Contributions to non-orthogonal multiple access techniques for massive communications

Manel Rebhi

DOCTORAT BRETAGNE SCIENCES LOIRE POUR L'INGENIEUR

Le Mans Université

UNIVERSITÉ DE NANTES

Doctoral Thesis of Le Mans University in conjunction with The University of Nantes

PhD school N° 602
Sciences pour l'Ingènieur
Specialization : *Acoustique*

Par
Manel REBHI

« Contributions to non-orthogonal multiple access techniques for massive communications »

Thesis presented and defended at «ENSIM - École Nationale Supérieure d'Ingénieurs du Mans», in « 26 April 2022 »
Research Units: LAUM UMR CNRS 6613 France and IETR UMR CNRS 6164
Thesis N° : « 2022LEMA1006 »

Rapporteurs before thesis defense:

Qassim NASIR	Professor at University Of Sharjah
Jean-Pierre CANCES	Professor at XLIM : Research Institute

Composition of the Jury :

Examiners :	Iyad DAYOUB	Professor at Polytechnic University of Hauts-de-France, IEMN-DOAE
	Soumaya CHERKAOUI	Professor at Université de Sherbrooke
Thesis director :	Kosai RAOOF	Professor at Le Mans University, LAUM
Thesis supervisor 2 :	Pascal CHARGE	Professor at Polytech Nantes, IETR
Thesis supervisor 1 :	Kais HASSAN	Associate professor at Le Mans University, LAUM

This thesis is dedicated to my parents.

For their endless love, support and encouragement

Your dream comes true Dad!

ACKNOWLEDGMENT

I would like to gratefully and sincerely thank my supervisor, Professor Kosai RAOOF, his support and guidance made my thesis work possible. He has been actively interested in my work and has always been available to advise me. I would like to thank my committee members for their interest in my work.

I would like to show my gratitude to my thesis reviewers: Professor Qassim Nasir and Professor Jean-Pierre Cances, and to the jury members: Professor Iyad Dayoubi and Professor Soumaya Cherkaoui for evaluating my PhD. work.

This work was possible because of the unconditional support provided by Dr. Kais Hassan. I am very grateful for his patience, motivation, enthusiasm, and immense knowledge that, taken together, makes him a great mentor. Thank you, for all your help and support.

I would like to thank Professor. Pascal Chargé for his supervisory role, technical support, time and insightful comments that helped me to improve my contributions.

A special thanks to my family. Words can not express how grateful I am to my parents, Mohamed and Mariem, for their faith in me and allowing me to be as ambitious as I wanted. It was under their watchful eye that I gained so much drive and an ability to tackle challenges head on.

I would like to express my thanks to my brother and sister Majd and Majdouline for their endless love, help and encouragement to finish my education.

Finally, and most importantly, I would like to thank my love and my life partner by my side, Ibrahim. You are the most amazing person in my life, you take care of me beyond my dreams and expectations.

To all my friends, thank you for your understanding and encouragement in my many, many moments of crisis. Your friendship makes my life a wonderful experience. I cannot list all the names here, but you are always on my mind.

Abstract

Multiple access techniques present many challenges and opportunities for the design of massive wireless networks. Therefore, substantial research efforts were devoted to the problem of serving users of the same network equally and simultaneously with some shared resources (time and/or frequency). Thus, the improvement of multiple access techniques of the next generations of mobile communications deserves a thorough study, which is the main objective of this thesis. The research work presented in this thesis focuses on sparse code multiple access (SCMA) and it is organized into two main parts. First, we study the adaptation of SCMA according to the users' needs in terms of energy, bandwidth and quality of service. The proposed adaptive SCMA architecture not only rightfully addresses the differences in users requirements, but also allows a more realistic use of the knowledge of transmission channels by customizing the codebook of each group of users which are clustered based on their channel state information. The second part concerns the application of deep learning techniques in the aim of efficiently decoding the received SCMA signal. Thus, a two-stage deep learning based SCMA detector was proposed under the assumption of additive white Gaussian noise. The first stage consists in denoising the signal using a denoising autoencoder before decoding it afterwards based on a deep neural network which allows to simultaneously estimate all the transmitted bits in one-shot. The complexity of SCMA detector and its performance in terms of bit error rate were evaluated. However, the performance of this method is slightly worse than that of the conventional SCMA detector, being rather iterative. Nevertheless, this comparison is not fair, since our firstly proposed detector, contrary to the conventional one, assumes that the SCMA codebook is unknown at the receiver. That is why, we propose a new distance-based deep neural network detector under the assumption that the codebook is known. The second proposed detector can be fairly compared to the conventional SCMA one, our proposition provides better performances.

Résumé

Les techniques d'accès multiple présentent de nombreux défis et possibilités pour la conception de réseaux sans fil massifs. Par conséquent, d'importants efforts de recherche ont été consacrés au problème de la distribution égale et simultanée des ressources partagées (temps et/ou fréquence) entre des utilisateurs d'un même réseau. Ainsi, l'amélioration des techniques d'accès multiple des prochaines générations de communications mobiles mérite une étude approfondie, ce qui est l'objectif principal de cette thèse. Le travail de recherche présenté dans cette thèse se concentre sur l'accès multiple par code parcimonieux. Dans un premier temps, nous étudions l'adaptation du SCMA en fonction des besoins des utilisateurs en termes d'énergie, de bande passante et de qualité de service. L'architecture SCMA adaptative proposée non seulement prend en compte les différences entre les besoins des utilisateurs, mais permet également une utilisation plus réaliste de la connaissance des canaux de transmission en personnalisant le livre de codes de chaque groupe d'utilisateurs qui sont regroupés en fonction de leurs informations sur l'état du canal. La deuxième partie concerne l'application de techniques d'apprentissage profond dans le but de décoder efficacement le signal SCMA reçu. Ainsi, un détecteur SCMA en deux étapes basé sur l'apprentissage profond a été proposé dans l'hypothèse d'un bruit blanc Gaussien additif. La première étape consiste à débruiter le signal à l'aide d'un autoencodeur de débruitage avant de le décoder ensuite sur la base d'un réseau neuronal profond qui permet d'estimer simultanément tous les bits transmis en une seule fois. Les performances du détecteur SCMA en termes de taux d'erreur binaire et sa complexité ont été évaluées. Cependant, les performances de cette méthode sont légèrement inférieures à celles du détecteur SCMA conventionnel, étant plutôt itératif. Néanmoins, cette comparaison n'est pas juste, car le détecteur que nous avons proposé en premier lieu, contrairement au détecteur conventionnel, suppose que le livre de code SCMA est inconnu au niveau du récepteur. C'est pourquoi nous proposons un nouveau détecteur à réseau neuronal profond basé sur la distance en supposant que le livre de codes est connu. Le deuxième détecteur proposé peut être comparé équitablement au détecteur conventionnel de SCMA et surpasse ses performances.

List of Acronyms

1G	First-Generation
2G	Second-Generation
3G	Third-Generation
3GPP	Third-Generation Partnership Project
4G	Fourth-Generation
5G	Fifth-Generation
6G	Sixth-Generation
ADAM	Adaptive Moment
AWGN	Additive White Gaussian Noise
B5G	Beyond Fifth-Generation
BER	Bit Error Rate
CD-NOMA	Code-Domain NOMA
CDMA	Code-Division Multiple Access
CCE	Categorical Cross Entropy
CCI	Co-Cell Interference
CNN	Convolutional Neural Network
CSI	Channel State Information
DAE	Denoiser AutoEncoder
DL	Deep Learning
DNN	Deep Neural Network
eMBB	enhanced Mobile BroadBand
FDMA	Frequency-Division Multiple Access

IDMA Interleave Division Multiple Access

IoT Internet of Things

IrSCMA Irregular SCMA

LDPC Low-Density Parity Check

LDS-CDMA Low-Density Spreading Code Multiple Access

LDS-OFDM Low-Density Signature Orthogonal Frequency-Division Multiplexing

LLR Log Likelihood Rate

LTE Long-Term Evolution

LSTM Long Short-Term Memory

MA Multiple Access

MC Mother Constellation

mMTC massive Machine Type Communication

ML Machine Learning

MPA Message Passing Algorithm

MSE Mean Squared Error

MUSA Multi-User Shared Access

NOMA Non-Orthogonal Multiple Access

NR New Radio

OFDMA Orthogonal Frequency-Division Multiple Access

OMA Orthogonal Multiple Access

PD-NOMA Power-Domain NOMA

PDMA Pattern-Division Multiple Access

QFI QoS Flow Identifier

QI relative QoS Importance

QoS Quality of Service

RAN	Radio Access Network
RE	Resource Element
RNN	Recurrent Neural Network
RSMA	Resource-Spread Multiple Access
SAMA	Successive Interference Cancellation Amenable Multiple Access
SC	Superposing Coding
SCMA	Sparse Code Multiple Access
SDMA	Spatial-Division Multiple Access
SIC	Successive Interference Cancellation
SNR	Signal-to-Noise Ratio
TDMA	Time-Division Multiple Access
URLLC	Ultra-Reliable Low-Latency Communications

NOMENCLATURE

SCMA codebook parameters

$\boldsymbol{C}_j$	Codebook of user j
$\mathbf{x}_j^{(m)}$	Codeword m of user j
J	Number of users
K	Number of subcarriers = Length of the codeword/codebook
M	Size of codebook = Number of alphabets (points) of $\mathcal{C}_{mc}$
N	Sparsity degree of a codebook = Number of dimensions of $\mathcal{C}_{mc}$

User's parameters

$\mathbf{h}_j$	Channel coefficient of user j
$\mathbf{T}_j$	User j-specific transformation matrix
$\mathbf{V}_j$	Mapping matrix of user j
M_j	Size of the codebook of user j
N_j	Sparsity degree of the codebook of user j

Regular SCMA system

$\boldsymbol{C}_{mc}$	Mother constellation matrix
$\mathbf{F}$	Factor graph matrix
$\mathbf{H}$	Channel matrix
$\mathbf{n}$	Additive zero-mean white circularly complex Gaussian noise
$\mathbf{y}$	SCMA received signal
ϕ	SCMA system
d_f	Degree of signal superposition on a given RE
d_s	Degree of RE occupation per user

E_b/N_0 Normalized SNR measure

λ Overloading factor

Irregular SCMA system

α Power scaling effect

β Interleaving which define the phase between different dimensions

ϕ_{ir} Irregular SCMA system

$\theta_{j,n}$ n^{th} Quantified rotation angle of user j

$M_i^{(g)}, i = 1, 2, 3$ Size of a codebook belonging to the i^{th} group

$N_i^{(g)}, i = 1, 2, 3$ Sparsity degree of a codebook belonging to the i^{th} group

DL-based SCMA system

$\mathbf{z}_d$ Detected codeword

$\mathbf{z}_t$ Transmitted codeword

$\mathcal{CC}$ Codewords combination set with cardinality M^J

$\mathcal{L}_{\text{CCE}}$ CCE loss function

$\mathcal{L}_{\text{MSE}}$ MSE loss function

$\mathcal{R}(\cdot)$ DAE-DNN-SCMA detecting process

$\hat{\ddot{\mathbf{y}}}$ Estimated value of $\ddot{\mathbf{y}}$

$\mathbf{r}^{(b)}$ $J\log_2(M)$ estimated decoded bits vector

$\mathbf{t}^{(b)}$ $J\log_2(M)$ transmitted bits vector

Mathematical tools

$\mathbb{P}(.)$ Probability of an event

$\sim \mathcal{CN}(0, \sigma)$ When a variable follows a zero-mean Gaussian distribution with variance σ

$\mathrm{d}_E^{(\min)}$ Minimum Euclidean distance

$\mathrm{d}_P^{(\min)}$ Minimum product distance

d_E Euclidean distance

d_P Product distance

D A design criterion

$Im(\cdot)$ Imaginary part of a complex number

$Re(\cdot)$ Real part of a complex number

$Q(.)$ Tail distribution of the standard normal distribution

Other notations

$\mathbf{I}_K$ Identity matrix of size K

$R_i^{(d)}, i = 1, 2, 3$ Radius of i^{th} ring of star-QAM constellation

List of Figures

1.2.1 The families of 5G usage scenarios and the associated services and applications. 38

1.3.1 Multiple access techniques through different generations of wireless communication systems: (a) FDMA for 1G (a) TDMA for 2G (c) CDMA for 3G (d) OFDMA for 4G. 39

1.3.2 Theoretical bandwidth efficiency of PD-NOMA compared to that of OFDMA for 2 users with respective rates R_1 and R_2 when both of user's signals are successively decoded based on SIC. $R_1^{(max)}$ and $R_2^{(max)}$ are fixed based on *Shannon capacity* for zero-error transmission. P_1 and P_2 denote respectively the power allocated to user 1 and user 2. For OMA, P_1 and P_2 supposed to be fixed for each subcarrier of OFDMA. $R_1^{(sim)}$ and $R_2^{(sim)}$ denote the maximum rates that could be allocated to user 1 and user 2 when using NOMA (simultaneous transmission). 40

1.3.3 An overview of exiting multiple access schemes 40

1.3.4 Illustration of downlink PD-NOMA principles. User A (near user) is with better channel conditions, while user B (far user) is with poorer channel conditions. 41

1.3.5 Basic principles of MUSA. 43

1.3.6 Performance comparison of LDS-CDMA, MUSA, PDMA and SCMA, in terms of BER, through Rayleigh fading channel : the number of orthogonal REs is 4 and the number of users is 6. 44

2.2.1 Representation of the encoder of a regular SCMA system. 55

2.3.1 The factor graph corresponding to encoder of the SCMA system presented in Figure 2.2.1. 56

2.3.2 A block diagram of SCMA codebook design procedure. 57

2.3.3 Low-projection constellation : an example of QAM SCMA constellation points of size $M = 4$ with two non-zero REs, labeled based on Gray coding. First step rotates the constellations to ensure a maximum product distance between symbols which enhances the detection process. The second step could better reduce the complexity of the receiver since some constellation points collide over each RE, for instance the constellation points corresponding to 00 and 11 in the M-sized constellation collide over the first subcarrier, however, they have maximum distance over the second one which makes them separable using M_p-QAM constellation while $M_p \leq M$. . 59

2.3.4 Illustration of two examples of 2-dimensional constellations with 4-codewords, i.e. T4QAM [86] and 4LQAM [87]: (a) projection of T4QAM over first dimension, (b) projection of T4QAM over second dimension, (c) projection of 4LQAM over first dimension and (d) projection of 4LQAM over second dimension. 61

2.3.5 An example of four-rings star-QAM mother constellation for the design of a SCMA codebook of size $M = 4$ and sparsity degree $N = 2$ where α and β are 2 reel design parameters. 62

2.3.6 Illustration of the six 2-dimensional constellations with 4-codewords, generated by applying unitary rotation transformation, as described in equation (2.11), on the T4QAM mother constellation [1]: the projections of the constellation of each user on its associated two REs are depicted. 62

2.3.7 Performance evaluation of uplink SCMA system with different codebook designs through Rayleigh fading channel : the number of orthogonal REs is 4 and the number of users is 6. 64

2.4.1 Block diagram of SCMA system model. 65

2.4.2 Performance comparison of MPA, Log-MPA and MAX-Log-MPA variations through Rayleigh fading channel : the number of orthogonal REs is 4 and the number of users is 6. 68

2.4.3 MPA performance comparison for different number of iterations through AWGN channel and Raleigh fading channel : the number of orthogonal REs is 4 and the number of users is 6. 69

2.4.4 Block diagram of an uplink MIMO-SCMA system with $N_t = 2$ transmit antennas and $N_r = 2$ receive antennas. 71

2.4.5 Uncoded uplink MIMO-SCMA compared to SISO-SCMA based on QPSK modulation when using maximum likelihood detection under AWGN channel [128]. 72

2.4.6 Structure of the SCMA encoder/detector: $\text{DNN}_j, j = 1 \cdots J$, with an input layer of m nodes and output one with $2K$ nodes (standing for the K-dimensional complex codeword), represent the SCMA mapping process for one single user, namely User j. Multiple DNNs are stacked together to form the SCMA DNN encoder, another DNN is employed as a SCMA detector. . 73

3.2.1 Presentation of the SCMA encoder for (a) regular system (b) irregular system with different constellation sizes (c) irregular system with different codebook sparsity degrees. 77
3.3.1 Four-rings star-QAM mother constellation for regular (a) and irregular (b) SCMA codebook design of size $M = 4$ and sparsity degree $N = 2$ when α and β are respectively reel and complex. 81
3.3.2 Rotation angles of different users are initialized to be assigned successively as described in Step 1 of Algorithm 3. Here, in order to simplify the illustration we considered a special case of two rings with no phase interleaving ($\alpha = 1$) and β of modulus 1.) 84
3.4.1 Role of QoS in de-cluttering flows [135] 85
3.4.2 Users are divided into three groups based on their channel conditions. A specific sparsity degree is assigned to each group, i.e. $N_1^{(g)} = 3$ for bad channels and $N_3^{(g)} = 1$ for excellent ones. 87
3.4.3 Irregular uplink SCMA system with near and far users using different sparsity degrees, $N = 1$ and $N = 3$ respectively, based on their channel states [134]. 88
3.4.4 Performance evaluation of uplink SCMA in [84] when users have different channel states (CSs) for different values of SNR level fluctuation $\delta = 0.5, \delta = 2, \delta = 4$ 89
3.4.5 Performance comparison between regular uplink SCMA and adaptive uplink SCMA with different sparsity degrees ($N_1^{(g)} = 3, N_2^{(g)} = 2, N_3^{(g)} = 1$) when the different channel states are taken into consideration. 90
3.4.6 Users are divided into three groups based on their channel conditions. A specific codebook size is assigned to each group, i.e. $M_1^{(g)} = 2$ for bad channels and $M_3^{(g)} = 8$ for excellent ones. 91
3.4.7 Performance comparison between regular uplink SCMA and adaptive uplink SCMA with different constellation sizes ($M_1^{(g)} = 2, M_2^{(g)} = 4, M_3^{(g)} = 8$) when the different channel states are taken into consideration. 91

4.3.1 Denoising autoencoder architecture: Example of a DAE applied on images from the modified national institute of standards and technology (MNIST) handwritten digit database [148]. 96
4.3.2 ReLU activation function. 97

4.3.3 TanH activation function. 97
4.3.4 Sigmoid activation function. 98
4.4.1 Block diagram of the proposed DAE-DNN-SCMA detector. 100
4.4.2 The graph of our configured DAE 101
4.4.3 Illustration of the impact of the presence of noise on transmitted constellation for different E_b/N_0 values: (a) multiplexed constellations before transmission ($\ddot{\mathbf{y}}$), (b) $\mathbf{\Omega}(\ddot{\mathbf{y}})$ when $E_b/N_0 = 2$, (c) $\mathbf{\Omega}(\ddot{\mathbf{y}})$ when $E_b/N_0 = 9$, (d) $\mathbf{\Omega}(\ddot{\mathbf{y}})$ when $E_b/N_0 = 15$ and (e) $\mathbf{\Omega}(\ddot{\mathbf{y}})$ when $E_b/N_0 = 20$ 103
4.4.4 MSE as a function of E_b/N_0 values (a) the same DAE is trained with signals of different E_b/N_0 values, (b) different DAEs are trained separately with signals of fixed E_b/N_0 value per DAE. 103
4.4.5 Mean squared error over 100 training epochs of DAE-SCMA when the used samples of SCMA received signal are corrupted over AWGN channel with $E_b/N_0 \in [6, 7, 8, 9]$. 104
4.4.6 The graph of our configured DNN model 106
4.4.7 The Accuracy as a function of the number of epochs when the E_b/N_0 level is low, medium and high. 107
4.5.1 Example of noise-free received signal compared to the M^J possible combinations of codewords in $\mathcal{CC}$ over the $K = 4$ subcarriers; $M = 4$ and $J = 6$ 108
4.5.2 Block diagram of the proposed fair SCMA detector. 109
4.6.1 Categorical cross entropy loss versus E_b/N_0 for DAE-DNN-SCMA detector as introduced in section 4.4 after 100 epochs to train the DAE-SCMA first and 3000 to train the whole system DAE-DNN-SCMA using the estimated denoised SCMA signal samples. 110
4.6.2 Accuracy versus E_b/N_0 for DAE-DNN-SCMA detector as introduced in section 4.4 after 100 epochs to train the DAE-SCMA first and 3000 to train the whole system DAE-DNN-SCMA using the estimated denoised SCMA signal samples. 111
4.6.3 The impact of the choice of E_b/N_0 values on the training of DAE-DNN-SCMA detector when using the configuration of DNN-SCMA model presented in Table 4.4.2. The several DAE-DNN-SCMA scenarios' models, DAE-DNN-SCMA_S1, DAE-DNN-SCMA_S2, DAE-DNN-SCMA_S3 and DAE-DNN-SCMA_S4, are compared to MPA-SCMA detector in [33] and the proposed DL-SCMA in [80]. 111
4.6.4 The loss function versus the depth after only 100 epochs of training when the SCMA signal is transmitted over AWGN channel using $E_b/N_0 \in [6, 7, 8, 9]$. 112

