected neural network, where the proposed CNN is shown to outperform its deep learning counterparts and achieve a much lower BER. The CNN is also compared to conventional SIC in a series of tests comprising of BER Vs SNR comparisons in different transmission environments. In general, the machine learning-based CNN suggested in this study demonstrates a notable level of performance, indicating that the concurrent attainment of offline training of NN channel properties and signal mapping is feasible.

[115] proposes a novel type of Active User Detection (AUD) based on a DNN for a grant-free NOMA system in an uplink scenario. The proposed AUD scheme aims to learn the non-linear mapping between the received NOMA signal and the indices of active users; hence, the proposed scheme can identify between active and idle users in the network, which is a significant problem for grant-free access NOMA systems. The system model consists of one BS equipped with a single antenna that receives information from multiple (N) machine-type devices with a single antenna.

During the transmission process, the mobile user chooses a frequency sub-band randomly from a set of free sub-bands. In this work, the overloaded scenario is considered, where the number of users in a cell is higher than the number of frequency resources. The BS is required to identify the active users from the inactive users since each user transmits packets freely without scheduling first. Active users transmit to the BS using device specific non orthogonal sequences, namely LDS, being that every codeword used for signal spreading contains lots of zeros. Each user signal is first modulated then the signal is spread using sparse low density codewords. As a result of using sparse codewords, each user has a more unique spread signal, thereby reducing inter-user interference drastically. The codewords are generated using a predesigned set of codewords from a codebook. The channels experience independent Rayleigh fading along with AWGN.

The proposed AUD scheme uses a deep neural network with one input layer, one output layer, and L hidden layers. The are many results obtained for this work. The probability of success comparison with SNR is presented to evaluate the performance of the proposed algorithm. The simulation is based on the grant-free NOMA transmission in OFDM systems. The probability of success as a function of SNR is presented, where it is shown that the proposed deep learning AUD algorithm outperforms the comparison benchmarks; MaxMSE-block orthogonal matching pursuit (BOMP), LS-BOMP, and approximate message passing (AMP) algorithm. The probability of success is further tested under various overloading factors as well as compared against varying device activity levels. The hyperparameter tuning process is also presented, including different factors such as, depth of hidden layers, width of hidden layers, batch size, dropout probability, optimiser, and activation function. Finally, the probability of success is presented as a function of SNR in a multi-antenna scenario.

In [116], the issue of pilot contamination attack (PCA) on NOMA in mm-Wave and massive MIMO in 5G communications and beyond is introduced. PCA detection is confronted with novel issues due mainly, to the novel characteristics of NOMA including but not limited to superposed signals with multiple users. Two efficient PCA detection techniques for NOMA are proposed for both static and dynamic conditions by making expert use of the sparse nature of NOMA as well as the statistics of mm-Wave and massive MIMO virtual channel. In the case of the dynamic channel condition, to differentiate between the normal state and the contaminated state, the statistics of the peak in the virtual channel is influenced. In the case of the static channel condition, as a binary hypothesis test of virtual channel sparseness, the problem of PCA detection is developed. To reach high detection performance, a machine learning, and peak detection algorithm-based technique are presented.

The system model being considered for this work consists of one BS and multiple users, some of which are pilot contamination attackers, in an uplink NOMA scenario. The simulation results obtained for this work showcase the performance of the proposed system under different scenarios, that being a varying number of PCA attackers. The case with one attacker achieving 92.1% accuracy and the one with two and three attackers reaching 96.15% and 96.85% respectively. Also, the detection rate can reach 100% with a very low false alarm rate in the case of the static conditions and it can reach higher than 95% in dynamic conditions under varying system parameters settings. For future research directions though, there are a number of issues with this work that can be addressed. Issues such as, the selection of the optimal threshold and the distance problem of attackers. Under the condition of legitimate users being very spatially close to the attackers, because of the correlation between the two existing, the performance of the virtual channel needs to be examined further. Also, deciding the optimal detection threshold is an issue to be addressed with the proposed machine learning algorithm.

[117] introduces a SIC technique incorporating the K-means clustering algorithm tailored for a Standard Single Mode Fibre (SSMF) integrated with multiple user On-Off Keying modulated

PD-NOMA systems. The adoption of NOMA in this study aims to augment user capacity. Empirical analyses conducted within this research underscore the efficacy of the advanced direct detection scheme for transmitting PD-NOMA signals at 20 Gbps at a wavelength of 1.55 um. Furthermore, simulation outcomes reveal that the unsupervised DL based direct detection receiver is capable of producing a BER substantially below the FEC threshold, as noted in [111] and [112]. This is evident for data rates of 5 Gbps, 5.5 Gbps, and 10 Gbps spanning a 50 km SSMF distance in the context of the far user scenario. The peak observed power penalty for the cumulative 20 Gbps per channel data rate, when employing the DL methodology, registers at 2.7 dB at a BER of 10^{-3}, again pertaining to the far user scenario. However, the proposed technique does not include any compensation method for handling fibre induced damage while still detecting the data for the near and far user, which is usually 50 km away from the BS. This can be strongly considered an avenue for future research direction this work can be extended to.

In [118], the use of deep learning is explored in the context of uplink Multi-User Detection (MUD) for NOMA technology, specifically focusing on the Welch bound equality Spread Multiple Access. Multiple non-cooperative users are allocated individual NOMA signature sequences, which are then sent over the same resource. The signature sequences stated above exhibit a limited degree of correlation among themselves, hence facilitating user separation at the receiver during MUD. Several key subtasks are associated with MUD, including combining, slicing, interference cancellation, and signal reconstruction. The NN offers a one-shot estimation for these modules, while also replacing the clearly defined receiver blocks with a singular opaque entity. As a result, the performance of the proposed neural networks is evaluated in comparison to conventional receivers. This work explores two distinct implementations of supervised feed forward neural networks for MUD: a deep learning based NN and a 2D-convolutional NN. The performance of the two proposed neural networks is compared with their traditional equivalents in terms of SER.

[118] focuses on a system model that includes a single BS and K users in an uplink NOMA scenario. Each of the K users modifies its sent symbol by a preassigned spread-vector before to transmitting it over the shared uplink channel. The signal received at the access point can be described as a noisy representation of the combined transmit vectors from all users, who share the same communication resources based on the principle of NOMA. The neural network used for MUD consists of two distinct phases: a training phase and a testing phase. The objective

of the training phase is to adequately construct and optimise the NN in anticipation of the subsequent test phase. Furthermore, it is during the testing phase that the neural network's function as a detector is actualized. Prior to being inputted into the neural network, the received signal and channel estimations undergo pre-processing in both phases. In this scenario, the process of pre-processing entails the manipulation and combination of the received signal and the channel estimations in a suitable manner to conform to the input dimension of the neural network, which can be either a vector or a matrix. The neural network being examined in this study is presumed to have the capability to process inputs that are expressed as real numbers. Hence, the conversion of complex values into real-valued numbers constitutes the ultimate stage of pre-processing.

The NN consists of multiple hidden layers, each containing trainable parameters known as biases and weights. These parameters are optimised during the training phase. There are multiple layers in the system, each characterised by a distinct number of nodes. Additionally, there are forward connections established between nodes belonging to different layers. The architecture of a NN is primarily concerned with the manner in which its layers are interconnected, configured, and organised to establish a coherent framework. The input is sequentially processed through the fully connected layer, the activation layer, and the Batch-Normalisation (BN) layer. The BN layer ensures appropriate scaling of the input flowing through the NN, while the activation layer introduces non-linearity and does not require any trainable parameters. The Rectified Linear Unit (ReLU) is a commonly employed activation function for hidden layers. It operates by generating an element-wise output for a given input, where the output is determined by the largest value between the input and zero. Several activation layers, such as BN, ReLU, and fully connected layers, are iteratively applied in a sequential fashion until the output is obtained. In the context of a classifier, the selection of the activation function for the output layer is often either a softmax or a sigmoid function. Simultaneously, the determination of the number of nodes in the output layer is dependent on the specific functionality of the NN.

The findings of this study indicate that by careful selection of NN parameters, the black box estimation can achieve superior and more efficient performance compared to standard MUD approaches. Furthermore, the performance of the system attains a SER that is nearly identical to the ultimate symbol error rate achieved by the sophisticated maximum likelihood-based detectors.

[119] starts with a preliminary examination of DL for signal detection for multiple users in

a NOMA wireless communication system. Typical in NOMA systems, where multiple users' messages are decoded sequentially, the SIC process is carried out at the receiver. However, due in large part to the effects of error propagation, the correct detection of previous users will decide the detection accuracy of the entire system. Designed to detect and decode user messages for multiple users in one-shot process, a DL-based NOMA receiver is designed, without approximating channels in an explicit manner.

Jointly carrying out signal detection and channel approximation, the DL-based NOMA receiver is represented by a DNN. The transmission process is divided into two phases, an offline training phase and an online deployment phase. During the offline training phase, the DNN is trained using simulation data that is based on channel statistics. During the online deployment phase, the simulation data provided in the previous phase is used to recover the transmitted symbols in a direct manner. The system model under consideration in this work is compose of two users in an uplink NOMA scenario in an OFDM system. Both users are sharing the same frequency bands according to the NOMA principle to transmit data in a simultaneous manner. In this work, the DNN for MUD is comprised of 5 layers: an LSTM layer, a SoftMax layer, an input layer, a fully connected later, and a classification layer. Able to make expert use of data time-dependencies, the LSTM is the core component of the DNN and is also a kind of RNN that is usually deployed to handle sequences and time series data for classification purposes. An LSTM base network has the ability to preserve relevant information as well as being able to learn new information between time steps of sequence data. In the LSTM-based DL model for this work, the time steps are equivalent to subcarriers in the OFDM system. The DNN can be trained to realize MUD for an arbitrary subcarrier by focusing on the one time-step module in the LSTM layer.

The preliminary results obtained for [119] demonstrate that the DL based approach to signal detection has the capability to achieve superior performance when compared with traditional pilot-based channel approximation techniques. It is also shown to be more robust to the number of pilot symbols. Furthermore, the DNN is shown to be capable of handling the likely event of error propagation as a result of SIC occurring in the receiver. The DL solution to signal detection can even outperform the ML detector, which does not factor in the effects of interference, in the case of high inter-symbol interference. For future research directions, the proposed DL algorithm for signal detection can be further tested under dynamic channel conditions.

Although NOMA has great potential as a prospective solution for wireless communications in the future, its inherent non-orthogonality poses a challenge in terms of achieving its performance limit alone through conventional communication theoretic techniques. In [120], for end-to-end optimisation, deep multi-task learning is sought as solution through regarding the overlapped transmissions as multiple distinctive but correlated learning tasks. A unified multi-task DNN framework for NOMA is first created that is aptly called DeepNOMA. The DeepNOMA system comprises three main modules: DeepMUD, a MUD module; a channel module; and DeepMAS, a Multiple Access Signature (MAS) mapping module. DeepMAS and DeepMUD are trained using an automated approach that relies on data-driven methods.

A multi-task balancing method is then proposed to ensure fairness amongst tasks and to avoid the local optima. Furthermore, in order to make expert use of the advantages of the communication-domain expertise, a constellation shape prior is introduced as well as an inter-task interference cancellation structure. These modules are then inserted into the DeepMAS and DeepMUD design structure. The aforementioned advanced designs aid in reducing the deployment complexity of the proposed system and algorithms without negatively impacting the DNN's universal function estimation property, which, in turn, makes DeepNOMA a widely applicable universal transceiver optimisation method. Deep multi-task learning achieves inductive migration among multiple associated tasks, as opposed to traditional DL methods which optimise a specific and single task. One task may supply inductive bias into the system to other tasks, which encourages the parameters to converge while maintaining a better level of generalisation. Due to its improvements in system performance, deep multi-task learning is now universally deployed in the machine learning community in order to acquire better collaboration amongst tasks [121].

The findings of this study demonstrate that DeepNOMA can achieve improved transmission accuracy and reduced computing complexity concurrently across different channel types. It is also shown to outperform its conventional counterparts when it comes to operative in a dynamic environment. For future research directions, the proposed work can be extended for the multi-antenna case and MIMO. For this case the parameter sharing and recurrent network structure could be made expert use of in order to lower the overall computational complexity.

2.9 Methodology for Research

This thesis has presented key works of literature concerning the subject of NOMA in the next generation of wireless communication systems. Every relevant piece of work was reviewed in a concise manner, outlining at the end of each topic covered the challenges facing the area of NOMA under review. Three areas were chosen as the focus for the review as well as future research: 1) NOMA in cooperative relaying, 2) NOMA with EXIT chart analysis, and 3) NOMA with deep learning techniques.

2.9.1 NOMA in Cooperative Relaying Wireless Communication Systems

Over the course of the material covered, the following challenges for cooperative relaying in NOMA were identified:

Spectral Efficiency Optimisation for User-to-User Relaying

In order to conduct cooperative relaying effectively without experiencing inter-user interference, cooperative transmissions are divided into a transmission phase and a cooperative phase. Having to assign the relaying phase it's own slot out of the available bandwidth has networks employing cooperative schemes requiring more total bandwidth to function. One solution suggested by the literature is to deploy a composite scheme that combines cooperative relaying with full duplex technology, which refers to a communication system that carries out transmission and reception simultaneously between two devices, in a NOMA enabled network. The idea is to have the cooperative relaying occur simultaneously with the normal direct phase transmissions, thus eliminating the system's need for more bandwidth as the cooperation occurs within the same allotted bandwidth as both phases now share the resource. Implementing this scheme, however, requires answering the self-interference problem of FD, which is planned to be the subject of a future collaborative work.

2.9.2 NOMA with EXIT Chart Analysis

User Fairness Analysis

EXIT chart analysis can be utilized to measure the level of user fairness for power allocation in NOMA networks by observing the convergence behaviour in the inner/outer decoder convergence graph. If the inner and outer decoders converge at unity gain then the system is considered fair and balanced in terms of power allocation.

Convergence Behaviour

EXIT charts have evolved into a valuable tool for analysing the convergence characteristics of models by employing iterative decoding techniques [30–32]. The attainment of an exceptionally low BER at the SNR can be predicted by the utilisation of EXIT charts, eliminating the need for Monte Carlo Simulations [30, 31]. The implementation of EXIT charts in the NOMA system involves the exchange of information between the input/output module of MUD.

2.9.3 NOMA with Deep Learning Techniques

Power Allocation

To safely exploit the features of the NOMA system, the distribution of power with restricted resources becomes a crucial concern. The issue of achieving optimal power allocation has been demonstrated to be NP-hard, indicating that a comprehensive examination of all possible channel assignment alternatives is necessary to get an optimal solution. However, this approach is impracticable due to its high computational complexity and associated costs. Consequently, a multitude of methods have been proposed by researchers to effectively tackle this matter. The proposed solutions encompass power allocation techniques for a downlink Single-Input Single-Output (SISO) NOMA system with two users [27]. Additionally, power allocation methods for achieving optimal user fairness [28] and maximising energy efficiency [29] are also considered. Nevertheless, a considerable number of these proposed solutions have been proven to be less than ideal, thus requiring the implementation of deep learning methodologies.

Signal Detection

Until recently, SIC has been widely employed as the prevailing technique for detecting NOMA signals. In the context of a downlink scenario, a transmission signal is generated by SC while adhering to a predetermined power constraint at the BS. Subsequently, the impaired signal will employ the SIC technique to decipher its message, contingent upon the sequence of channel gains of the user UE within a certain group of users. The SIC process will subsequently be executed in accordance with the NOMA principle. From an information theoretic standpoint, SIC is considered an ideal strategy for detecting multiple access in terms of the attainable area of capacity for many users, both in the uplink and the downlink.

Examples of works that have utilised SIC for signal detection include, joint user activity and data detection based on structured compressive sensing for NOMA [70], generalised spatial modulation-based multi-user and signal detection scheme for terrestrial return channel with NOMA [71], and joint user identification, CE, and signal detection for grant-free NOMA [72].Instead of SIC however, deep learning techniques are pro-posed as a viable substitute. Multi-layer NNs in particular are proposed to achieve signal detection in NOMA systems.

User Clustering

User clustering is a significant problem in NOMA systems. Users would be grouped together according to a clustering algorithm. Every individual user is then served by a NOMA beam. In such a system, the problem of clustering the users in an optimal way becomes evident. The aim being to cluster users in such a way that users in one group have fully occupied a resource with minimal to no overlapping.

Illustrating Multi-cell NOMA uplink resource allocation by using optimisation algorithm to efficiently cluster users for each resource block at the base-station side [32] then transmitting over highly correlated channels while, simultaneously, having less correlation with the other users in the network. This will lead to an efficient use of the available resources since the system will experience much reduced interference while simultaneously optimising the throughput.

Previous solutions to user clustering in NOMA include, dynamic user grouping in order to achieve better system throughput with better BER using a joint resource allocation algorithm[49], and user clustering for a downlink NOMA system that remarkably outperforms others techniques proposed in [50]. The clustering problem, especially with a large number of users,is a combinatorial problem which is specifically why the application of deep learning is required to solve it optimally.

2.10 Summary

Incorporating other techniques with NOMA in order to address some of its shortcomings was thoroughly discussed in this chapter. It has been shown that, while each combination of techniques results in improved performance, each technique, when combined with NOMA, addresses an issue facing both NOMA and the next generation of wireless communications.

For example, NOMA with cooperative relaying was shown to be able to serve cell edge users more efficiently, NOMA with EXIT chart analysis has better user fairness during the power allocation stage, and NOMA with deep learning techniques has the potential to result in more optimal power allocation, more efficient user clustering, and more reliable signal detection with reduced receiver complexity.

Chapter 3

Cooperative Non-Orthogonal Multiple Access for 5G Networks and Beyond

3.0.1 Introduction

In addressing the pressing requirements for enhanced data rates and augmented user capacity, NOMA emerges as a promising solution for forthcoming wireless communication architectures. This is attributed to its superior diversity gains and potential for vast connectivity. This work introduces a cooperative relaying approach designed to augment the overall data rates and diversity gains of the NOMA-centric system. Concurrently, EXIT charts are employed to probe the user fairness of the proposed system and to assess its performance under IRregular Convolution Coding (IRCC). Notably, the EXIT chart incorporating IRCC offers insights into the convergence behaviour of the aforementioned system. This research harnesses EXIT charts for convergence optimization by leveraging power optimization as well as SIC for combined data rates, thereby enriching the system's fairness evaluation. The simulations performed underscore the efficacy of the proposed cooperative NOMA framework in attaining markedly superior data rates and diversity gains.

As a result of the work carried out in this chapter, the following paper(s) were published:

- A. Ahmed, Z. Elsaraf, F. A. Khan and Q. Z. Ahmed, "Cooperative Non-Orthogonal Multiple Access for Beyond 5G Networks," in IEEE Open Journal of the Communications Society, vol. 2, pp. 990-999, 2021.

3.0.2 System Model

The user data is passed to the QAM, which maps the information bits. These bits are spread using predefined codewords and assigned a power in accordance with NOMA principles. This information is conveyed across a Rayleigh fading channel to the receiver. The MUD on the receiver side compensates for channel-caused interference. As the receiver knows the predesigned code, the signal is despread. As depicted in Fig.3.1, MRC is applied following demodulation of the user bits.

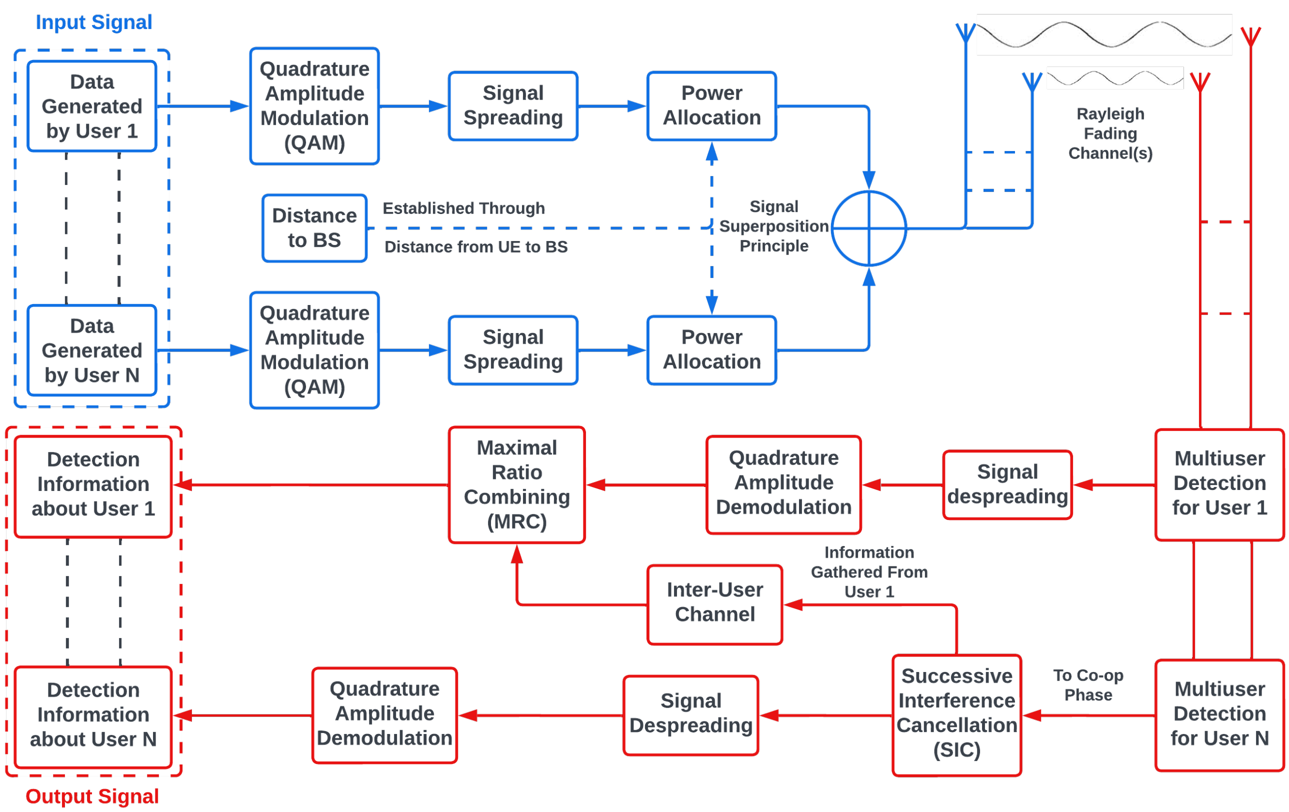


Figure 3.1: Proposed cooperative NOMA system

Fig.3.1 depicts the cooperative communication for NOMA. It can be seen from the figure that there are two distinct sorts of users. There are N NU positioned in close proximity to the BS and one Far User (FU) in the system. Phase I and Phase II are the two phases of communication.

4-QAM (or QPSK) was used to carry out signal modulation in this system model. This specific modulation scheme was chosen because of the following factors:

- **Robustness in Noisy Environments:** In terms of BER performance, 4-QAM outperforms higher-order QAM configurations such as 16-QAM or 64-QAM when they operate at an equivalent SNR. This characteristic renders 4-QAM appropriate for scenarios where the SNR may be low or inconsistent.

- **Affinity with PD-NOMA:** In PD-NOMA, superposition coding is utilized, allowing for the concurrent transmission of signals intended for distinct users at varying power levels. The inherent simplicity and resilience of 4-QAM facilitate the differentiation and decoding of overlapping signals at the receiving end, particularly when a pronounced power gap exists between those signals.

The NOMA cooperative communication operation is depicted in Fig.3.2. It can be seen from the figure that there are two distinct categories of users: N number of NUs who are located closer to the BS, and FUs who are located further away. There are two phases to the telecommunications procedure: phase I and phase II.

- **Phase-I:** As depicted in Fig.3.2, the BS transmits the signal to every NU and the FU. This signal is composed of the superimposed signals of the NUs in accordance with the NOMA principles. The nth NU user signal is represented by:

$$r_{n,1} = h_{n,1}x + \sqrt{P_{FU}}h_{n,1}x_{FU} + w_{n,1}, \tag{3.1}$$

  $\sqrt{P_{FU}}h_{n,1}x_{FU}$ represents the FU signal, where P_{FU} denotes the allocated power to the FU and x represents the total composite transmit signal that contains all the users', NUs and FU, messages. The composite signal is structured as follows:

$$x = \sqrt{P_1}x_1 + \sqrt{P_2}x_2 + ... + \sqrt{P_N}x_N, \tag{3.2}$$

  $w_{n,1}$ represents the AWGN for each channel that naturally occurs as a result of transmission over a fading channel. The Rayleigh fading of the utilised channel is represented in this model by $h_{n,1}$. This type of fading is experienced by every user in the network.

  The power allocation process during this phase is carried out by first assigning the distance of the n-th user away from the BS, where one user, FU, is placed far away from the BS

(10m) and the other user, NU, is placed very close to the BS (1m). In this model, the user farthest away from the BS is allocated the most power while the user closest to the BS is allocated the least amount of power. The aim of this protocol is to allocate power to users based on their subjective needs to receive their messages successfully. The channel conditions of the proposed NOMA system are presented as:

$$|h_{FU,1}|^2 < |h_{1,1}|^2 < |h_{2,1}|^2 < ... < |h_{N,1}|^2 . \tag{3.3}$$

The power levels in the proposed system are consequently presented as:

$$P_{FU} > P_1 > P_2 > ... > P_N, \tag{3.4}$$

and

$$P_{FU} + \sum_{n=1}^{N} P_n = 1. \tag{3.5}$$

- **Phase-II:** This phase is sectioned into N time slots by assigning each NU a time slot to transmit a superposed signal to the FU. All NUs make use of their allocated time slot to relay data to the FU. The order regarding which NU transmits in negligible, as long as all NUs have transmitted to the FU by the last time slot. In order to avoid and thereby eliminate IUI, every transmission is carried out in a separate fashion.

  This phase begins by having the n^{th} NU broadcast the composite signal it received at the end of phase-I to the FU in the n-th time slot. The purpose being to transmit the data of the FU from different sources thereby increasing the diversity gains experienced at the FU since the FU does not need to carry out SIC and can decode its message directly as a result of being allocated more, if not most, of the available transmission power. This is where the cooperative relaying takes place.

  After this process is finished, the signal experienced by the FU as a result of the cooperative relaying will be presented as:

$$r_{n,2} = \sqrt{P_{n,2}} h_{n,2} x_{FU} + w_{n,2}, n = 1, 2, ..., N. \tag{3.6}$$

The power allocated by the n^{th} user is denoted by $P_{n,2}$ while the Rayleigh fading experienced as a result of the signal passing through the inter-user channels is denoted by $h_{n,2}$ and the AWGN is denoted as $w_{n,2}$.