4.6.5 The accuracy versus the depth after only 100 epochs of training when the SCMA signal is transmitted over AWGN channel using $E_b/N_0 \in [6, 7, 8, 9]$. 113
4.6.6 The required training time versus the depth after only 100 epochs of training when the SCMA signal is transmitted over AWGN channel using $E_b/N_0 \in [6, 7, 8, 9]$. 113
4.6.7 Training and validation accuracy for different values of width, when depth = 6, after only 100 epochs of training when the SCMA signal is transmitted over AWGN channel using $E_b/N_0 \in [6, 7, 8, 9]$. 114
4.6.8 Performance comparison of MPA-SCMA detector [33], DL-SCMA [80], our proposed joint DAE-DNN-SCMA detector, our proposed Fair-DNN and Fair-DAE-DNN detectors over AWGN channel when $J = 6, M = 4$ and $K = 4$. 115

List of Figures in Appendix

A.1 The 2-dimensional codebooks with 4-codewords, generated for $J = 6$ users, as described in [1]. 121

List of Tables

2.3.1 Two examples of 2-dimensional constellations with 4-codewords where $x_n^{(m)}$ is the n^{th} entry of the m^{th} codeword m, that is $x_n^{(m)}$ belongs to dimension n. 60

4.4.1 Comparison between deep autoencoder and convolutional autoencoder in terms of complexity and training time for fixed number of nodes and fixed depth after 100 epochs of training. 101

4.4.2 Summary of the experiment configurations 105

List of Tables in Appendix

B.1 Review of existing SCMA codebook designs 123

D.1 Key parameters in complexity analysis 126

D.2 Review of existing SCMA detector designs 127

Table of Contents

Acknowledgment **5**

Abstract **7**

Résumé **9**

General Introduction **27**

1 Background and Problem Statement **33**

1.1 Introduction 34

1.2 An Overview of 5G and beyond Technology 34

1.2.1 Evolution from 4G to 5G and beyond 34

1.2.2 The Technologies of 5G and beyond 35

1.2.3 Standardization Organizations 36

1.2.4 The Families of 5G Usage Scenarios 37

1.3 Non-Orthogonal Multiple Access 38

1.3.1 From OMA to NOMA 39

1.3.2 NOMA's Domains 40

1.3.3 NOMA Progress of Standardization with 3GPP 44

1.4 Machine Learning for Communication Systems 45

1.4.1 What is Machine Learning? 46

1.4.2 Recent Advances on Machine Learning 46

1.4.3 How Machine Learning is Applied in Wireless Networks? 47

1.4.4 Machine Learning for NOMA 48

1.4.5 Machine Learning for SCMA 49

1.5 Focus of the Thesis 50

2 Comprehensive Study of SCMA **53**

2.1 Introduction 54

2.2 SCMA System Architecture 54

2.2.1 Basic Principles of SCMA 54

2.2.2 SCMA Signal Model 55

2.3 SCMA Codebook Design 55

2.3.1 Codebook Design Procedure . . . 56
2.3.2 Mother Constellation Design . . . 57
2.3.3 Transformation Operators Design . . . 62
2.4 SCMA Detector Design . . . 65
2.4.1 Message Passing Algorithm . . . 65
2.4.2 Detectors for Channel-Coded SCMA . . . 69
2.4.3 MIMO-SCMA Detectors . . . 70
2.4.4 Machine Learning based Detectors . . . 72
2.5 Conclusions . . . 73

3 Adaptive SCMA Scheme 75
3.1 Introduction . . . 76
3.2 Irregular SCMA . . . 76
3.2.1 Non-regular Structure . . . 76
3.2.2 Review of Existing Irregular SCMA Schemes . . . 77
3.3 Irregular SCMA Codebook Design . . . 78
3.3.1 Proposed Mother Constellation Design . . . 78
3.3.2 Proposed Transformation Operators Design . . . 82
3.4 Adaptation of Irregular SCMA . . . 83
3.4.1 Adaptation Scenario I : Different Quality of Service Levels . . . 85
3.4.2 Adaptation Scenario II : Fairness Among Users . . . 87
3.4.3 Adaptation Scenario III : Higher Overall Throughput . . . 90
3.5 Conclusions . . . 91

4 Machine Learning for SCMA Detector and Channel Denoiser 93
4.1 Introduction . . . 94
4.2 Motivation and Objectives . . . 94
4.3 Machine Learning Configuration . . . 95
4.3.1 Employed ML's Techniques . . . 95
4.3.2 Training and Hyper-Parameters . . . 96
4.4 Joint Denoiser and Detector of DL-based SCMA (DAE-DNN-SCMA) . . . 99
4.4.1 Motivation behind the Proposed Solution . . . 99
4.4.2 Denoising SCMA Signals over AWGN Channel . . . 100
4.4.3 Deep Neural Network based SCMA Decoder . . . 104
4.5 Fair SCMA Detector based Deep Learning (Fair-DNN-SCMA) . . . 107
4.5.1 Proposed Solution . . . 108
4.5.2 Pre-processing . . . 109
4.6 Analysis of Results . . . 109
4.6.1 Impact of Training Data with Different Noise Intensity Levels . . . 110
4.6.2 Impact of Hyper-Parameters . . . 112

4.6.3 Performance Evaluation of Proposed Detectors 113
4.6.4 Complexity Study of the Proposed Models 115
4.7 Conclusions 115

General Conclusions and Future Works **117**

Appendices of the Thesis **121**
A Codebook Example 121
B Review of Existing SCMA Codebook Designs 122
C Data-aided-CSI Estimation Algorithm 124
D Review of Existing SCMA Detector Designs 125

Bibliography **129**

General Introduction

General Context and Motivations

The next decades will encounter many emerging applications which will require to massively connect new devices. For instance, the aim of the internet of things (IoT) is to massively connect various types of physical and virtual objects into a dynamic global network infrastructure and to enable applications such as connected cars, industrial IoT, smart cities, connected healthcare, etc [2]. The major targets of the fifth-generation (5G) mobile communication systems and beyond are the massive connectivity, the ultra-high data rates, the ultra-low latency, as well as flexible air interface design to support several user requirements [3]. Early in 2014, the 5G was expected to increase the connectivity density by 10 (at least $10^6/\text{km}^2$) [4]. Later, in 2021, there was more than 10 billion active IoT devices. This number is projected to surpass 75 billion by the end of 2025 [5]. Thus, it will be very difficult to satisfy this huge demand on data traffic and massive connectivity without proposing new enhanced communications techniques especially enhanced multiple access schemes.

One key of the evolution of wireless communication systems over their different generations was to propose new multiple access techniques including frequency-division multiple access (FDMA) for first-generation (1G), time-division multiple access (TDMA) for second-generation (2G), code-division multiple access (CDMA) for third-generation (3G), and orthogonal frequency-division multiple access (OFDMA) for fourth-generation (4G) [6], [7]. All these techniques are orthogonal, that is the wireless resources are orthogonally shared among multiple users in different domains as time, frequency and code. The orthogonality of the allocated resource elements (REs) reduces the multiple access interference and had enabled low-complexity and efficient receivers. The aforementioned orthogonal multiple access (OMA) techniques are not sufficient to reply to the increasing connectivity demand of the future generations since the number of supported users is limited by the number of available orthogonal REs. In order to enable the massive access, non-orthogonal multiple access (NOMA) mechanisms were introduced [8]. The idea was to further exploit the REs by employing a non-orthogonal access, for instance, multiple users can share the same narrow frequency band over the same time slot which allows to serve more users at the expense of increasing the receiver's complexity. The NOMA techniques are divided into two main categories according to the type of allocated resources, namely: power-domain NOMA (PD-NOMA) and code-domain NOMA (CD-NOMA). However, other existing techniques can also assign resources belonging to

multiple domains to multiple users in a non-orthogonal manner. Hence, the objectives of 5G and beyond motivated, and will continue to motivate, the research evolution of NOMA techniques despite the fact that the third-generation partnership project (3GPP) decided not to include them in 5G.

In the last decade, trying to incorporate PD-NOMA in 5G networks have gained attention of researchers around the globe [9]–[11]. The main idea of PD-NOMA is to exploit a new dimension, that is the power. On the other hand, CD-NOMA was inspired by the classic CDMA systems in which multiple users share the same RE by adopting a specific signature per user. The key difference between CD-NOMA techniques and CDMA ones is that the spreading sequences of the former are restricted to non-orthogonal low cross-correlation sparse sequences.

Sparse code multiple access (SCMA) is among the CD-NOMA techniques that are considered by the new radio (NR) study in the 3GPP long-term evolution (LTE)-advanced, since it was among the most reliable multiple access candidates for 5G [12], [13]. However, as 5G networks are being rolled out in many different countries nowadays, the investigations on how to upgrade and expand them toward the 6^{th} generation of wireless communications (6G) had started. In this context, the international telecommunication union radio-communication sector (ITU-R) published its first release of the international mobile telecommunications standard (*IMT-2020*) recommendation in February 2021 [14], and initiated works for future technology trends report and vision recommendation for *IMT towards 2030 and beyond 6G*, which are expected to complete around June 2022 and June 2023, respectively. After just recently approving its release 18 package for 5G-Advanced in December 2021, the 3GPP technical specification group radio access networks (TSG-RAN), in accordance with the ITU-R vision recommendation and other proponents, will start preparing the technical specifications of 6G. As for the first commercial roll-out of 6G services, it is expected to be around 2030. An interconnection of everything as well as the development of a ubiquitous intelligent mobile world for intelligent life is expected to be realized. Artificial intelligence (AI) has the potential to be the foundation of beyond 5G and as of 6G, that is, it will be native and ubiquitous to make every component of the future network, intelligent. At the beginning of the actual year, one of the initial works listing the challenges of 6G, [15], confirmed that AI will be embedded everywhere, not only in the network including the core, but also down to the air-interface design. In addition, AI will become a part of new services provided by the network in order to lead to a more autonomous, secure and flexible network as well as more customized and privatized applications. Thus, the objective of this research work is to contribute to fulfilling the massive connectivity requirements for beyond 5G (B5G) and 6G relying especially on AI techniques.

Thesis Contributions

This thesis hopes to participate in developing a more realistic and powerful non-orthogonal multiple access techniques such that their integration in the upcoming generations of wireless systems can be more probable. The main contributions of this thesis are summarized as follows:

- Chapter 1 shows why multiple access techniques are important for the enhancement of the massive access scenarios of B5G like massive machine-type communications (mMTC), ultra-reliable low-latency communications (URLLC) and enhanced mobile broadband (eMBB). Also, digging into the existing contributions and the various discussions and decisions led by 3GPP, ITU and internet engineering task force (IETF) about 5G and B5G explains our decision to focus on a specific technique namely the SCMA.

- It is required to comprehensively study SCMA with the objective to explore its architecture, the design of its encoder and detector, its interplay with other 5G technologies such as millimeter wave (mmWave) communications and multiple-input multiple-output (MIMO) systems. To the best of our knowledge, this effort is yet to be conducted in the state-of-the-art, that is why we start by filling this gap in Chapter 2.

- We propose an adaptive SCMA scheme to fit with more realistic scenarios of massive communication where users have different channel conditions in addition to several business requirements such as data rate, quality of service (QoS), and network priority. Unlike the traditional SCMA, the sparsity degree, the constellation size, and the allocated power per user, can be adapted to each requirement in order to increase the overall performance of SCMA system. Several proposition were introduced in Chapter 3. This had inevitably required to adapt the irregular structure of SCMA using our new proposed design of irregular SCMA codebooks.

- It is vital to push on with developing SCMA detectors that can outperform the performance of the conventional message passing algorithm (MPA) based detector in terms of both complexity and performance. In this Ph.D work, a new approaches to improve SCMA detection performance using deep learning methods are explored. The first contribution in this domain is to propose to jointly design and train a denoising autoencoder (DEA) and deep neural network (DNN) to decode SCMA signals over an additive white Gaussian noise (AWGN) channel.

- Despite the fact that the proposed DEA-DNN system has shown similar performance to that of a conventional SCMA iterative detector, i.e. the MPA, with a complexity lesser than that of the latter, nevertheless, this comparison is not totally fair. It was

necessary to introduce a new deep-learning based detector under the assumption that the SCMA codebook is known at the receiver as in the case of an MPA one. The proposed detector can be fairly compared to MPA, and simulations confirmed that its performance is better.

Publications

The following is a list of publications in refereed journals and conference proceedings produced during my Ph.D. candidature.

Journal papers

1. **M. Rebhi**, K. Hassan, K. Raoof and P. Chargé, "Sparse Code Multiple Access: Potentials and Challenges," in IEEE Open Journal of the Communications Society, vol. 2, pp. 1205-1238, 2021.

Conference papers

1. **M. Rebhi**, K. Hassan, K. Raoof, and P. Chargé, "An adaptive uplink SCMA scheme based on channel state information," URSI France 2020, Future Network: 5G Beyond Workshop, Paris, France, Mars 2020, pp. 1–7.
2. **M. Rebhi**, K. Hassan, K. Raoof, and P. Chargé, "Deep Learning for a Fair Distance-based SCMA Detector", IEEE Wireless Communications and Networking Conference(WCNC), Austin-Texas,USA, April 2022.

Structure of the document

This thesis is divided into four Chapters. In the first Chapter, we start by presenting the necessary notions concerning the requirements of B5G wireless systems before focusing on the multiple access technique in general and more specifically the non-orthogonal ones
.
In the second Chapter, we will present a comprehensive study of SCMA by focusing on its basic architecture and then by highlighting and comparing the several proposed methods to optimize both its encoder and detector.
Later, in the third Chapter, we turn our attention to the irregular structure of SCMA and we propose an adaptive design that fits to B5G user's needs in some realistic scenarios especially when the channel state information is considered. Three main scenarios are investigated.
At the last Chapter, the application of deep learning techniques on SCMA detection problem is introduced. We propose a fair distance-based SCMA detector in addition to

evaluating the impact of the used hyper-parameters on the performance of our proposed full system.

Chapter 1

Background and Problem Statement

Contents

1.1 Introduction **34**

1.2 An Overview of 5G and beyond Technology **34**

1.2.1 Evolution from 4G to 5G and beyond 34

1.2.2 The Technologies of 5G and beyond 35

1.2.3 Standardization Organizations 36

1.2.4 The Families of 5G Usage Scenarios 37

1.3 Non-Orthogonal Multiple Access **38**

1.3.1 From OMA to NOMA 39

1.3.2 NOMA's Domains 40

1.3.3 NOMA Progress of Standardization with 3GPP 44

1.4 Machine Learning for Communication Systems **45**

1.4.1 What is Machine Learning? 46

1.4.2 Recent Advances on Machine Learning 46

1.4.3 How Machine Learning is Applied in Wireless Networks? . . 47

1.4.4 Machine Learning for NOMA 48

1.4.5 Machine Learning for SCMA 49

1.5 Focus of the Thesis **50**

1.1 Introduction

With the ever increasing demands of users for ubiquitous mobile services, the 5G of wireless communication systems, the B5G and even the 6G are facing distinguishable challenges motivated by the need for extremely high connection density with a large number of simultaneous services. One key of the evolution of wireless communication systems over their different generations was to enhance the multiple access (MA) techniques to guarantee better use of the radio spectrum and to enable far more devices to access the network at the same time. In this Chapter we will present an overview of 5G, its characteristics and the new technologies attached to it. Based on the orthogonality of the allocated wireless resources shared among multiple users in different domains as time, frequency and code, MA schemes are classified into two categories: (i) OMA and (ii) NOMA. Here, our focus will be on the NOMA schemes designed either in the power or code domain. Finally, we will examine how machine learning (ML) is generally employed to enhance wireless networks performance and especially that of the MA techniques.

1.2 An Overview of 5G and beyond Technology

In this section, we will highlight the challenges imposed by the 5G when it comes to MA techniques. The aim is to continue to enhance existing MA schemes in order to fulfill the several needs of massive connected devices over 5G and beyond networks. In the next subsections, we present characteristics, technologies and services of the 5G and B5G.

1.2.1 Evolution from 4G to 5G and beyond

5G is the fifth-generation of mobile telecommunications technology promoted to support faster data rate and to provide higher connection density and much lower latency which presents significant advances when compared to the forth generation of wireless communication, 4G, [3]. The potential throughput of 5G can reach around 20 Gb/s which exceeds traditional wire-line network speeds [16]. Also, 5G increases the connectivity density to cover at least 10^6 devices per km^2 [4]. The latency refers to the time it takes for a packet of information to be transmitted over a frequency band. 4G latency is about 60 to 98 milliseconds (ms) [17], while 5G drops to less than 1 ms which is about 100 times lower. However, 6G latency is estimated to be instantaneous which will help developing real-time applications. In addition, the 5G has the potential to be a more secure cellular network than its predecessors thanks to its built-in security [18]. For all these reasons, 5G became more attractive for several applications such as IoT.
Also, regarding the exploited spectrum, 4G works basically in the sub-3 GHz radio spectrum. While both 5G and 6G employ higher frequencies to transmit massive data rapidly. For instance, 5G uses the sub-6 GHz frequencies, so-called low 5G band, and above 24.25

GHz, known as high 5G band [19]. Regarding 6G, it should operate at 95 GHz to 3 terahertz (THz) which will allow 6G to be around 1000 times faster than 5G [20]. Furthermore, 6G aims to enhance the massive connectivity on land, under the sea, and even in space, enabling a smart infrastructure for the sustainable society [21].

1.2.2 The Technologies of 5G and beyond

The 5G radio access network (RAN) is based on a *dense heterogeneous* network which contains several tiers, from macro base stations to small cell stations located on light poles and building roofs, each station transmits signals over the 5G NR interface. One key technology for 5G is mmWave communications which exploit the spectrum between 30 GHz to 300 GHz. The short wavelength of mmWave makes their application possible for small cells in densely populated areas and backhaul communications to core network such that 5G works better. However, the millimeter waves only travel short distances and are susceptible to weather conditions and obstacles such as buildings, windows, walls and leaves. That is why the line-of-sight transmissions are privileged. On the other hand, for less dense areas, 5G networks have to use lower frequency bands such as low-band and mid-band spectrum.

An essential way to increase the capacity of cellular networks is to increase the number of antennas at base stations and mobile terminals which is called *massive* MIMO. However, this will not be possible without the recent advances in complementary metal oxide semiconductor (CMOS) circuits which allow to pack more antennas in a smaller physical size.

Early in 2014, the 5G was expected to increase the connectivity density by 10 [4]. In 2021, the number of IoT devices and connected mobile users exceeded 50 billion of active hosts [5]. The number of active IoT devices is expected to explode in the upcoming decades. It will be very difficult to satisfy this huge demand on data traffic and massive connectivity without proposing new enhanced communications techniques especially enhanced multiple access schemes.

Furthermore, to ensure its implementation, the 5G architecture is based on *virtualization* such that networking functionalities can be managed through software rather than hardware. This attractive feature enables 5G architecture to be flexible in order to provide the access, anytime and anywhere, for served users. Motivated by the need to accommodate with the huge variety of applications and services enabled by 5G, and by the need to fit with QoS requirements, slicing network architecture, evolved from the concept of RAN sharing [22], is developed to play a central role in 5G mobile networks. The idea is to enable the multiplexing of independent virtualized networks over the same physical network infrastructure so that each network slice is an isolated end-to-end network adapted to lay the several needs solicited by a particular application. However, the increasing number of

used radio base stations is making the software architecture of the slice-based networks more complex. This can come at the expense of required delay of fractions of a millisecond for some applications such as those for self-driving cars or remote surgery. The solution is to mainly automate the processes in the central network management.

1.2.3 Standardization Organizations

Here, our aim is to highlight the reliability of 5G and beyond through the work proposed by the most popular organizations of standardization on the filed of communication in all over the world.

(a) **The 3rd Generation Partnership Project :**
3GPP is an international organization with communications-focus composed by a group of telecommunications standard development bodies known under the name of the organizational partners. For 5G, 3GPP continue to produce and publish the technical specifications as it had done for 3^{rd} and 4^{th} generations of mobile networks. However, the organization was charged with formulating the 5G technical specifications, which ultimately become standards. By the end of 2017, the first 5G specifications was proposed by 3GPP in its Release 15, including a set of tasks and checkpoints to guide ongoing studies about 5G architecture and NR network at that time, and also to discuss number of independent and autonomous radio access technologies. The use of 5G with enhanced mobile broadband, ultra-reliability and low latency, frequency ranges, and the importance of forward compatibility in radio and protocol design was also a main panel for 3GPP Release 16.

(b) **Internet Engineering Task Force :**
IETF is a large international community around networking and is the premier Internet standardization body that develops open standards through open processes. The main mission of the IETF is to make the Internet works better by producing relevant technical documents in order to manage the way ongoing works will design and use the Internet. 5G affects IETF proceedings, for instance, the time-critical communication and low-latency applications is one of the 5G use cases that is also a component technology on which IETF is working in order to define mechanisms to guarantee deterministic delays for some flows across a network. Furthermore, thanks to IETF, the key specifications for virtualization functions are coming up. The idea is to evolve internet protocols (IPs) to support network virtualization. Moreover, new technologies under development by IETF are proposed including routing-related testing, protocols for distributed networking, segment routing, and path computation in order to meet the constraints of the 5G NR. Finally, IETF had collaborated with 3GPP on the development of some of 5G technologies.

(c) **International Telecommunication Union :**
ITU is the United Nations agency specialized in information and communication technologies. It is founded to facilitate international connectivity in communications networks. Its major mission is to allocate the global radio spectrum and satellite orbits, and to develop technical specifications in order to ensure the interconnect between networks and technologies. Start working on 5G in 2015, the ITU identified three spectrum bands that will be used for this new generation of telecommunication networks. Later in 2016, the criteria for selecting radio interface to be adopted in 5G technologies was refined. By the end of the same year, in order to meet higher 5G's performances, ITU concluded a preliminary study including a focus on network architecture, network management requirements and framework.

1.2.4 The Families of 5G Usage Scenarios

This subsection briefly introduces the three families of usage scenarios as defined by ITU-R IMT for 2020 and beyond [4], namely eMBB, mMTC and URLLC. These are seen as the use case categories that together cover the overall technical needs for the existing and future services and applications as illustrated in Figure 1.2.1.

(a) **enhanced Mobile BroadBand**
Also known as extreme mobile broadband, the eMBB is enabled for data-driven use cases that require high data rates across a wide coverage area. The aim behind is to improve broadband access, mainly in densely populated areas, by enhancing indoor and outdoor coverage in high-rise buildings and crowded city centers. The eMBB supports stable connections with very high peak data rates, as well as moderate rates for hundreds of cell-edge users in environments with heavy data traffic.

(b) **massive Machine Type Communication**
As a new category of cellular services, mMTC could support extremely high connection density of devices which requires to employ efficient multiple access strategies in order to support the massive IoT networks. Its main aim is to provide connectivity to a huge number of devices ($\approx$ 1 million devices per km^2), that transmit and receive only small data blocks using low bandwidth channels. The mMTC usage case is also latency-tolerant.

(c) **Highly reliable low latency communication**
URLLC is a very essential usage scenario for 5G as it ensures the transmission of data in few milliseconds and, as its name implies, it provides an ultra-high network reliability which is more than 99.999% [23]. This category is mainly used in critical-mission use cases such as a remote surgery, the intelligent transport system

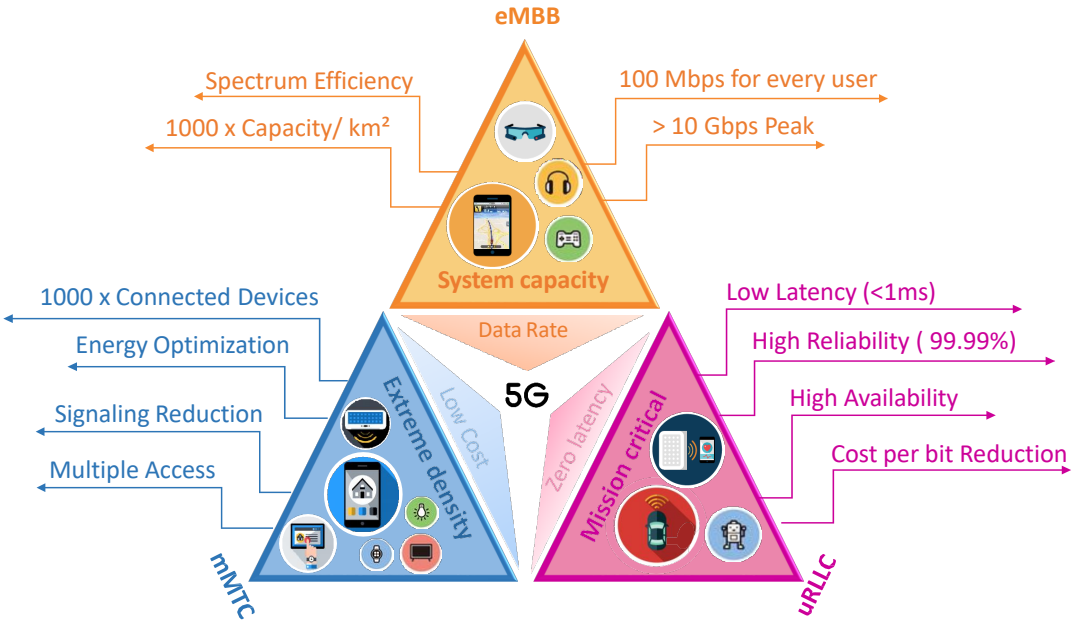


Figure 1.2.1: The families of 5G usage scenarios and the associated services and applications.

and especially for smart grids where latency-sensitive applications will require wide coverage.

1.3 Non-Orthogonal Multiple Access

As explained in the previous section, the massive connectivity is one of the main requirements of the 5G and beyond. One key to fulfill this objective is to allow several devices to efficiently access the same resources simultaneously, this approach is called multiple access. Historically, proposing new MA techniques over different generations was a key of the evolution of wireless communication systems, this includes FDMA for 1G, TDMA for 2G, CDMA for 3G. However, the most advanced schemes of multiple access that was proposed for 4G LTE, such as the OFDMA [24] and the single-carrier FDMA (SC-FDMA) [25], can no longer meet the current needs of the new generations of wireless communications in terms of data traffic and massive connectivity. Hence, it was legitimate to propose new enhanced communications techniques especially enhanced multiple access schemes which gave the birth of the so-called NOMA.
In this section, we will explain the NOMA concept, and review briefly its state-of-the-art, before presenting the progress of its standardization.

1.3.1 From OMA to NOMA

Based on how the resources are shared among multiple users, two types of multiple access could be distinguished: OMA and NOMA. It is obvious from Figure 1.3.1 that users in FDMA, TDMA and OFDMA based systems can be easily separated since they did not occupy the same frequency band at the same time. While CDMA users employ orthogonal codes and a simple correlation operation is sufficient to detect their signals. The non-orthogonality means that users can exploit simultaneously resources belonging to two different domains at least. For instance, different users could be sharing the same time slot and the same subcarrier, so-called RE, for uplink or downlink communications while still be able to decode the data of each user despite the existing inter-user interference and the channel conditions.

By the end of 2013, in its Release 12, the 3GPP initiated further evolution of LTE in order to face the expected increasing demands of future networks especially supporting large scale heterogeneous traffic and users. That is why a new modulations and multiple access schemes started to be developed to meet this growing demand. In 2017, 3GPP LTE-advanced proposed NOMA as a promising technology for addressing the aforementioned challenges by accommodating several users within the same RE [7]. By doing so, significant bandwidth efficiency enhancement was expected to be attained over conventional OMA techniques as presented in Figure 1.3.2 . Regarding to B5G, NOMA is recognized as the promising candidate for the future generations of radio access technologies.

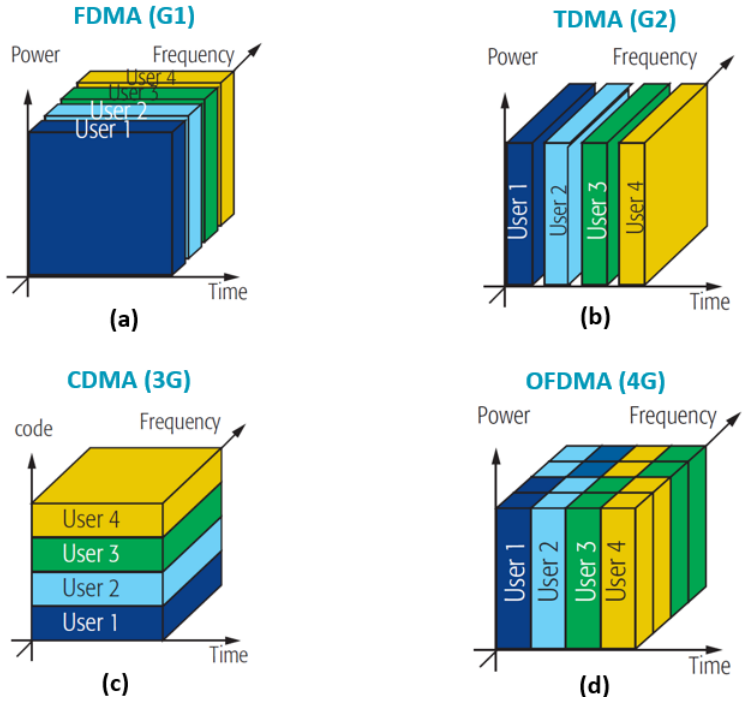


Figure 1.3.1: Multiple access techniques through different generations of wireless communication systems: (a) FDMA for 1G (a) TDMA for 2G (c) CDMA for 3G (d) OFDMA for 4G.

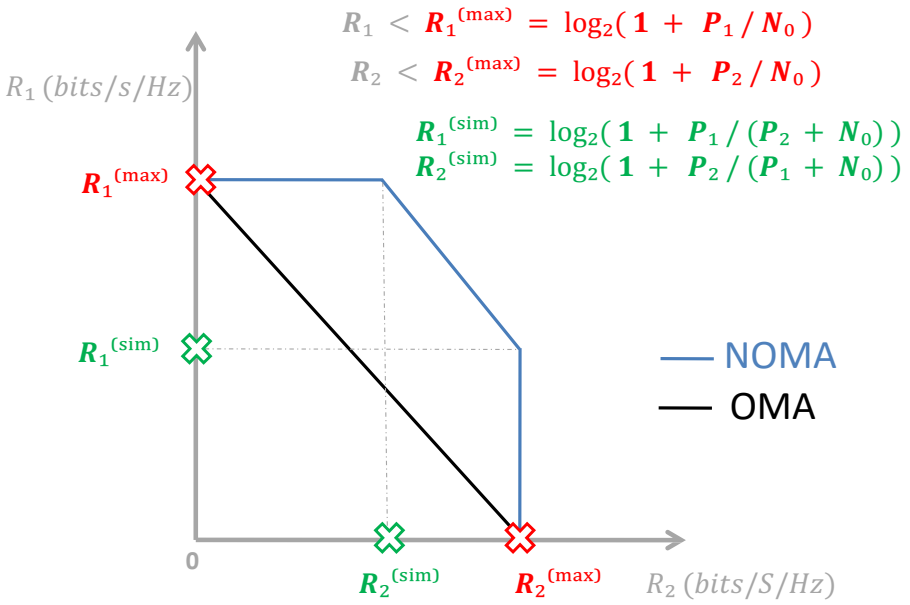


Figure 1.3.2: Theoretical bandwidth efficiency of PD-NOMA compared to that of OFDMA for 2 users with respective rates R_1 and R_2 when both of user's signals are successively decoded based on SIC. $R_1^{(max)}$ and $R_2^{(max)}$ are fixed based on *Shannon capacity* for zero-error transmission. P_1 and P_2 denote respectively the power allocated to user 1 and user 2. For OMA, P_1 and P_2 supposed to be fixed for each subcarrier of OFDMA. $R_1^{(sim)}$ and $R_2^{(sim)}$ denote the maximum rates that could be allocated to user 1 and user 2 when using NOMA (simultaneous transmission).

1.3.2 NOMA's Domains

According to how users are non-orthogonally multiplexed at available REs, NOMA schemes are divided into two categories: PD-NOMA and CD-NOMA [26]. Apart from these two categories, other existing alternative NOMA schemes was also proposed in the literature. In this thesis, we will briefly present the key NOMA technologies by addressing only the domains of power and code. Figure 1.3.3 gives a classification of existing multiple access schemes.

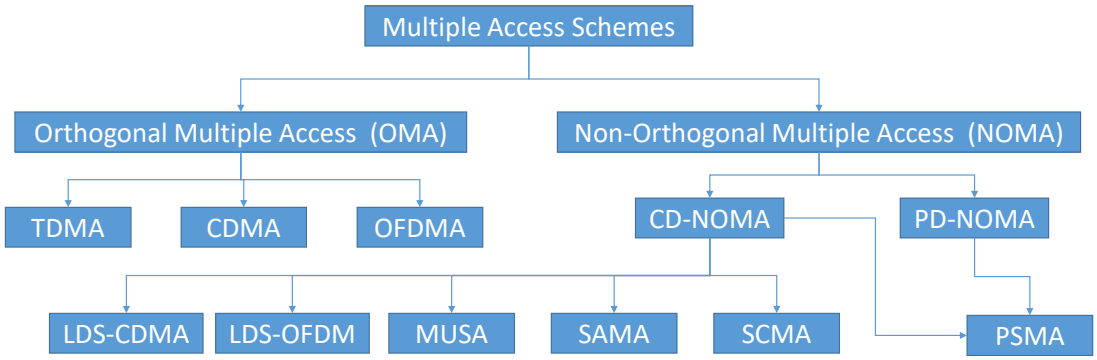


Figure 1.3.3: An overview of exiting multiple access schemes

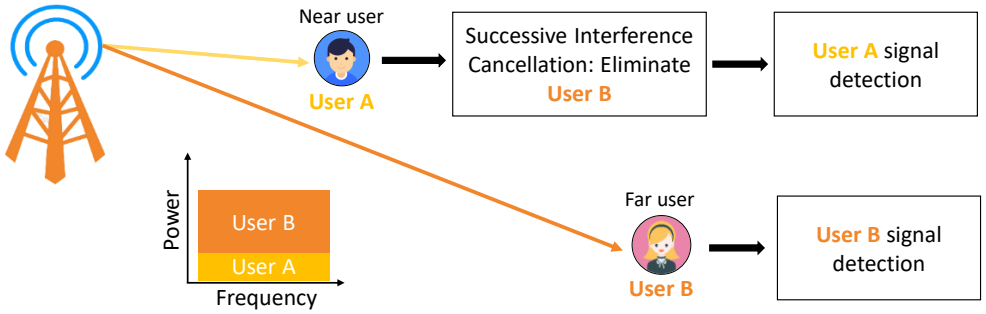


Figure 1.3.4: Illustration of downlink PD-NOMA principles. User A (near user) is with better channel conditions, while user B (far user) is with poorer channel conditions.

Power-Domain NOMA

The basic idea of PD-NOMA, is to allow supporting more multiple access by employing a new dimension: the signal's power [9]. PD-NOMA is based on combining two key techniques namely, superposition coding at the transmitter, and mainly successive interference cancellation (SIC) at the receiver's side to segregate user information.

(a) **Superposition Coding (SC) :** The fundamental concept of superposition coding is to guarantee higher detection reliability by superimposing several users at the same RE with a power allocation among users based on the near-far propriety. More power is allocated when the normalized channel gain is smaller. In other words, when the user is far from the base station (BS). Multiplexing users in PD-NOMA consists of superimposing the constellation diagram of different users such that each user's information is modulated and summed among users with appropriate power allocations so that the resulting signal still form higher-order constellation.

(b) **Successive Interference Cancellation (SIC) :** At the receiver side, SIC is employed to cancel interference. The detection process start from the strongest user to the weakest one such that each user can detect its message without substantial interference imposed by the stronger users whose signals are already detected and removed from the superposed signal. In this way, the message of the user with the most allocated power is decoded where the messages from other ones are considered as noise, and thus the decoded message is least contaminated. This mechanism is depicted in Figure 1.3.4.

Code-Domain NOMA

The concept of CD-NOMA has been inspired by the classic CDMA systems which are primarily built based on the idea of allowing multiple users to share the same RE by employing unique user-specific spreading sequence that presents its signature. Users are then separated by exploiting the differences among their different spreading codes.

Unlike CDMA systems, the separation of NOMA users at the receiver may require sophisticated techniques, such as SIC and MPA, at the cost of an increased complexity. According to the different designs of code, we can distinct several CD-NOMA schemes, the main existing ones are presented in what follows.

(a) **Low Density Spreading Code-Division Multiple Access (LDS-CDMA) :** The most basic scheme of CD-NOMA is the LDS-CDMA [27], [28] which directly extends the CDMA. In CDMA systems, each user is associated to one among non-sparse quasi-orthogonal spreading sequences. Each spreading sequence is composed of a number of time slots so-called *chips*. The symbol of all users are spread over different chips and superimposed before transmission. At the receiver, a simple low-complexity correlation detector is sufficient to cancel the inter-sequence interference thanks to the orthogonality of sequences, but this comes at the expense of the number of connected users. LDS-CDMA aims to increase the number of users by spreading their information over a small number of chips rather than all of them which consequently limits the interference on each chip. Hence, low-density and consequently non-orthogonal sequences are required. The MPA detector is needed at the receiver even when the AWGN assumption is considered. Hence, LDS-CDMA has the drawback of high multi-user detection complexity.

(b) **Low Density Signature Orthogonal Frequency-Division Multiplexing (LDS-OFDM) :** LDS-OFDM can be interpreted as a transformed version of LDS-CDMA where each user's symbol is spreaded across a carefully selected number of subcarriers and superimposed in the frequency domain which makes LDS-OFDM more adapted for strong frequency-selective channels [29].