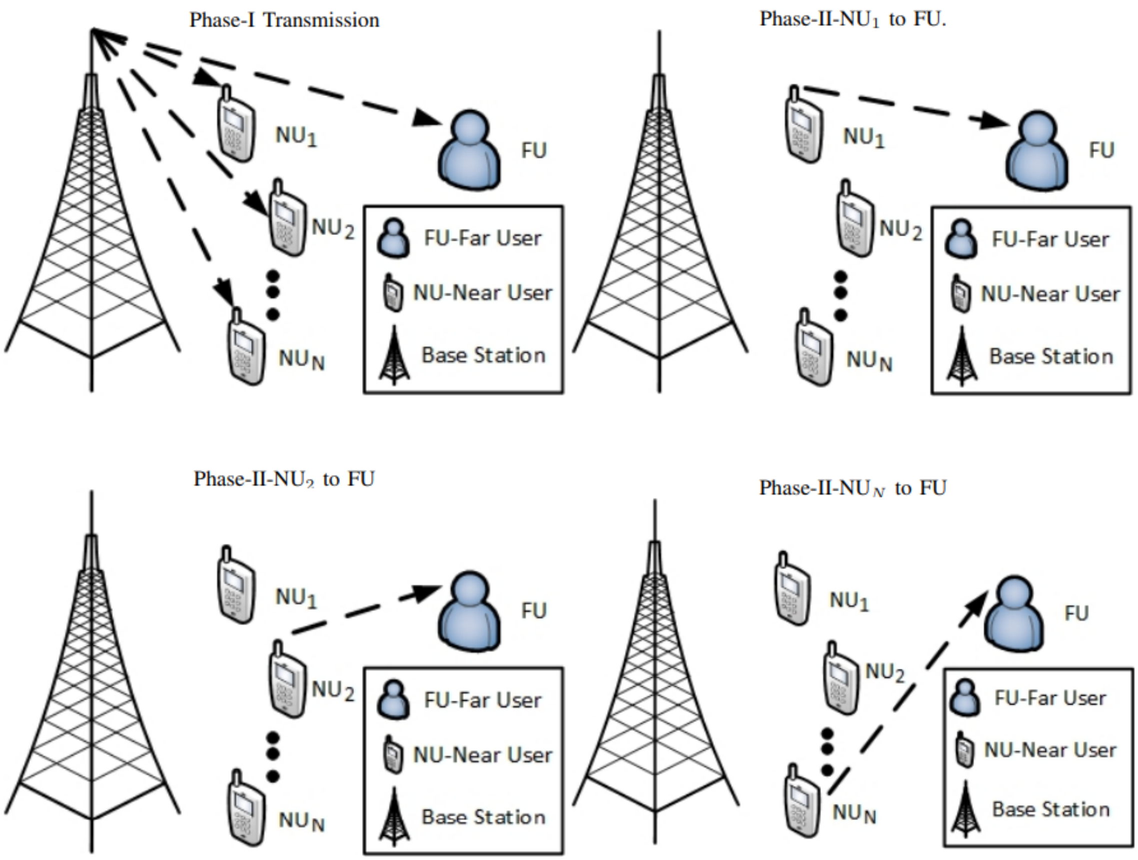


Figure 3.2: The cooperative relaying NOMA system comprises two distinct phases, namely Phase-I and Phase-II.

Both Phases, I and II, are represented visually by Fig.3.2.

Factors Affecting Processing Time for Various Cooperative Relay Models

The processing time for various cooperative relay models in a wireless communication system can vary significantly based on multiple factors, including the specific model, system parameters, and hardware used. Cooperative relay models involve additional signal processing steps beyond direct communication between the source and destination, which can introduce additional processing delays. Here are some common cooperative relay models and factors affecting their processing time:

1. **Amplify-and-Forward Relay:**
    - In AF relay, the relay amplifies the received signal and forwards it to the destination.
    - Processing time depends on the complexity of amplification and forwarding, as well as the hardware used.
    - Typically, AF relay processing introduces minimal delay, especially if the amplification is straightforward.
2. **Decode-and-Forward Relay:**
    - In DF relay, the relay decodes the source's signal, re-encodes it, and then forwards it to the destination.
    - Processing time is influenced by the complexity of decoding and encoding, and it may introduce more significant delays compared to AF relays.
3. **Compress-and-Forward Relay:**
    - Compress-and-Forward relay models involve data compression before forwarding.
    - The processing time depends on the compression algorithm and the hardware used.
    - Complex compression algorithms may lead to longer processing times.
4. **Network Coding Relay:**
    - Network coding relays combine multiple source signals and transmit a coded signal to the destination.
    - The processing time depends on the encoding and decoding of network-coded signals.
    - The complexity of network coding can vary, affecting processing time.

Factors influencing processing time:

- **Hardware and computational resources:** The processing time can be influenced by the computational capabilities of the relay nodes. More powerful hardware may reduce processing time.
- **Transmission rates:** Higher data rates may require more processing time for coding and decoding operations.

- **Coding complexity:** The complexity of the coding and modulation schemes used in the relay models can impact processing time.

- **Relay locations:** The distance between the source, relay, and destination nodes can affect processing time, as it determines the propagation delay for signals between nodes.

- **QoS requirements:** Different applications may have varying requirements for processing time. Low-latency applications, such as real-time multimedia, demand shorter processing times.

In practice, the processing time for cooperative relay models is typically minimized to ensure efficient communication. Various optimizations, including hardware acceleration and efficient coding techniques, can be applied to reduce processing delays. The specific processing time for a cooperative relay model should be determined through simulations, measurements, or performance analysis, taking into account the system's parameters and the application's requirements.

List of Assumptions

Designing a cooperative NOMA system involves several considerations, and to create a tractable analysis or efficient design, certain assumptions might need be made. When designing the system model for this work, the following assumptions were made:

1. **Channel State Information (CSI):** It was assumed the BS and users have perfect knowledge of channel conditions. This assumption was made in order to simplify the power allocation process.

2. **Decoding Capability:** It was assumed that all users in the network are well equipped to carry out signal detection in a NOMA system by performing SIC.

3. **Power Allocation:** Since the power allocation is carried out based on distance from the BS for each user and in order to take full advantage of the NOMA scheme, the difference of power levels between both users (FU and NU) was designed to be wide.

4. **Cooperative Relaying:** It was assumed that the NU, when acting as a relay in the second phase of transmission, can perfectly decode and forward the signal of the FU.

5. **Time Synchronization:** It was assumed that all users are perfectly synchronized in time since timing synchronization is critical for NOMA to operate successfully.

6. **Gaussian Noise:** All background noise was modeled after AWGN.

7. **Antenna Devices:** It was assumed that both the BS and both users in the network are equipped with a single antenna.

8. **Number of Users:** The number of users in the proposed system was fixed to 2. This was done in order to simplify the SIC decoding process.

3.0.3 Performance Analysis

In this section, we operate under the assumption that ideal detection occurs during both of the communication phases. The overall performance of the system is measured through the evaluation of the data rates experienced at each user and the total system throughput. EXIT chart analysis is also utilised here to evaluate the convergence behaviour of the system and the system performance is further measured by diversity gains, normalisation throughput, and online evaluating complexity in MUD.

- **Data Rate:** The data rates experienced at the end of phase-I at the NUs after applying SIC is presented as:

$$R_n = \log_2\left(1 + \frac{P_n\left|h_{n,1}\right|^2}{\sum_{j=n+1}^{N} P_j\left|h_{n,1}\right|^2 + N_{0_{n,1}}}\right) \tag{3.7}$$

  The noise variance is denoted by $N_{0_{n,1}}$ and takes effect at the end of phase-I.

  By applying MRC at the receiver, the expression for the data rate of the FU at the end of phase-II can be presented as:

$$R_n = \log_2\left(1 + \frac{P_{FU}\left|h_{FU,1}\right|^2}{\sum_{n=1}^{N} P_n\left|h_{FU,1}\right|^2 + N_{0_{FU_1}}} + \frac{1}{N}\sum_{n=1}^{N}\frac{P_{n,2}\left|h_{n,2}\right|^2}{N_{0_{FU_2}}}\right). \tag{3.8}$$

  The data rate experienced by the FU relies on the addition of two data rates from two different time slots. The data rate experienced in the second time slot is normalised by N as a result of passing through the MRC first. $h_{n,2}$ represents the data relayed to the FU during the second phase (cooperative phase) and is the signal that the NU relays through the inter-user channel that contains the data of the FU. As opposed to the signal received by the FU at the end of the first phase, which contains the data of both the NU and FU that is superposed together, $h_{n,2}$ contains only the data of the FU.

- **System Throughput:** The system throughput is described as the collection of data rates received by each user. The system throughput is a significantly imperative metric to measure as it can comment on the overall efficiency of the proposed system. As per its definition, the system throughput is presented as:

$$S = \sum_{n=1}^{N} R_n + R_{FU} \tag{3.9}$$

- **EXIT Chart Analysis:** EXIT charts employ iterative decoding to analyse and monitor the system's convergence analysis [69].

  EXIT charts provide a means to bypass the need for Monte Carlo Simulations, since they enable the straightforward prediction of the BER for a specific SNR [69], [70]. EXIT charts facilitate the sharing of information between the MUD module by making modifications to the channel conditions. Furthermore, the MUD module calculates the probability of the joint alphabet. The computation of these probabilities can be performed offline by both the base station and the user.

  Interleavers are employed for the purpose of emulating EXIT charts. In an ideal scenario, the utilisation of shorter interleavers in any given system would result in a trajectory that does not align with the inner and outer arcs of the EXIT chart [66], [68], [71]. Hence, interleavers of increased lengths are employed.

  In Figure 3.2, it is observed that there are N NU users and 1 FU. The channel encoder employed in this scenario utilises Recursive Systematic Convolution (RSC) codes. The aforementioned data is transmitted to the signal spreader, whereupon it is encoded utilising the principles of spread spectrum.

  The data is transferred to the interleavers, where it is mapped by the signal mapper, and subsequently broadcast via the antenna. The receiver obtains the mapped signal information for each user, and subsequently the MUD computes the Log Likelihood Ratio

(LLR). The evaluation of the extrinsic LLR of the output bit from the nth user's MUD is conducted in the following manner:

$$L_{ex}\left(i^{n}\right)=\ln\frac{\sum_{r\in v(0)}P\left(v\mid r\right)P\left(r\right)}{\sum_{Y\in v(1)}P\left(v\mid r\right)P\left(r\right)}-L_{apr}\left(i^{(n)}\right),$$

Let $v\in\{0,1\}$. The *apriori* LLR for the n^{th} user can be represented by $L_{apr}\left(i^{(n)}\right)$

It's important to highlight that, in the MUD context, the initial probabilities for bits being zero and one are equally likely. Therefore, initial values in the soft registers are set to zero, as pointed out in [68] and [71]. The *apriori* symbol is symbolized by P_r, signifying the multi-user probability. Consequently, the codeword can be represented as $\mathbf{r}=\left[r^{(0)},r^{(1)},...,r^{(N)}\right]^{T}$.

The channel probability relative to the *apriori* $\mathbf{r}$ is denoted by $P\left(v\mid r\right)$, which assesses the cost function for the MUD, defined as:

$$f\left(r\right)=P\left(v\mid r\right)P\left(r\right), \tag{3.10}$$

and

$$f\left(r\right)=\exp\left(-\left\|v-PHr\right\|^{2}\right)P(r). \tag{3.11}$$

Let P denote the aggregate power allocated, and H symbolize the channel conditions for all network participants. Prior to delivering the *apriori* LLRs to the channel decoder, a despreading process is executed. This ensures that the extrinsic LLRs for each user are ascertained through the deinterleaving action applied to the user input sequence.

The channel decoder produces bit-centric *aposteriori* LLRs. These are subsequently forwarded to the signal spreader, where further iterative processes occur between the interleavers and the MUD. Following multiple cycles involving both the Decoding (DEC) and the MUD-Despreading/Spreading (DES) stages, the decoder converges to a point where the receiver derives estimations for the information bits pertaining to each user.

- **System Performance:** This section will examine the diversity gain, normalisation throughput, and complexity associated with the MUD approach in the NOMA model.

1. **Diversity Gain:** It equals the number of given slots and is stated as:

$$\eta_{DG} = N + 1, \tag{3.12}$$

Let N represent the cumulative count of NUs within the system. Furthermore, the system's diversity gain, which plays a role in ensuring fair treatment among users, is expressed as:

$$\eta_{IG} = N. \tag{3.13}$$

2. **Normalization Throughput:** The normalization throughput of the MUD system is calculated as:

$$NormalizationThroughput = \frac{R \cdot B}{(N+1) \cdot SF} \tag{3.14}$$

let $B = \sum_{i=0}^{N} b^{(n)}$ represent the total number of bits per MUD, with R denoting the coding rate and SF=2 indicating the Spreading Factor (SF). Given that the MUD's inner decoder encompasses the despreader and the deinterleaver, the system capacity can be represented as:

$$C = \frac{1}{N+1}\left(B - \frac{1}{2^B}\sum_{i=0}^{2^B-1} E\left[\log_2\left\{\sum_{j=0}^{2^B-1}\exp\left(\varphi^{(i,j)}\right)\right\}\right]\right).$$

where $\varphi^{(i,j)}$ is presented as:

$$\varphi^{(i,j)} = -\left\|PH + W\right\|^2 + \left\|W\right\|^2 \tag{3.15}$$

Where W denotes the noise matrices in the system.

3. **Online Evaluating Complexity in MUD:** The complexity in the MUD system emerges during an iteration of the Maximum A Posteriori (MAP) detector when the Cost Function (CF) is computed over the number of bits developed in the cycle. The calculation is performed as follows:

$$N^{(MAP)} = \frac{\prod_{n=1}^{N} 2^{b^{(n)}}}{\sum_{n=1}^{N} b^{(n)}} \tag{3.16}$$

let $b^{(n)}$ denote the number of bits per codeword in the MUD designated for the n^{th} user. Factors such as the count of users accessing the channel, the quantity of transmitting antennas, and the number of receiving antennas play pivotal roles in the real-time assessment for calculating the CF, thereby affecting the MUD's complexity. Furthermore, the complexity evaluations of both the $SISOMUD$ and the Hard-Input Hard-Output are represented as $O\left(\sqrt{N^{MAP}_{(CFEs/bits)}}\right)$ based on [71].

3.0.4 Simulation Results

Command and Control Information

- **Number of Relaying Slots:** Theoretically, the system model is designed to accommodate up to 3 users sharing one frequency band according to the NOMA principle. In practice, however, in order to minimise inter-user interference, the system model consists of one near user and one far user. Thus, one relaying slot is required since only the near user will participate in cooperative relaying.
- **Transparency with the Receiver:** The frame structure of the cooperative NOMA system is designed to include a predefined number of slots for relaying. This frame structure is adhered to by all network nodes, ensuring that both the transmitter and receiver know when and how many relaying slots are available.
- **Duration of Transmission:** The total duration of transmission has been calculated over a range of SNR values.

  Fig.3.3 shows that the total transmission time improves drastically with an increasing SNR value. This is due to there existing less interference and thereby less time is required to decode and detect each user signal. As expected, the far user has a much lower total transmission time since its signal gets detected easliy by treating all other user signals as noise. However, it can be noted that at 33 dBs, the total transmission time for the near user drops lower than the far user. This is due to the interuser interference that is experienced at high SNR values.

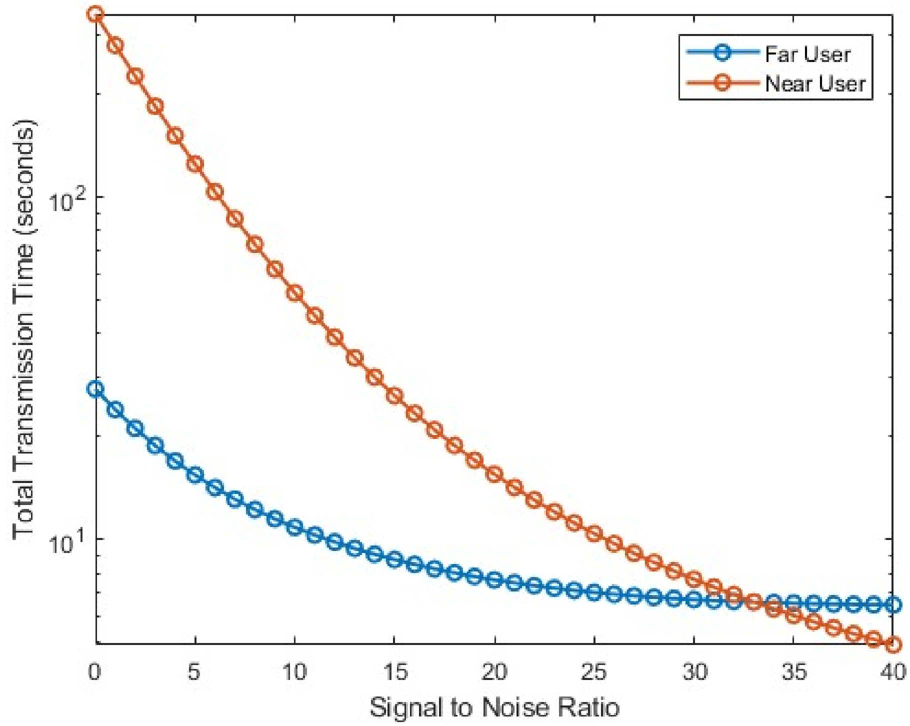


Figure 3.3: Total transmission time for the proposed cooperative relaying NOMA system for a range of SNR values

Figure 3.4 illustrates a comparison between conventional and cooperative NOMA techniques, specifically in terms of their SNR against BER performance. In this particular instance, the analysis focused solely on one FU and one NU. The graph demonstrates a positive correlation between the SNR and BER performance of the system, indicating that an increase in SNR leads to an improvement in system performance. The graph presented herein illustrates that the utilisation of cooperative communication in the FU yields a notably superior performance compared to FU systems that do not apply cooperation. Diversity is attained by means of the iterative reception of information from the NU throughout the Phase-II transmission process, leading to an enhancement in performance.

Figure 3.5 illustrates the contrast between traditional NOMA and cooperative NOMA in relation to data rate and system power allocation. α_1 represents the power splitting ratio between the NU and FU ranging from $0 \leqslant \alpha \leqslant 1$. In the context of typical NOMA deployment, it can be observed from Figure 3.5 that the system necessitates 3.5 bits for broadcasting at a power allocation of 0.1. This assertion is substantiated through an analysis of the intersection between FU and NU. In contrast, the utilisation of NU

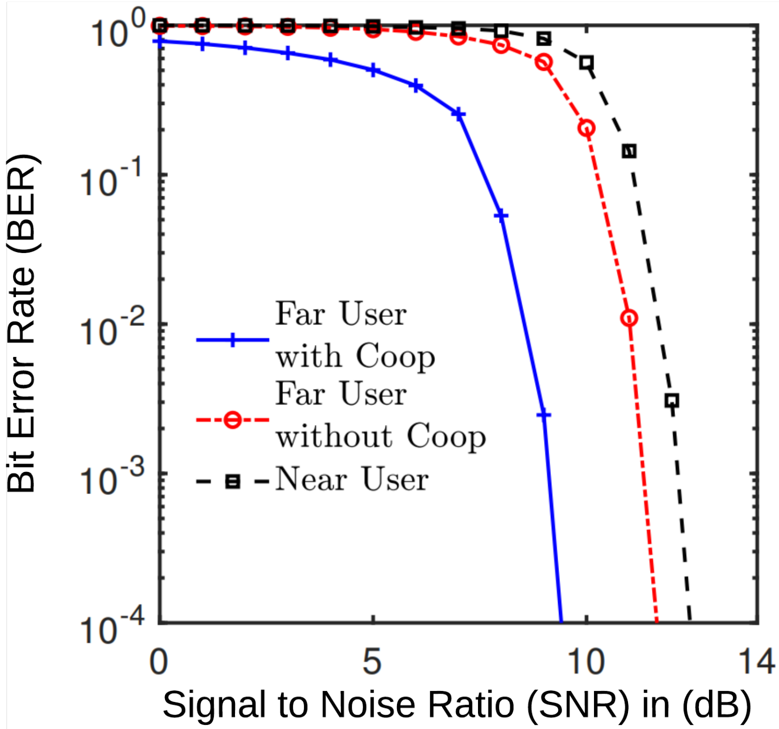


Figure 3.4: BER Vs SNR for the proposed system comprising of the performance of the NUs and the FU with and without cooperative relaying

and FU in a cooperative NOMA framework necessitates 1.5 bits for broadcasting at an equivalent power allocation. This enables the cooperative NOMA system to achieve superior bandwidth utilisation compared to a conventional NOMA system.

The efficacy of cooperative NOMA may be observed through the comparison of the data rates achieved by the system with cooperation and the system without cooperation. It is evident that the system employing cooperation exhibits superior performance in terms of data rate.

In Fig.3.6, both the NU and FU are depicted as moving away from the BS. Notably, when both FU and NU are in close proximity to the BS and are allocated equivalent power levels, the system's efficacy significantly diminishes due to the SIC's inability to decode the stronger user effectively. However, as the FU starts distancing itself from the BS, there is a marked improvement in performance. Given that both the NU and FU need to maintain identical decoding rates, it becomes essential to allocate more power to the symbol information, especially when distanced from the BS, to substantially boost the cooperative model's performance. Based on the simulation results, it is recommended

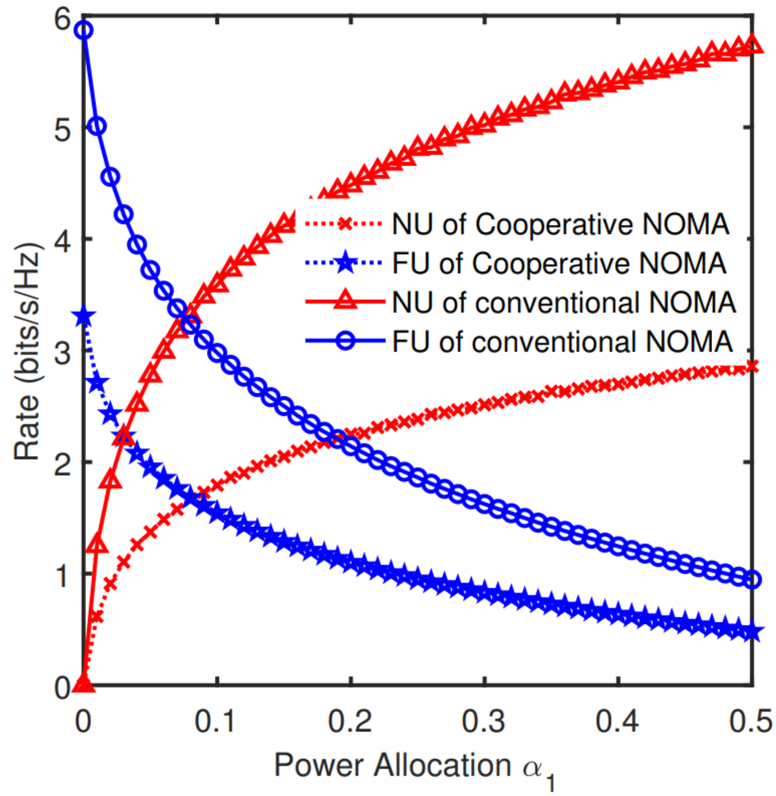


Figure 3.5: Allocation of power within the NOMA framework and its correlation with decoded data speeds in both cooperative and traditional NOMA setups

to position the FU and NU at an increased separation from each other.

Fig.3.7 illustrates the cooperative NOMA system using EXIT chart analysis. The inner decoder is predicated on MI, while the outer decoder leverages channel decoding/despreading data. With the SF set at 2 and $R = 1/2$, the joint spread stands at 0.25. Given the initialization of both ones and zeros with an equal probability of 0.5, an initial loss becomes apparent from the chart. It's discernible that an open tunnel materializes around a 10 dB SNR. Provided there are adequate iterations between the decoder and despreader, transmission can achieve error-free outcomes. This visualization suggests that EXIT charts can be effectively employed for cooperative NOMA, contingent on the optimal number of iterations between MUD_{DEC} and MUD_{DES}.

In order to analyse the capabilities of the MUD, a $SF = 2$ is utilised by exploiting the IRCC codes as can be observed from Fig.3.7.

Fig.3.8 exhibits a narrow pathway between the inner and outer curves at an $SNR = -10$dB within the EXIT chart. A convergence of the system is discerned when these

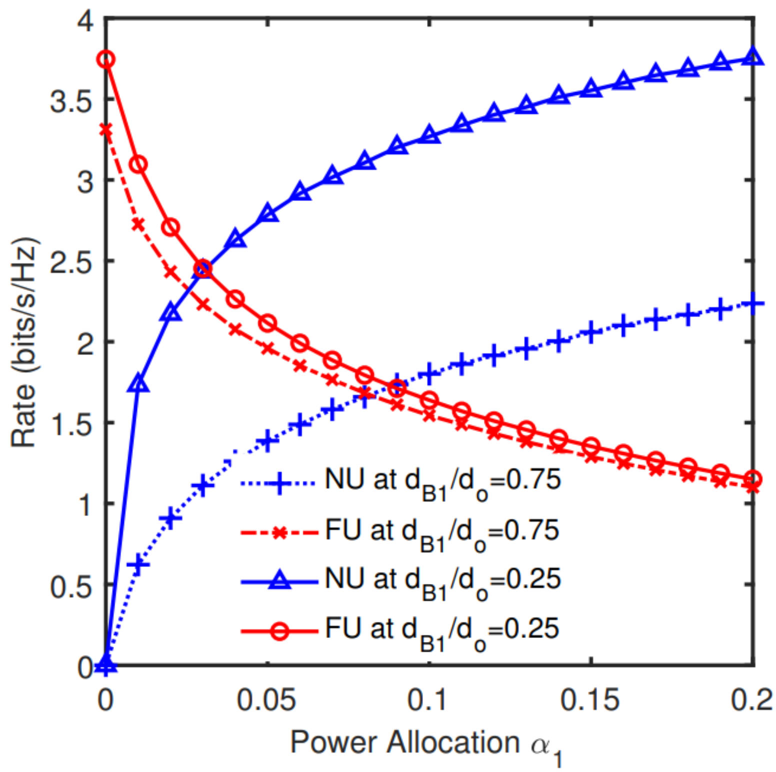


Figure 3.6: The association between power distribution and decoded data speeds when both the NU and FU receive equivalent power allocations

trajectories align, manifesting an "open tunnel". This point of convergence, denoted as [1, 1], is featured in the top right quadrant of Fig.3.8. For the inner decoder, the MUD employs a Gaussian distribution to formulate the a priori LLRs. In this context, the inner curve overshoots for the MAP in the interval $0 < IMUDapr < 1$, as illustrated in Fig. 3.7. An IRCC generates the outer curve of the EXIT chart, aligning with the inner curve of the EXIT chart for the MAP MUD, as delineated in Fig.3.8.