(c) **Multi User Shared Access (MUSA) :** The main idea of MUSA is that users are divided into several groups. In each group, the signal of each user is multiplied by its specific power scaling coefficient. Next, the signals of all the users of a given group are superimposed to be spreaded through a specific sequence [30] as presented in Figure 1.3.5. The group-specific sequences are designed to be orthogonal such that inter-group interference is easily eliminated at the receiver by using a low-complexity correlator. Finally, a SIC detector can be employed to solve the intra-group interference problem.

(d) **SIC Amenable Multiple Access (SAMA) :** Unlike MUSA, SAMA divides users into several groups based on a joint design of the system signature matrix and iterative multi-user interference cancellation technique such that a diversity gain is obtained [31]. At the receiver side, an iterative MPA is used in an effective way. The manner in which users are organized into groups accelerates the convergence of the iterative detector at the receiver by first eliminating the most reliable users

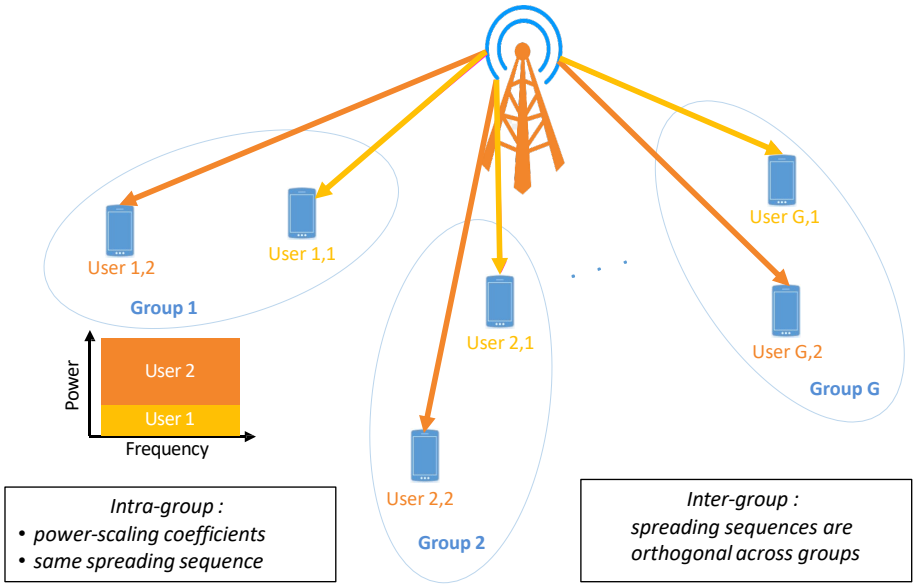


Figure 1.3.5: Basic principles of MUSA.

(the ones whose data is spreaded on more subcarriers) at each iteration which will consequently facilitate decoding the less reliable users. The authors in [31] considered that the above-described design possesses a convergence-amenable property.

(e) **Pattern-Division Multiple Access (PDMA) :** Motivated by the error propagation problem caused by SIC in PD-NOMA, PDMA uses a specific sparse pattern to map transmitted data onto small part of a group of resources [32]. The PDMA pattern is a binary vector whose length is the number of resources in the group, the binary value of each one among its entries determines if the user is mapped to the corresponding resource. The patterns are selected to maximize the diversity order while minimizing the overlapping among users. Belief propagation (BP) based iterative detection and decoding is adopted at the receiver in an uplink scenario, while BP or SIC method is applied in downlink scenario.

(f) **Sparse Code Multiple Access (SCMA) :** The idea of SCMA is to combine together the bit to constellation mapping and low-density spreading. Indeed, this technique attracts a lot of attention since it employs multi-dimensional constellations, so-called codebooks, such that bits are directly mapped to different sparse SCMA codewords. This leads to a constellation shaping gain and consequently to better spectral efficiency [33]. However, some work is still needed to propose more efficient codebook designs.
Among all existing CD-NOMA techniques, SCMA scheme is shown to achieve a better link level performance [34]. Figure 1.3.6 illustrates a performance comparison between SCMA and some other CD-NOMA techniques, namely LDS-OFDM, MUSA and PDMA, in terms of bit error rate (BER), through Rayleigh fading chan-

nel. Simulations results confirm that SCMA scheme provides better performances when compared to others. Comparison study between PD-NOMA and SCMA in [35] showed that SCMA outperforms PD-NOMA when comparable resources allocation strategies for heterogeneous cellular networks are employed. This gain in performance is at the expense of more complexity where SIC method is used to separate signals in the PD-NOMA case while the more sophisticated MPA one is used for SCMA systems, the later costs more in terms of operations and materials [36]. That is why a lot of research work was conducted to reduce the complexity of SCMA detectors as it will be shown in Chapter 2, however additional future contributions are expected to further enhance this aspect.

1.3.3 NOMA Progress of Standardization with 3GPP

In this subsection our aim is to review all technical outcome and specification set proposed by the 3GPP, especially by its technical specification group TSG-RAN, to investigate NOMA for LTE and 5G under Release 15 and Release 16. In April 2016, a discussion on different multiple access techniques for both downlink and uplink was proposed in [37] where OMA schemes, namely TDMA, FDMA and spatial-division multiple access (SDMA), and NOMA ones, namely MUSA, LDS-CDMA, SCMA and resource-spread multiple access (RSMA), were compared in order to provide candidates for the three different types of services for 5G networks whose requirements vary according to supported functionalities: mMTC (improved link budget, low device complexity, low energy con-

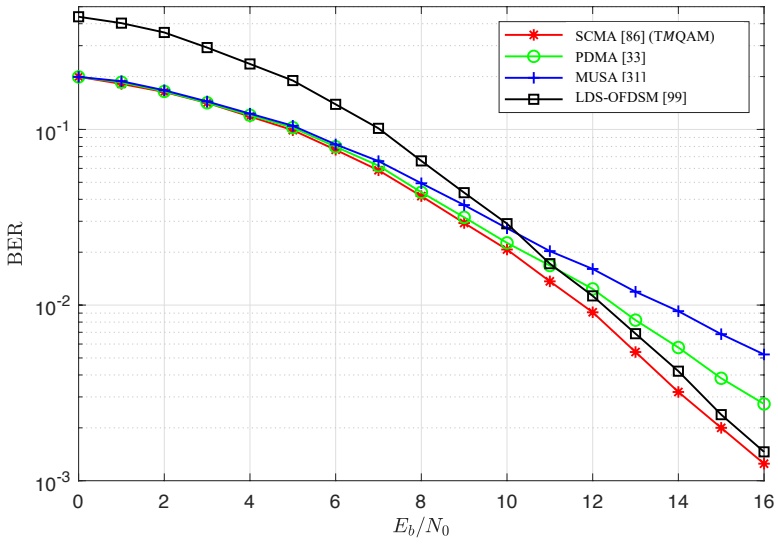


Figure 1.3.6: Performance comparison of LDS-CDMA, MUSA, PDMA and SCMA, in terms of BER, through Rayleigh fading channel : the number of orthogonal REs is 4 and the number of users is 6.

sumption and high density device deployment), eMBB (low latency and higher spectral efficiency/throughput) and URLLC (low packet error rate and low latency). Then, by the end of 2016, the activity on NOMA as work item, was put on hold to give way for more basic 5G functionalities. Many NOMA schemes were evaluated in LTE Release 14 and the decision was to continue with NOMA as study item for 5G Release 15 at least for mMTC. Later, due to the strong interest in both academia and industry [38], NOMA study was revived in Release 15 for both URLLC and eMBB scenarios [39]. That is NOMA was studied under four main aspects: I) transmission side NOMA schemes, II) receiver algorithms, III) NOMA related procedures and IV) preliminary performance evaluations. Furthermore in Release 16, 3GPP TSG RAN, during their 85^{th} meeting in May 2016, considered (a) 3 categories of CD-NOMA based on how data is spreaded over allocated resources: symbol-level spreading such as MUSA, bit-level scrambling/interleaving such as the interleave-division multiple access (IDMA), and joint modulation and spreading such as SCMA, and (b) 3 major typical multi-users receivers: minimum mean square error (MMSE) hard interference cancellation, elementary signal estimator with soft-input soft-output decoder, and expectation propagation algorithm (EPA). Several proposed NOMA schemes in 3GPP studies were categorized and compared, in a qualitative manner, in [40]. Recently, the authors in [41] had also highlighted the final conclusions presented in NOMA study-items of 3GPP Release 15. No clear gain from NOMA over Release 15 and Release 16 mechanisms was observed in all studied scenarios. Mainly, in a large number of cases, link-level results from many companies showed no gain, and the system-level simulations do not provide conclusive performance enhancement. Currently, NOMA techniques are not deployed in the 5G, and are not mentioned explicitly in the planning of 3GPP Release 17 and Release 18, where their content was largely decided at the December 2021. However, NOMA is expected to be reconsidered in very high density scenarios of beyond 5G generations [41].

1.4 Machine Learning for Communication Systems

ML has recently attracted a lot of attention and is applied in many domains. In this section, we start by presenting the definition of the concept of ML. Then, we will highlight recent advances on ML before explaining how it is applied to wireless communication networks. In the rest of this section, we will turn our focus to ML-based PD-NOMA and CD-NOMA techniques. Finally, it is interesting to study specifically the ML application for SCMA since the latter is the main topic of this thesis.

1.4.1 What is Machine Learning?

ML is a field of study in artificial intelligence that is devoted to data analysis with an aim to create, in an automatic way, trained models which can analyze and draw inferences from patterns in data. This model can then be exploited on new data to make decisions. Learning algorithms can be categorized according to the learning mode they employ. Among the ML methods, two are the most widely adopted, namely: (i) the supervised and (ii) the unsupervised learning. The supervised learning which refers to when the data, which are employed to train the algorithm, are already labeled with correct answers. We can divide the supervised learning problems into regression and classification ones. Both problems have an objective which is to build a model that can predict from the attribute variables, the value of the dependent attribute. Usually, the dependent attribute is numerical when the problem is a regression one and categorical when the problem is a classification one.
Unlike the supervised learning, and as its name implies, the unsupervised learning is when the input data is not labeled. Indeed, the objective of this type of learning is to automatically find some similarities among its input data. The two essential techniques of unsupervised ML are: (i) Clustering and (ii) Association. Firstly, a clustering-based system is able to find groups from the trained data, so-called clusters, and then classify the predicted data according to the auto-defined new clusters. While an association-based system consists of finding the adequate rules in order to discover interesting representations and relationships which are hidden in large datasets.

1.4.2 Recent Advances on Machine Learning

ML techniques, including artificial neural networks, exist in the literature since several decades. However, ML had known its more recent revival around 2006 thanks to some in advances in deep learning (DL) which is one of the branches of ML. Here, *"deep"* refers to the depth of the neural network which is employed in the learning process. In other words, deep learning refers to the use of neural networks with a lot of layers and a lot of nodes per layer such that higher-level features can be progressively extracted from raw inputs. Deep neural networks can be used for supervised, semi-supervised and unsupervised learning. For instance, since the early 2000s, DNNs have been used to implement language models for natural language processing such as for Google's neural machine translation system [42].
Convolutional neural networks (CNNs) are widely applied to image and video recognition, recommendation systems and natural language processing [43]–[45]. CNN, called also ConvNet, is a type of feed-forward artificial neural network which uses convolution in place of general matrix multiplication in at least one of its layers, this approach was inspired by the visual cortex of animals and allows processing data that has a known

grid-like topology. As for recurrent neural network (RNN), as its name implies, it is an artificial neural network with recurrent connections. It processes sequences by iterating through the sequence elements and maintaining a state containing information relative to what it has seen so far. In effect, an RNN is a type of neural network that has an internal loop. This type of ML is mainly used for sequences of variable length, for instance time-series and text. Such as an RNN can determine whether a phrase is positive or negative which is also known as sentiment analysis: the network connects a positive or negative answer to certain word sequences it has seen in training examples [46]. As special kind of RNN, so-called the long short-term memory (LSTM), attracts more attention since it is designed to be capable to learn long-term dependencies. In other words, LSTM is the smart use of RNNs where it is conceived to remember things that have happened in the past and to look for patterns across time in order to make sense to its next guesses. An important approach to improve its performance was proposed in [47] by learning an ensemble of LSTMs rather than learning a single one, however, the performance of the ensemble depends on the accuracy and diversity of individual LSTMs.
Autoencoders (AEs) are employed for unsupervised learning which aim is to create a compressed representation of raw data. Thanks to its bottleneck architecture, AE is composed of an encoder and a decoder sub-models. The encoder compresses the input into its most reliable features, and the decoder role is to recreate the input from its compressed version provided by the encoder. AEs has many types with several uses, such as the denoising AE, the sparse AE, the variable AE and the contractive AE [48]–[51]. Hence, AE is a basic tool for representation learning.
Most of the above-mentioned concepts are known for a while, however, the ML learning field suffered from a major slow-down throughout the 1990s and 2000s. The main reason behind that is data and hardware since this field is guided by experimental findings rather than by theory, algorithmic advances only become possible when appropriate data and hardware are available to try new ideas or to scale up old ideas (as it is often the case).

1.4.3 How Machine Learning is Applied in Wireless Networks?

Further than its application in wide-ranging fields such as computer vision, healthcare, speech and others, ML contributed also to advances in some aspects of wireless communication systems [52]. In addition to supervised and unsupervised learning as presented in subsection 1.4.1, other learning paradigms, such as semi-supervised learning and reinforcement learning, are also exploited in wireless communication systems.
For instance, DL-based solutions are employed for signal detection in [53]–[55]. Some other research works study channel encoding and decoding [56], [57]. Other works focus on channel estimation, prediction, and compression as are [58], [59]. In addition, end-

to-end communications and semantic communications were also considered in [60], [61]. Finally, resource allocation problem was investigated in [62], [63]. All these works had pointed out that the performance of wireless networks does improve if ML is properly exploited.
For instance, in [56] a deep neural network is used for MIMO detection where both block-fading and variant channels are considered. Later, a neural network based estimation of the channel state information (CSI) was proposed, firstly for OFDM systems [58], and recently for 5G wireless communications [59] under the assumption of complicated channel distortion and interference, where the aim is to reduce the high computational complexity that makes the regular channel estimation mechanism unsuitable for future communication networks. Newly, an application of ML to secure vehicular communications for internet of vehicles was proposed in [61].

1.4.4 Machine Learning for NOMA

Since the end of 2018, several ML based NOMA schemes have been observed in the literature [64]–[71] with the aim of enhancing multiple access performance. Such approach can be reasonably considered, not only thanks to ML breakthrough, but also because of the improvement of the computational performance of wireless systems.
Hence, several schemes of NOMA were proposed by using DNN models [65], [66], [72]. For instance, in [66], a non-convex DNN-based optimization for resources allocation is used to minimize the total transmit power of a joint downlink of PDMA system. Similarly in [65], a two user PD-NOMA scenarios were considered, a significant performance improvement of a DNN receiver over a conventional SIC one is provided.
Recently, a CNN is considered for symbol-level multi-user detection in Welch bound equality spread based NOMA system [71], this NOMA scheme employs low-correlation spreading signatures. The supervised learning based solution matches almost the conventional detector performance. Also in [73], a CNN based approach to estimate the channel of NOMA based mmWave hybrid systems was introduced. The idea is that users are grouped into different clusters based on their channel gains, then a beamforming operation is applied on each cluster. Afterwards, a coarse estimation of the channel is made from the received signal, this estimate is finally used as the input of the CNN model.
The application of LSTM networks for NOMA systems was also studied. Gui *et al.* proposed, in their work [64], a supervised LSTM based PD-NOMA where channel states of multiple users are automatically tracked by the LSTM network which already learned the environment via offline training. It was proven that the LSTM-based framework is very suitable for user activity and data detection of PD-NOMA systems. The authors in [70] were motivated by proposing a joint precoding (on the transmitter side) and decoding (on the receiver one) for downlink MIMO-NOMA, an AE was employed. The idea is to

extract the important features from the transmission and reception processes in order to provide a significant complexity reduction compared to the conventional process.

1.4.5 Machine Learning for SCMA

In this thesis, we are more interested in how deep learning was applied to design SCMA systems. Very recently, this problem attracted some attention [74]–[80]. The main aim is to design a detector by offline training a DNN such that one shot online non-iterative decoding is performed with a relative low-complexity.

In [74], denoising autoencoders (DAEs) are used to jointly design the encoder and the detector. Hence, one DNN generates the codebook automatically and must learn how to efficiently map symbols to a complex constellation, another fully-connected DNN decodes the received vector to detect the symbols. The training is based on an end-to-end objective function which must be minimized in order to minimize the BER. The training data is randomly generated with a given noise level. A similar approach is also studied in [80]. However, the aforementioned detectors can not be used with any given codebook. That is why the proposition in [76] studied only the detection problem independently from how codebook is designed and generated. A sparsely connected DNN was designed such that the propagation between two of its layers is calculated based on how messages are passed between subcarrier nodes and user nodes in traditional iterative MPA. The idea is to unfold the MPA factor graph iterations into the network layers which results in sparse DNN. The simulated training data are generated dynamically to have a different noise level at each step. The same unfolding principle was studied in [75] in order to propose a joint detection and decoding of channel-coded SCMA systems. The channel coding is randomly generated and is included in the training procedure which makes this proposition more suitable for practical applications. The limitation of these methods is that they utilize a fixed constellation mapping and they are only applied in the case of regular SCMA codebooks.

A hybrid multiple access scheme was proposed in [79] where OFDMA orthogonal resources are used for near users, and non-orthogonal SCMA based access is used for far ones. Two deep learning-based detectors were proposed, one DNN for near users detection, and the other is for far ones detection. In the two cases, a standard DNN structure was offline trained via a simulated data.

The unique research work to employ CNN based detector for SCMA systems is proposed in [77] where a blind detector is designed using unitary filters to fit with uni-dimensional (1D) input instead of the common and known use of CNNs with two-dimensional (2D) inputs. The drawback of this proposition is that its performance is limited in low signal-to-noise ratio (SNR) regime when compared to that of the MPA.

Data in NOMA based wireless systems will be more and more complex and heteroge-

neous which requires to employ powerful tools in order to inspect them and to detect the transmitted bits. This means that the deep learning based solutions will be increasingly attractive. However, there are still some challenges that must be overcome before the deep learning based NOMA systems become reality. Firstly, despite the fact that existing propositions enhance the NOMA performance but this come at the expense of high computational complexity which represents a major issue. Moreover, the optimal neural network, in terms of architecture, depth, number of parameters and training data and methods, is still a subject of research.

1.5 Focus of the Thesis

Multiple access techniques are very crucial for the evolution of different generations of wireless communication systems from 1G to 5G. NOMA is the privileged multiple access technique for beyond 5G. In this thesis, we are more interested by CD-NOMA techniques since we think that they are more effective in terms of performance and flexibility. SCMA mechanisms are designed to facilitate the support of massive connectivity and are considered as potentially promising multiple access candidate for future generations based on their capacity to achieve a better link level performance. However, there are still numerous challenging problems to be solved. In this Ph.D. work, we will focus on SCMA with the aim to enhance the design of its codebook and its detector.
In regular SCMA, the users are uniformly served which it is not optimal since all users don't have the same business requirements, and their physical layer needs depend on many factors such as the maximum allowed delay, the required rate and the targeted quality of service. It will be more practical if the SCMA systems can be differently adapted to users needs and their channel states for instance. Some works on adaptive or irregular SCMA do exist, however this area must be further explored to propose some tailored codebook designs and/or tailored detectors for adaptive SCMA systems.
Traditional SCMA suffers from high complexity due to the use of sophisticated iterative decoders. One solution for SCMA codebook design and/or signal detection is to exploit ML techniques and more specifically deep learning which proposes several efficient tools. For the time being, this axis of research was rarely studied [74]–[80], that is why it is deemed as promising and attractive.

One way to better clarify the focus of our thesis is by presenting its main contributions, as follows,

- It is required to comprehensively study SCMA with the objective to explore its architecture, the design of its encoder and detector, its interplay with other 5G technologies such as mmWave and MIMO. To the best of our knowledge, this effort is yet to be conducted in the state-of-the-art, that is why we start by filling this gap.

- In practice, users have different channel conditions and business requirements such as data rate, QoS, and network priority. Here, an adaptive SCMA scheme is proposed to tackle this problem. The idea is to divide users into different groups such as an adaptive designed codebook is allocated to each group. Unlike the traditional SCMA, the sparsity degree, the constellation size, and the allocated power per user, can be adapted to each group requirements in order to increase the overall performance of SCMA system. Simulation results show that the proposed scheme outperforms the traditional regular SCMA.

- It is vital to push on with developing SCMA detectors that can outperform the performance of the conventional MPA based detector in terms of both complexity and bit error rate. In this Ph.D, a new approaches to improve SCMA detection performance using deep learning methods are explored. First, we propose to jointly design and train a DAE and DNN to decode SCMA signals over an additive white Gaussian noise channel. Then, we propose a new distance-based DNN detector under the assumption that the codebook is known at the receiver. The proposed detector can be fairly compared to MPA, and simulations confirmed that its performance is better than that of MPA.

In the following, a comprehensive study of SCMA is presented in Chapter 2. Then, a new adaptive design of SCMA codebooks is introduced in Chapter 3. Finally, Chapter 4 concentrates on the application of deep learning techniques to SCMA detection problem.

Chapter 2

Comprehensive Study of SCMA

Contents

2.1 Introduction **54**

2.2 SCMA System Architecture **54**

2.2.1 Basic Principles of SCMA 54

2.2.2 SCMA Signal Model 55

2.3 SCMA Codebook Design **55**

2.3.1 Codebook Design Procedure 56

2.3.2 Mother Constellation Design 57

2.3.3 Transformation Operators Design 62

2.4 SCMA Detector Design **65**

2.4.1 Message Passing Algorithm 65

2.4.2 Detectors for Channel-Coded SCMA 69

2.4.3 MIMO-SCMA Detectors 70

2.4.4 Machine Learning based Detectors 72

2.5 Conclusions **73**

2.1 Introduction

In this Chapter, we will present the SCMA system architecture by presenting its basic principles and its signal model. Then, existing methods for SCMA codebooks design will be reviewed. Finally, we will explain how SCMA signal can be detected at the receiver either using the traditional MPA or other more sophisticated methods. This will help understanding why SCMA is considered as a promising massive multiple access candidate for future generations of wireless communication systems, and will highlight its capacity to fit with their requirements.

2.2 SCMA System Architecture

In this section, we want to highlight the principles of SCMA system in both uplink and downlink links in order to explain how it can provide multiple access.

2.2.1 Basic Principles of SCMA

In this thesis, we consider a synchronous SCMA system with a BS and J separate users so-called layers that receive and send their data from and for the BS over K OFDM subcarriers, so-called REs. A SCMA transmitter encodes $\log_2(M_j)$ data bits of user j and maps them into a K-dimensional codeword, $\mathbf{x}_j^{(m)}$, which is selected from a distinct codebook $\boldsymbol{C}_j$ of size M_j. Each codebook presents the signature of the corresponding user. The codebooks are built based on multi-dimensional constellation, $\mathcal{C} = \{\boldsymbol{C}_j, 1 \leq j \leq J\}$. As to the codewords of SCMA, they are sparse, i.e. only $N_j \ll K$ of their entries are non-zero and the rest are zeros. We denote N_j the *codebook sparsity degree*. The sparsity key of SCMA that all codewords corresponding to the j^{th} SCMA layer have a unique location of non-zero entries at the same $(K - N_j)$ positions.
Based on the different parameters of SCMA system, for instance, the size of codebook for each user M_j and the codebook sparsity degree N_j, we can distinguish two kinds of SCMA system architectures : (i) regular SCMA and (ii) irregular SCMA.
A regular SCMA system is defined by $N_j = N, 1 \leq j \leq J$ and $M_j = M, 1 \leq j \leq J$, an example is presented in Figure 2.2.1 where all users employ a codebook of size 4 and their signals are spread over two REs, i.e. $M_j = 4, N_j = 2, 1 \leq j \leq 6$. The maximum degree of user superposition on a given RE is denoted d_f, and the overloading factor, λ, is defined by the ratio of number of users to number of REs, $\frac{J}{K}$. Regrading the system in Figure 2.2.1, $d_f = 3$ and $\lambda = 150\%$.
The irregular architecture, as its name indicates, is designed such that the codebooks are allocated differently according to the different needs of users. In Chapter 3, we will focus notably on existing irregular SCMA structures before proposing a new adaptive method

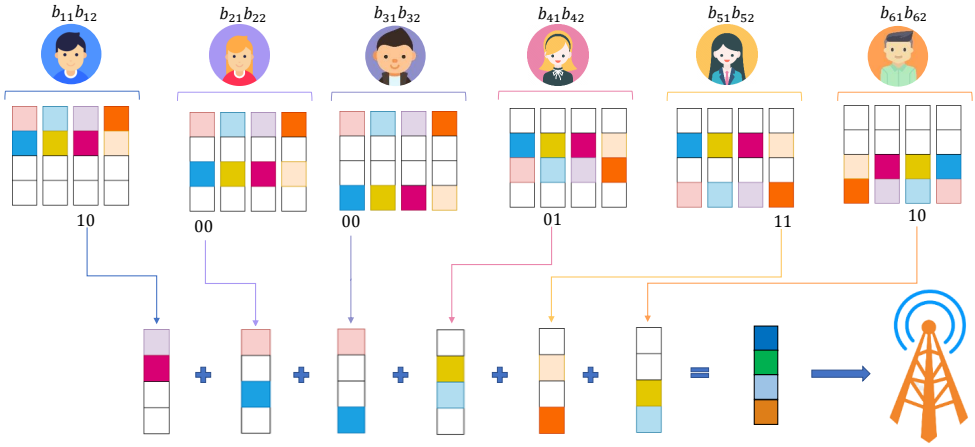


Figure 2.2.1: Representation of the encoder of a regular SCMA system.

for irregular SCMA codebook design.
In the rest of this manuscript, a simple "SCMA" will refer implicitly to regular SCMA.

2.2.2 SCMA Signal Model

The codewords of all users, $\mathbf{x}_j^{(m)}, 1 \leq j \leq J$, are superimposed and exchanged over the K REs. For an uplink scenario based SCMA, the K-dimensional uplink received vector is given by,

$$\mathbf{y} = \sum_{j=1}^{J} \mathbf{H}_j \mathbf{x}_j^{(m)} + \mathbf{n}, \tag{2.1}$$

where $\mathbf{y} = (y_1, \cdots, y_K)^T$, $\mathbf{x}_j^{(m)} = \left(x_{j,1}^{(m)}, \cdots, x_{j,K}^{(m)}\right)^T$, $\mathbf{H}_j = \text{diag}(\mathbf{h}_j)$ and $\mathbf{h}_j = (h_{j,1}, \cdots, h_{j,K})^T$ is the $K \times 1$ channel gain vector of user j. The $K \times 1$ vector $\mathbf{n}$ corresponds to the additive zero-mean white circularly complex Gaussian noise with variance N_0; i.e. $\mathbf{n} \backsim \mathcal{CN}(0, N_0\mathbf{I}_K)$, where $\mathbf{I}_K$ is the identity matrix of size K. However, in a downlink scenario based SCMA, the Equation (2.1) is transformed to express the K-dimensional downlink received vector of user u, $\mathbf{y}_u$, as follows,

$$\mathbf{y}_u = \mathbf{H}_u \sum_{j=1}^{J} \mathbf{x}_j^{(m)} + \mathbf{n}. \tag{2.2}$$

2.3 SCMA Codebook Design

The design of SCMA codebook for each user j is usually based on several steps, a description of each one among them is given in this section. The idea is that the constellation function, associated with each user j generates a constellation set with M_j alphabets of length N_j. Then, the mapping matrix $\mathbf{V}_j$ maps the N_j-dimensional constellation points to SCMA codewords to form the codebook $\boldsymbol{C}_j$.

2.3.1 Codebook Design Procedure

We can describe a SCMA system by a $K \times J$ factor graph matrix $\mathbf{F} = (\mathbf{f}_1, \cdots, \mathbf{f}_J)$ whose columns define the positions of non-zero elements of each user. Thus, the matrix $\mathbf{F}$ is related to the codeword $\mathbf{x}_j^{(m)}$, in equation (2.1) and equation (2.2), by the fact that the structure of $\mathbf{F}$ defines where zeros are located in the codebook from which the codeword $\mathbf{x}_j^{(m)}$ is selected.

Another way to represent the system is to employ $\mathcal{V} = \{\mathbf{V}_j, 1 \leq j \leq J\}$ where $\mathbf{V}_j$ is the mapping matrix of each user j, it is worth noting that $\mathbf{f}_j = \mathbf{V}_j \mathbf{V}_j^T$. For instance, the system in Figure 2.2.1 is represented by,

$$\mathbf{F} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} \quad ==> \text{For instance, } \mathbf{V}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

A factor graph is depicted in Figure 2.3.1 where every circle represents a user (so-called variable node) and every block represents a subcarrier (so-called function node).

As to the SCMA codebook design, it is considered as a joint optimization problem to find both the optimum user-to-RE mapping matrices $\mathcal{V}^*$ and the optimum multi-dimensional constellation $\mathcal{C}^*$, which can be defined as,

$$\mathcal{V}^*, \mathcal{C}^* = \arg\max_{\mathcal{V},\mathcal{C}} D\left(\phi(\mathcal{V}, \mathcal{C}; J, \{M_j\}, \{N_j\}, K)\right) \tag{2.3}$$

where D is a design criterion and ϕ is the SCMA system as it was described above. However, when a SCMA system is fully loaded, one mapping matrix solution is possible and hence it is automatically the optimal one.

Existing SCMA codebooks designs simplified this optimization problem into a suboptimal multi-stage approach [81], such that the design of SCMA codebook is performed in three

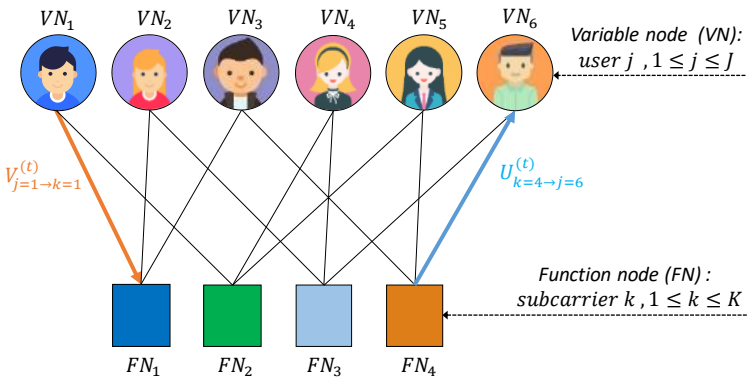


Figure 2.3.1: The factor graph corresponding to encoder of the SCMA system presented in Figure 2.2.1.

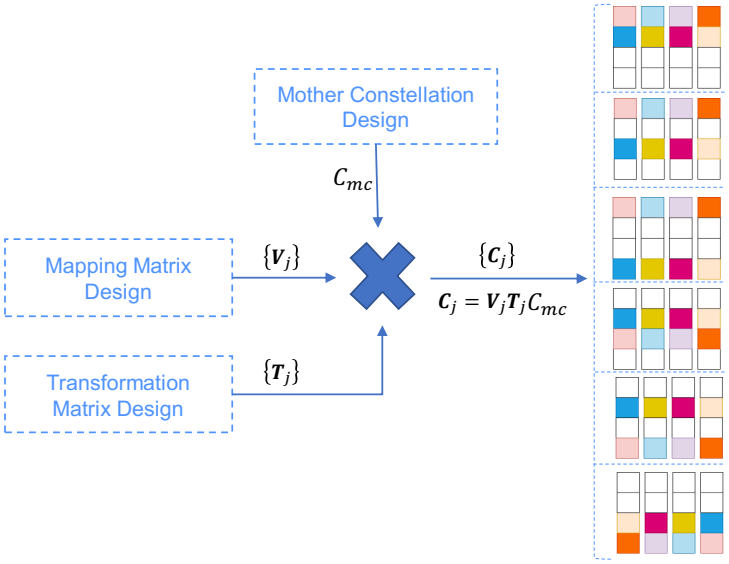


Figure 2.3.2: A block diagram of SCMA codebook design procedure.

main steps: firstly a mother constellation, $\boldsymbol{C}_{mc}$, is designed, then user-specific transformation matrices, $\mathbf{T}_j$, are used to generate user-specific multi-dimensional constellations which are finally spread to generate the J codebooks. Hence, the optimization problem in (2.3) is reformulated as follows,

$$\{\mathbf{T}_j^*\}, \boldsymbol{C}_{mc}^* = \arg \max_{\{\mathbf{T}_j\}, \boldsymbol{C}_{mc}} D(\phi(\mathcal{V}^*, \{\mathbf{T}_j \boldsymbol{C}_{mc}\}; J, \{M_j\}, \{N_j\}, K)) \tag{2.4}$$

such that the j^{th} codebook is calculated by,

$$\boldsymbol{C}_j = \mathbf{V}_j^* \mathbf{T}_j^* \boldsymbol{C}_{mc}^*. \tag{2.5}$$

In the following parts of this section and inspired by the codebook design procedure, illustrated in Figure 2.3.2, we present the major keys to design the mother constellation and the appropriate transformation operators.

2.3.2 Mother Constellation Design

The mother constellation matrix of regular SCMA consists of N rows (or dimensions) of size M which allows to represent each $\log_2(M)$ bits with a constellation point of N entries, the i^{th} entry is a complex value which belongs to the i^{th} dimension of the multi-dimensional constellation. The designed codebook must possess a good distancing property among the points of the overall multi-dimensional constellation according to the criterion D such that performance can be enhanced. Exiting design criteria, D, were reviewed in [81]. In the following, we present the definition of the most important ones.

Euclidean distance: The Euclidean distance between two constellation points, $\mathbf{x}_i^{(u)}$ and $\mathbf{x}_j^{(m)}$, $1 \leq u \leq M$, $1 \leq m \leq M$, of user i and j respectively, $1 \leq i \leq J$, $1 \leq j \leq J$, is calculated by,

$$\mathrm{d}_E\left(\mathbf{x}_j^{(m)}, \mathbf{x}_i^{(u)}\right) = \|\mathbf{x}_j^{(m)} - \mathbf{x}_i^{(u)}\| \tag{2.6}$$

A classic design criterion is the minimum Euclidean distance of a multi-dimensional constellation [82], [83], it is defined as,

$$\mathrm{d}_E^{(\min)} = \min_{\substack{1\leq u,m\leq M \\ 1\leq i,j\leq J}} \left\{ d_E\left(\mathbf{x}_j^{(m)}, \mathbf{x}_i^{(u)}\right) \right\} \tag{2.7}$$

This criterion is more useful for evaluating the design of $\boldsymbol{C}_{mc}$ when all users are observing the same fading channel coefficients over their REs.

Euclidean kissing number: The key here is to count the number of distinct constellation point pairs, when the Euclidean distance separating each pair is equal to the minimum Euclidean distance employed by the mother constellation to separate its points.

Product distance: The product distance between two N-dimensional complex constellation points, $\mathbf{x}_j^{(m)} = \left(x_{j,1}^{(m)}, \cdots, x_{j,N}^{(m)}\right)^T$ and $\mathbf{x}_i^{(u)} = \left(x_{i,1}^{(u)}, \cdots, x_{i,N}^{(u)}\right)^T$, is expressed as,

$$\mathrm{d}_P\left(\mathbf{x}_j^{(m)}, \mathbf{x}_i^{(u)}\right) = \prod_{\substack{1\leq n\leq N \\ x_{j,n}^{(m)} \neq x_{i,n}^{(u)}}} |x_{j,n}^{(m)} - x_{i,n}^{(u)}| \tag{2.8}$$

The minimum product distance of a multi-dimensional constellation is given by,

$$\mathrm{d}_P^{(\min)} = \min_{\substack{1\leq u,m\leq M \\ 1\leq i,j\leq J}} \left\{ d_P\left(\mathbf{x}_j^{(m)}, \mathbf{x}_i^{(u)}\right) \right\} \tag{2.9}$$

This criterion is preferred when evaluating the design of $\boldsymbol{C}_{mc}$ in strong fading channel case, i.e., when channel coefficients over employed subcarriers are different.

Product kissing number: It is the number of distinct constellation point pairs with product distance equal to the minimum product distance.

Hence, having several dimensions gives SCMA additional degrees of freedom, this results in an inherent shaping gain which is defined by the ratio of the minimum distance between the codewords of normalized multi-dimensional constellation of size M to the minimum distance between the points of a normalized traditional one-dimensional constellation of the same size. For example, the shaping gain of 4-point two-dimensional constellation as proposed in [33], [82] over quadrature phase shift keying constellation, in terms of

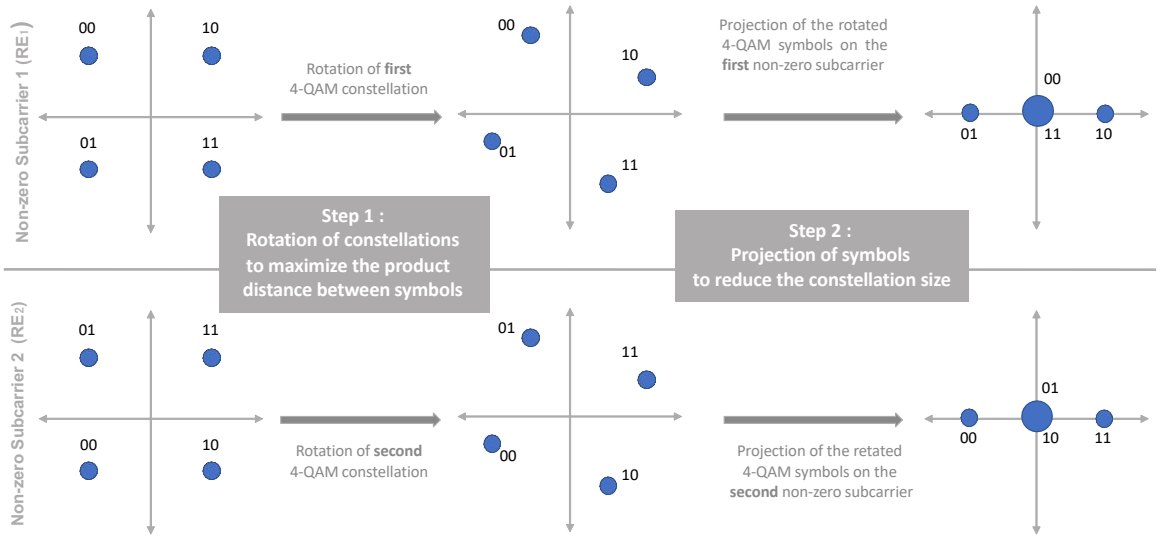


Figure 2.3.3: Low-projection constellation : an example of QAM SCMA constellation points of size $M = 4$ with two non-zero REs, labeled based on Gray coding. First step rotates the constellations to ensure a maximum product distance between symbols which enhances the detection process. The second step could better reduce the complexity of the receiver since some constellation points collide over each RE, for instance the constellation points corresponding to 00 and 11 in the M-sized constellation collide over the first subcarrier, however, they have maximum distance over the second one which makes them separable using M_p-QAM constellation while $M_p \leq M$.

Euclidean distance, is up to 1.25 decibel (dB). The gain on the minimum distance is translated in terms of better system performance.