All participants transmit over the accessible subcarriers at a rate given by $R/SF = 0.4$, where $SF = 2$. The decoding pathway navigates through the open tunnel depicted in the EXIT chart, ultimately converging to a value of $I_{DES=DEC;ex} = 0.96$ after undergoing 1000 iterations between $MUD_{DES/DEC}$. This process culminates in a notably reduced BER.

The BER performance of the MUD system is illustrated in Fig. 3.9. From the figure, it can be discerned that after a span of 100 decoding iterations, the system is 1.08 dB off from the Shannon channel capacity. This observation is corroborated by the BER

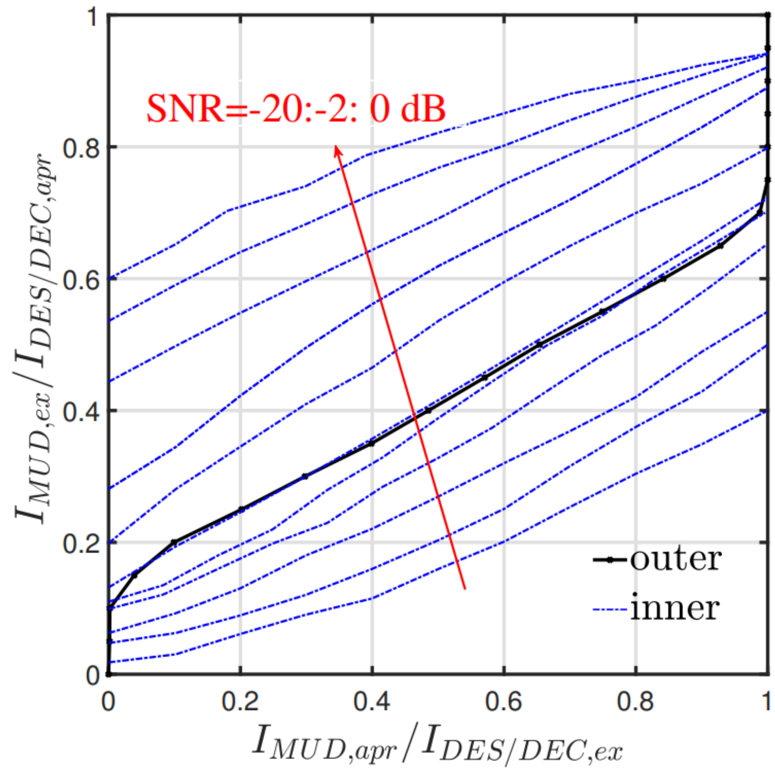


Figure 3.7: The NOMA system's EXIT chart, the outer decoder employs an RSC code with a SF of 2, while the inner decoder is adapted for multiple SNR values within the MUD framework.

line situated at $= 10^{-5}$, where the SNR registers at -0.86 dB. Such a deviation can be attributed to a duo of reasons: initially, the constraints of the interleaver, which has a length of 2048 bits per user, and subsequently, the external coding mechanism which fails to meet the $IDES = DEC; ex = 1$ benchmark.

It is evident that conducting over 100 iterations can lead to the attainment of optimal Shannon capacity performance. This suggests that our proposed model would exhibit improved performance if a greater number of decoding iterations were executed within the MUD model.

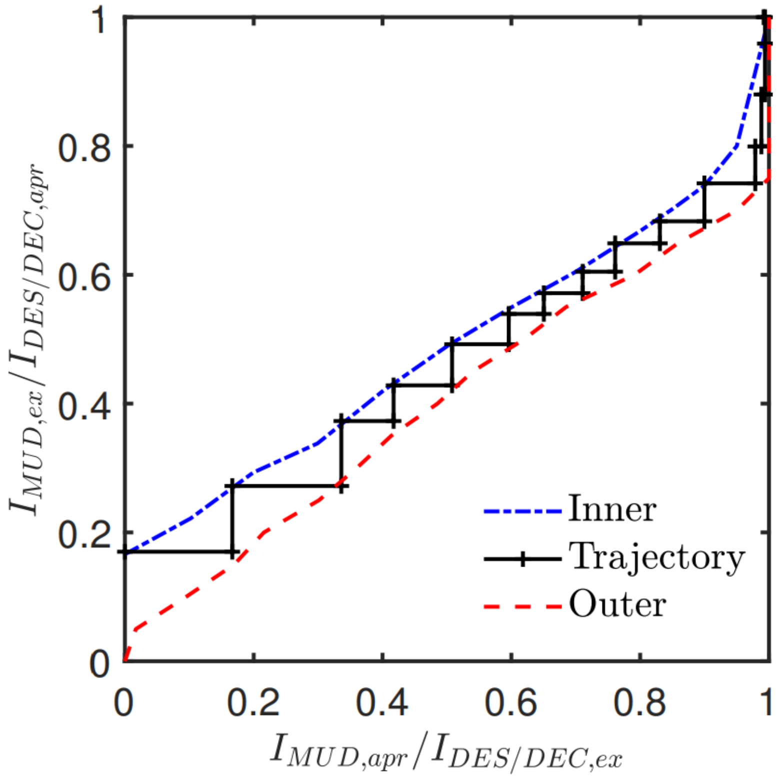


Figure 3.8: EXIT chart analysis using IRCC codes for the proposed NOMA system

3.1 Summary

In this chapter, NOMA was implemented into a cooperative relaying scenario. The system model consisted of four users and a base station. Three users were placed in close proximity to the base station, while one user was designated as the cell edge user and placed at a greater distance away.

The primary objective of this simulation was to demonstrate that applying NOMA to a cooperative relaying scenario would result in improved performance for all users and the ability to effectively support the far cell edge user, despite it having significantly higher QoS requirements than the nearby users.

The results demonstrate that both the far cell edge user and the users in close proximity to the BS were provided with acceptable service with extremely low error rates as a result of efficient power allocation according to the NOMA concept.

An EXIT chart analysis was also performed on the proposed cooperative NOMA system. Its results are able to show the SNR needed for the system to achieve a very low BER. It

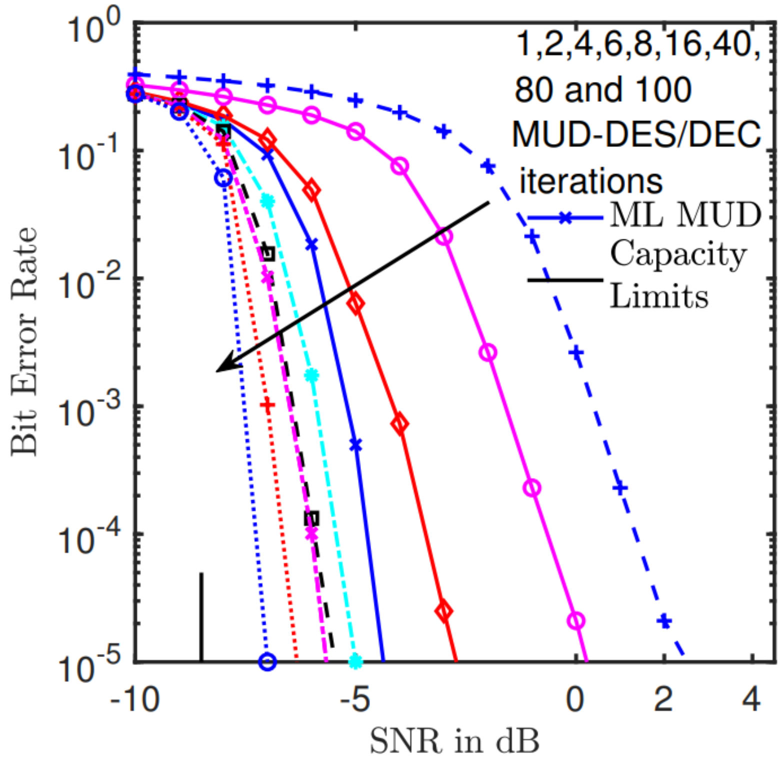


Figure 3.9: BER Vs SNR for the MUD system, where various number of MUD DES/DEC iterations are utilised when SF=2

can also be observed that, by increasing the number of decoding and despreading iterations between the inner and outer decoders, the BER experienced by the MUD significantly decreases. These results showcase the importance of EXIT chart analysis as a tool for performance enhancement and evaluation.

Chapter 4

EXIT Chart Analysis in Cooperative NOMA

4.1 Introduction

The employment of SIC is a crucial element in facilitating signal acquisition within a NOMA framework. This is due to its ability to provide an opportunity for users with relatively weaker power levels to effectively retrieve their transmitted messages. The process of SIC involves the initial identification of the signal emanating from the higher power user while considering the remaining users as noise. Subsequently, the identified signal is extracted from the received, superimposed signal, leading to the retrieval of the signal belonging to the low power user [6, 8]. A drawback associated with PD-NOMA is its tendency to prioritise users with inferior channel conditions by allocating them with a larger proportion of the available transmission power, while users with superior channel conditions receive a significantly lower amount of power, which may not be sufficient to meet their requirements in certain scenarios. As a result, the performance of individual users may be negatively impacted, as reported in [6, 8, 23]. This chapter introduces EXIT charts as a potential solution to address the aforementioned issue.

As a result of the work carried out in this chapter, the following paper(s) were published:

- Elsaraf, Z., Ahmed, A., Khan, F.A., Q. Z. Ahmed, "Cooperative Non-Orthogonal Multiple Access for Wireless Communication Networks by Exploiting the EXIT Chart Analysis," J Wireless Com Network 2021, 79 (2021).

- Z. Elsaraf, A. Ahmed, F. A. Khan and Q. Z. Ahmed, "EXIT Chart Analysis of Cooperative Non-Orthogonal Multiple Access for Next Generation Wireless Communication Systems," 2020 European Conference on Networks and Communications (EuCNC), Dubrovnik, Croatia, pp. 281-285, 2020.

4.2 System Model

It is important to note that the system model utilised here is identical to the one presented in chapter 3. Concisely, this system model operates by transforming the information bits according to 4-QAM (or QPSK) [8, 23], the user's data are modified. In the process, the signal bits undergo transmission to the signal-spreading component. Within this component, the mapped bits undergo multiplication with a receiver-acknowledged, pre-established code. Following the foundational principle of power allocation intrinsic to the NOMA mode, the data derived from the signal spread is allocated a specific power level. Subsequent to its navigation through the Rayleigh fading channel, the interference signal is then captured at the receiving end. At the receiver juncture, the interference introduced by the channel is attenuated using the MUD iterative method. Herein, the optimal sequence for SIC detection is ascertained by prioritizing the detection of the most robust user, followed by the least robust.

Based on the specified SIC-detection order, each user has the capability to detect its own information while minimising the adverse impact on other users with lower normalised channel gains. Subsequently, the received signal undergoes despreading using a pre-established code, followed by demodulation at the receiver. In order to accurately identify the appropriate components at the receiver, the utilisation of the MRC technique [27] is employed. This approach is applicable to a wide range of antennas, including those utilised in BS [28-31], due to the general nature of our system model.

4.3 Performance Analysis

The evaluation of the proposed system's effectiveness is contingent upon the attainable rate at each individual user and the total throughput of the system. The achievable rate for each user is in accordance with the SIC chain. The determination of the user's fairness is contingent upon the analysis of the EXIT chart.

4.3.1 Achievable Rate Analysis

The achievable rate is typically measured in bits/s/Hz. At the conclusion of the first phase of transmission, the attainable rate of the FU is calculated as follows:

$$R_{1,FU} = \log\left(1 + \frac{P_{1,FU}\left|h_{1,FU}\right|^2}{\sum_{j=1}^{N} P_{1,j}\left|h_{1,j}\right|^2 + N_{0_{1,FU}}}\right) \tag{4.1}$$

Thus, the NU_j achievable rate after the first phase of transmission is denoted as:

$$R_{1,NU_j} = \log\left(1 + \frac{P_{1,j}\left|h_{1,j}\right|^2}{N_{0_{1,NU_j}}}\right), j = 1, 2, ..., N. \tag{4.2}$$

As shown in Fig. 2, the ultimate FU attainable rate following the cooperative phase of transmission can be represented as:

$$R_{FU} = \log = \left(1 + \frac{P_{1,FU}\left|h_{1,FU}\right|^2}{\sum_{j=1}^{N} P_{1,j}\left|h_{1,j}\right|^2 + N_{0_{1,FU}}} + \sum_{j=1}^{N} \frac{P_{2,j}\left|h_{2,j}\right|^2}{N_{0_{j,FU}}}\right). \tag{4.3}$$

4.3.2 System Throughput Analysis

For the purpose of simulation analysis, the system throughput is delineated as the aggregate of the data rates experienced by individual users. This metric serves as a pivotal performance indicator, instrumental in assessing the holistic efficacy of the proposed system. The throughput of the system can be represented as:

$$S = \sum_{j=1}^{N} R_{1,NU_j} + R_{FU} \tag{4.4}$$

EXIT chart analysis

The encoded bit stream is subjected to a signal mapper to facilitate the conveyance of information to a receiver characterized by noise. Consequently, when accommodating N parallel decoder chains, the MUD [34, 36] is employed to compute the LLR for the pertinent bits. The extrinsic LLR corresponding to the nth user's z^{th} bit, emanating from the MUD, is assessed in the context of the reception of the m^{th} sub-carrier as:

$$L_{MUD,ex^{(i_l^{(n)})}} = \ln \frac{\sum_{Y\in v(m,z,0)} P\left(Z_q \mid Y\right) P\left(Y\right)}{\sum_{Y\in v(m,z,1)} P\left(Z_q \mid Y\right) P\left(Y\right)} - L_{MUD,apr(i_l)^{(n)}}, \tag{4.5}$$

In this context, z pertaining to the nth user's data is aggregated, where both v and n are elements of their respective sets. The *apriori* LLR of the nth user's zth bit can be denoted as $L_{MUD,apr}(i_l^{(n)})$. It is imperative to acknowledge that within the MUD framework, the initial probabilities of zero and one bits are equal. Consequently, as delineated in [34,38,39], the inaugural values within the soft registers are initialized to zero. Additionally, the symbol for the *apriori* is represented by $P(Y)$, signifying the multi-user probability, with the codeword being articulated as $Y =$. Furthermore, the channel probability associated with Y is conveyed as $P(Z_q \mid Y)$, and the cost function (CF) is defined in accordance with [34,40].

$$F_{CF}(Y) = P\left(Z_q \mid Y\right)\left(P\left(Y\right)\right) = \exp\left(-\left|\left|Z_q - PH_nWY\right|\right|\right)^2 \times P\left(Y\right). \tag{4.6}$$

In this framework, H_n symbolizes the channel states encompassing all users within the model, while W denotes the quantity of transmitter and receiver antennas that are equitably matched. Prior to forwarding the *apriori* LLRs to the channel decoder, a despreading operation is executed, facilitating the identification of the extrinsic LLRs for each user, as determined by the deinterleaving applied to the user's input sequence. The channel decoder subsequently yields bit-oriented a posteriori LLRs, which are channeled to the signal spreading, wherein supplementary iterations transpire between the interleavers and the MUD. Following a predetermined count of DEC and MUD-Despreading/Spreading iterations, labeled as 'I', the decoder engages in a rigorous selection process.

4.4 Simulation Results

Figure 4.1 illustrates the BER performance of the NU and FU in relation to SNR. As anticipated, owing to its elevated power level (approximately 90 % of Pt) attributed to its extended distance from the BS, specifically at a distance of 10 meters, the FU manifests a diminished BER in comparison to the NU. The NU is allocated a minimal power, approximately 1% of Pt, determined by its proximity to the BS, being merely 1 meter away.

Figure 4.2 assesses the rate at which data is transferred for each individual user. Empirical evidence suggests that as the SNR increases, the data rate of the FU reaches a ceiling, whereas the data rate of NU systems surpasses it. The rise in channel SNR directly corresponds to the increase in IUI implemented by the NU onto the FU. The observable impact of cooperative

NOMA can be seen in Figures 4.1 and 4.2. The implementation of cooperative relaying significantly improves the data rate and BER in a NOMA network, as compared to a NOMA network that does not employ cooperative relaying.

Figure 4.3 illustrates the overall system throughput of both the cooperative and non-cooperative systems in the proposed model. The aggregate throughput of the system is determined by the combined data transfer speeds of the NU and FU.

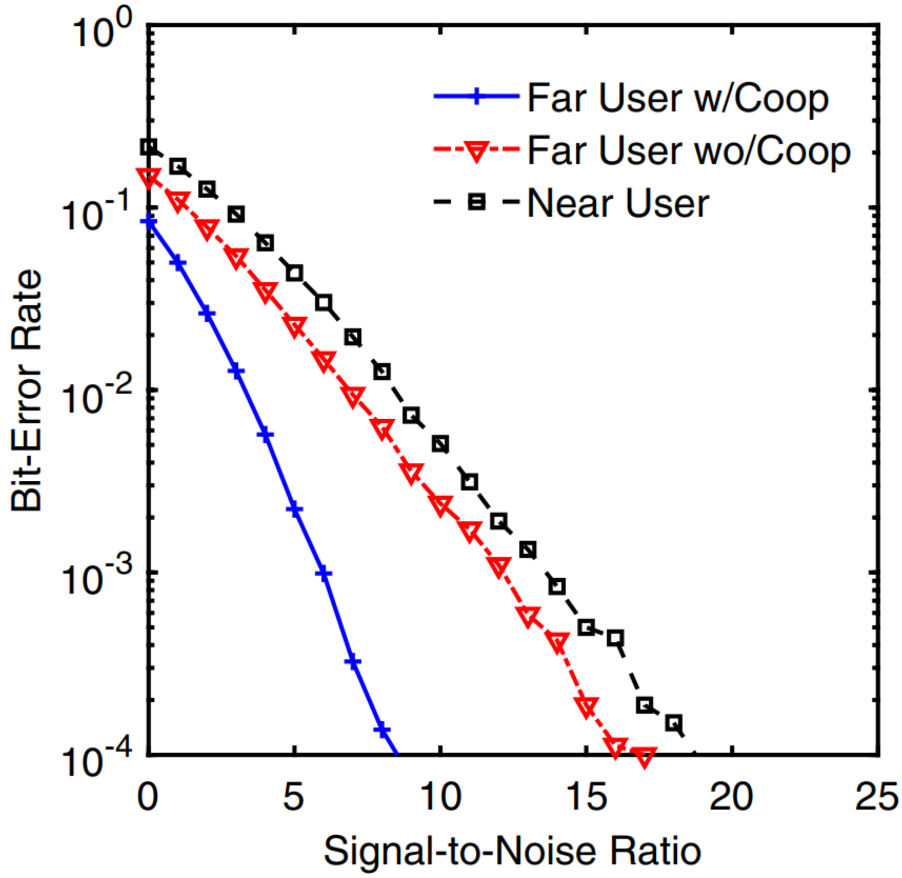


Figure 4.1: Comparison of BER performance between near and far users

Cooperative relaying dramatically increases the system's throughput, as the cooperative relaying outcome exceeds its noncooperative counterpart. At 0 dBs, the cooperative system's system throughput begins to be much higher than that of its non-cooperative counterpart, and this difference persists until 40 dBs. As the SNR value for the channel climbs and the system approaches its capacity limit, the performance gap narrows. According to Shannon's capacity rule [34, 40], however, the performance reaches a limit at approximately 40 dBs.

Figure 4.4 illustrates the EXIT chart analysis, where the inner decoder utilises the MI of the MUD, while the outer decoder utilises the information obtained from despreading and

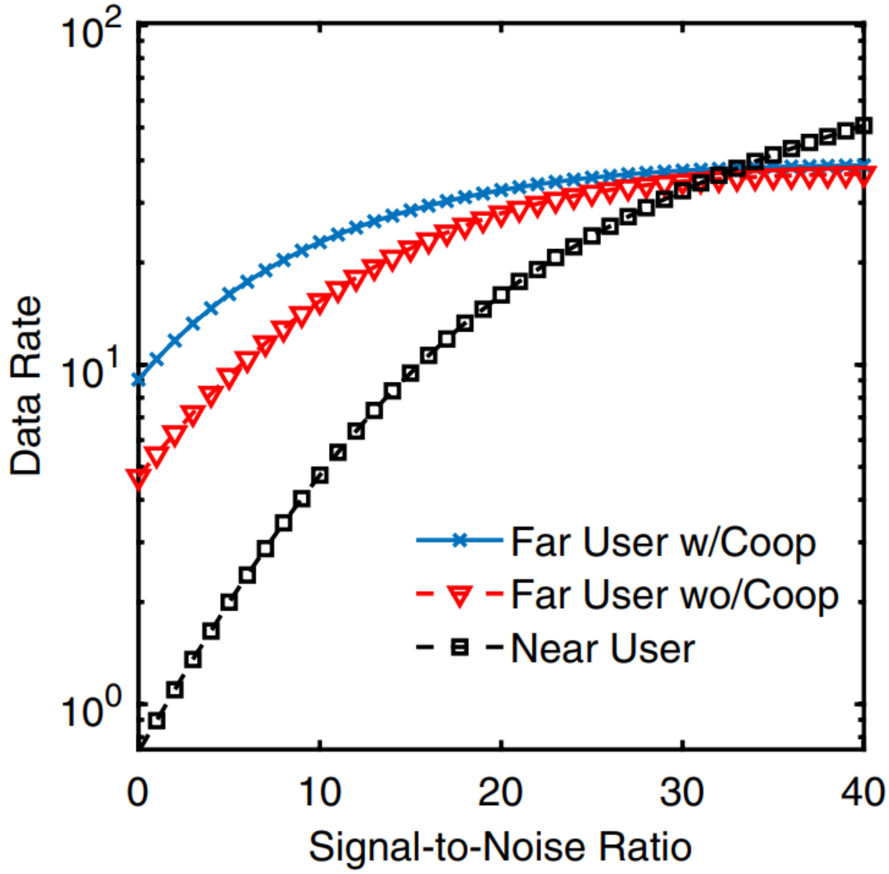


Figure 4.2: Near and far users' data rates

channel decoding. The only further calculation performed in the MUD involves the offline determination of the joint probability of the alphabet at the receivers of both the user and the BS. The aforementioned objective is accomplished by the use of a fair assortment of codewords in combination with the predetermined parameters of our systems' interleavers.

The MUD decoder produces the bit-oriented a posteriori LLRs for the corresponding codewords. These are then channeled to the DES spreading, followed by the interleavers, and subsequently recirculated to the MUD for further iterations. As a result, the coding rate for the combined spreading is expressed as R/SF, which, in the context of this study, stands at 0.25. This corresponds to a SF of 2 and R of 0.5. Drawing insights from Fig. 4.4, it is inferred that the MUD necessitates a minimum SNR of 12 dB to achieve the intended output. Hence, by integrating the RSC code with a repetition factor of SF=2, it becomes feasible to attain error-free transmission, contingent upon the MUD-DES and MUD-DEC undergoing a suitable number of iterations.

The suboptimal performance exhibited by the MUD can be attributed to the initialization

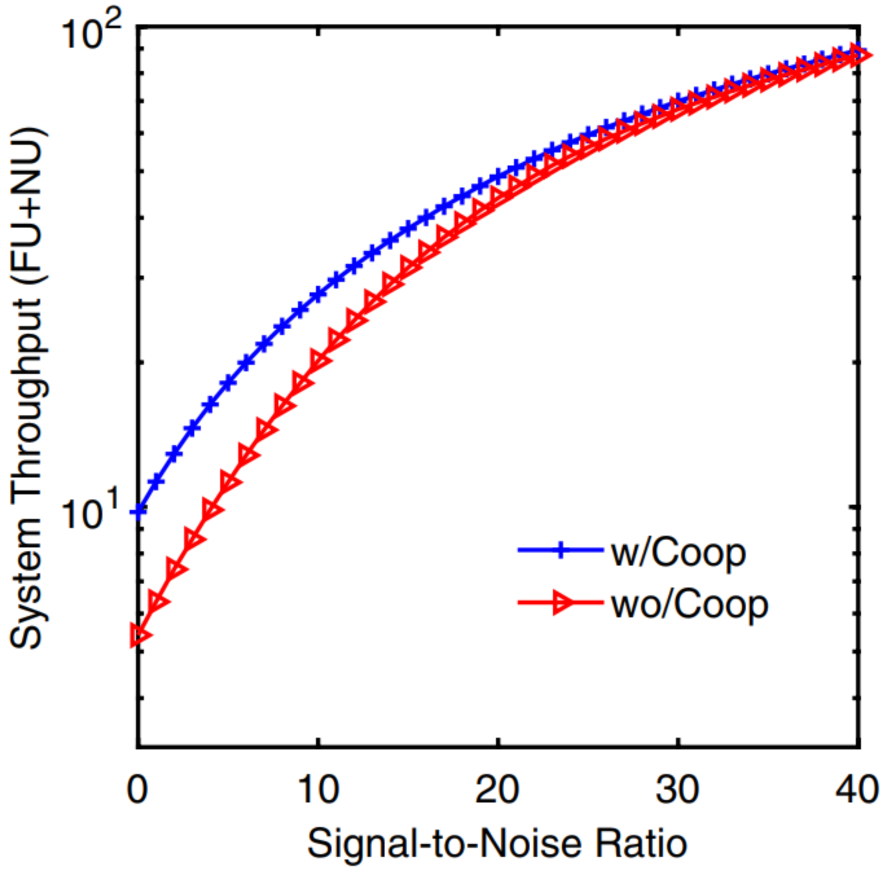


Figure 4.3: System throughput for proposed NOMA system

of MI for $IMUD, apr$ as zero, given the equivalent likelihood of ones and zeros at the inception. To generate the 'open tunnel' delineated between the inner and outer curves, a higher SNR becomes indispensable. As illustrated in Fig. 4.5, when the external code is synergized with the MUD architecture, a value of $I_{DES/DEC,ex} = 1$ is achieved at relatively diminutive $I_{DES/DEC,apr}$ levels.