With the aim to optimize the design of the mother constellation, $\boldsymbol{C}_{mc}$, several works was proposed. For instance, in [82] the M-Beko mother constellation can be generated by minimizing the average alphabet energy for a given minimum Euclidean distance between any two constellation points. Similarly, the M-Peng scheme [83] is designed based on maximizing the minimum Euclidean distance between constellation points of different users. In [84], a complex constellation building is proposed. Such that its imaginary part is independent of its real one, which can help to reduce the decoding complexity, a shuffling operation can be employed to separate the imaginary and the real part of the complex constellation.

One approach to further reduce the complexity of the receiver is to conceive a low-projection mother constellation [84]–[88], that is to employ a lesser number of colliding constellation points over each dimension, however the codewords of each user can still be decoded since they are distinct on other dimensions as shown in Figure 2.3.3. This will decrease the complexity at the detector, for instance, MPA complexity can be reduced from M^{d_f} to $M_p^{d_f}$, where M_p is the size of the constellation after projection. The best employed criterion to design the low-projection constellations is *product distance* which has to be adjusted to be as low as possible without degrading the performance in the high

Table 2.3.1: Two examples of 2-dimensional constellations with 4-codewords where $x_n^{(m)}$ is the n^{th} entry of the m^{th} codeword m, that is $x_n^{(m)}$ belongs to dimension n.

Codeword m	T4QAM		4LQAM	
	$x_1^{(m)}$	$x_2^{(m)}$	$x_1^{(m)}$	$x_2^{(m)}$
1 (00)	$+\frac{3}{\sqrt{10}}$	$+\frac{1}{\sqrt{10}}$	$-\frac{\sqrt{2}}{2}$	$-\frac{\sqrt{2}}{2}i$
2 (01)	$-\frac{1}{\sqrt{10}}$	$+\frac{3}{\sqrt{10}}$	$-\frac{\sqrt{2}}{2}$	$+\frac{\sqrt{2}}{2}i$
3 (10)	$+\frac{1}{\sqrt{10}}$	$-\frac{3}{\sqrt{10}}$	$+\frac{\sqrt{2}}{2}$	$-\frac{\sqrt{2}}{2}i$
4 (11)	$-\frac{3}{\sqrt{10}}$	$-\frac{1}{\sqrt{10}}$	$+\frac{\sqrt{2}}{2}$	$+\frac{\sqrt{2}}{2}i$

SNR zone.

However, optimizing spreading codes and constellations is not a new problem, different approaches with their associated criteria exist in the literature [89]–[94]. These works inspired a variety of multi-dimensional SCMA constellation designs. Based on the quadrature amplitude modulation (QAM), the TMQAM scheme was proposed in [86] where the design of the proposed constellation uses a shuffling method which establishes the N-dimensional complex constellation from the Cartesian product of two N-dimensional real symbols with a specific Euclidean distance. Then, a rotation is applied to maximize the minimum product distance of both N-dimensional constellations. Similar to TMQAM, the MLQAM [87] technique applies the shuffling method but its constellation has a low number of projections. Also, the authors in [95] proposed a constellation for SCMA systems over Rayleigh fading channels based on a criterion derived from cutoff rate of MIMO systems, their proposition was denoted as M-Bao. A spherical coding was also proposed in [96] by the same authors of [95], however they shown that M-Bao codebooks outperform the spherical coding based ones. The TMQAM, MLQAM and M-Bao are all based on Cartesian product of two $\log_2(M)$-QAM which constitutes the M corners of a $\log_2(M)$-dimensional hyper-cube. The authors in [97] proposed the MHQAM scheme based on an optimization of rotation angles of the hyper-cube, this leads to a reduction of MPA complexity from M^{d_f} to $(\log_2(M))^{d_f}$. For a better illustration of mother constellation design, two examples of 2-dimensional ones with 4-codewords are provided in Table 2.3.1 and Figure 2.3.4.

Some multi-dimensional constellations are constructed by exploiting multi-radius rings [87], [98]–[100], i.e. the constellation points of a given dimension are not randomly placed but they belong to concentric rings. The M-point circular constellation MCQAM [87] is based on the analysis of the signal space diversity for MIMO systems over Rayleigh fading channels with a low number of projections for each complex dimension. MPA complexity is then reduced from M^{d_f} to $(M-1)^{d_f}$. A multi-dimensional SCMA codebook design based

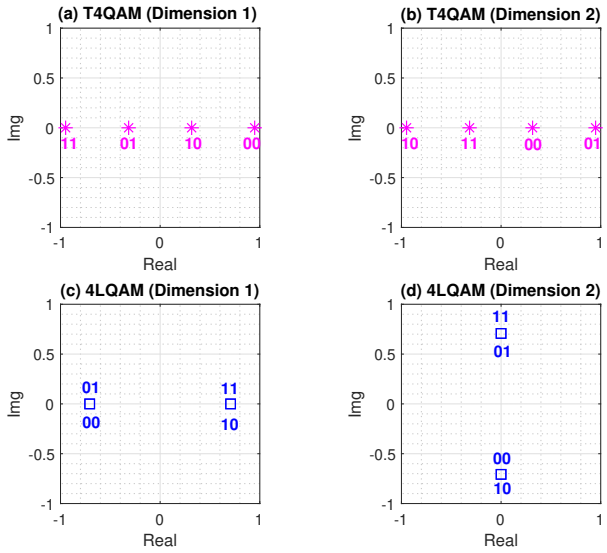


Figure 2.3.4: Illustration of two examples of 2-dimensional constellations with 4-codewords, i.e. T4QAM [86] and 4LQAM [87]: (a) projection of T4QAM over first dimension, (b) projection of T4QAM over second dimension, (c) projection of 4LQAM over first dimension and (d) projection of 4LQAM over second dimension.

on star-QAM constellations was proposed for uplink SCMA systems in [98], [99]. That is the first dimension of mother constellation is constructed using a M-dimensional star-QAM constellation, then, the other dimensions are obtained by scaling and permuting the points of the first one, an example is illustrated in Figure 2.3.5. The mother constellation parameters are obtained through computer search inspired by the approaches in [89], [93]. This approach is powerful for designing codebooks with large size and/or high dimension. The optimization objective in [99] was to minimize the pairwise error probability between two transmitted codewords $\mathbf{x}^{(a)}, \mathbf{x}^{(b)}$ which is given by,

$$\mathbb{P}(\mathbf{x}^{(a)}, \mathbf{x}^{(b)}|\mathbf{H}) = Q\left(\sqrt{\frac{\|\mathbf{H}\left(\mathbf{x}^{(a)} - \mathbf{x}^{(b)}\right)\|^2}{2N_0}}\right). \tag{2.10}$$

This criterion can be interesting since it is applied directly on the generated codewords of the J codebooks and not on the mother constellation.
Multi-stage optimization of another ring-based approach was proposed in [100]. Each ring is composed of uniformly spaced phase shift keying (PSK) points such that the whole rings form an amplitude and phase shift keying (APSK) constellation which is capable of outperforming the classic square shaped QAM constellation in peak-power-limited systems. The first dimension is designed by maximizing the coded modulation capacity, then the other dimensions of the mother constellation are deducted using optimized permutations. Once the mother constellation is designed, optimized and evaluated based on one of the above-presented design criteria, the applied transformations which are used to generate the J codebooks must be designed to preserve the characteristics of the mother constel-

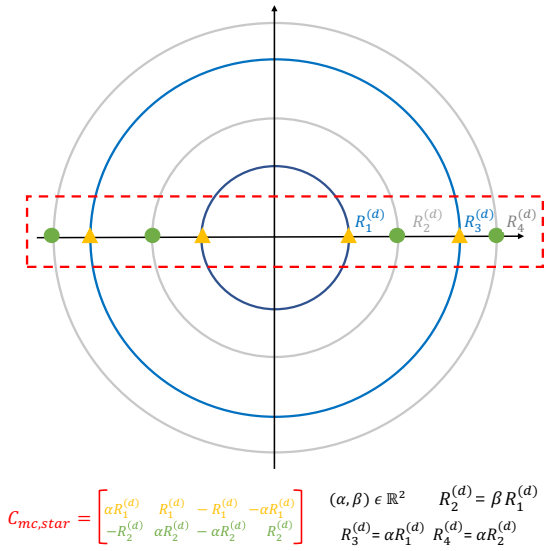


Figure 2.3.5: An example of four-rings star-QAM mother constellation for the design of a SCMA codebook of size $M = 4$ and sparsity degree $N = 2$ where α and β are 2 reel design parameters.

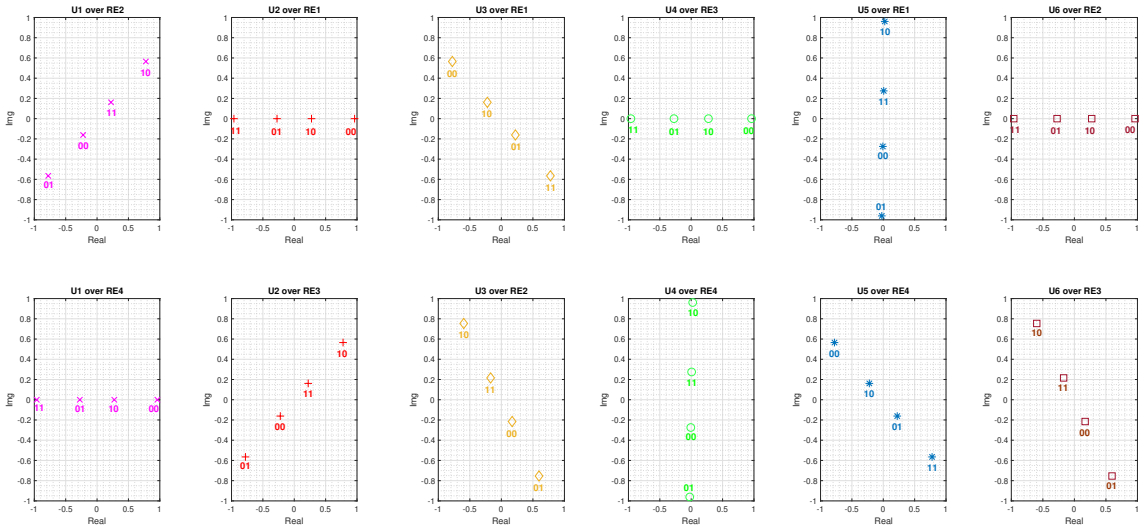


Figure 2.3.6: Illustration of the six 2-dimensional constellations with 4-codewords, generated by applying unitary rotation transformation, as described in equation (2.11), on the T4QAM mother constellation [1]: the projections of the constellation of each user on its associated two REs are depicted.

lation.

2.3.3 Transformation Operators Design

Once having the mother constellation of N-dimension, $\boldsymbol{C}_{mc}$, different sets of operators can be applied on top of it to build multiple sparse codebooks for several layers of SCMA.

As one of the fundamental steps in the design of SCMA codebooks, the transformation operators design was investigated in recent researches works [99]–[105]. A transformation operator, $\mathbf{T}_j$ (see equation (2.5)), could be either one or a combination of typical operators such as complex conjugate, rotation operator, interleaving and vector permutation. In most works, preserving an adequate Euclidean distance profile of the mother constellation is based on applying a unitary rotation matrix [1], [33], [82]–[84], [86], [106]–[108]. The authors in these research works have combined the mapping matrix and the transformation one to form a new rectangular matrix, $\mathbf{F}_T$. An example, for $K = 4, N = 2$ and $M = 4$, is given by [1],

$$\mathbf{F}_T = \begin{bmatrix} 0 & \varphi_1 & \varphi_2 & 0 & \varphi_3 & 0 \\ \varphi_2 & 0 & \varphi_3 & 0 & 0 & \varphi_1 \\ 0 & \varphi_2 & 0 & \varphi_1 & 0 & \varphi_3 \\ \varphi_1 & 0 & 0 & \varphi_3 & \varphi_2 & 0 \end{bmatrix} \tag{2.11}$$

where $\varphi_1 = e^{j\theta_1}, \varphi_2 = e^{j\theta_2}$ and $\varphi_3 = e^{j\theta_3}$. Traditionally, $\theta_1 = 0, \theta_2 = \frac{\pi}{3}$, and $\theta_3 = \frac{2\pi}{3}$. In this circumstance, the codebook of user 1, for instance, is calculated based on the following mapping and transformation matrices,

$$\mathbf{V}_1 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{T}_1 = \begin{bmatrix} e^{j\theta_2} & 0 \\ 0 & e^{j\theta_1} \end{bmatrix}.$$

The matrix $\mathbf{F}_T$ respects the Latin criterion where not only non-zero elements in each row are distinct but also those in each column which leads to controlling both dimensional dependency and power variation of the multi-dimensional constellation, while keeping the Euclidean distance unchanged [84], [106]. Figure 2.3.6 shows the normalized 2-dimensional constellations with 4-codewords, generated by applying unitary rotation transformation, as described in equation (2.11), on the T4QAM mother constellation, as presented in Table 2.3.1. The projections of the constellation of each user on its associated two REs are depicted in subfigures. The complete codebooks, as introduced in [1], are given in Figure A of Appendix A.

In [95], multi-user codebooks are obtained via specific computer-designed rotation matrices instead of the unitary rotation ones. Furthermore, a two-dimensional specific user rotation is applied to the mother constellation proposed in [87] to generate the different SCMA codebooks. In [103], the SCMA design and the security of the link are combined, the codebooks are generated by rotating the mother constellation with random angles extracted from channel phases, this requires to know the CSI. These encrypted codebooks protect exchanged information with low complexity.

An optimization of transformation operator was proposed in [102] based on a novel cri-

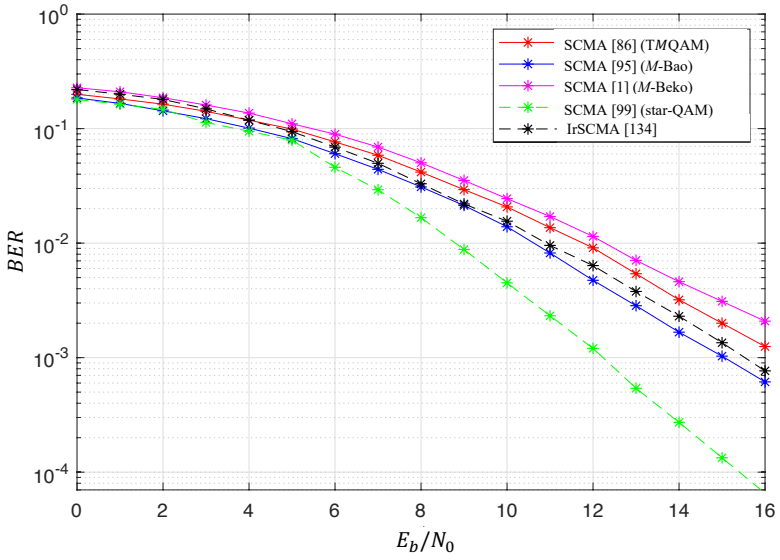


Figure 2.3.7: Performance evaluation of uplink SCMA system with different codebook designs through Rayleigh fading channel : the number of orthogonal REs is 4 and the number of users is 6.

terion to select the most appropriate permutation set in order to improve the probability of reliable detection of the first user which largely improves the performance of MPA detector. The proposed criterion tries to maximize the sum of distances among dimensions of interfering codewords multiplexed on each RE. A permutation-based SCMA scheme is proposed in [109], the idea is that the codebook of user j is designed by mapping the encoded codeword to N among K REs whose positions are defined according to the values of data bits to be transmitted, i.e. different non-zero locations of encoded complex vectors are assigned to different codewords. This approach is different from the majority of SCMA designs where the fixed positions of non-zero elements of each user are determined by the columns of the factor graph matrix, $\mathbf{F}$, as illustrated in subsection 2.3.1. The proposed permutation approach increases the spectral efficiency without a complexity overhead when compared to traditional SCMA. A combination of matrix permutation and rotation is employed as transformation operator in [98], [99].

It is worth mentioning that labeling affects the performance of SCMA system in terms of BER, hence it is important to employ the appropriate labeling. For instance, once the multi-dimensional constellation is designed in [100], its labeling is optimized such that the slope of the extrinsic information transfer (EXIT) chart is well adjusted.

A performance evaluation of uplink SCMA system with different codebooks through Rayleigh fading channel is shown in Figure 2.3.7, it is obvious that the design of codebook has an important effect on the system performance. A summary of existing SCMA codebook designs is presented in Table B.1 of Appendix B. Furthermore, a review of SCMA multi-dimensional constellations design was proposed in [81] in which further details on some of the aforementioned SCMA codebooks can be found.

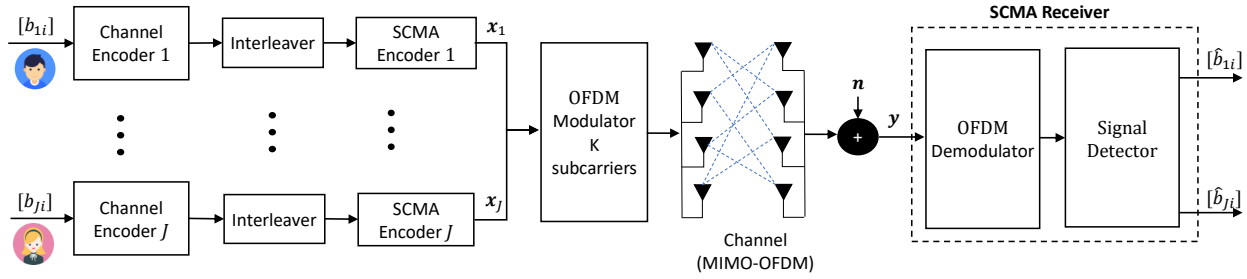


Figure 2.4.1: Block diagram of SCMA system model.

2.4 SCMA Detector Design

A block diagram of classic SCMA system is shown in Figure 2.4.1, after the OFDM demodulator, the transmitted signals must be estimated which requires to have some knowledge of the communication channel, to segregate the SCMA codewords and finally to decode the channel encoding after a deinterleaving operation.

In this section, our objective is to highlight existing algorithms and mechanisms employed to enable the receiver to separate the superimposed codewords which were selected from one of the above presented SCMA codebooks. Each RE of the SCMA systems is occupied by several users which results in smaller distances among the constellation points and could degrade the link performance.

Most existing works on SCMA detectors employ MPA or one of its variations, or a combination of MPA and other methods. Hence, it is straight-forward to start by presenting MPA in the next subsection before reviewing other techniques. Most existing methods propose to sequentially apply the aforementioned steps, however some research works aim to design joint detection methods for MIMO-SCMA systems. Deep learning is also employed in SCMA systems. A review of existing MIMO-SCMA detectors and deep-learning based ones is provided. Also, in the Table D.2 of Appendix D, we summarize the SCMA detection techniques by showing their approach, their assumptions and their complexity order, based on the key parameters of SCMA system as presented in Table D.1 of Appendix D.

2.4.1 Message Passing Algorithm

Traditional MPA

MPA is defined as an iterative parallel decoding technique based on passing the extrinsic information from function nodes (FNs) to variables nodes (VNs) and vice versa [27], [110] as shown in Figure 2.3.1. In each iteration, each FN of the factor graph computes

its outgoing message to a given VN depending on the incoming messages received from the rest of VNs. Then, each VN will send back its message depending on the received messages from the reminder of FNs. Finally, one iteration is considered as completed when one outgoing message has passed in both directions along every edge. The log-likelihood-rates (LLRs) of each coded bit are calculated after few iterations in order to estimate the bites of each user j. We can present MPA process based on three main steps as shown in Algorithm 1 for a SCMA system of J users using M constellation points through K orthogonal REs.