As shown in Fig.4.5, the suggested NOMA model achieves user fairness because the inner/outer curves reach the [1 1] point without crossing at any given position. Fig. 4.6 demonstrates that when the MUD model is deployed, the inner curve has a steeper slope than the outer curve. In addition, the decoding trajectory of our proposed system is dependent on the iteration between the MUD and the despreader. Moreover, at $SNR = 3dB$, the open EXIT tunnel exists between the inner and outer curves after 1000 iterations in $MUD_{DES/DEC}$, where the inner and outer curves intersect at $I_{MUD,ex} = 0.89$ and $I_{DES/DEC,ex} = 1$, as depicted in Fig.4.5.

In order to determine the normalised throughput of a single-carrier and multi-user carrier

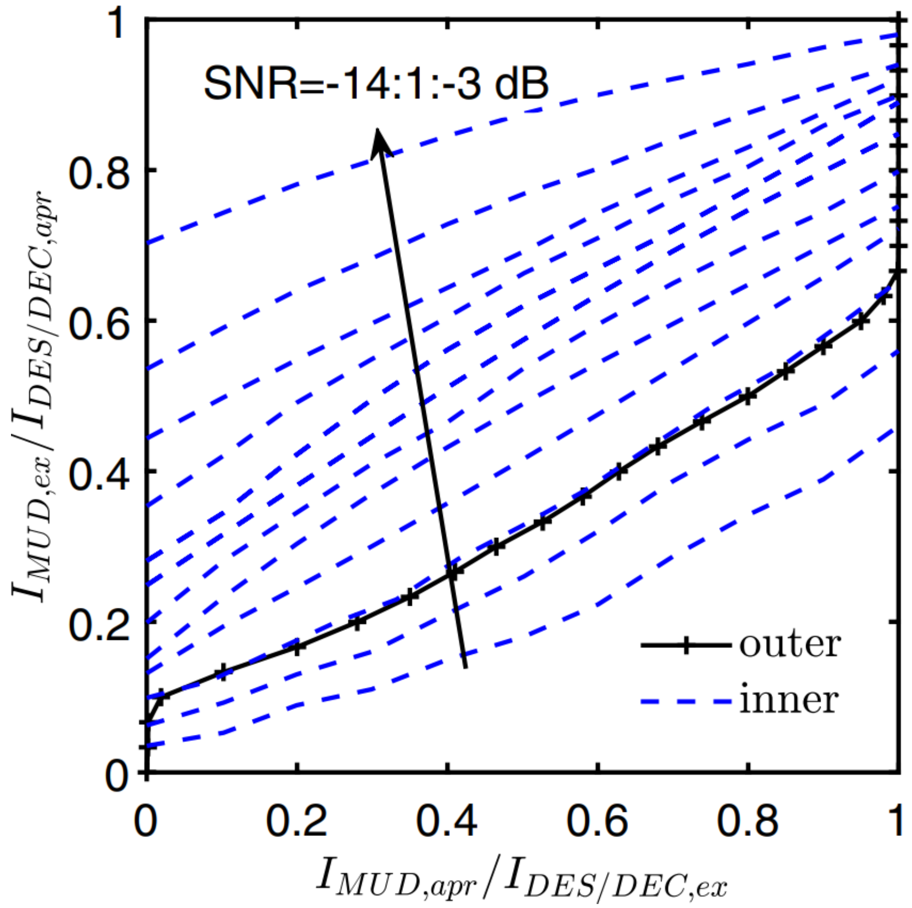


Figure 4.4: EXIT diagram for a NOMA system with the outer decoder utilizing RSC code at an SF=2, while the inner decoder functions across diverse SNR values within the MUD framework

model when the number of users is $U = 1$ and $U = 2$ in non-dispersive Rayleigh channels, Fig.4.6 is generated. Fig.4.6 illustrates that when used in the decoding paradigm, both multi-user and single-carrier users achieve a higher throughput gain simultaneously. In addition, the throughput gain in the single carrier system is greater than in the multi-user model, indicating that more transmitting antenna combinations are utilised in the single carrier system.

In a single carrier system, ten transmitting antennas are utilised, but only two are used in a multi-user type. This work demonstrates that increasing the number of antenna array combinations for a single user decreases the correlation between the code word used to decode the message, hence enabling the uniform selection of antenna array combinations across the single-user model.

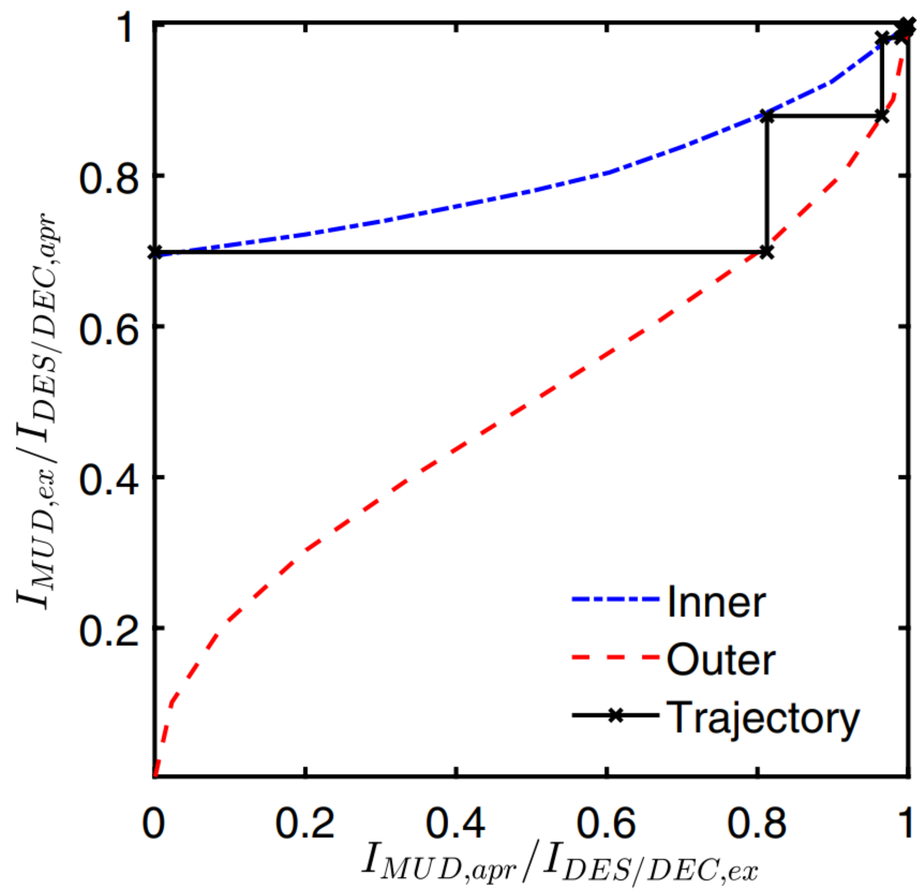


Figure 4.5: EXIT chart for the proposed NOMA system

4.5 Summary

This chapter applies NOMA to a cooperative relaying scenario and performs an EXIT chart analysis on the network. According to the EXIT chart analysis, information was exchanged between the inner and outer decoders over a thousand times until an infinitesimally small BER was achieved. The analysis of the EXIT chart also revealed that the simulation reached unity gain, where convergence is observed after one thousand iterations and the open EXIT tunnel between the inner and outer curves can be observed. The results show that unity gain was achieved at point [1,1] without the inner and outer curves intersecting at any previous point. Therefore, the system was able to achieve optimal user fairness during power allocation through the application of NOMA. The normalised throughput was also shown to increase significantly when a single carrier system is implemented, as it allows for more antenna combinations to be utilised for transmission and reception. As opposed to multi-carrier systems, where antenna combinations are significantly lower.

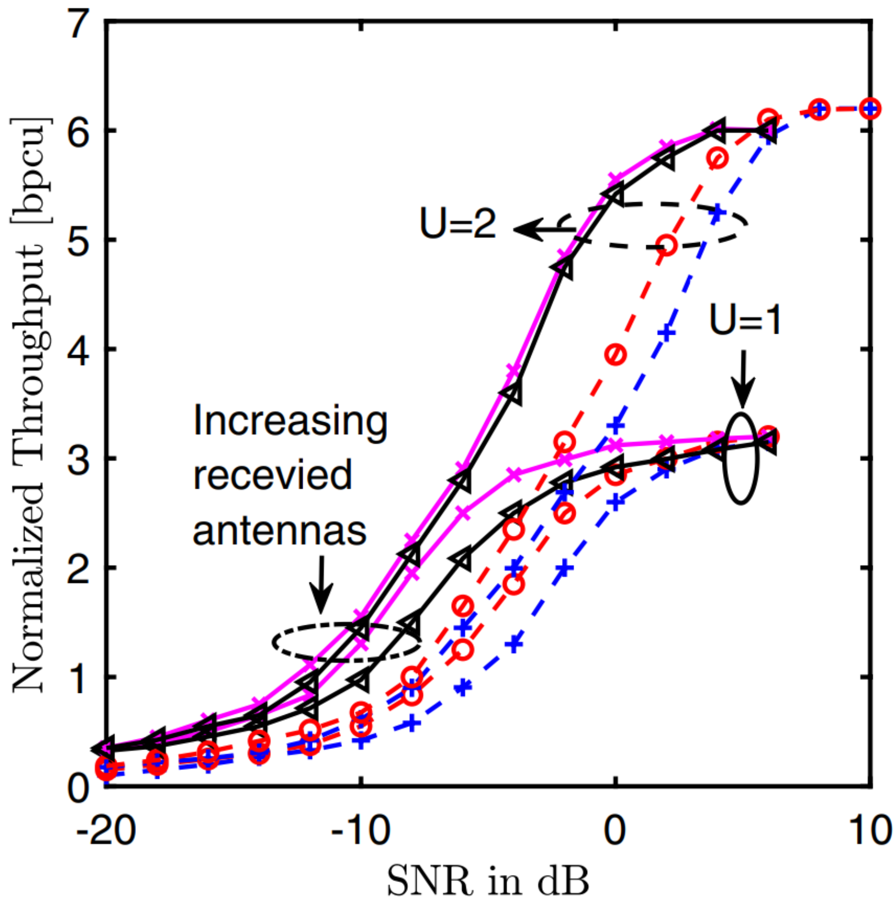


Figure 4.6: Normalised throughput analysis for MUD model

Chapter 5

Enhancing the Performance of NOMA Systems Using Deep Learning Techniques

5.1 Introduction

Deep learning is a key driver behind numerous AI applications and services, which enhance automation by enabling analytical and physical tasks to be performed without the need for human intervention. The technology of deep learning underlies commonplace products and services, including but not limited to digital assistants, voice-activated television remotes, and credit card fraud detection, as well as nascent technologies such as autonomous vehicles.

This chapter will begin by introducing the concept of deep learning before focusing on the research performed on signal detection in a conventional NOMA system using deep learning assisted SIC instead of conventional SIC. The aim of the research is to reduce the receiver complexity inherent in SIC without sacrificing receiver reliability.

As a result of the work carried out in this chapter, the following paper(s) were published:

- Z. Elsaraf, F. A. Khan and Q. Z. Ahmed, "Deep Learning Based Power Allocation Schemes in NOMA Systems: A Review," 2021 26th International Conference on Automation and Computing (ICAC), Portsmouth, United Kingdom, 2021, pp. 1-6.

5.2 Signal Detection in a Downlink NOMA system using Deep Learning

5.2.1 Introduction

In order to highlight the importance of adopting a deep learning approach to signal detection in future wireless communication systems (5G and beyond), a signal detection using deep learning for both a NOMA and OFDM system is proposed here.

The aim is to adequately present the significant advantages a network adopting both NOMA and deep learning, for its transmission and reception of data, has in terms of efficient and less complex signal detection. The system model will be outlined and discussed first before presenting and discussing the acquired results.

5.2.2 System Model

General Block Diagram

Fig.5.1 illustrates the general relationship between functions. The squares represent the scripts, the circles represent the functions and the dashed circle represents a private internal function in the testData.m script. The arrows represent the transfer of data from one entity to another. It is obvious that the transfer of data between scripts is unidirectional as opposed to bidirectional when it comes to functions. A function receives arguments when it is called and then returns certain outputs hence the bi-directionality. We note that the parameters exchanged between each entity are not specified in order to have a clean figure. We should also note that the size of each entity does not reflect its importance nor the size of its code.

As mentioned, there are three scripts (trainData, trainNN and testData) as well as seven functions (getFeaturesAndLabels, allocatePower, dataTransmissioReception, detectML, channelEstimation, symbolDecodeSIC, symbolDecodeDL) and an private internal function for the testData script (getRhh).

This deep learning technique follows the standard approach. First the creation of the training data using the trainData script, followed by creating and training a neural network using the trainNN script then finally testing the network using the testData script. As the name suggest the trainData script is used to generate the data used. It takes no inputs whatsoever, following

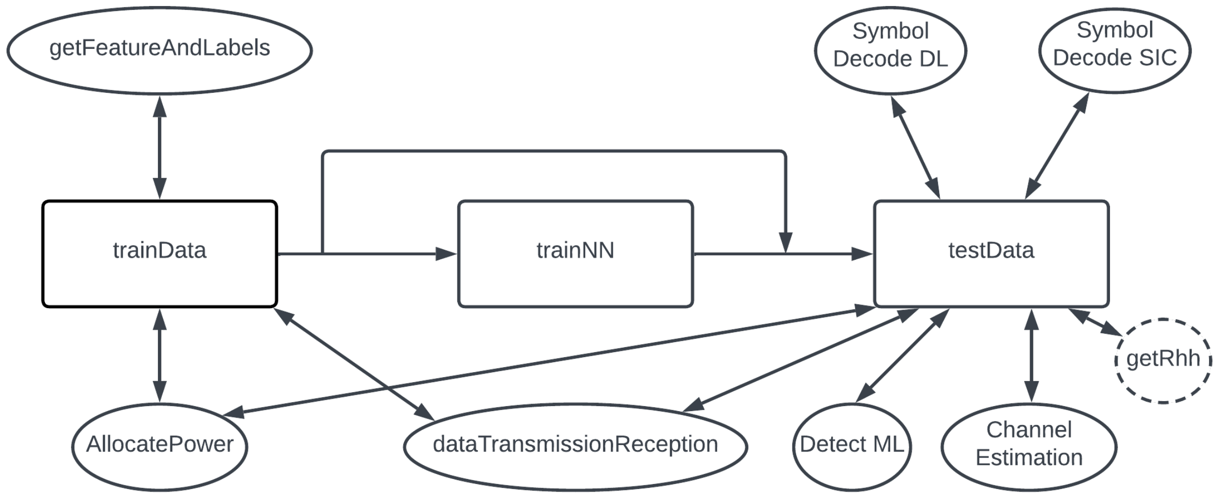


Figure 5.1: General diagram for the proposed deep learning model

the code it calls allocatePower, dataTransmissionReception and getFeaturesAndLabels in this order to finally generate the training data and some essential parameters.

The trainNN script is a conventional neural network training script. It used the training data generated by the trainData script to train a LSTM-based neural network and then returns at the end the trained network. This code does not use any external functions.

The testData script generates the test data using the trained network as well as some parameters generated by the trainData script and almost all the functions in the following order getRhh, allocatePower, dataTransmissionReception, detectML, channelEstimation, symboleDecodeSIC, SymboleDecodeDL. The getFeatureAndLabel function is called by the symbolDecodeDL function.

Signal Modulation

QPSK was used for signal modulation in this work. It was chosen for its simplicity in implementation as as well as its robustness in Low to mid range SNRs, making it a favorable choice of modulation schemes in environments with varying conditions. Using QPSK instead of other more complex modulation schemes is beneficial for DL system design since lower complexity leads to faster training for the neural network.

Power Allocation

The allocatePower function's role is to allocate transmit power to the users based on their channel gain in order for each user to achieve the targeted SNR. This function takes as arguments the signal power per symbol (symPower), the static channel gain (gainH), the targeted SNR for each user (targetSNR-1,targetSNR-2) and the noise variance (nVar) then returns the resulted power factor(powerFactor) and the decoding order of the users (decOrder). When the function is called the difference between the channels' gain is calculated, depending on the result low and high gain levels are assigned for each user. Using the gain levels, the targeted SNR and the noise variance low and high power factors are computed which results in the general power factor. Using the power factor, the decoding order of the users can be found. Algorithm 1 shows the power allocation process in pseudo-code. (Figure A.1 in the Appendix shows the power allocation process in the form of a flowchart)

Channel Estimation (CE)

The channelEstimation function performs least square (LS) and minimum mean square error (MMSE) CE. It takes as input arguments the received data (rData, the pilot frame (pilotFrame), the power factor (powerFactor), the pilot start (pStart), the channel covariance matrix (RHH), the noise variance (nVar), the number of pilot subcarrier (numPSC) and the perfect channel frequency response (H-perf) then returns the LS estimation (H-LS) and the MMSE (H-MMSE) CE. The process begins by computing the pilot spacing, then, for each user, the pilot symbol, the power factor, and the data symbol are extracted and the power of the interfering users is computed. Then for each packet of this user, the LS and MMSE CEs are computed. When all is done the relevant values are returned. Algorithm 2 showcases the CE process in pseudo-code.(Figure A.2 in the Appendix illustrates the CE in the form of a flowchart)

Data transmission and reception

The dataTransmissionReception function's role is to model the signal received after being transmitted using the OFDM method. A key point in OFDM is that the receiver will receive a superposition of the signals from every user. This function takes as arguments the transmitPacket, the power factor (powerFactor), the length of the Cyclic Prefix (CP) (lengthCP), the channel response (h), the noise variance (nVar) and returns the received packets (receivePacket) and the

Algorithm 1 Power Allocation

```
function ALLOCATEPOWER(symPower, gainH, targetSNR_1, targetSNR_2, nVar)
    (numUE, numSC) ← dimensions of gainH
    // Channel gain levels
    gainDiff ← difference of gainH along first dimension
    for each element of gainDiff do
        if element < 0 then
            highGain[element] ← 1
        else
            highGain[element] ← 2
            lowGain[element] ← 1
        end if
    end for
    for each element of gainDiff do
        if element > 0 then
            lowGain[element] ← 1
        else
            lowGain[element] ← 2
        end if
    end for
    Transpose highGain
    Transpose lowGain
    //Calculate power allocation factor
    Initialize powerFactor with zeros of size (numSC, numUE)
    for sc = 1 to numSC do
        lowPower ← (targetSNR_2×nVar) / (symPower×gainH[highGain[sc], sc])
        highPower ← (targetSNR_1×(lowPower×symPower×gainH[highGain[sc], sc]+nVar)) / (symPower×gainH[lowGain[sc], sc])
        highPowerFactor ← (highPower) / (highPower+lowPower)
        lowPowerFactor ← (lowPower) / (highPower+lowPower)
        powerFactor[sc, highGain[sc]] ← lowPowerFactor
        powerFactor[sc, lowGain[sc]] ← highPowerFactor
    end for
    Initialize decOrder with zeros of size (numSC, numUE)
    decOrder[:, 1] ← index of maximum value of powerFactor along the second dimension
    decOrder[:, 2] ← index of minimum value of powerFactor along the second dimension
    return powerFactor, decOrder
end function
```

Algorithm 2 Channel Estimation (CE)

function CHANNELESTIMATION(rData, pilotFrame, powerFactor, pStart, RHH, nVar, numPSC, H_perf)
    (numUE, numSC, numPacket) $\leftarrow$ dimensions of pilotFrame from the 2nd, 3rd, and
    4th dimensions
    pilotSpacing $\leftarrow \frac{\text{numSC}}{\text{numPSC}}$
    Initialize H_LS as zeros of size (numSC, numUE, numPacket)
    Initialize H_MMSE as zeros of size (numSC, numUE, numPacket)
    for $u = 1$ **to** numUE **do**
        // Pilot symbols and data symbols
        pilot_sc $\leftarrow$ sequence starting from $pStart[u]$, incrementing by pilotSpacing, up to numSC
        pilot $\leftarrow$ extract the specific pilotFrame values based on u, pilot_sc, and u for all packets
        $pF \leftarrow$ extract the specific powerFactor values based on pilot_sc, u, and all packets
        data $\leftarrow$ extract the specific rData values based on u, pilot_sc, and all packets
        // Power of interfering user
        intfPower $\leftarrow$ extract powerFactor values excluding the current u
        for $p = 1$ **to** numPacket **do**
            $pl \leftarrow \sqrt{\text{pF for packet } p} \times$ pilot for packet p
            $hLS \leftarrow$ data for packet $p \div pl$
            $hLS\_interp \leftarrow$ interpolate hLS using spline method from 1 : pilotSpacing : numSC to 1 : numSC
            Transpose hLS_interp
            H_LS for packet p and $u = hLS\_interp$
            H_MMSE for packet p and $u = RHH \times$ (inverse of $(RHH + nVar \times$
            diagonal of intfPower for packet $p) \times hLS\_interp$
        end for
    end for
    return H_LS, H_MMSE
end function

random phase. The process is simple and is divided into two parts, transmission then reception. First generate relevant variables like the number of symbols per packet (numSym), the number of subcarriers (numSC), the number of user equipment (numUE) and the number of packets (numPacket) based on trasmitPacket. These values could have been sent to the function as arguments however, this can cause some problems due to the high number of arguments that would have to be written (8 arguments).

The transmission begins by generating a random phase shift for each packet of each user. The signal of each user is then multiplied by the square root of its power factor and the inverse Discrete Fourier transform is applied to the result. Then a CP is inserted between each subsequent symbol in the time domain. The result is transposed (parallel to serial) and the convolution with the multipath channel response is computed then the phase shift is introduced. The signals of the users are added together hence the superposition and a Gaussian noise is added to this sum. At the receiver, the received signal is transposed (serial to parallel), the CP is removed and the discrete Fourier transform is applied (DFT). The function then returns the received packet and the random phase. Algorithm 3 shows this process in pseudo-code. (Figure. A.3 in the Appendix illustrates this process in the form of a flowchart)

Deep learning signal detection

The detectML function performs ML detection for two users while assuming perfect CE. It takes as inputs the channel frequency response (H), the random phase, the Quadrature Phase Shift Keying (QPSK) modulation constellation (constQPSK), the power factor (pF), the receive data (rData), the decoding order indices (idx-1, idx-2), the transmitted packets (tData) and the QPSK symbols combination class (symClass) then returns the error rate between the transmitted labels and the estimated labels (errorML)as well as the received labels (rLabel).

When the function is called the symbols labels are computed followed by all possible transmitted data and by extension all possible received data. Having both received data and estimated received data, the MSE of the two is calculated then the transmitted signals for the minimum MSE are found. Following all that the labels of the estimated symbols and the transmitted symbols are found and the error rate is computed. The function then returns the error rate (errorML) and the received labels. Algortihms 4 and 5 showcase this process in pseudo-code. (Figure A.4 in the Appendix illustrates this process in the form of a flowchart)

Algorithm 3 Data Transmission and Reception

```
function dataTransmissionReception(transmitPacket, powerFactor, lengthCP, h, nVar) → receivePacket, randomPhase
    // Extract dimensions of transmitPacket
    (numSym, numSC, numUE, numPacket) ← dimensions(transmitPacket)
    // Reshape and duplicate powerFactor
    powerFactor ← reshape(powerFactor, 1, numSC, numUE, numPacket)
    powerScale ← duplicate(powerFactor, numSym, 1, 1, 1)
    // Generate randomPhase matrix
    randomPhase ← generateComplexExponentialMatrix(numUE, numPacket)
    for u = 1 to numUE do
        for p = 1 to numPacket do
            // IFFT for the subcarrier
            x1 ← Inverse FFT(sqrt(powerScale_{u,p}) × transmitPacket_{u,p})
            // Inserting CP
            x1_CP ← addCyclicPrefix(x1)
            // Parallel to serial
            x ← convertToColumnVector(x1_CP)
            // Multipath channel convolution
            y_conv ← convolve(x, h_{u,p})
            y ← randomPhase_{u,p} × y_conv[1 : length(x)]
        end for
    end for
    // Superpose signals from 2 users
    y_total ← sum(y, 2)
    // Add AWGN to channel
    sigLength ← length(y_total)
    nFre ← generateGaussianNoise(numPacket, numSC)
    nTime ← Inverse FFT(nFre, sigLength)
    // Receiver
    y_total ← y_total + nTime
    Y ← emptyMatrix(numPacket, numSym, numSC)
    for p = 1 to numPacket do
        // Serial to parallel
        block ← reshape(y_total_p, numSC + lengthCP, numSym)
        // Removing CP
        y_block ← block[:, lengthCP + 1 : lengthCP + numSC]
        // FFT
        Y_p ← FFT(y_block, 2)
    end for
    receivePacket ← permuteDimensions(Y)
     return receivePacket, randomPhase
end function
```

Algorithm 4 Deep Learning Signal Detection

```
function DETECTML(H, randomPhase, constQPSK, pF, rData, idx_1, idx_2, tData, symClass)
    // Initialize sizes and labels
    (numUE, numPacket) ← Dimensions of H
    numLabel ← Length of constQPSK^2
    for each element in symClass do
        if element == constQPSK[1] then
            symLabel of that element ← 1
        else if element == constQPSK[2] then
            symLabel of that element ← 2
        else if element == constQPSK[3] then
            symLabel of that element ← 3
        else if element == constQPSK[4] then
            symLabel of that element ← 4
        end if
    end for
    Reshape symLabel to size(numLabel, numUE)
    // Compute all possible transmitted data
    allSym ← Copy of 1/√2 × symClass repeated numPacket times
    powerFactor ← Reshape pF with dimensions (1, numUE, numPacket)
    powerFactor ← Repeat powerFactor with dimensions (numLabel, 1, 1)
    allData ← allSym times the square root of powerFactor
    // Compute all possible received data
    Reshape H to dimensions (1, numUE, numPacket)
    H_all ← Repeat H for numLabel times
    phase_all ← Reshape randomPhase to dimensions (1, numUE, numPacket)
    phase_all ← Repeat phase_all for numLabel times
    restoreData ← allData × H_all × phase_all
    restoreDataSum ← Sum of restoreData along second dimension
    // Compute mean square error
    Y ← Permute rData to dimensions (2, 1)
    Y ← Repeat Y for numLabel times
    err ← Square of (Y − restoreDataSum)
    // Find the transmitted signals for the minimum mean square error
    idx ← Indices of Minimum values of err along first dimension
    // Obtain labels for estimated symbols
    estLabel ← Initialize zeros matrix of size (numPacket, numUE)
    for p = 1 to numPacket do
        estLabel at (p, :) ← symLabel at (idx(p), :)
    end for
    // Labels for estimated symbols for given indices
    estLabel_1 and estLabel_2 ← Initialize zeros vector of size 1 × numPacket
    for p = 1 to numPacket do
        estLabel_1[p] ← estLabel at (p, idx_1[p])
        estLabel_2[p] ← estLabel at (p, idx_2[p])
    end for
end function(cntd...)
```

Algorithm 5 Deep Learning Signal Detection cntd

// Labels for transmitted symbols
$tLabel \leftarrow$ Initialize zeros matrix of the same size as $tData$
for each c in constQPSK **do**
    Replace $tLabel$ elements where $tData == (1/\sqrt{2} \times \text{constQPSK}[c])$ with c
end for
// Compute Error rate
$errorNum\_1 \leftarrow 1 - (\text{Sum of } (estLabel\_1 \text{ equals } tLabel[1,:]))/numPacket$
$errorNum\_2 \leftarrow 1 - (\text{Sum of } (estLabel\_2 \text{ equals } tLabel[2,:]))/numPacket$
$errorML \leftarrow \text{Vector } [errorNum\_1, errorNum\_2]$
$rLabel \leftarrow \text{Matrix } [estLabel\_1; estLabel\_2]$
return $(errorML, rLabel)$
end function

Acquiring the feature and label

The getFeatureAndLabel function's purpose is to construct a real-valued feature vector and corresponding labels for training from complex data. The function takes as arguments the real (realData) and the imaginary part (imagData) of the data, the labels (labelData) and the target label (targetLabel) then returns the feature vector (feature), the label (label) and the indices (idx) where the data label matches the target label.

When the function is called the indices where the data labels match the target label are found. Then a label vector (label) is created which contains simply the target label. After that, a real data and an imaginary data feature vectors are created which are then combined into a single feature vector (feature). Feature, label and idx are then returned. Algorithm 6 showcases this process in pseudo-code. (Figure A.5 in the Appendix illustrates this process in the form of a flowchart).

Symbol Decoding Using Deep Learning

The symbolDecodeDL function uses a trained neural network to detect received symbols for two users simultaneously. It takes as input arguments the classes of labels (labelClass), the received packets (receivePacket), the labels of the transmitted symbols (dataLabel), the trained neural network (net), the decoding order (decOrder-sc), the QPSK symbols combination class (symClass), the QPSK modulation constellation (constQPSK) then returns the error rate (numErr) and the received labels.

When the function is called two index variables are created for each user based on the initial

Algorithm 6 Get Feature and Label

function GETFEATUREANDLABEL($realData, imagData, labelData, targetLabel$) $\rightarrow$ *feature, label, idx*
  Extract dimensions of $realData \rightarrow (numSym, numSC, \_)$
  $dimFetureVec \leftarrow numSym \times numSC \times 2$
  $idx \leftarrow$ indices in $labelData$ where value $= targetLabel$
  $numSample \leftarrow$ length(idx)
  // Labels
  $label \leftarrow$ array($numSample$, $targetLabel$)
  // Real-valued feature vectors
  $RealCollection \leftarrow$ extract data from $realData$ using idx
  Permute dimensions of $RealCollection$
  Reshape $RealCollection \rightarrow (numSC \times numSym, numSample)$
  // Collect real and imaginary parts of training data as feature vectors
  $ImagCollection \leftarrow$ extract, permute, and reshape $imagData$ similarly
  $feature \leftarrow$ zero matrix of size $(dimFetureVec, numSample)$
  Fill odd rows of $feature$ with $RealCollection$
  Fill even rows of $feature$ with $ImagCollection$
  return *feature, label, idx*
end function

decoding order. Then a feature vector (XTest) and estimated label vector (YTest) are created using the getFeatureAndLabel function for all possible labels. Having created the feature vector, the trained network can be used for signal detection. Based on the result given by the network, missclassified packets are found and then corrected. The error rate for each user are computed as well as the received constellations. The number of errors (numErr) and the received labels are returned. Algorithms 7 and 8 showcase this process in pseudo-code. (Figure A.6 in the Appendix illustrates this process in the form of a flowchart)

Symbol decode Using SIC

The symbolDecodeSIC function performs SIC at the receiver when having a two-user NOMA system. The signal received is a superposition of two signals from two users one considered weak and the other strong. During the SIC process, the signal of the strong user is decoded first and then subtracted from the received signal leaving only the signal of the weak user that is then decoded. The function takes as input the received data (rData), the channel frequency response (H), the decoding order (decOrder), the power factor (powerFactor), QPSK constellation (const), the transmitted packet (tData) then returns the error rate (numErr) and the received labels (rLabel).

Algorithm 7 Symbol Decoding Using Deep Learning

```
function SYMBOLDECODEDL(labelClass, receivePacket, dataLabel, net, decOrder_sc, symClass, constQPSK) return numErr, rLabel
    // Extract dimensions from decOrder_sc
    (numUE, numPacket) ← Dimensions of decOrder_sc
    idx_1 ← First row of decOrder_sc
    idx_2 ← Second row of decOrder_sc
    // Construct feature vectors and estimated labels
    Initialize XTest as an empty list with length numPacket
    Initialize YTest as a zero list with length numPacket
    for each n in labelClass do
        (feature, label, idx) ← getFeatureAndLabel with parameters (real part of receivePacket, imaginary part of receivePacket, dataLabel, n)
        Convert feature matrix to list and store in XTest at idx
        Store label in YTest at idx
    end for
    Convert XTest to column vector
    Convert YTest to categorical column vector
    // Signal detection (prediction)
    YPred ← classify signals using net and XTest
    wrongPred ← Values of YPred where YPred is not equal to YTest
    Convert wrongPred to numbers
    numWrongPred ← Length of wrongPred
    // Identify misclassified packets
    wrongPacket ← Packet indices where YPred is not equal to YTest
    // Calculate correct prediction for each user
    correctPred_1 ← Count where YPred equals YTest
    correctPred_2 ← Count where YPred equals YTest
    for each n from 1 to numWrongPred do
        correctLabel ← dataLabel at wrongPacket[n]
        correctSym ← symClass at correctLabel
        correct_1 ← correctSym at idx_1[wrongPacket[n]]
        correct_2 ← correctSym at idx_2[wrongPacket[n]]
        decodeSym ← symClass at wrongPred[n]
        decodeSym_1 ← decodeSym at idx_1[wrongPacket[n]]
        decodeSym_2 ← decodeSym at idx_2[wrongPacket[n]]
        if correct_1 is equal to decodeSym_1 then
            Increment correctPred_1
        end if
        if correct_2 is equal to decodeSym_2 then
            Increment correctPred_2
        end if
    end for
    // Calculate error rate per user
    numErr_1 ← 1 − correctPred_1 / numPacket
    numErr_2 ← 1 − correctPred_2 / numPacket
    numErr ← Vector [numErr_1, numErr_2]
end function(cntd...)
```

Algorithm 8 Symbol Decoding Using Deep Learning cntd

```
// Obtain received constellation for each user
for each symbol in symClass do
    if symbol is equal to constQPSK[1] then
        Set corresponding value of symLabel to 1
    else if symbol is equal to constQPSK[2] then
        Set corresponding value of symLabel to 2
    else if symbol is equal to constQPSK[3] then
        Set corresponding value of symLabel to 3
    else if symbol is equal to constQPSK[4] then
        Set corresponding value of symLabel to 4
    end if
end for
Reshape symLabel to size (length of labelClass, numUE)
Convert YPred to a list of numbers named estimateLabel
Initialize rLabel as a zero matrix of size (numPacket, numUE)
for each p from 1 to numPacket do
    rLabel[p, :] ← symLabel at estimateLabel[p]
end for
Transpose rLabel
return (numErr, rLabel)
end function
```

When the function is called two index variables are created for each user based on the initial decoding order. Then zero-forcing is applied to user number 1 who is considered to be the strong user. Zero-forcing aims to restore the signal received after passing by the channel by applying the inverse of the frequency response of said channel. The result is then hard decoded. Then the decoded signal is extracted from the received signal to find the signal of user 2, which is considered the weak user and then zero-forcing is applied followed by hard decoding. The labels for the transmitted symbols are found followed by those of the detected symbols and the error rate between the two is calculated. All relevant values are then returned by the function. Algorithm 9 showcases this process in pseudo-code. (Figure A.7 in the Appendix illustrates this process in the form of a flowchart)

Test data

The testData script is the third script to be run which loads the output of both previous scripts and during which every function is called whether directly or indirectly. The process begins by loading essential parameters generated by the trainData script which are the channel response (h), the number of pilot subcarrier (numPSC), the length of the CP (lengthCP), the index

Algorithm 9 Decoding Symbols Using Conventional SIC

function SYMBOLDECODESIC($rData, H, decOrder, powerFactor, const, tData$)
 Extract the dimensions of $tData$ as $(numUE, numPacket)$
 Squeeze the dimensions of matrix H
 // Indices for strong user and weak user
 $idx_1 \leftarrow$ first row of $decOrder$
 $idx_2 \leftarrow$ second row of $decOrder$
 // Strong user
 $zfSym_1 \leftarrow$ array of zeros of size $(1, numPacket)$
 for $p = 1$ to $numPacket$ **do**
  $zfSym_1[p] \leftarrow rData[p]/H[idx_1[p], p]$
 end for
 // Hard decoding
 $decSym_1 \leftarrow \frac{1}{\sqrt{2}} \times$ complex(sign(real($zfSym_1$)), sign(imag($zfSym_1$)))
 // Weak user
 $zfSym_2 \leftarrow$ array of zeros of size $(1, numPacket)$
 for $p = 1$ to $numPacket$ **do**
  $resData \leftarrow rData[p] - H[idx_1[p], p] \times \sqrt{powerFactor[idx_1[p], p]} \times decSym_1[p]$
  $zfSym_2[p] \leftarrow resData/H[idx_2[p], p]$
 end for
 // Hard decoding
 $decSym_2 \leftarrow \frac{1}{\sqrt{2}} \times$ complex(sign(real($zfSym_2$)), sign(imag($zfSym_2$)))
 Combine $decSym_1$ and $decSym_2$ to form matrix $decSym$
 // Obtain labels for transmitted symbols
 $tLabel \leftarrow$ matrix of zeros of size of $tData$
 for each c in $const$ **do**
  $tLabel[tData = \frac{1}{\sqrt{2}} \times c] \leftarrow$ index of c in $const$
 end for
 // Obtain labels for detected symbols
 $rLabel \leftarrow$ matrix of zeros of size of $decSym$
 for each c in $const$ **do**
  $rLabel[decSym = \frac{1}{\sqrt{2}} \times c] \leftarrow$ index of c in $const$
 end for
 // Error rate
 $numErr \leftarrow$ array of zeros of size $(numUE, 1)$
 for $u = 1$ to $numUE$ **do**
  $numErr[u] \leftarrow 1 - \frac{\text{sum}(rLabel \text{ row } u = tLabel \text{ row } u)}{numPacket}$
 end for
 return $numErr, rLabel$
end function

of subcarrier (idx-sc), and the fixed pilot symbols (fixedPilot) then load the trained neural network (net) that resulted from running the trainNN script. Some system parameters are set then the QSPK modulation parameters are generated from which the labels and combination class are extracted. After that, the target SNR are set and the noise is computed. The channel covariance matrix can either be generated by calling the (getRhh) function or by loading in the beginning, a previously generated one. We will not go into detail on how (getRhh) works.

Then two successive loops are used, the first looping over the number of Monte-Carlo iterations and the second over the size of the noise variance. In the second loop, we first create the transmitted packets for which we collect the labels. Following that the power is allocated by calling the allocatePowerfunction and the packets are received by calling the dataTransmission-Reception function. ML detection is applied using the proper function and then LS and MMSE CE are done using the channelEstimation function. After that, the symboleDecodeSIC function is applied to decode based on the LS and MMSE CE. Then the deep learning approach is used by calling the symboleDecodeDL function. After all iterations of both loops finish the error rate for LS, MMSE, and ML are computed and comparative figures are plotted. Algorithms 10 and 11 showcase this process in pseudo-code. (Figure A.8 in the Appendix illustrates this process in the form of a flowchart)

Training data

The trainData script is the first script that needs to be run. It creates the data that will be used to train the neural network. No preexisting variables are needed to be loaded at the beginning and the result saved is the training data and some essential parameters that are required to maintain the same conditions.

In the beginning, the random seed is set to reproduce the static channel. Then some system parameters are initialized like the length of the CP (lenghtCP), the number of pilot subcarriers (numPSC), the number of user equipment (numUE), the number of subcarriers (numSC) and many more. Then all parameters for the QSPK modulation are created, from the constellation to the symbol combination class. Following that the noise is computed, the target SNR are set and the static channel is created with a channel response h.

The allocatePower function is called to calculate the power allocation factor and generate the decoding order. After that, the number of OFDM labels per packet is set and the pilot

Algorithm 10 Testing Data

Clear all variables
// Load training data and essential parameters
`Load` $h, numPSC, lengthCP, idx\_sc, fixedPilot$ `from` trainData.mat
// Load neural netwrok
`Load` net `from` NN.mat
// System parameters
$[numPath, numUE] \leftarrow$ `dimensions of` h
$numSC \leftarrow 64$
$numPSym \leftarrow numUE$
$numDSym \leftarrow 1$
$numSym \leftarrow numPSym + numDSym$
$pilotSpacing \leftarrow \frac{numSC}{numPSC}$
$pilotStart \leftarrow [1, 1]$
// QPSK modulation
$constQPSK \leftarrow [1 - 1j, 1 + 1j, -1 + 1j, -1 - 1j]$
$[a, b, c, d] \leftarrow constQPSK$
// Labels
$symClass \leftarrow$ `combinations of` a, b, c, d
$labelClass \leftarrow 1$ `to size of` $symClass$
// Testing data size
$numPacket \leftarrow 1000$
`Replicate` $fixedPilot$ `for` $numPacket$ `times`
// Power allocations
$targetSNR_1, targetSNR_2 \leftarrow 12$
`Convert` $targetSNR_1, targetSNR_2$ `to linear scale`
$H \leftarrow$ `FFT of` h `along first dimension`
`Compute gain of` H
// Noise computation
$EsN0\_dB \leftarrow$ `sequence` 4 `to` 28 `with step` 2
$EsN0 \leftarrow$ `conversion of` $EsN0\_dB$
$symRate \leftarrow 2$, $Es \leftarrow 1$
$sigPower \leftarrow Es \times symRate$
$symPower \leftarrow \frac{sigPower}{numUE}$
$N0 \leftarrow \frac{sigPower}{EsN0}$
$bw \leftarrow 1$, $nPower \leftarrow N0 \times bw$
$nVar \leftarrow \frac{nPower}{2}$
// Generate channel covariance matrix
$Rhh \leftarrow$ `getRhh`$(numPath, numSC, 1e5)$
// Testing stage
$ITER \leftarrow 1$
`Initialize matrices` $numErr\_ML, numErr\_LS, numErr\_MMSE, numErr\_DL$ `with zeros...`

Algorithm 11 Testing Data cnt

for $it \leftarrow 1$ `to` $ITER$ **do**
    for $snr \leftarrow 1$ `to size of` $nVar$ **do**
        // Transmit packets
        `Initialize` $pilotFrame$ `with zeros`
        `Fill` $pilotFrame$ `with random complex values`
        `Override` $pilotFrame$ `values with` $fixedPilot$ `at specific positions`
        `Initialize` $dataFrame$ `with random complex values`
        `Initialize` $transmitPacket$ `with zeros`
        $transmitPacket \leftarrow pilotFrame$
        `Assign modified` $dataFrame$ `to` $transmitPacket$
        // Collect labels for transmitted data symbols
        `Determine` $tLabel$ `based on` $dataFrame$ `and` $symClass$
        // Allocate Power
        `Calculate` $powerFactor$ `and` $decOrder$
        `Replicate` $h, powerFactor$, `and` $decOrder$ `for` $numPacket$ `times`
        // Received packets
        $[receivePacket, randomPhase] \leftarrow$ `dataTransmissionReception function`
        $[receivePilot, receiveData] \leftarrow$ `extract from` $receivePacket$
        // ML detection with perfect channel estimation
        `Compute` $decOrder\_sc, idx\_1, idx\_2, H\_sc, pF\_sc, rData$, `and` $tData$
        $[numErr\_ML, rLabel\_ML] \leftarrow$ `detectML function`
        // LS and MMSE estimation
        $[H\_LS, H\_MMSE] \leftarrow$ `channelEstimation function`
        $[numErr\_LS, rLabel\_LS] \leftarrow$ `symbolDecodeSIC function with` H_LS
        $[numErr\_MMSE, rLabel\_MMSE] \leftarrow$ `symbolDecodeSIC function with` H_MMSE
        // DL detection
        $numErr\_DL \leftarrow$ `symbolDecodeDL function`
    end for
end for
`Average errors and plot results`

symbol is fixed for all packets. The code then loops over the number of classes i.e. over numLabel, to generate the training data for each class. In the loop, the pilot frame is created followed by the data symbols which are then transmitted and received by calling the dataTransmissionReception function. After that, the feature vector and the label vector are created by calling the getFeatureAndLabel function. After the condition of the loop is fulfilled the training data are saved and the program ends. Algorithms 12 and 13 showcase this process in pseudo-code. (Figure A.9 in the Appendix illustrates this process in the form of a flowchart.)

The following are the main components of the dataset definition used in the train data process:

- **System Parameters:** The code starts by defining various system parameters like the number of pilot subcarriers (numPSC), the number of users (numUE), and other relevant parameters to simulate a NOMA system.

- **Symbol Modulation:** The symbols are modulated using QPSK (Quadrature Phase Shift Keying). The constQPSK variable contains the 4 possible QPSK symbols. A combination of symbols that can be sent by two users is stored in symComb.

- **Noise Computation:** The code calculates the energy per symbol (Es) and the signal power (sigPower). Based on a given SNR (E_sN_0dB), it calculates the noise variance (nVar).

- **Channel and Power Allocation:** There's a static channel realization generated with numPath number of paths. The power allocation for the users and the decoding order are determined using the allocatePower function.

- **Training Data Generation:** For each class (combination of symbols from symComb), the code generates training samples. For each sample the pilot symbols are generated, random data symbols are created, the symbols are the target subcarrier (idx_sc) are then replaced with the current data combination (symComb(n,:)), and the packet, which combines the pilot and data symbols, undergoes data transmission and reception. The received packet is then used to construct the feature vector and labels.

- **Feature Extraction:** The processing of this stage is detailed in "Acquiring the Feature and Label".

- **Data Storage:** The extracted features ('XTrain') and labels ('YTrain') are stored in the 'trainData.mat' file, as shown in the pseudocode in algorithm 13.

Train NN

The trainNN script is the second script to be used during which the training data is loaded and a neural network based on LSTM is created and trained then returned. This script is a conventional neural network training script where the training data is loaded then some training parameters like the number of batches and the number of epochs are set. Following that the structure of the network is set and the network is created, trained and finally saved. No external function is called by this script. Algorithms 14 and 15 showcase this process in pseudo-code. (Figure A.10 in the Appendix illustrates this process in the form of a flowchart).

SNR Range

A range of SNR values was chosen to test this proposed system. The range chosen was from 0 to 30 dBs. This range of SNRs was chosen for the following reasons:

- **Medium SNR range (0 to 20 dB):** This range typically represents urban and suburban environments, which allows for the proposed system to be tested in moderate noise levels that represent most real-world application cases.

- **High SNR range (20 to 30 dB or more):** This range represents good channel conditions, where the Line of Sight (LoS) is typically unobstructed. This range was chosen in order to ensure the deep learning signal detection model does not produce errors in ideal scenarios and to test just how far the tested system can get to its theoretical limits in near perfect conditions.

Algorithm 12 Training Transmitted Data

Clear all variables
Close all figures
Initialize s as random stream with 'mt19937ar' and Seed 1921164231
Set s as the global random stream
// Define parameters
$lengthCP \leftarrow 20$
$numPSC \leftarrow 64$
$numUE \leftarrow 2$
$numSC \leftarrow 64$
$numPSym \leftarrow numUE$
$numDSym \leftarrow 1$
$numSym \leftarrow numPSym + numDSym$
$pilotSpacing \leftarrow numSC/numPSC$
$pilotStart \leftarrow [1, 1]$
// Define QPSK constellation
$constQPSK \leftarrow [1 - 1j, 1 + 1j, -1 + 1j, -1 - 1j]$
$a \leftarrow$ first element of $constQPSK$
$b \leftarrow$ second element of $constQPSK$
$c \leftarrow$ third element of $constQPSK$
$d \leftarrow$ fourth element of $constQPSK$
// Generate combinations of symbols
$symComb \leftarrow [a\ a,\ a\ b,\ a\ c,\ a\ d,\ b\ a,\ b\ b,\ b\ c,\ b\ d,\ c\ a,\ c\ b,\ c\ c,\ c\ d,\ d\ a,\ d\ b,\ d\ c,\ d\ d]$
$labelClass \leftarrow$ sequence from 1 to number of rows in $symComb$
$numLabel \leftarrow$ length of $labelClass$
// Define SNR parameters
$EsN0\_dB \leftarrow 40$
$EsN0 \leftarrow 10^{(EsN0\_dB/10)}$
$symRate \leftarrow 2$
$Es \leftarrow 1$
$sigPower \leftarrow Es \times symRate$
$symPower \leftarrow sigPower/numUE$
$N0 \leftarrow sigPower/EsN0$
$bw \leftarrow 1$
$nPower \leftarrow N0 \times bw$
$nVar \leftarrow nPower/2$
$targetSNR\_1 \leftarrow 12$
$targetSNR\_2 \leftarrow 12$
$targetSNR\_linear\_1 \leftarrow 10^{(targetSNR\_1/10)}$
$targetSNR\_linear\_2 \leftarrow 10^{(targetSNR\_2/10)}$
// Generate channel response
$numPath \leftarrow 20$
$h \leftarrow (1/\sqrt{2}/\sqrt{numPath}) \times$ complex(random numbers with Gaussian distribution of size $numPath \times numUE$)
$H \leftarrow FFT$ of h with size $numSC$ along the first dimension
$gainH \leftarrow$ square magnitude of H transposed

Algorithm 13 Training Transmitted Data cntd

$[powerFactor, decOrder]$ ← allocatePower($symPower$, $gainH$, $targetSNR\_1$, $targetSNR\_2$, $nVar$)
$numPacketClass \leftarrow 3 \times 10^4$
$fixedPilot \leftarrow$ zeros of size $[numPSym, numPSC, numUE]$
$fixedPilot$ for user 1 ← complex numbers with random bipolar values of size $[1, numPSC, 1]$
$fixedPilot$ for user 2 ← complex numbers with random bipolar values of size $[1, numPSC, 1]$
$fixedPilotPacket$ ← replicate $fixedPilot$ with size $[numPSym, numPSC, numUE, numPacketClass]$
$idx\_sc \leftarrow 20$
$XTrain \leftarrow$ empty list
$YTrain \leftarrow$ empty list
Start timer
for $n = 1$ $numLabel$ **do**
    Initialize $pilotFrame$ with zeros of size $[numPSym, numSC, numUE, numPacketClass]$
    $pilotFrame$ for user 1 ← $1/\sqrt{2}$ multiplied by complex random bipolar values of size $[1, numSC, 1, numPacketClass]$
    $pilotFrame$ for user 2 ← $1/\sqrt{2}$ multiplied by complex random bipolar values of size $[1, numSC, 1, numPacketClass]$
    $pilotFrame$ for user 1 at $pilotSpacing$ intervals starting from $pilotStart(1)$ ← $fixedPilotPacket$ for user 1
    $pilotFrame$ for user 2 at $pilotSpacing$ intervals starting from $pilotStart(2)$ ← $fixedPilotPacket$ for user 2
    $dataFrame$ ← $1/\sqrt{2}$ multiplied by complex random bipolar values of size $[numDSym, numSC, numUE, numPacketClass]$
    $currentData$ ← replicate $symComb$ nth row across all packet classes
    reshape $currentData$ to size $[1, 1, numUE, numPacketClass]$
    $dataFrame$ at subcarrier idx_sc ← $1/\sqrt{2}$ multiplied by $currentData$
    $hAll$ ← replicate h for all packet classes
    $powerFactorAll$ ← replicate $powerFactor$ for all packet classes
    $decOrderAll$ ← replicate $decOrder$ for all packet classes
    Initialize $transmitPacket$ with zeros of size $[numSym, numSC, numUE, numPacketClass]$
    $transmitPacket$ for pilot symbols ← $pilotFrame$
    $transmitPacket$ for last symbol ← $dataFrame$
    $[receivePacket, \ ]$ ← dataTransmissionReception($transmitPacket$, $powerFactorAll$, $lengthCP$, $hAll$, $nVar$)
    $dataLabel$ ← n replicated for $numPacketClass$ times
    $[feature, label, \ ]$ ← getFeatureAndLabel(real part of $receivePacket$, imaginary part of $receivePacket$, $dataLabel$, n)
    Convert $feature$ matrix into a cell array with one column per sample
    Append $feature$ cell array to $XTrain$
    Append $label$ array to $YTrain$
end for
Stop timer
Transpose $XTrain$
Convert $YTrain$ to categorical and transpose
Save $XTrain$, $YTrain$, h, $numPSC$, $lengthCP$, idx_sc, $fixedPilot$ into 'trainData.mat'
End Function

Algorithm 14 Training the Neural Network

function TrainNeuralNetwork
    // Initialization
    Clear all variables.
    Close all graphical interfaces.
    // Load data
    $(XTrain, YTrain) \leftarrow$ LoadData($'trainData.mat'$)
    // Parameter Initialization
    $numSC \leftarrow 64$
    $miniBatchSize \leftarrow 4000$
    $maxEpochs \leftarrow 50$
    $inputSize \leftarrow 2 \times numSC \times 3$
    $numHiddenUnits \leftarrow 128$
    $numHiddenUnits2 \leftarrow 64$
    $numHiddenUnits3 \leftarrow numSC$
    $numClasses \leftarrow 16$
    // Define Neural Network Layers
    $layers \leftarrow$ CreateLayers($inputSize, numHiddenUnits, numClasses$)
    // Training Options
    $options \leftarrow$ SetTrainingOptions($maxEpochs, miniBatchSize$)
    // Train the Neural Network
    $startTime \leftarrow$ GetCurrentTime
    $net \leftarrow$ TrainNetwork($XTrain, YTrain, layers, options$)
    $endTime \leftarrow$ GetCurrentTime
    // Save the Trained Network
    SaveModel('NN.mat', net)
end function

Algorithm 15 Training the Neural Network (Supporting Functions)

```
function LoadData(fileName)
    return data from given fileName
end function
function CreateLayers(inputSize, numHiddenUnits, numClasses)
    return a list containing:
    Sequence Input Layer of size inputSize
    LSTM Layer with numHiddenUnits and output mode set to 'last'
    Fully Connected Layer with numClasses
    Softmax Layer
    Classification Layer
end function
function SetTrainingOptions(maxEpochs, miniBatchSize)
    return training options with the following parameters:
    Optimizer: 'adam'
    Initial Learning Rate: 0.01
    Execution Environment: 'auto'
    ... (and other provided parameters)
end function
function TrainNetwork(XTrain, YTrain, layers, options)
    return trained network using given data, layers, and options
end function
function SaveModel(fileName, model)
    Save given model to specified fileName
end function
```

5.2.3 Simulation Results

Fig.5.2 demonstrates that the second user (User 2) does not perform as well as the first user (User 1) in terms of efficiency. This is to be expected due to the nature of NOMA signal superposition, which provides one user with more power than is allocated to the other user based on need. Because it was determined that User 1 required more assistance than User 2, User 1 was granted more authority. The requirement was established after considering the amount of interference that each user was anticipated to experience. This was determined by employing strategies for deep learning, such as the MMSE approximation.