Variations of MPA

Despite being a referent decoder for SCMA, the complexity evaluation of MPA reveals that it relies on a large number of exponential calculus which are of high complexity. With the challenge to reduce this complexity and to fit with critical requirements of future wireless networks, several variations of MPA were proposed, among them we present here, the Max-Log-MPA [111], [112] and Log-MPA [112], [113] methods.

Max-Log-MPA : It is a simplified version of MPA based on a mathematical simplification which approximates the logarithm of a sum of exponential operations into a maximum operation. The key purpose is to move the iterative decoding process into logarithmic domain which eliminates the exponential terms in MPA by employing the simplified formula of *Jacobean logarithm*,

$$\log\left(\exp\left(a_1\right)+\cdots+\exp\left(a_n\right)\right)\approx\max\left(a_1,\ldots,a_n\right) \tag{2.12}$$

Thus, passing numerous messages from FN to VN, and vice versa, will be very less expensive in term of complexity. Based on (2.12), the expression of $\mathrm{LLR}(b_i)$ presented in Algorithm 1 is modified as follows,

$$\mathrm{LLR}(b_i)=\max_{\{\mathbf{x}_j^{(m)}\in\boldsymbol{C}_j|b_i=0\}}\left(\log\left(\mathbb{P}(\mathbf{x}_j^{(m)})\right)\right)-\max_{\{\mathbf{x}_j^{(m)}\in\boldsymbol{C}_j|b_i=1\}}\left(\log\left(\mathbb{P}(\mathbf{x}_j^{(m)})\right)\right) \tag{2.13}$$

Log-MPA : The approximation of the *Jacobean logarithm formula* as presented in (2.12) makes the Max-Log-MPA a sub-optimal solution and results in a performance degradation. To mitigate this issue, a correction term was added by using another *Jacobean logarithm formula*. The adopted approximation is given by,

$$\log\left(\exp\left(a_1\right)+\cdots+\exp\left(a_n\right)\right)=a_j+\log\left(1+\sum_{i\in\{1\ldots n\}\setminus j}\exp\left(-|a_j-a_i|\right)\right) \tag{2.14}$$

where $a_j=\max\left(a_1,\ldots,a_n\right)$. Hence, the LLRs are further updated to be as below,

Algorithm 1: Message Passing Algorithm

Input: $\mathbf{y}, N_0, \boldsymbol{C}_j, \mathbf{h}_j, j = 1, \cdots, J.$

Result: Estimate the value of coded bits for each user j, $j = 1, \cdots, J$.

Definitions

VNs represent users, FNs represent resources or subcarriers,

$\mathcal{U}(k) = \{j, 1 \leq j \leq J \mid \text{VN}_j \text{ is connected to FN}_k\}$,

$\mathcal{R}(j) = \{k, 1 \leq k \leq K \mid \text{FN}_k \text{ is connected to VN}_j\}$.

Step1: Initialization

The prior probability for each codeword is given by:

$V_{j\to k}^{0}(\mathbf{x}_j^{(m)}) = \mathbb{P}(\mathbf{x}_j^{(m)}) = \frac{1}{M}, j = 1, \cdots, J, k \in \mathcal{R}(j)$

Step2: Iterative message passing along edges

while $t \leq N_{iter}$ **do**

1. FN update: the message to be passed from FN_k to one of its neighbors VN_j, $k = 1, \cdots, K, j \in \mathcal{U}(k)$, for a given codeword $\mathbf{x}_j^{(m)} \in \boldsymbol{C}_j, m = 1, \cdots, M$, is calculated as,

$$U_{k\to j}^{t}(\mathbf{x}_j^{(m)}) = \sum_{\mathbf{x}_i^{(m)} | i \in \mathcal{U}(k)\setminus j} \exp\left\{ -\frac{1}{N_0} \left\| y_k - \sum_j h_{j,k} x_{j,k}^{(m)} \right\|^2 \right\} \prod_{i \in \mathcal{U}(k)\setminus j} V_{i\to k}^{t-1}(\mathbf{x}_i^{(m)})$$

2. VN update: the message to be passed from VN_j to one of its neighbors FN_k, $j = 1, \cdots, J, k \in \mathcal{R}(j)$, for a given codeword $\mathbf{x}_j^{(m)} \in \boldsymbol{C}_j, m = 1, \cdots, M$, is given by,

$$V_{j\to k}^{t}(\mathbf{x}_j^{(m)}) = \frac{\prod_{i \in \mathcal{R}(j)\setminus k} U_{i\to j}^{t-1}(\mathbf{x}_j^{(m)})}{\sum_{\mathbf{x}_j^{(l)} \in \boldsymbol{C}_j} \prod_{i \in \mathcal{R}(j)\setminus k} U_{i\to j}^{t-1}(\mathbf{x}_j^{(l)})}.$$

Normalization is necessary to keep the algorithm numerically stable.

end

Step 3 : Taking a decision (LLR at each VN)

1. For each layer j, $j = 1, \cdots, J$, the posteriori probability of codeword $\mathbf{x}_j^{(m)}, m = 1, \cdots, M$ is defined as,

$$\mathbb{P}(\mathbf{x}_j^{(m)}) = \prod_{k \in \mathcal{R}(j)} U_{k\to j}^{N_{\text{iter}}}(\mathbf{x}_j^{(m)})$$

2. Log-Likelihood-Rate for each coded bit, $b_i, 1 \leq i \leq \log_2(M)$, is represented by,

$$\text{LLR}(b_i) = \log\left(\frac{\mathbb{P}(b_i = 0)}{\mathbb{P}(b_i = 1)}\right) = \log\left(\frac{\sum_{\{\mathbf{x}_j^{(m)} \in \boldsymbol{C}_j | b_i = 0\}} \mathbb{P}(\mathbf{x}_j^{(m)})}{\sum_{\{\mathbf{x}_j^{(m)} \in \boldsymbol{C}_j | b_i = 1\}} \mathbb{P}(\mathbf{x}_j^{(m)})}\right)$$

Finally, each bit is decided according to the value of its LLR as following,

$$\hat{b}_i = \begin{cases} 1 \text{ if } \text{LLR}(b_i) \leq 0 \\ 0 \text{ otherwise.} \end{cases}$$

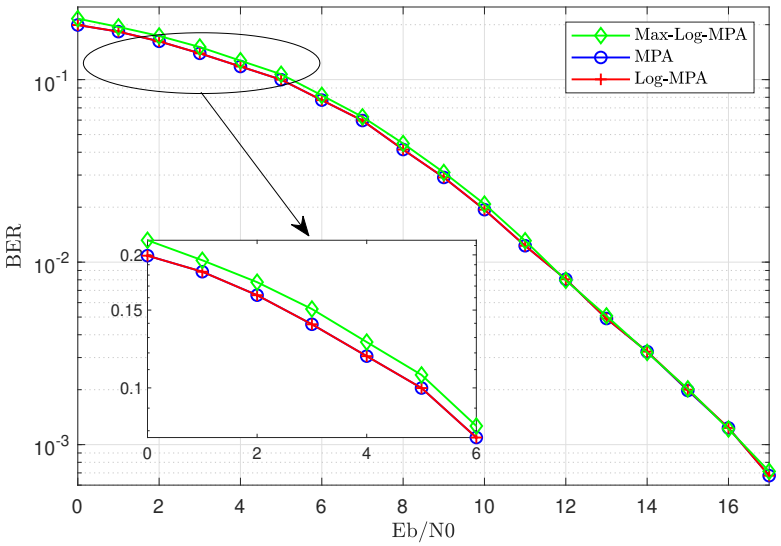


Figure 2.4.2: Performance comparison of MPA, Log-MPA and MAX-Log-MPA variations through Rayleigh fading channel : the number of orthogonal REs is 4 and the number of users is 6.

rather than as in (2.13),

$$\begin{aligned}\mathrm{LLR}(b_i) = &\left[\max_{\{\mathbf{x}_j^{(m)}\in \boldsymbol{C}_j|b_i=0\}}\left(\log\left(\mathbb{P}(\mathbf{x}_j^{(m)})\right)\right)+\right.\\ &\left.\log\left(1+\sum_{m'\in\{1\ldots M\}\setminus m}\exp\left(-|\log\left(\mathbb{P}(\mathbf{x}_j^{(m)})\right)-\log\left(\mathbb{P}(\mathbf{x}_j^{(m')})\right)|\right)\right)\right]\\ &-\left[\max_{\{\mathbf{x}_j^{(m)}\in \boldsymbol{C}_j|b_i=1\}}\left(\log\left(\mathbb{P}(\mathbf{x}_j^{(m)})\right)\right)+\right.\\ &\left.\log\left(1+\sum_{m'\in\{1\ldots M\}\setminus m}\exp\left(-|\log\left(\mathbb{P}(\mathbf{x}_j^{(m)})\right)-\log\left(\mathbb{P}(\mathbf{x}_j^{(m')})\right)|\right)\right)\right] \quad (2.15)\end{aligned}$$

Figure 2.4.2 presents the performance, in terms of BER, obtained by the three previously mentioned variations of MPA, namely MPA, Log-MPA and Max-Log-MPA, through Rayleigh fading channel. It is obvious that, contrary to Max-Log-MPA, Log-MPA can achieve near-optimum performance when compared to MPA due to the aforementioned correction term which compensates for the performance loss of Max-Log-MPA. Furthermore, the performances of the three methods converge when SNR increases. However, the computational complexity of Log-MPA is still a big challenge for some practical implementations especially for energy-sensitive user equipment's in the downlink scenario. As the computation complexity of the above-mentioned detection algorithms increases exponentially with d_f, the detection still takes considerable time. The parameter d_f must be designed to be very small, which largely limits the choice of codebooks.

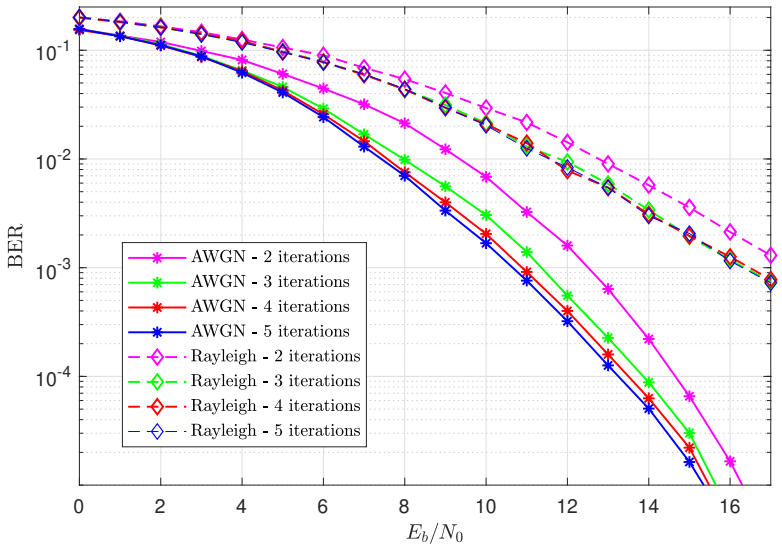


Figure 2.4.3: MPA performance comparison for different number of iterations through AWGN channel and Raleigh fading channel : the number of orthogonal REs is 4 and the number of users is 6.

Most MPA variations are based on a fixed number of iterations which is not optimal since a large number of iterations will considerably increase the complexity, while a low one will lead to less performing detectors. Figure 2.4.3 evaluates the performance of MPA, in terms of BER, for different number of message passing iterations when the channel is assumed to be AWGN or Rayleigh distributed. In the two cases, the performance improves as the number of iterations increases, nevertheless no significant improvement can be observed beyond a certain limit. Moreover, the authors in [112] affirmed that the same convergence behavior can be observed for Log-MPA and Max-Log-MPA. Therefore, a reasonable number of iterations can be set to 4. A flexible number of iterations was allowed in [124] by supervising the convergence rate of each codeword probability which will guarantee to reduce the complexity without compromising the performance.

2.4.2 Detectors for Channel-Coded SCMA

Each wireless link requires an error-correction channel coding, the turbo codes were used in 4G systems, and the low-density parity-check (LDPC) codes and polar codes were adopted by the 5G NR standard [114], [115]. Some SCMA detectors include the error-correction codes structure into their design which gains a lot of attention in the literature. For instance, the LDPC code was employed as the channel coding scheme for the SCMA system in [116], [117] while a polar-coded SCMA was studied in [118]–[120].
In [116], the sparse factor graphs of LDPC coding and SCMA are combined into a joint sparse graph in order to jointly perform the decoding and the detection on one graph.

Then, MPA was simplified by using partial message passing based on a joint trellis representation. Simulations showed that the joint approach is compatible with delay-sensitive applications since it reduces the processing latency compared with disjoint turbo approaches. Another approach was studied in [117] by optimizing the interface between the SCMA detector and the LDPC decoder which made a hardware solution feasible. The authors proposed a MMSE with parallel interference cancellation (MMSE-PIC) detection in the FN operations, and a bit LLR values for message passing. In order to minimize the hardware overhead and to reduce the processing latency, both the proposed algorithm and very large scale integration architecture are jointly designed.
Further in [118], a simple combination of MPA detector for SCMA and a soft-input soft-output successive cancellation for polar coding was studied. The architecture of the receiver was adjusted by re-encoding the soft information of the polar codeword. The reconstructed information is fed into the iterative SCMA detection procedure which results in an additional coding gain. A joint factor graph of SCMA detector and polar decoder are employed in [119], and hence a joint iterative message updating operation was introduced.
The EXIT chart based analysis provided a more understating of how to optimize the polar-coded SCMA system. More precisely, a weight factor was conceived and optimized to mitigate the effect of the correlation among the soft outputs of polar decoder.
As for the CSI, it is not evident how to estimate its real value for some practical mMTC applications where short packets are essentially exchanged. That is why it is interesting to propose a joint channel estimation and decoding for polar-coded SCMA. In [120], the joint detection and decoding scheme is based on traditional Max-Log-MPA detector for SCMA and soft-successive cancellation list (S-SCL) algorithm for polar decoding. The S-SCL provides the prior symbol probability for the SCMA detector while SCMA detector calculates the prior information of the polar decoder. The joint detection and decoding scheme is serialized with a frozen bit error rate mapper such that the CSI is updated at each iteration. The iterative joint channel estimation and decoding scheme is initialized with a sparse Bayesian learning based channel estimation.

2.4.3 MIMO-SCMA Detectors

A straight-forward extension of SCMA factor graph to multiple-antenna case makes the number of resource nodes proportional to the number of antennas, this means that the complexity will respectively increase exponentially. Several research works tried to overcome this difficulty [121]–[127]. Figure 2.4.4 represents a block diagram of an uplink MIMO-SCMA system when spatial multiplexing techniques is applied.
In [128] a maximum likelihood based detection for MIMO-SCMA signal was proposed and compared to SISO-SCMA when using quadrature phase shift keying (QPSK)-mother

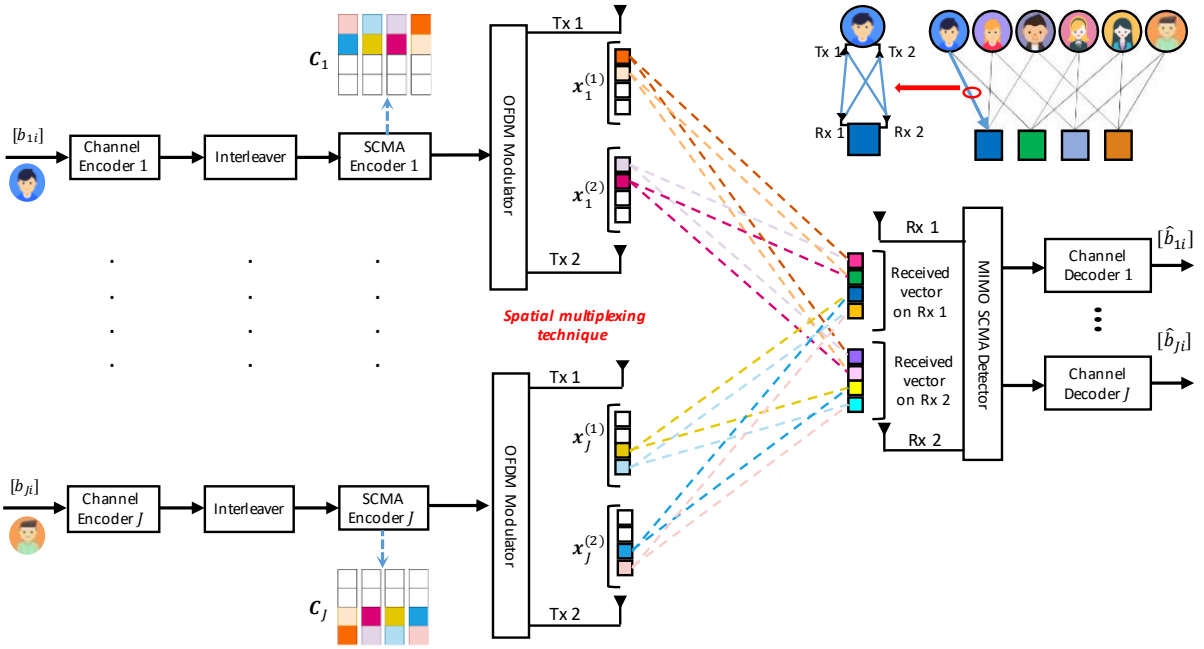


Figure 2.4.4: Block diagram of an uplink MIMO-SCMA system with $N_t = 2$ transmit antennas and $N_r = 2$ receive antennas.

constellation under AWGN channel. Simulations, confirmed that the interplay between MIMO and SCMA can improve the performance of SCMA system in term of binary error. For instance, a gain of 2 dB is achieved when BER= 10^{-5} as presented in Figure 2.4.5. In [123] partial-decoding MPA was proposed, this method tries to remove redundant combinations at FNs to reduce decoding complexity, however, it still suffers from the exponential term in complexity. [125] introduced a stretched factor graph representation for MIMO-SCMA systems which enables the design of a hybrid belief and expectation propagation receiver that includes a channel decoder. Despite the relative low complexity of this algorithm, its main shortcoming is that it can not be applied to user terminals which are equipped with multiple antennas. In [124], a minimum singular value estimation of the channel matrix is used to separate the resources into two categories according to the channel condition at each resource. Then, a low-complexity jointly Gaussian algorithm is applied to the well-conditioned resources, while the ML approach is used with the resources with bad-conditioned channel. This will allow a certain trade-off between performance and complexity when compared to the MPA detector. A similar approach based on an antenna-subcarrier subset selection was introduced in [121], the selection method is based on the channel norms. Furthermore, the MIMO-SCMA factor graph can be partitioned into subgraphs by a QR decomposition of the channel matrix such that the number of resource nodes depends only on the number of users sharing each resource, d_f, this can reduce the complexity of messages calculation [122]. Unlike the later research work, the authors in [127] associated the QR decomposition with EPA, and consequently they proposed a sparse-channel-based EPA (SC-EPA) method which intends to reduce

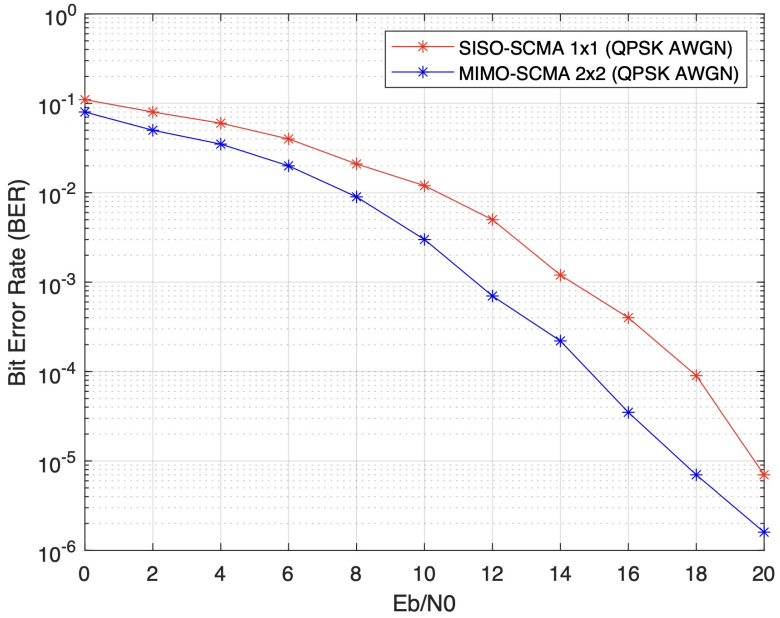


Figure 2.4.5: Uncoded uplink MIMO-SCMA compared to SISO-SCMA based on QPSK modulation when using maximum likelihood detection under AWGN channel [128].

the complexity by modifying the factor graph to create resource clusters, by exploiting the channel sparsity and by proposing a high-parallelism message passing technique. The effect of channel estimation errors on the SC-EPA performance was studied, it seems as robust as some of existing methods.