Another observation is that, for both users, the deep learning technique for ML for data regression has performed better than the other deep learning techniques, which are LS and MMSE. This is a conclusion that can be drawn from the results of the comparison. This demonstrates that these data regression procedures, while comparable, are not identical, and that, for our proposed system, adopting the ML methods provides the highest performance.

The training procedure for our conceptual neural network is depicted in Fig.5.3 in regard to the system accuracy and the amount of data lost. It is clear from looking at the figure that the precision of the system continues to rise with each new iteration and epoch that is added. As was to be expected, the accuracy of the system approaches one hundred percent (about 99.78%) by the time the fourth epoch or the 400th iteration has passed. Each iteration and epoch that is performed brings the system's data loss closer and closer to being at a point where there is no data loss at all. But, by the 1000th iteration, or the beginning of the 9th epoch, the system data loss has decreased to a value very close to zero (0.32). Based on these findings, we are able to deduce that there is room for improvement in the data loss of our proposed system. Specifically, there is room for improvement so that the system can achieve near-zero data loss at an earlier stage in a manner that is congruent with the system accuracy reaching one hundred percent. The goal of this is to reduce the total number of necessary iterations so that the proposed neural network can be trained to its full potential and then deployed. This is an opportunity for new works in the future.

The training procedure for the neural network that is depicted in Fig.5.4 was carried out for a total of 30 epochs, or 3600 iterations. The processing depicted in this figure is comparable to that depicted in the previous figure; however, there are very slight drops in performance twice, once at the 20th and 21st epochs and again at the 28th and 29th epochs, as shown by

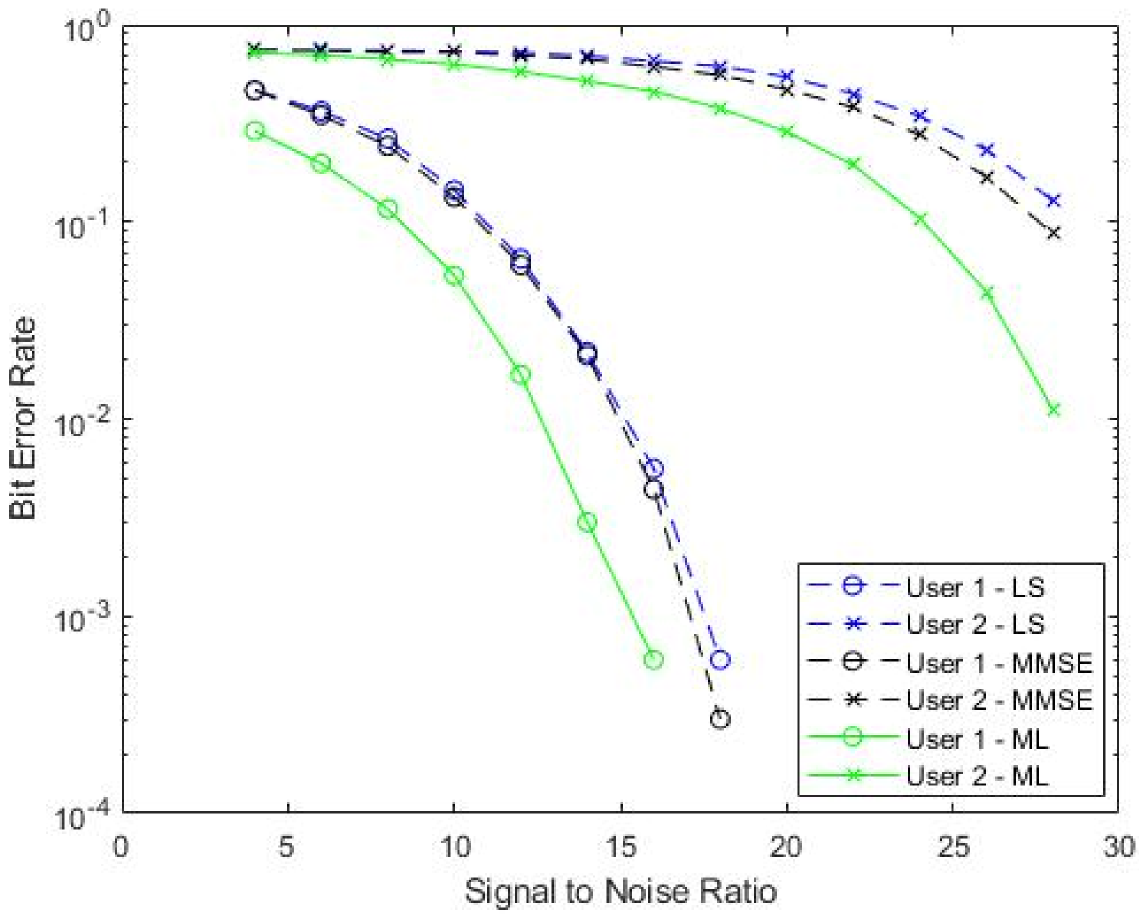


Figure 5.2: BER Vs SNR graph for 2 users using different signal detection techniques

the system accuracy graph and the system data loss graph, respectively. Because the code was being executed on an out of date machine, this has been determined to be a problem with the system. Because the performance drops are fixed in the subsequent epochs and iterations and do not have an effect on the performance of the system as a whole, the error has been assessed to be of minimal significance.

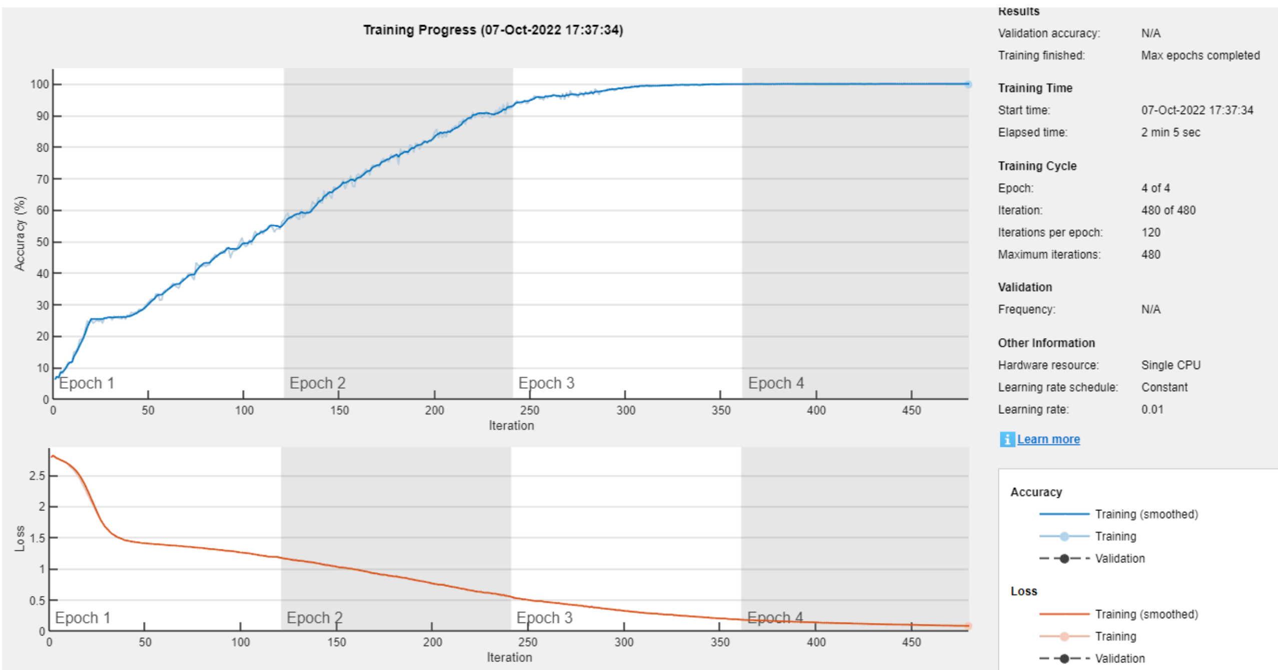


Figure 5.3: Training neural network in terms of system accuracy and data loss

The Effects of Spectral Bandwidth

The spectral bandwidth in a NOMA system can affect deep learning performance for signal detection in several ways. Spectral bandwidth refers to the range of frequencies or the amount of frequency spectrum allocated to a communication system. In the context of signal detection in NOMA, here's how spectral bandwidth can impact deep learning performance:

- **SNR Improvement:** A wider spectral bandwidth typically allows for more power to be transmitted over the channel. This can lead to improved SNR, which is a crucial factor in signal detection. Deep learning models used for signal detection can benefit from higher SNR as it leads to cleaner and more distinguishable signals, making it easier to classify and detect them accurately.

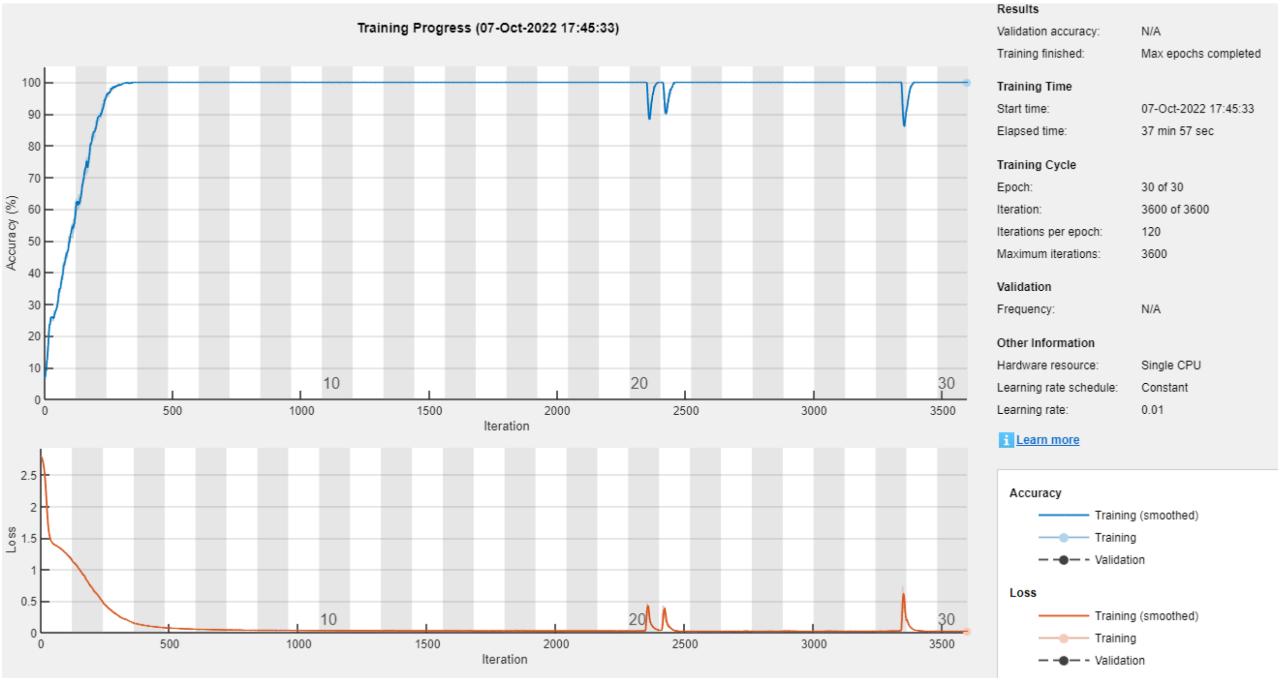


Figure 5.4: Training neural network in terms of system accuracy and data loss for 30 epochs

- **Increased Complexity:** With wider bandwidths, the complexity of the communication system increases. Deep learning models can handle complex systems, but they require more parameters, layers, and computational resources to effectively model and detect signals in wider bandwidths. Training and deploying deep learning models for NOMA signal detection in a high-bandwidth environment can be computationally intensive.

- **Multipath and Interference:** In wireless communication, wider bandwidths may introduce more multipath propagation and interference sources. Deep learning models need to account for these additional complexities when detecting signals. Training deep learning models to cope with a wider range of multipath and interference conditions may require more diverse and extensive training datasets.

- **Data Availability:** The availability of training data is crucial for deep learning models. In NOMA systems with wider spectral bandwidths, obtaining labeled training data for various scenarios and channel conditions can be more challenging. Adequate and diverse training data is essential for deep learning models to generalize well to real-world conditions.

- **Computational Overhead:** Processing wide bandwidths requires more computational power, both during training and inference. Deep learning models used for signal detection may need to be optimized and run on more powerful hardware to handle the increased computational demands.

- **Latency and Real-Time Processing:** In some NOMA systems, especially those targeting low-latency applications, the processing time for signal detection is critical. Wider bandwidths may require faster and more efficient deep learning models and hardware to maintain low latency.

In summary, while wider spectral bandwidths can potentially improve SNR and capacity in NOMA systems, they also introduce challenges for deep learning-based signal detection due to increased complexity, data requirements, and computational demands. To optimize deep learning performance in such systems, a careful balance between spectral bandwidth and the capabilities of the deep learning model, along with the availability of suitable training data, should be considered. Additionally, advanced techniques like transfer learning and model compression can be explored to make deep learning models more efficient for signal detection in NOMA systems with varying spectral bandwidths.

5.3 Summary

In this chapter, Deep Learning was used to enhance the signal detection capabilities of a standard NOMA system. The system model includes two users who are served by NOMA and OMA. Multiple deep learning algorithms for signal detection were tested via MATLAB simulation. The algorithms for signal detection that were tested were LS, ML, and MMSE.

The transmission data is first generated and then trained, where power allocation occurs, after which it is passed to the subsequent stage, where the neural network is first generated and then trained. This process is carried out over two thousand epochs, and the resulting data is then tested in a transmission and reception environment during the final stage. The testing data process is also executed over two thousand epochs until the optimal performance is attained, as demonstrated by the test results.

The results reveal a significant performance disparity between OMA and NOMA, with NOMA exhibiting significantly superior performance. Comparing the various signal detection algorithms, it is demonstrated that the ML protocol achieves the best performance in detecting the NOMA signal with minimal to no errors.

Chapter 6

Conclusion and Future Works

In this chapter, the thesis will conclude by presenting a thorough summary of the carried out work. In order to effectively summarise the work presented in the previous chapters, the conclusion will be comprised of a number of sections.

Answers to Research Questions: in this section, the research questions posed at the introduction chapter will be answered in order to show the progress made throughout the research.

How the Aims and Objectives were Met: the exact methodology the research was carried out with will be presented.

Contribution of Research: The results from the original work that was carried out over the course of the research will be mentioned here.

Future works: areas for possible future research for each research area conducted will be presented.

6.1 Answers to Research Questions

1. Enhancing the diversity gains of a cooperative relaying system by utilising key NOMA features.

    - NOMA techniques were applied to a cooperative relaying system and the results show a significant improvement in overall system performance in regards to BER vs SNR.

2. Investigating the effects of employing EXIT chart analysis on a cooperative NOMA system.

    - The effects of applying EXIT chart analysis on a cooperative NOMA system were investigated theoretically and practically.

3. Improving the performance of a NOMA system by employing a deep learning model.

    - The application of deep learning in communication systems was investigated as well as its deployment in a NOMA system in regards to signal detection.

6.2 How the Aims and Objectives were Met

The methodology undertaken to approach each research topic was as follows:

1. Conduct a thorough literature review in order to build a strong background in the respective area of research and identify areas where a possible original work can be carried out.

2. Conduct a review of already established and published works in order to ensure that the work(s) decided upon in the previous step are wholly original.

3. Build a rudimentary block diagram in order to represent the system model in a visual state.

4. Build a mathematical system model in order to represent the system model in a theoretical state.

5. Use both the block diagram constructed in step 3 and the mathematical model in step 4 to construct the code on MATLAB.

6. Debug the code to ensure there are no conflicts or errors.

7. Run the code and draw conclusions on the propsed system model from the results.

6.3 Contribution of research

The findings of all the original research will be summarised here, along with their overall significance to communication technologies.

6.3.1 Cooperative Relaying with NOMA

Three observations were made concerning cooperative relaying with NOMA. First was the BER versus SNR comparison, which compared a distant user's transmission without cooperative relaying, a distant user's transmission with cooperative relaying, and a nearby user's transmission with cooperative relaying. Note that regardless, all users utilised NOMA. Despite the fact that all users received good service, the result demonstrated that the cell edge users at greater distances received superior service. Among the two distant cell edge users, the one who utilised cooperative relaying in conjunction with NOMA had the best performance. This result demonstrates the possibility of using NOMA and cooperative relaying in tandem to provide service to users with higher QoS.

The second comparison was between power allocation and data rate, which illustrated the number of bits required for transmission at a given power allocation factor. A conventional NOMA system requires 3.5 bits to broadcast at 0.1 power allocation factor, while a cooperative NOMA system only requires 1.5 bits to transmit at the same power allocation factor. This improved performance results in a reduction of required data bits by more than 50 percent while maintaining the same power allocation factor. This demonstrates that cooperative relaying with NOMA uses transmission resources significantly more efficiently than a conventional NOMA system.

The EXIT chart analysis for the cooperative NOMA system revealed that the system initialises at a loss, as both ones and zeros are initialised with a probability of 0.5. This initial loss can be mitigated by repeatedly decoding and despreading between the inner and outer decoders. This result demonstrates that EXIT charts can be useful for cooperative relaying with NOMA, provided a certain number of iterations are performed to eliminate the initial loss and achieve error-free transmission.

6.3.2 EXIT Chart Analysis in Cooperative NOMA

The design and analysis of a cooperative NOMA system produced its data rate and system throughput. The system was subsequently analysed using EXIT chart analysis.

The data rate illustrates the rates attained by each user in three transmission scenarios: far user with NOMA but without cooperation, far user with NOMA and cooperation, and near user with NOMA and cooperation. The result demonstrates that the far user with NOMA and cooperation attained the highest data rate, while the near user with NOMA and cooperation attained the lowest data rate. However, it can also be observed that this performance disparity becomes negligible in cases with high SNR (more than or equal to 32 dBs). This phenomenon is known as the error floor and is a characteristic of NOMA. The error floor can be offset by either switching to OMA in the high SNR range or employing more transmit and receive antennas (MIMO).

The system throughput is equal to the total data transfer rate of all network users. The system throughput is simulated in two transmission scenarios, one with NOMA and the other without. The results mirror those observed in the comparison of data rates, where it was determined that the system throughput with cooperative relaying is significantly higher than that without it. This further demonstrates that NOMA with cooperative relaying has the potential to serve users in scenarios requiring a higher QoS.

Additionally, the normalised throughput is measured for the single and multiple user cases. It is demonstrated that both the single-user and multi-user cases simultaneously achieve higher throughput gains. When ten antennas are utilised, the correlation between the code word used to decode the message decreases. It can be deduced from this result that increasing the number of antenna array combinations improves performance.

Since the initial loss for the system is zero, the EXIT chart analysis demonstrates that the minimum SNR required for error-free transmission is 12 dBs. To achieve a very small BER, the data must be transferred from the MUD decoder to the MUD despreader and back for a specific number of iterations (approximately 1000). Finally, the EXIT chart analysis demonstrates that the user fairness at the power allocation stage was optimal, as the lines for the inner and outer decoders converge at unity gain without intersecting at any earlier point.

6.3.3 Deep Learning for Signal Detection in NOMA

Two observations have been made regarding the use of deep learning techniques to improve signal detection in a NOMA system. First, the performance of the first user is superior to that of the second user. This enhanced performance was anticipated due to the nature of NOMA signal superposition, which allocates more transmission resources to users with greater requirements. The first user was deemed to have a greater need than the second user because the amount of interference the first user was expected to experience was significantly greater than that of the second user. Using MMSE approximation, the precise level of interference was determined.

Three deep learning (data regression) signal detection algorithms were then evaluated and compared. The signal detection algorithms evaluated were ML, LS, and MMSE. ML was shown to be the most effective algorithm for signal detection when compared to LS and MMSE. This leads to the conclusion that, despite the fact that these algorithms are comparable, they are not identical.

Regarding system accuracy and data loss, it is evident from the results that network accuracy increases exponentially with the number of iterations and epochs. At the 400th iteration (or fourth epoch), the system accuracy approaches 100 percent (approximately 99.78 percent), as shown by the results. By increasing the number of iterations further, the proposed system can achieve an error rate that is approximately 0.22 percent. While the error rate is extremely low, it is not zero; therefore, it stands to reason that the proposed system can be improved by reducing the number of training iterations required to achieve an operable error rate.

The accuracy and system data loss graphs illustrate the proposed system's training progress. As the number of iterations and epochs increases, the graphs depict the system's accuracy and data loss steadily improving. However, between the 20th and 21st epoch and again between the 28th and 29th epoch, the performance decreases slightly for a brief period of time before resuming its exponential growth. The performance decreases are infrequent and insufficient to affect the overall performance; as a result, they were deemed a system error caused by operating on an outdated machine.

6.4 Future Works

There are a number of research directions for possible future work that were discovered. These include but are not limited to cancelling out inter-user interference for NOMA based systems where the inter-user interference that is inherent in NOMA systems as a result of conducting its SIC process during the signal detection phase remains an issue and conducting hardware based experiments where factors like signal penetration and fading can be further taken into account by simulating a real life-like scenario.

6.4.1 NOMA with Cooperative Relaying

The following are examples of possible avenues for future research in NOMA systems that employ cooperative relaying:

1. **Multi-user scenario:** It still remains to be seen how the NOMA system will adapt to an increasing number of users as the IUI problem becomes more apparent and relevant with a higher number of users.

2. **Relay nodes as energy harvesters for SWIPT:** it is possible to utilise inactive or idle users in a network as energy harvesting nodes for the application of SWIPT in the network. Doing so will allow for inactive or idle user to play a supporting role in providing service to far away or cell edge users. This scheme might even increase the overall broadcast range of a network.

6.4.2 NOMA with EXIT Chart Analysis

The following are examples of possible research direction(s) for NOMA with EXIT chart analysis:

1. **Increasing the number of inner and outer decoders:** Increasing the number of inner and outer decoders will allow for the EXIT chart analysis to be carried out more efficiently as when the information passes through more inner and outer decoders, the result will be an even lower BER for the system's signal detection. This will allow for EXIT chart analysis to assist in high interference circumstances such as NOMA's inherent inter-user interference.

6.4.3 NOMA with Deep Learning Techniques

The following are examples for future work in the area of power allocation in NOMA using deep learning:

1. **Dynamic mm-Wave NOMA:** The majority of techniques listed for optimal power allocation do not take into account more advanced reinforcement and online learning procedures that update the partition according to dynamic mm-Wave NOMA scenarios; this is a potential area for future research.

2. **Smart Jamming:** some of the presented works can be extended to include more practical applications for NOMA broadcasts, notably with the inclusion of smart jamming, where a programmable jammer takes use of radio equipment to choose jamming policies in a flexible manner.

3. **Multiple Antennas and MUs:** Some of the planned works can be expanded to accommodate users with multiple antennas, as well as MUs for whom the power allocation and beam selection must be updated in real time.

Other probable future research directions include data set collection, model selection, learning mechanism and performance analysis, DRL for the wireless physical layer, and model compression for deep learning-based 5G and beyond.

The following are examples for future work in the area of signal detection in NOMA using deep learning:

1. **SER Error Floor:** by considering a deeper architecture, increasing the number of training epochs, and increasing the number of filters, the SER floor may be reduced.

2. **Channel Profile:** the works can also be extended for practical applications by considering a dynamic channel profile as well as evaluating further the effects of channel fading on the system; this can be done with the aim to further test the sturdiness of the DL model to dynamic/unknown channel conditions.

3. **System Security:** issues to be addressed regarding security issues are, the selection of the optimal threshold as well as the distance issue of the attackers, i.e., dynamically changing the distance of the attackers to the BS. Attackers may also correlate with other

users if they are close enough. The performance of the algorithms needs to be evaluated under the aforementioned condition.

4. **Signal Detection in Different Types of NOMA:** the proposed algorithms can be implemented to detect signals in other types of NOMA systems, including, SCMA, pattern division multiple access, and MUSA.

Furthermore, the works can be extended to consider evaluating the performance under varying channel conditions and with the implementation of multiple clusters. Signal detection utilising RNNs can be further explored. CNNs can also be applied to the proposed system on account of its potential for signal detection. Implementing some of the proposed works for cases with varying transmission parameters and channel conditions can also be considered. The algorithms' performance can also be improved for the high SNR regime. Some of the proposed works can also be extended for the multi-antenna case as well as a number of interesting applications, such as, mmWave CE, MIMO detection, and direction of arrival estimation.

Overall, drawing from the observations made from all simulation results, it has been made abundantly clear that, while NOMA is an effective candidate for meeting the needs of the next generation of wireless communications, in order to overcome NOMA's issues, such as inter-user interference, it is imperative for it to be combined with other communication technologies. NOMA has been shown to have a great affinity for working alongside other communication technologies.

Bibliography

[1] Ding, Z., Yang, Z., Fan, P., Poor, H. V. (2017), "NOMA technology for 5G and beyond," IEEE Communications Standards Magazine, 1(2), 70-75.

[2] S. Parkvall, E. Dahlman, A. Furuskar and M. Frenne, "NR: The New 5G Radio Access Technology," in IEEE Communications Standards Magazine, vol. 1, no. 4, pp. 24-30, Dec. 2017.

[3] L. Chettri and R. Bera, "A Comprehensive Survey on Internet of Things (IoT) Toward 5G Wireless Systems," in IEEE Internet of Things Journal, vol. 7, no. 1, pp. 16-32, Jan. 2020.

[4] N. Al-Falahy and O. Y. Alani, "Technologies for 5G Networks: Challenges and Opportunities," in IT Professional, vol. 19, no. 1, pp. 12-20, Jan.-Feb. 2017.

[5] Liu, Y., Qin, Z., Elkashlan, M., Ding, Z., Nallanathan, A., Hanzo, L. (2017). Non-orthogonal multiple access for 5G and beyond. Proceedings of the IEEE, 105(12), 2347-2381.

[6] Z. Elsaraf, F. Khan, Q. Ahmed "Code Domain Non-orthogonal Multiple Access Techniques for 5G mobile communications", Emerging Technology Conference Conference (EMiT19), University of Huddersfield, January 2019.

[7] C. -Y. Chang, S. -K. Lee and M. -C. Lin, "Performance Comparison of GDMA, LDS-CDMA, and SCMA for Transmissions over Rayleigh Fading Channels," 2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall), Norman, OK, USA, 2021, pp. 1-5.

[8] T. Huang, J. Yuan, X. Cheng, and W. Lei, "Design of degrees of distribution of LDS-OFDM," in Proc. Int. Conf. Signal Process. Commun. Syst., Cairns, QLD, Australia, Dec. 2015, pp. 1-6.

[9] D. Roque and C. Siclet, ”Performances of Weighted Cyclic Prefix OFDM with Low-Complexity Equalization,” in IEEE Communications Letters,” vol. 17, no. 3, pp. 439-442, March 2013.

[10] M. Vameghestahbanati et al.,”Multidimensional Constellations for Uplink SCMA Systems: A Comparative Study,” in IEEE Communications Surveys and Tutorials, vol. 21, no. 3, pp. 2169-2194, third quarter 2019.

[11] D. Cai, P. Fan, X. Lei, Y. Liu and D. Chen, ”Multi-Dimensional SCMA Codebook Design Based on Constellation Rotation and Interleaving,” 2016 IEEE 83rd Vehicular Technology Conference (VTC Spring), Nanjing, China, 2016, pp. 1-5.

[12] Z. Yang, J. Cui, X. Lei, Z. Ding, P. Fan, and D. Chen, ”Impact of factor graph on average sum rate for uplink sparse code multiple access systems,” IEEE Access, vol. 4, pp. 6585-6590, 2016.

[13] L. Yang, Y. Liu, and Y. Siu, ”Low complexity message passing algorithm for SCMA system,” IEEE Commun. Lett., vol. 20, no. 12, pp. 2466-2469, Dec. 2016.

[14] Ding et al., ”Application of non-orthogonal multiple access in LTE and 5G networks,” IEEE Commun. Mag., vol. 55, no. 2, pp. 185-191, Feb. 2017.

[15] I. Budhiraja et al., ”A Systematic Review on NOMA Variants for 5G and Beyond,” in IEEE Access, vol. 9, pp. 85573-85644, 2021.

[16] Z. Wei, J. Yuan, D. W. K. Ng, M. Elkashlan, and Z. Ding, ”A survey of downlink non-orthogonal multiple access for 5G wireless communication networks,” ZTE Commun., vol. 14, no. 4, pp. 17-26, Oct. 2016.

[17] X. Chen, R. Jia and D. W. K. Ng, ”On the Design of Massive Non-Orthogonal Multiple Access With Imperfect Successive Interference Cancellation,” in IEEE Transactions on Communications, vol. 67, no. 3, pp. 2539-2551, March 2019.

[18] Y. Du et al., ”Joint Channel Estimation and Multiuser Detection for Uplink Grant-Free NOMA,” in IEEE Wireless Communications Letters, vol. 7, no. 4, pp. 682-685, Aug. 2018.

[19] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan and V. K. Bhargava, "A Survey on Non-Orthogonal Multiple Access for 5G Networks: Research Challenges and Future Trends," in IEEE Journal on Selected Areas in Communications, vol. 35, no. 10, pp. 2181-2195, Oct. 2017.

[20] S. M. R. Islam, M. Zeng, O. A. Dobre and K. -S. Kwak, "Resource Allocation for Downlink NOMA Systems: Key Techniques and Open Issues," in IEEE Wireless Communications, vol. 25, no. 2, pp. 40-47, April 2018.

[21] Z. Chen, Z. Ding, X. Dai, and R. Zhang, "A mathematical proof of the superiority of NOMA compared to conventional OMA," IEEE Trans. Signal Process., 2016.

[22] P. Xu, Z. Ding, X. Dai, and H. V. Poor, "A new evaluation criterion for non-orthogonal multiple access in 5G software defined networks," IEEE Access, vol. 3, pp. 1633-1639, 2015.

[23] P. Wang, J. Xiao, and L. Ping, "Comparison of orthogonal and non-orthogonal approaches to future wireless cellular systems," IEEE Veh. Technol. Mag., vol. 1, no. 3, pp. 4-11, Sep. 2006.

[24] Anass Benjebbour et al., "Concept and practical considerations of non-orthogonal multiple access (NOMA) for future radio access, "Intelligent Signal Processing and Communications Systems (ISPACS) 2013 International Symposium on, pp. 770-774 Nov 2013.

[25] S. M. Raizul Islamet al., "Power-Domain Non-Orthogonal Multiple Access (NOMA) in 5G Systems: Potentials and Challenges," IEEE Communications Surveys and Tutorials, vol.PP, no.99, 25 Oct. 2016, pp. 1-1.

[26] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (NOMA) for cellular future radio access," in Proc. IEEE Veh. Technol. Conf., Dresden, Germany, Jun. 2013, pp. 1-5.

[27] J. Choi, "Power allocation for max-sum rate and max-min rate proportional fairness in NOMA," IEEE Commun. Lett., vol. 20, no. 10, pp. 2055-2058, Oct. 2016.

[28] Z. Ding, P. Fan, and H. V. Poor, "Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions," IEEE Trans. Veh. Technol., vol. 65, no. 8, pp. 6010-6023, Aug. 2016.

[29] J. Choi, ”Minimum power multicast beamforming with superposition coding for multiresolution broadcast and application to NOMA systems,” IEEE Trans. Commun., vol. 63, no. 3, pp. 791-800, Mar. 2015.

[30] W. Shin, M. Vaezi, B. Lee, D. J. Love, J. Lee, and H. V. Poor, ”Non-orthogonal multiple access in multi-cell networks: Theory, performance, and practical challenges,” IEEE Commun. Mag., vol. 55, no. 10, pp. 176-183, Oct. 2017.

[31] J. Men, J. Ge and C. Zhang, ”Performance Analysis of Nonorthogonal Multiple Access for Relaying Networks Over Nakagami-m Fading Channels,” in IEEE Transactions on Vehicular Technology, vol. 66, no. 2, pp. 1200-1208, Feb. 2017.

[32] Z. Ding, M. Peng and H. V. Poor, ”Cooperative Non-Orthogonal Multiple Access in 5G Systems,” in IEEE Communications Letters, vol. 19, no. 8, pp. 1462-1465, Aug. 2015.

[33] Y. Liu, Z. Ding, M. Elkashlan and H. V. Poor, ”Cooperative Non-orthogonal Multiple Access With Simultaneous Wireless Information and Power Transfer,” in IEEE Journal on Selected Areas in Communications, vol. 34, no. 4, pp. 938-953, April 2016.

[34] Q. Y. Liau and C. Y. Leow, ”Successive User Relaying in Cooperative NOMA System,” in IEEE Wireless Communications Letters, vol. 8, no. 3, pp. 921-924, June 2019.

[35] J. Kim and I. Lee, ”Non-Orthogonal Multiple Access in Coordinated Direct and Relay Transmission,” in IEEE Communications Letters, vol. 19, no. 11, pp. 2037-2040, Nov. 2015

[36] J. N. Laneman, D. N. C. Tse and G. W. Wornell, ”Cooperative diversity in wireless networks: Efficient protocols and outage behavior,” in IEEE Transactions on Information Theory, vol. 50, no. 12, pp. 3062-3080, Dec. 2004.

[37] Z. Ding and H. V. Poor, ”Cooperative energy harvesting networks with spatially random users,” IEEE Signal Process. Lett., vol. 20, no. 12, pp. 1211-1214, Dec. 2013.

[38] Z. Ding, I. Krikidis, B. Sharif, and H. V. Poor, ”Wireless information and power transfer in cooperative networks with spatially random relays,” IEEE Trans. Wireless Commun., vol. 13, no. 8, pp. 4440-4453, Aug. 2014.

[39] A. S. Marcano and H. L. Christiansen, "System-Level Performance of C-NOMA: A Cooperative Scheme for Capacity Enhancements in 5G Mobile Networks," 2017 IEEE 86th Vehicular Technology Conference (VTC-Fall), Toronto, pp. 1-6 ON, 2017.

[40] W. Mei and R. Zhang, "Cooperative NOMA for Downlink Asymmetric Interference Cancellation," in IEEE Wireless Communications Letters, vol. 9, no. 6, pp. 884-888, June 2020

[41] Q. Li, M. Wen, E. Basar, H. V. Poor and F. Chen, "Spatial Modulation-Aided Cooperative NOMA: Performance Analysis and Comparative Study," in IEEE Journal of Selected Topics in Signal Processing, vol. 13, no. 3, pp. 715-728, June 2019.

[42] D. J. Costello and G. D. Forney, "Channel coding: The road to channel capacity," in Proceedings of the IEEE, vol. 95, no. 6, pp. 1150-1177, June 2007.

[43] S. Ten Brink, "Designing iterative decoding schemes with the extrinsic information transfer chart," AEU Int. J. Electron. Commun, vol.54, no.6, pp. 389-398, 2000, Nov. 2001.

[44] Y. Zhang, H. M. Wang, T. X. Zheng, and Q. Yang, "Energy-efficient transmission design in non-orthogonal multiple access," IEEE Trans. Veh. Technol., vol. 66, no. 3, pp. 2852–2857, Mar. 2017.

[45] P. Ongsulee, "Artificial intelligence, machine learning and deep learning," 2017 15th International Conference on ICT and Knowledge Engineering (ICTKE), Bangkok, Thailand, pp.1-6, 2017.

[46] Ludovic Arnold, Sébastien Rebecchi, Sylvain Chevallier, Hélène Paugam-Moisy. "An Introduction to Deep Learning," European Symposium on Artificial Neural Networks (ESANN), Bruges, Belgium, Apr 2011.

[47] Juergen Schimdhuber, "Deep Learning," Scholarpedia, 10(11):32832, 2015.

[48] Nikolaus Kriegeskorte, Tal Golan, "Neural network models and deep learning," Current Biology, Volume 29, Issue 7, nPages R231-R236, ISSN 0960-9822, 2019.

[49] Ian Goodfellow, Yoshua Bengio, Aaron Courville, "Deep Learning" Cambridge, MA : MIT Press, 2017.

[50] Fan J, Ma C, Zhong Y., "A selective overview of deep learning," Stat Sci. 2021 May;36(2):264-290, PMID: 34305305; PMCID: PMC8300482, Apr 2020.

[51] LeCun, Y., Bengio, Y. Hinton, G., "Deep learning," Nature 521, 436–444, 2015.

[52] S. Albawi, T. A. Mohammed and S. Al-Zawi, "Understanding of a convolutional neural network," 2017 International Conference on Engineering and Technology (ICET), Antalya, Turkey, pp. 1-6, 2017.

[53] Z. Li, F. Liu, W. Yang, S. Peng and J. Zhou, "A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects," in IEEE Transactions on Neural Networks and Learning Systems, vol. 33, no. 12, pp. 6999-7019, Dec. 2022.

[54] Mehak, m., WHIG, P., "More on Convolution Neural Network CNN," International Journal of Sustainable Development in Computing Science, 4(1), 2022.

[55] Keiron O'Shea and Ryan Nash, "An Introduction to Convolutional Neural Networks," eprint 1511.08458, arXiv, 2015.

[56] Cheng, B., Titterington, D. M., "Neural Networks: A Review from a Statistical Perspective," Statistical Science, 9(1), 2–30, 1994.

[57] Picton, P., "What is a Neural Network?. In: Introduction to Neural Networks," Palgrave, London, 1994.

[58] A. K. Jain, Jianchang Mao and K. M. Mohiuddin, "Artificial neural networks: a tutorial," in Computer, vol. 29, no. 3, pp. 31-44, March 1996.

[59] Chris M. Bishop, "Neural Networks and their Applications", Review of scientific instruments, volume 65, issue 6, 1998.

[60] Minje Kim, Paris Smaragdis, "Bitwise Neural Networks", CoRR, volume 1601.06071, Aug 2018.

[61] Cunningham, P., Cord, M., Delany, S.J. , "Supervised Learning. In: Cord, M., Cunningham, P. (eds) Machine Learning Techniques for Multimedia," Cognitive Technologies, Springer, Berlin, Heidelberg, 2008.

[62] Hastie, T., Tibshirani, R., Friedman, J., "Overview of Supervised Learning. In: The Elements of Statistical Learning," Springer Series in Statistics, Springer, New York, NY, 2009.

[63] Liu, B., "Supervised Learning. In: Web Data Mining," Data-Centric Systems and Applications, Springer, Berlin, Heidelberg, 2011.

[64] Hastie, T., Tibshirani, R., Friedman, J., "Unsupervised Learning. In: The Elements of Statistical Learning," Springer Series in Statistics. Springer, New York, NY, 2009.

[65] H. U. Dike, Y. Zhou, K. K. Deveerasetty and Q. Wu, "Unsupervised Learning Based On Artificial Neural Network: A Review," 2018 IEEE International Conference on Cyborg and Bionic Systems (CBS), Shenzhen, China, pp. 322-327, 2018.

[66] Yuxi Li, "Deep Reinforcement Learning: An Overview," CoRR journal, volume 1701.07274, Aug 2018.

[67] K. Arulkumaran, M. P. Deisenroth, M. Brundage and A. A. Bharath, "Deep Reinforcement Learning: A Brief Survey," in IEEE Signal Processing Magazine, vol. 34, no. 6, pp. 26-38, Nov. 2017.

[68] Marco Wiering, Martin van Otterlo, "Reinforcement Learning: State of the Art," volume 12, 2012.

[69] Thomas M. Moerland, Joost Broekens, Aske Plaat and Catholijn M. Jonker , "Model-based Reinforcement Learning: A Survey", Foundations and Trends® in Machine Learning: Vol. 16: No. 1, pp 1-118, 2023.

[70] Vincent François-Lavet, Peter Henderson, Riashat Islam, Marc G. Bellemare and Joelle Pineau , "An Introduction to Deep Reinforcement Learning", Foundations and Trends® in Machine Learning: Vol. 11: No. 3-4, pp 219-354, 2018.

[71] S. ten Brink, "Designing iterative decoding schemes with the extrinsic information transfer chart," AEU Inernational Journal of Electronic and Communications, vol. 54, no. 6, pp. 389–398, 2000.

[72] M. El-Hajjar and L. Hanzo, "EXIT charts for system design and analysis," IEEE Communications Surveys Tutorials, vol. 16, no. 1, pp. 127–153, May 2014.

[73] S. Sugiura, S. Chen, and L. Hanzo, "Coherent and differential spacetime shift keying: A dispersion matrix approach," IEEE Transactions on Communications, vol. 58, no. 11, pp. 3219–3230, 2010.

[74] Watanachaturaporn, Pakorn and Varshney, P.K. and Arora, Manoj. "Evaluation of factors affecting support vector machines for hyperspectral classification". 2022.

[75] Koehrsen, W. "Overfitting vs. Underfitting: A Complete Example". towardsdatascience.com. https://towardsdatascience.com/overfitting-vs-underfitting-a-complete-example-d05dd7e19765. 2018.

[76] D."Cross-Validation". www.kaggle.com. $https://www.kaggle.com/code/dansbecker/cross-validation/notebook$. 2017.

[77] Sara G., "Gaussian Mixture Models part 2: Clustering". gsarantitis.wordpress.com. https://gsarantitis.wordpress.com/2020/09/13/gaussian-mixture-models-part-2-clustering/. 2020.

[78] Pedregosa et al., JMLR 12, pp. 2825-2830, 2011

[79] Osiński, B., Budek, K. "What is reinforcement learning? The complete guide". deepsense.ai. https://deepsense.ai/what-is-reinforcement-learning-the-complete-guide/. 2018.

[80] Bhatt, S. "Reinforcement Learning 101". towardsdatascience.com. https://towardsdatascience.com/reinforcement-learning-101-e24b50e1d292. 2018.

[81] A. Ahmed, P. Botsinis, S. Won, L. Yang and L. Hanzo, "EXIT Chart Aided Convergence Analysis of Recursive Soft m-Sequence Initial Acquisition in Nakagami-m Fading Channels," IEEE Transactions on Vehicular Technology, vol. 67, no. 5, pp. 4655-4660, May 2018.

[82] C. Berrou, A. Glavieux and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: turbo-codes (1)," Proc. IEEE International Conference on Communication (ICC), Geneva, Switzerland, pp. 1064-1070, May 1993.

[83] 1. Lodge, R. Young, P. Hoeher, and 1. Hagenauer, "Separable MAP 'filters' for the decoding of product and concatenated codes," Proc. IEEE International Conference on Communication (ICC), Geneva, Switzerland, pp. 1740-1745, May 1993.

[84] J. Hagenauer, "The exit chart - introduction to extrinsic information transfer in iterative processing," 2004 12th European Signal Processing Conference, pp. 1541-1548, 2004.

[85] S. ten Brink, "Convergence behaviour of iteratively decoded parallel concatenated codes," IEEE Trans. on Comm., vol. 49, Oct 2001.

[86] Schreckenbach, F; Gortz, N.; Hagenauer, J.; Bauch, G.: Optimized Symbol Mappings for Bit Interleaved Coded Modulation with Iterative Decoding. - In:2003 IEEE Global Telecommunications Conference (GLOBECOM 2003), San Francisco, December 2003.

[87] Y. Ba¸stanlar and M. Özuysal, Introduction to Machine Learning. Totowa, NJ, USA: Humana Press, pp. 105–128, 2014.

[88] O'Shea, T.; Timothy, J.; Hoydis, J. An Introduction to Machine Learning Communications Systems. arXiv:1702.00832, 2017.

[89] B. Di, L. Song, and Y. Li, "Sub-channel assignment, power allocation, and user scheduling for non-orthogonal multiple access networks," IEEE Trans. Wireless Commun., vol. 15, no. 11, pp. 7686–7698, Nov. 2016.

[90] Nachmani, E; Be'ert, Y; Burshtein, D. Learning to Decode Linear Codes Using Deep Learning. , arXiv:1607.04793, 2016.

[91] Gruber, T.; Cammerer, S.; Hoydis, J.; Brink, S.T. On Deep Learning-Based Channel Decoding. arXiv:1701.07738, 2017.

[92] Liang, F.; Shen, C.; Wu, F. An Iterative BP-CNN Architecture for Channel Decoding. IEEE J. Sel. Top. Sign. Process. 12, 144–159, 2018.

[93] F. Fang, H. Zhang, J. Cheng, and V. C. M. Leung, "Energy-efficient resource allocation for downlink non-orthogonal multiple access network," IEEE Trans. Commun., vol. 64, no. 9, pp. 3722–3732, Sep. 2016.

[94] J. Cui, Z. Ding, P. Fan and N. Al-Dhahir, "Unsupervised Machine Learning-Based User Clustering in Millimeter-Wave-NOMA Systems," in IEEE Transactions on Wireless Communications, vol. 17, no. 11, pp. 7425-7440, Nov. 2018.

[95] Sun, B.; Feng, H. Efficient Compressed Sensing for Wireless Neural Recording: A Deep Learning Approach. IEEE Signal Process. Lett. 24, 863-867, 2017.

[96] Xu, Y.; Li, D.; Wang, Z. A Deep Learning Method Based on Convolutional Neural Network for Automatic Modulation Classification of Wireless Signals. In Proceedings of the 2nd EAI International Conference on Machine Learning and Intelligent Communications (MLICOM 2017), Weihai, China, 5–6 August 2017.

[97] Wang, Z. The Applications of Deep Learning on Traffic Identification. Available online: $https://www.blackhat.com/docs/us-15/materias/us-15-Wang-The-Applications-Of-DeepLearning-On-TrafficIdentification-wp.df$ (2017).

[98] Aceto, G.; Ciuonzo, D.; Montieri, A.; Pescape, A. Mobile Encrypted Traffic Classification Using Deep Learning. Available online: $https://tma.ifip.org/2018/wp-contentuploads/sites/3/2018/06/tma2018_paper0.pdf$ (2019).

[99] Jeon, Y.; Hong, S.N.; Lee, N. Blind Detection for MIMO Systems With Low-Resolution ADCs Using Supervised Learning. In Proceedings of the IEEE International Conference on Communications (IEEE ICC 2017), Paris, France, 21–25 May 2017

[100] Wang, C. Research and Application of Traffic Sign Detection and = Recognition Based on Deep Learning. In Proceedings of the International Conference on Robots and Intelligent System (ICRIS2018), Amsterdam, The Netherlands, 21–23 February 2018.

[101] Ye, H.; Li, G.Y.; Juang, B.H. Power of Deep Learning for Channel Estimation and Signal Detection in OFDM Systems. IEEE Wirel. Commun. Lett. 7, 114–117, 2017.

[102] O'Shea, T.; Erpek, T.; Clancy, T.C. Deep Learning-Based MIMO Communications. arXiv:1707.07980v1, 2017.

[103] Gui, G.; Huang, H.; Juang, B.H.; Song, Y.; Sari, H. Deep Learning for an Effective Non-orthogonal Multiple Access Scheme. IEEE Trans. Veh. Technol. 67, 8440–8450, 2018.

[104] Lin C, Chang Q, Li X. A Deep Learning Approach for MIMO-NOMA Downlink Signal Detection. Sensors. 19(11):2526, 2019.

[105] C. L. Wang, J. Y. Chen, and Y. J. Chen, “Power allocation for a downlink non-orthogonal multiple access system,” IEEE Wireless Commun. Lett., vol. 5, no. 5, pp. 532–535, Oct. 2016.

[106] P. P. Shinde and S. Shah, ”A Review of Machine Learning and Deep Learning Applications,” 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, pp.1-6, 2018.

[107] S. Timotheou and I. Krikidis, “Fairness for non-orthogonal multiple access in 5G systems,” IEEE Signal Process. Lett., vol. 22, no. 10, pp. 1647–1651, Oct. 2015.

[108] Janiesch, C., Zschech, P. Heinrich, ”K. Machine learning and deep learning,” Electron Markets 31, 685–695, 2021.

[109] X. Zhou et al., ”Active Terminal Identification, Channel Estimation, and Signal Detection for Grant-Free NOMA-OTFS in LEO Satellite Internet-of-Things,” in IEEE Transactions on Wireless Communications, vol. 22, no. 4, pp. 2847-2866, April 2023.

[110] A. Almohamad, M. O. Hasna, S. Althunibat and K. Qaraqe, ”A Novel Downlink IM-NOMA Scheme,” in IEEE Open Journal of the Communications Society, vol. 2, pp. 235-244, 2021.

[111] N. Zhang, K. Cheng and G. Kang, ”A Machine-Learning-Based Blind Detection on Interference Modulation Order in NOMA Systems,” in IEEE Communications Letters, vol. 22, no. 12, pp. 2463-2466, Dec. 2018.

[112] C. Qing, B. Cai, Q. Yang, J. Wang and C. Huang, ”Deep Learning for CSI Feedback Based on Superimposed Coding,” in IEEE Access, vol. 7, pp. 93723-93733, 2019.

[113] Diederik P. Kingma and Jimmy Ba, ”Adam: A Method for Stochastic Optimization,” arXiv, eprint 1412.6980, 2017.

[114] P. Q. Thai, N. T. Long and H. H. Tin, ”De-Multiplexing of NOMA VLC Signals Using Probabilistic Neural Network,” 2022 IEEE Ninth International Conference on Communications and Electronics (ICCE), Nha Trang, Vietnam, pp. 18-22, 2022.

[115] W. Kim, Y. Ahn and B. Shim, "Deep Neural Network-Based Active User Detection for Grant-Free NOMA Systems," in IEEE Transactions on Communications, vol. 68, no. 4, pp. 2143-2155, April 2020.

[116] N. Wang, L. Jiao, A. Alipour-Fanid, M. Dabaghchian and K. Zeng, "Pilot Contamination Attack Detection for NOMA in 5G mm-Wave Massive MIMO Networks," in IEEE Transactions on Information Forensics and Security, vol. 15, pp. 1363-1378, 2020.

[117] D. P. Jana and S. Gupta, "Machine Learning Enabled Detection for QPSK-PD-NOMA System Employing Single Mode Fiber," 2020 National Conference on Communications (NCC), Kharagpur, India, pp.1-5, 2020.

[118] Krishna Chitti and Joao Vieira and Behrooz Makki, "Deep-Learning based Multiuser Detection for NOMA," arXiv, eprint 2011.11752, 2020.

[119] Ahmet Emir, Ferdi Kara, Hakan Kaya, Xingwang Li, "Deep learning-based flexible joint channel estimation and signal detection of multi-user OFDM-NOMA," Physical Communication, Volume 48, 101443, ISSN 1874-4907, 2021.

[120] N. Ye, X. Li, H. Yu, L. Zhao, W. Liu and X. Hou, "DeepNOMA: A Unified Framework for NOMA Using Deep Multi-Task Learning," in IEEE Transactions on Wireless Communications, vol. 19, no. 4, pp. 2208-2225, April 2020.

[121] Michael Crawshaw, "Multi-Task Learning with Deep Neural Networks: A Survey," arXiv, eprint 2009.09796, 2020.

Appendix A

Appendix

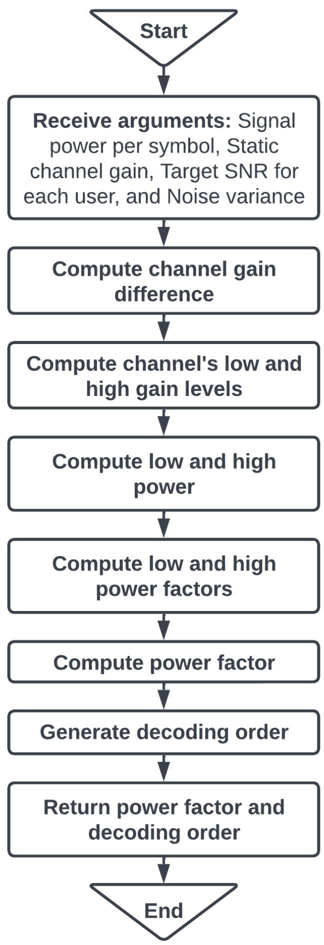


Figure A.1: Flowchart for the power allocation algorithm (allocatePower) for the proposed system

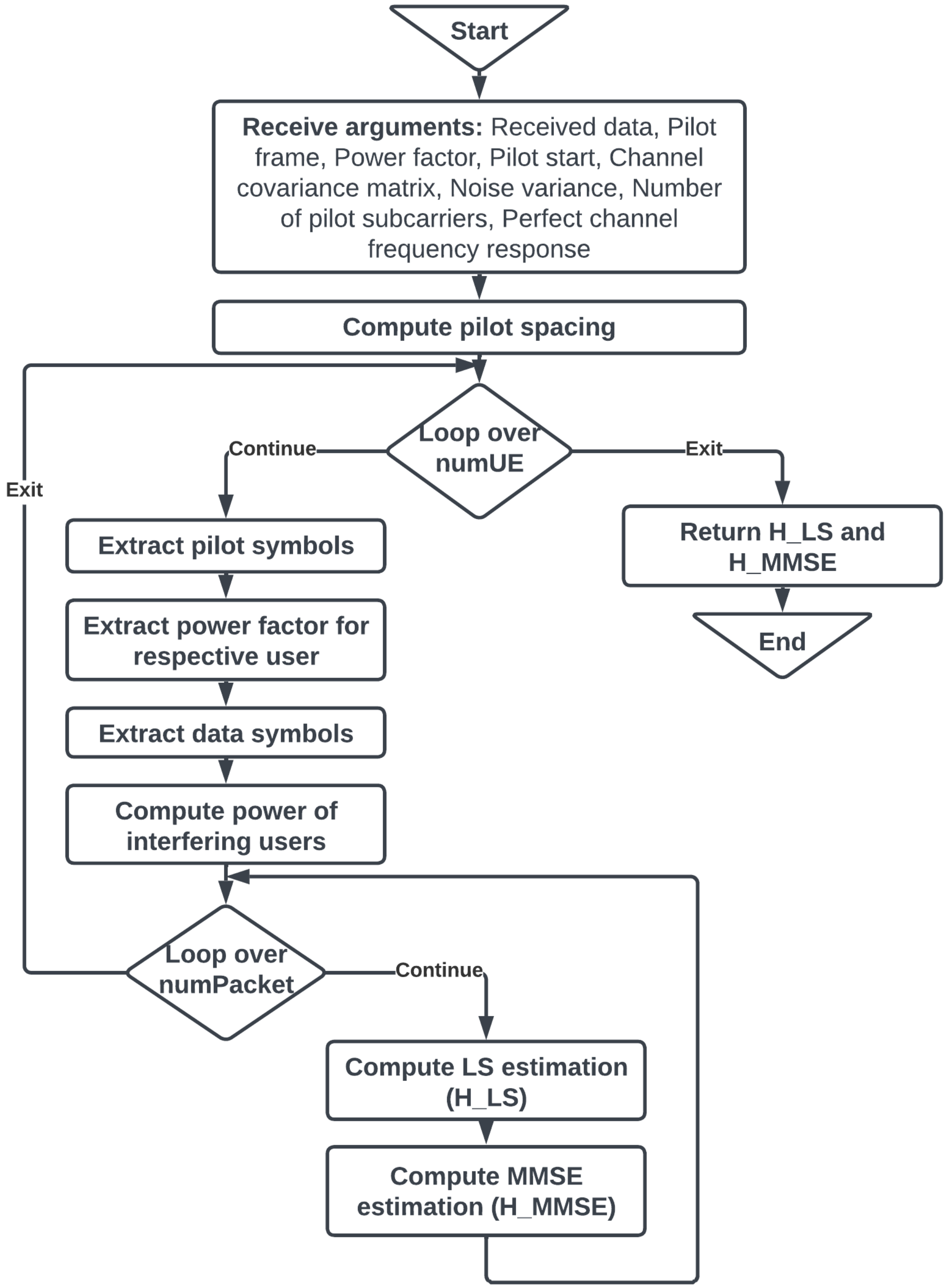


Figure A.2: Flowchart for the CE algorithm for the proposed system

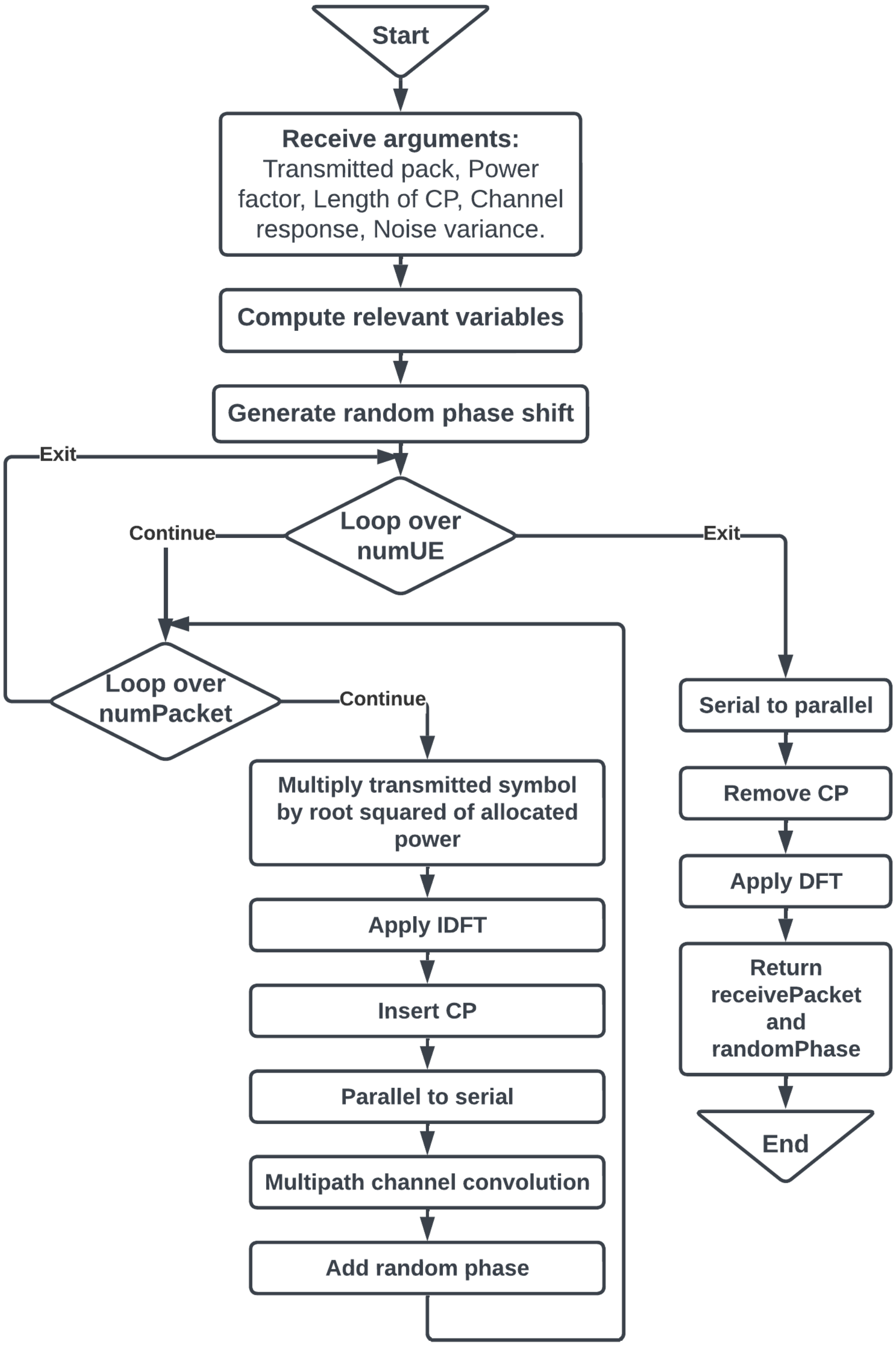


Figure A.3: Flowchart for the data transmission and reception algorithm for the proposed system

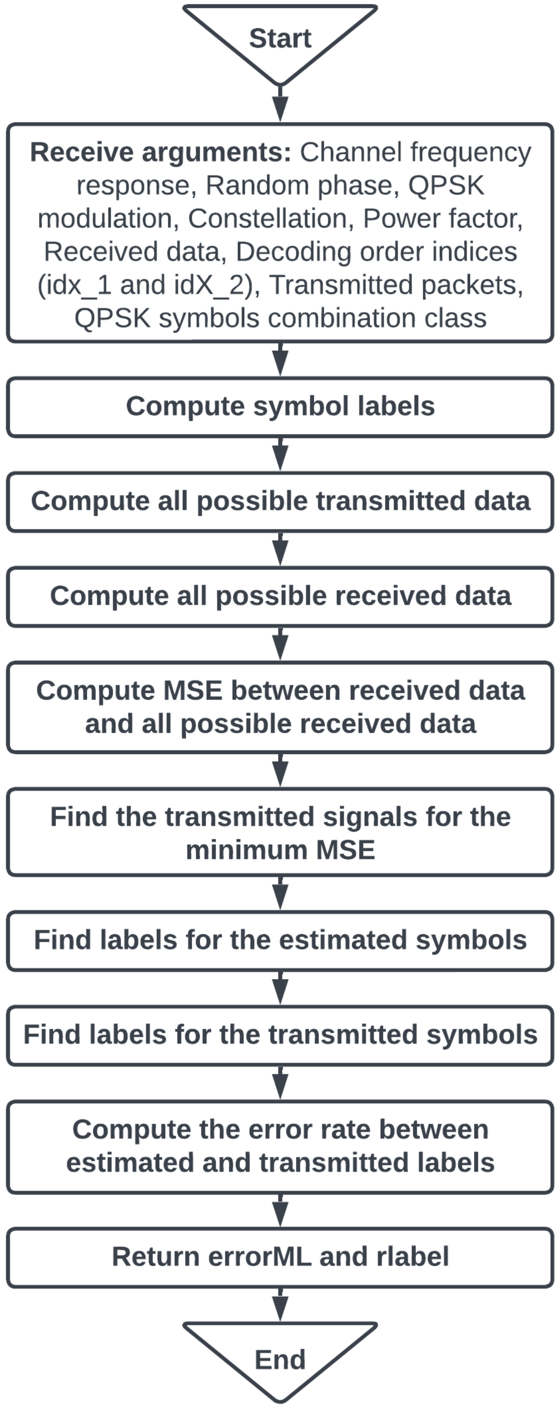


Figure A.4: Flowchart for the signal detection algorithm for the proposed system

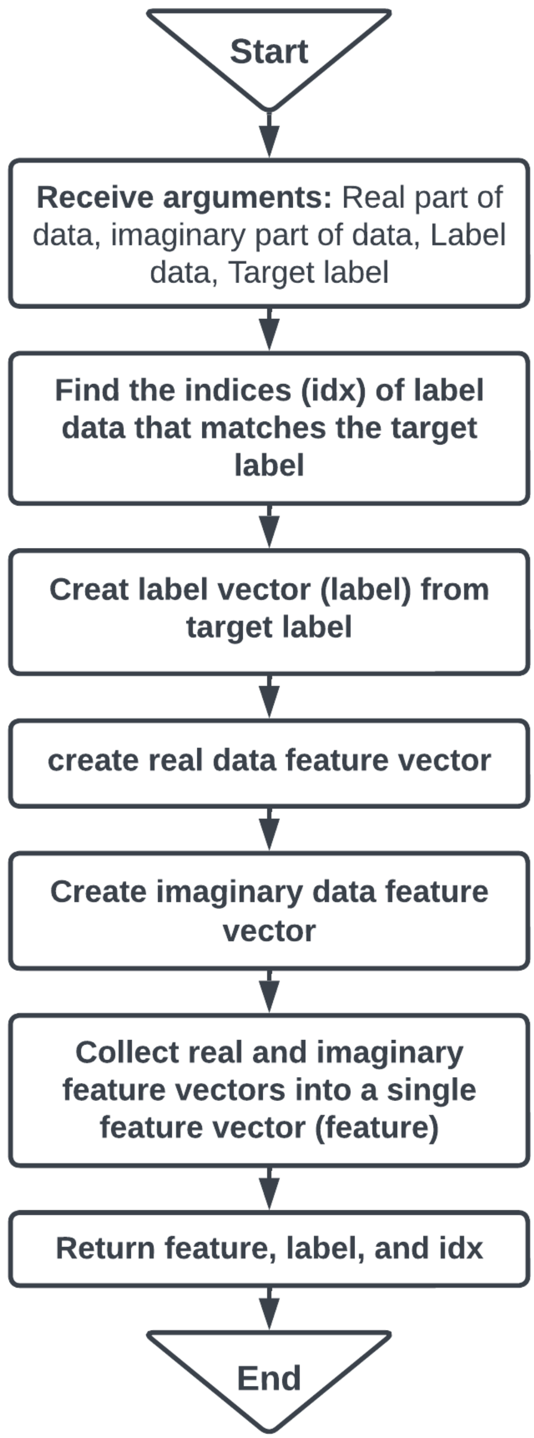


Figure A.5: Flowchart for acquiring the feature and label algorithm for the proposed system

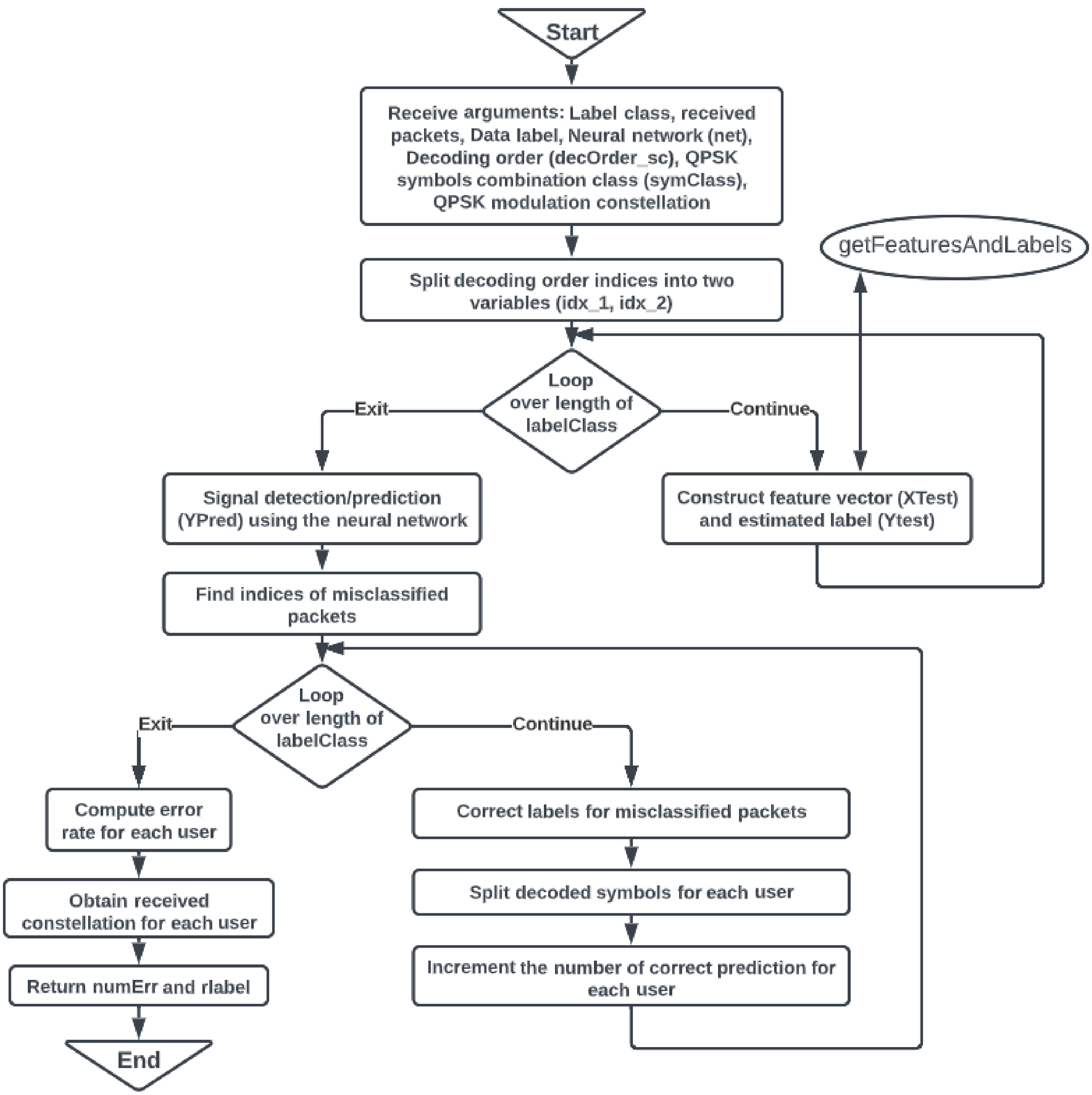


Figure A.6: Flowchart for decoding transmitted symbols for the proposed system

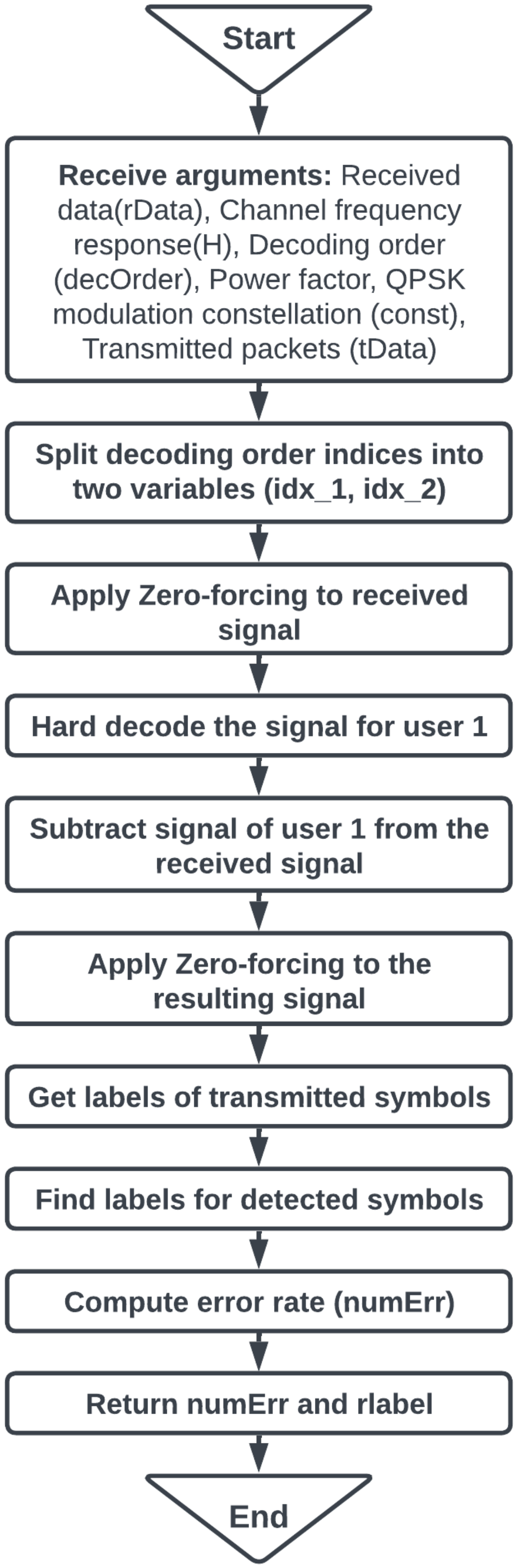


Figure A.7: Flowchart for decoding transmitted symbols using SIC for the proposed system

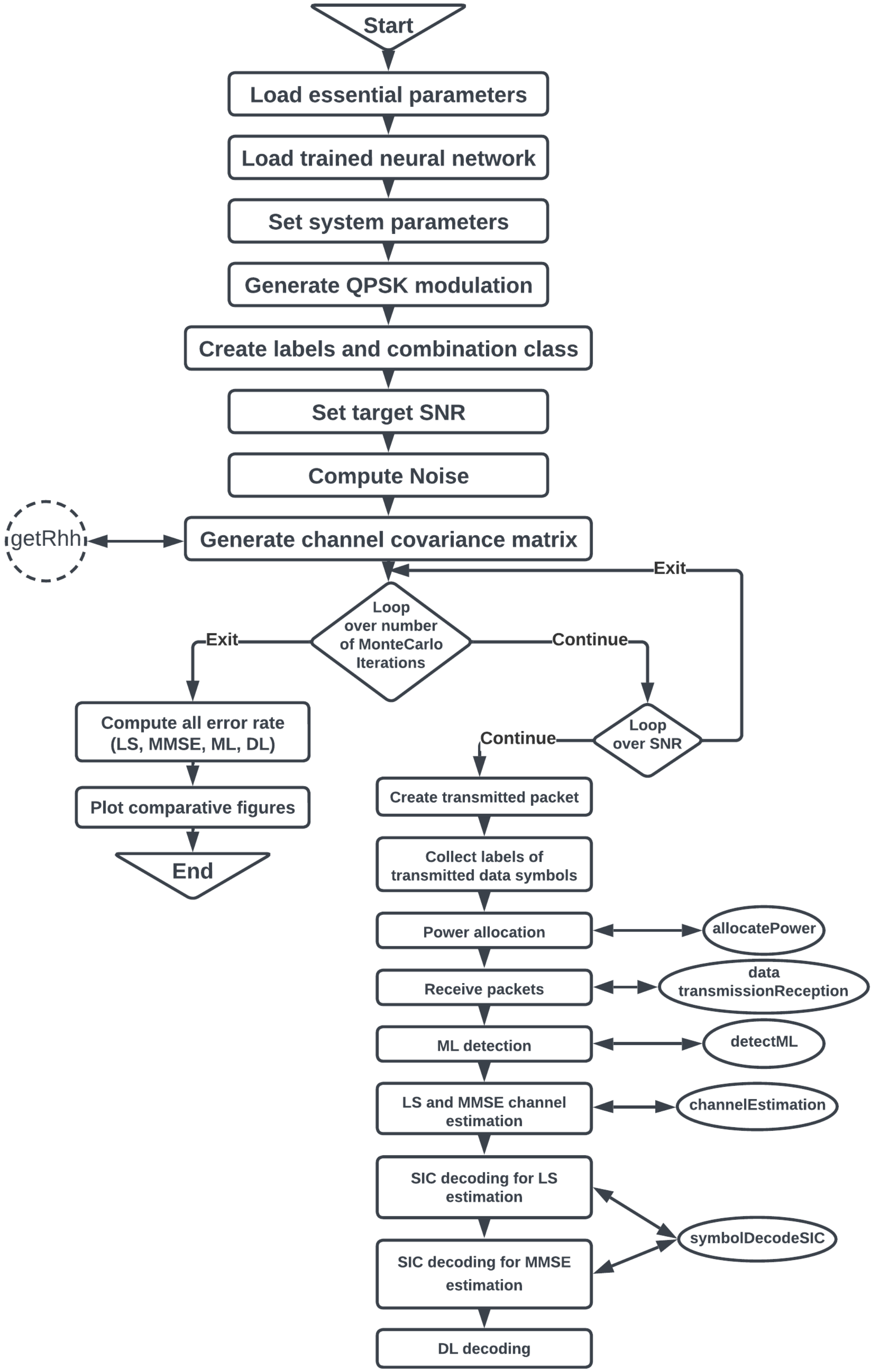


Figure A.8: Flowchart for Testing the transmitted data for the proposed system

Figure A.9: Flowchart for training the transmitted data of the proposed system

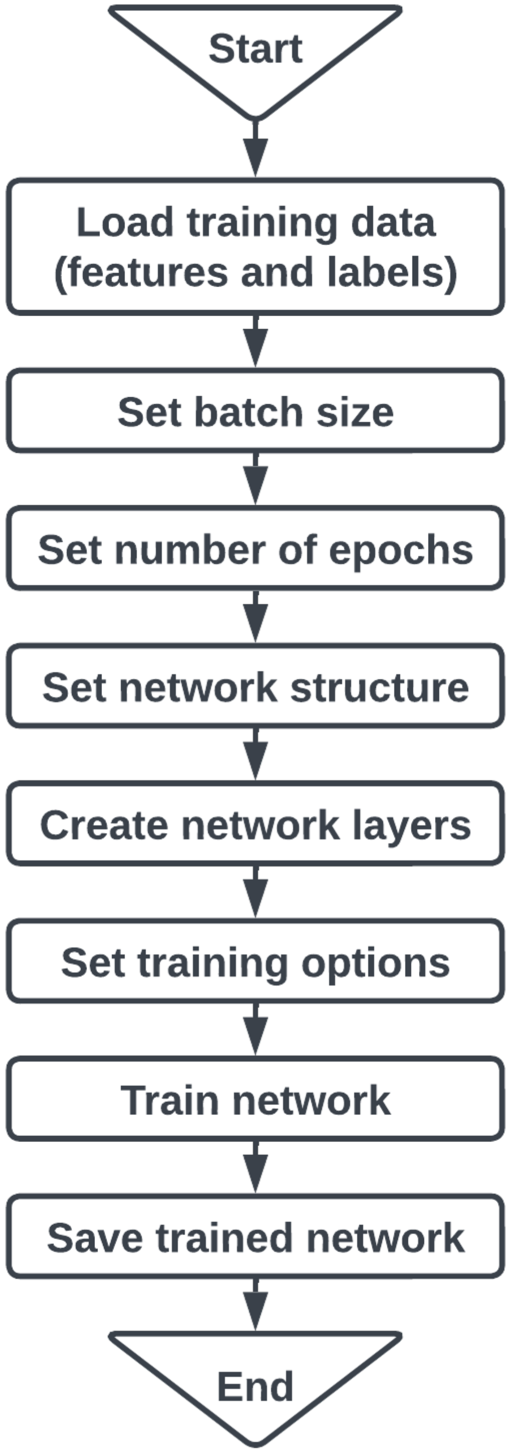


Figure A.10: Flowchart for training the neural network of the proposed system