2.4.4 Machine Learning based Detectors

In this subsection, we are interested in how machine learning was applied to design SCMA detectors with the aim to design a detector by offline training a model such that one shot online non-iterative decoding is performed with a relative low-complexity.
Since 2018, intelligent SCMA encoders and decoders were proposed in [74]–[76], [78]–[80] based on supervised and unsupervised learning. These research works were reviewed in Chapter 2 and could be divided into several categories based on the employed machine learning technique. These categories are DNN, CNN and AE based detectors.
For instance, a jointly design of SCMA encoder and decoder is shown in Figure 2.4.6 where a joint end-to-end objective function was employed to train two neural networks, the first is a DNN which generates the codebook automatically and must learn how to efficiently map symbols to a complex constellation, and the second is a fully-connected DNN which decodes the received vector and detects the symbols. This approach may reduce the BER. However, these techniques are less complex than conventional MPA but they can not yet outperform it in terms of BER.

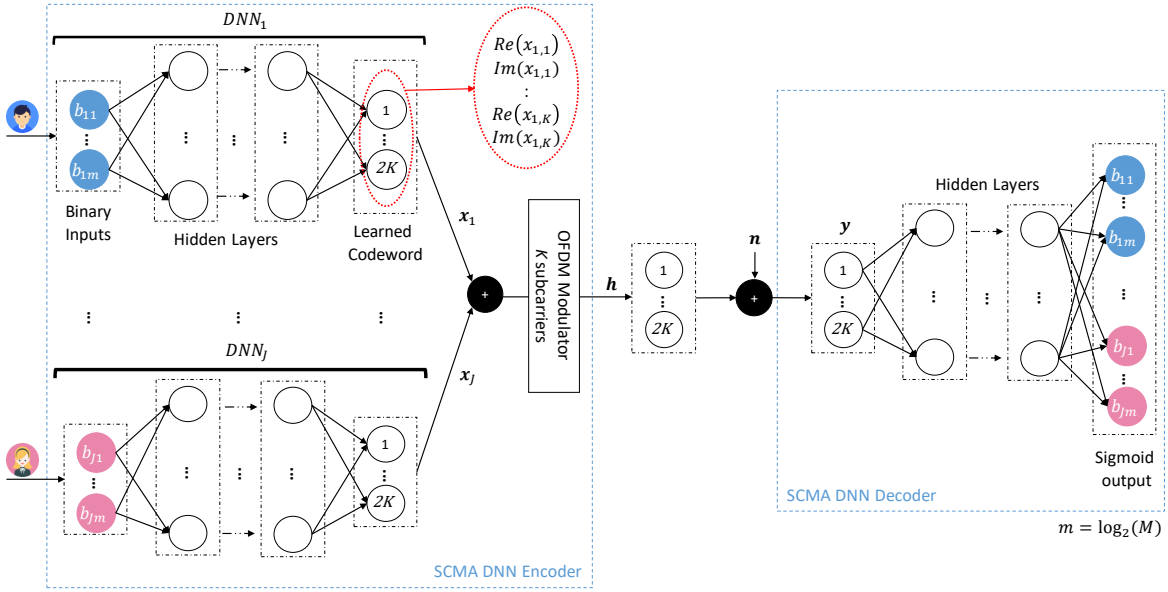


Figure 2.4.6: Structure of the SCMA encoder/detector: $\text{DNN}_j, j = 1 \cdots J$, with an input layer of m nodes and output one with $2K$ nodes (standing for the K-dimensional complex codeword), represent the SCMA mapping process for one single user, namely User j. Multiple DNNs are stacked together to form the SCMA DNN encoder, another DNN is employed as a SCMA detector.

2.5 Conclusions

In this Chapter, we presented the structure and basic principles of SCMA. Then, SCMA encoder and detector designs were reviewed through their most known techniques. In order to propose well targeted perspectives and well guided projections on this multiple access technique, we have reinforced this state-of-the-art by a simulations-based comparison study of different existing approaches of codebook design as well as of signal decoding and reception. These discussions highlighted that more work on irregular SCMA systems is needed (this problem will be studied in Chapter 3) and that some new propositions on deep-learning based SCMA detectors are required (this will be the subject of Chapter 4). On the other hand, we have proposed some comprehensive Tables, available in the Annexes of our thesis, that gather the majority of the codebook design techniques as well as the different SCMA detection methods. The referenced works were categorized according to some useful criteria. A complexity estimation of each technique was also elaborated. This will allow the reader to possess a well knowledge of the state-of-the-art. 4.

Chapter 3

Adaptive SCMA Scheme

Contents

3.1 Introduction **76**

3.2 Irregular SCMA **76**

3.2.1 Non-regular Structure 76

3.2.2 Review of Existing Irregular SCMA Schemes 77

3.3 Irregular SCMA Codebook Design **78**

3.3.1 Proposed Mother Constellation Design 78

3.3.2 Proposed Transformation Operators Design 82

3.4 Adaptation of Irregular SCMA **83**

3.4.1 Adaptation Scenario I : Different Quality of Service Levels . 85

3.4.2 Adaptation Scenario II : Fairness Among Users 87

3.4.3 Adaptation Scenario III : Higher Overall Throughput 90

3.5 Conclusions **91**

3.1 Introduction

As presented in the previous Chapter, in addition to the regular SCMA architecture, an irregular one was proposed in the state-of-the-art. However, it is important to highlight that most existing SCMA works are based on regular structure where users are treated equally. Nevertheless, this scheme can not be adapted to different users' business needs or conditions in realistic scenarios. In order to fix that issue, the irregular SCMA architecture is proposed.
In this Chapter, we will propose an adaptive design of SCMA codebooks based on an irregular architecture to fit some particular users' requirements for B5G networks.

3.2 Irregular SCMA

Several research works such as in [129]–[133] addressed how the irregular scheme of SCMA can be adapted to various users' needs in terms of rate, quality of service, priority and also, how different allocation mechanisms were exploited to achieve the fairness among users in the same network.

3.2.1 Non-regular Structure

As it was presented in the section 2.2, SCMA transmitter encodes the data bits of user j and maps them into a K-dimensional codeword, $\mathbf{x}_j^{(m)}$, only N_j entries of $\mathbf{x}_j^{(m)}$ are non-zero. The codebook of each user j, $\boldsymbol{C}_j$, is a constellation of M_j alphabets. In other words, M_j is the size of codebook and N_j is its sparsity degree.

Figure 3.2.1 illustrates the SCMA system under three scenarios. A regular SCMA is presented in Figure 3.2.1(a) where all users are spread over two REs ($N_j = 2, 1 \leq j \leq J$) and employs a codebook of size $M_j = 4, 1 \leq j \leq J$. The system can be described with the factor graph matrix $\mathbf{F}_{ss} = \left(\mathbf{f}_1^{(ss)}, \cdots, \mathbf{f}_J^{(ss)}\right)$ as given in Equation (3.1), it is obvious that all users have the same codebook sparsity degree.
The same factor graph matrix is valid when each user employs his specific constellation size M_j as shown in Figure 3.2.1(b). This scenario allows to serve users at different data rates. On the other hand, different sparsity degrees can be used such as each user sends $\log_2(M)$ bits over a specific number of subcarriers as depicted in Figure 3.2.1(c). An example of a factor graph with different sparsity degrees is given by $\mathbf{F}_{ds} = \left(\mathbf{f}_1^{(ds)}, \cdots, \mathbf{f}_J^{(ds)}\right)$,

$$\mathbf{F}_{ss} = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 \end{bmatrix} \quad , \quad \mathbf{F}_{ds} = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \end{bmatrix} \tag{3.1}$$

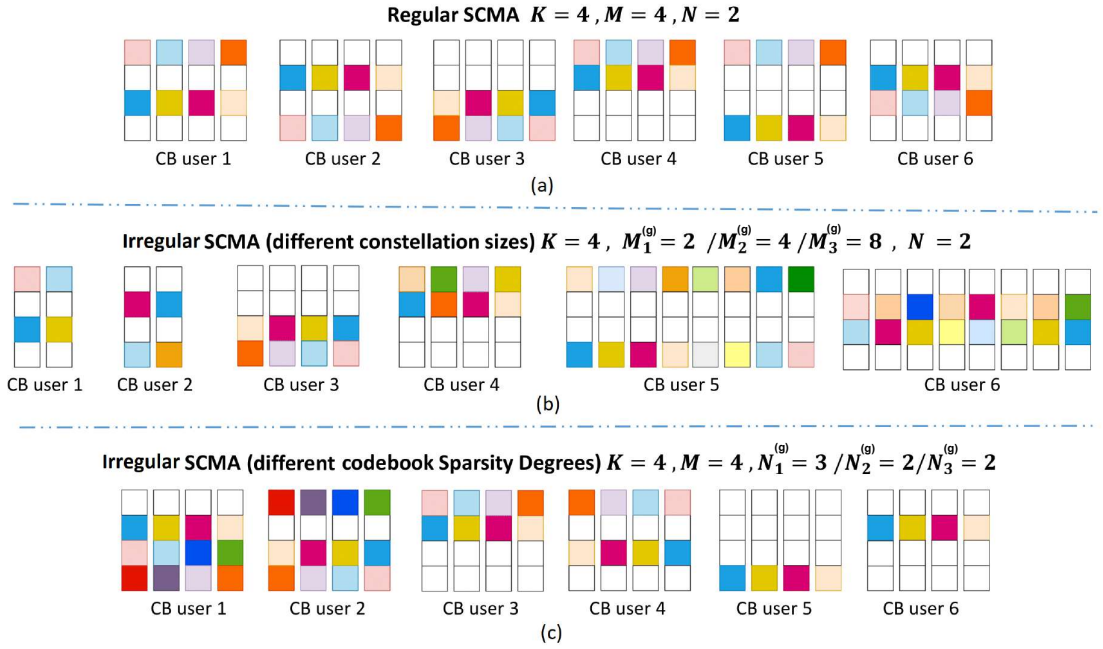


Figure 3.2.1: Presentation of the SCMA encoder for (a) regular system (b) irregular system with different constellation sizes (c) irregular system with different codebook sparsity degrees.

The codewords of all layers are then superimposed and transmitted over the K REs as it was introduced in Equations (2.1) and (2.2).

3.2.2 Review of Existing Irregular SCMA Schemes

Early in 2015, an irregular SCMA codebook design was proposed in [129] with the aim to assign different codebooks with various dimensions according to different users requirements. Simulations show that using an irregular codebook design does improve the system performance in terms of BER. However the authors did not take into consideration the impact of the correlation among users. One year later, another irregular SCMA design was proposed in [130], the idea was to employ different rotated angles to design different codebooks for several user needs, nevertheless the proposed codebooks are still far from being optimal.

Based on other approach and motivated by the same aim, the authors in [131] studied the resource allocation for different users in the same system by proposing a flexible resource scheduling scheme. Other contribution in [132] proposed an energy-saving algorithm for a joint codebook design and assignment, and power allocation for both uplink and downlink SCMA scenarios.

Further, in order to enhance SCMA mapping system, researchers in [133] proposed a more flexible uplink SCMA scheme by directly mapping a variable number of coded symbols from each user onto subcarriers. This proposition improved the overloading factor and

system performance at the expense of more complex signaling control.

3.3 Irregular SCMA Codebook Design

The steps to design a codebook for a regular SCMA system were presented in Chapter 2, these steps are maintained here. Hence, designing an irregular SCMA codebook requires to propose (a) a mother constellation design and (b) a transformation operators design. However, the design of irregular SCMA depends on which parameter represents its irregular character, this may concern the sparsity degree or the codebook size as shown in Figure 3.2.1. In the following subsections, our proposed design of mother constellations and transformation matrices of irregular SCMA will be introduced.

3.3.1 Proposed Mother Constellation Design

As in the regular case, our objective is to solve the following joint optimization problem, as it has been defined in Chapter 2,

$$\mathcal{V}^*, \mathcal{C}^* = \arg\max_{\mathcal{V},\mathcal{C}} D\left(\phi_{\text{ir}}(\mathcal{V}, \mathcal{C}; J, \{M_j\}, \{N_j\}, K)\right) \tag{3.2}$$

where D is a design criterion and ϕ_{ir} is the irregular SCMA system as described in the current chapter. Hence, our aim is to find the optimum user-to-RE mapping matrix $\mathcal{V}^*$ and the optimum multi-dimensional constellation $\mathcal{C}^*$.
The choice of the best design criterion is essential. Exiting design criteria, such as Euclidean distance, Euclidean kissing number, product distance, and others, were investigated in Chapter 2. Most existing criteria are applied directly to the mother constellation while hoping that the transformation operators will preserve its proprieties. Here, we will opt to employ a criterion that can be applied directly on the combination of codewords of different users, since this combination is the vector to be really transmitted such that we must be able to efficiently detect it. That is why we adopt the pairwise error probability between the transmitted codeword, $\mathbf{z}_t \in \mathcal{CC}$, and the detected codeword, $\mathbf{z}_d \in \mathcal{CC}$, which is expressed as,

$$\mathbb{P}(\mathbf{z}_t, \mathbf{z}_d|\mathbf{H}) = Q\left(\sqrt{\frac{\|\mathbf{H}\left(\mathbf{z}_t - \mathbf{z}_d\right)\|^2}{2N_0}}\right) \tag{3.3}$$

where $\mathbf{H}$ is the channel matrix as presented in section 2.3, and $\mathcal{CC}$ is the codewords combination set which is given by,

$$\mathcal{CC} = \left\{\mathbf{z} \in \mathbb{C}^K \text{ where } \mathbf{z} = \sum_{j=1}^{J} \mathbf{x}_j^{(m_j)}, \mathbf{x}_j^{(m_j)} \in \boldsymbol{C}_j, 1 \leq m_j \leq M, 1 \leq j \leq J\right\}, \tag{3.4}$$

the cardinality of $\mathcal{CC}$ is M^J. Then, the error rate when $\mathbf{z}_i$ is transmitted can be given by

$$\mathbb{P}\left(\mathbf{z}_i|\mathbf{H}\right) = \sum_{\substack{j=1 \\ i\neq j}}^{M^J} \mathbb{P}\left(\mathbf{z}_i, \mathbf{z}_j|\mathbf{H}\right) \tag{3.5}$$

Our aim is to minimize the error rate, $\mathbb{P}\left(\mathbf{z}_i\right), 1 \leq i \leq M^J - 1$. Therefore, the design criterion D, as defined in 3.2, is given by,

$$D = 1 - \mathbb{P}\left(\mathbf{z}_i|\mathbf{H}\right) \tag{3.6}$$

Moreover, for the AWGN channel, the pairwise error probability in (3.3) is reduced to,

$$\mathbb{P}(\mathbf{z}_t, \mathbf{z}_d|\mathbf{H}) = Q\left(\sqrt{\frac{\mathrm{d}_E^2\left(\mathbf{z}_t, \mathbf{z}_d\right)}{2N_0}}\right) \tag{3.7}$$

Hence, maximizing the design criterion D can be seen as maximizing the minimum square Euclidean distance among the points of the codewords combination set, $\mathcal{CC}$.
A ring-based approach, APSK, was proposed in [100] where each ring is composed of uniformly spaced PSK points. The star-QAM constellation adds different phases between adjacent rings compared to APSK [98], [99]. This makes the constellation points more uniformly distributed in order to ensure a higher minimum distance between them. However, the inter-ring ratio must be also optimized in the case of star-QAM as in the case of APSK. Our below-presented proposition is inspired by the work in [99] and extends it to the case of irregular SCMA.
We have two distinct cases, namely systems with different codebook sparsity degrees and those with different constellation sizes.

Case 1 : Irregular SCMA With Different Sparsity Degrees

Firstly, we study a SCMA system with three distinct groups of users as presented in Figure 3.2.1(c). That is, the idea that each group will be characterized with a different *sparsity degree* and all of them use equal-sized codebooks. For instance, $N_1^{(g)} = 3$, $N_2^{(g)} = 2$, $N_3^{(g)} = 1, M = 4$ in the system to be studied in the following.
The mother constellation of regular star-QAM design is given by the following $N \times M$ matrix,

$$\boldsymbol{C}_{mc,\text{star}} = \begin{bmatrix} \alpha R_1^{(d)} & R_1^{(d)} & -R_1^{(d)} & -\alpha R_1^{(d)} \\ -R_2^{(d)} & \alpha R_2^{(d)} & -\alpha R_2^{(d)} & R_2^{(d)} \end{bmatrix} \tag{3.8}$$

Here, users have different N values, that is why a mother constellation of size $N = \max(N_1, \cdots, N_J)$ will be designed. However, only N_j dimension are employed to generate the codebook $\boldsymbol{C}_j$ of user j since the mother constellation is multiplied with the mapping

matrix, $\mathbf{V}_j$, as expressed in Equation (2.5). The adopted factor graph matrix in this case is $\mathbf{F}_{ds}$ as defined in (3.1). For instance, we have,

$$\mathbf{V}_1 = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \mathbf{V}_3 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \quad \mathbf{V}_5 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

The employed mother constellation for irregular SCMA with different sparsity degrees is expressed as,

$$\ddot{\boldsymbol{C}}_{mc} = \begin{bmatrix} \alpha R_1^{(d)} & R_1^{(d)} & -R_1^{(d)} & -\alpha R_1^{(d)} \\ -R_2^{(d)} & \alpha R_2^{(d)} & -\alpha R_2^{(d)} & R_2^{(d)} \\ \alpha R_3^{(d)} & R_3^{(d)} & -R_3^{(d)} & -\alpha R_3^{(d)} \end{bmatrix} \tag{3.9}$$

where $R_2^{(d)} = \beta R_1^{(d)}$ and $R_3^{(d)} = \beta R_2^{(d)}$. Hence, the mother constellation is defined by three parameters, α, β and the radius of the inner ring, $R_1^{(d)}$. α denoted the power scaling effect that is specific for each group and β denoted the interleaving which define the phase between the different dimensions. That is, each dimension is generated from the previous one by adding an interleaving and power scaling effect such as inter-layer interference can be eliminated more easily.

Here, the work in [99] is extended by adopting a more powerful approach to optimize the mother constellation of our irregular SCMA scheme. We consider $\alpha, \beta \in \mathbb{C}$ in contrast to [99] which allows to have some inter- and intra-dimensional rotations in addition to the power variation, thanks to the complex pattern of α and β. The optimization consists of finding the values of α and β which maximizes the codebook design criterion.

The average energy per symbol of employed 6-rings constellation of irregular SCMA when $N_1^{(g)} = 3$, $N_2^{(g)} = 2$ and $N_3^{(g)} = 1$ is expressed as,

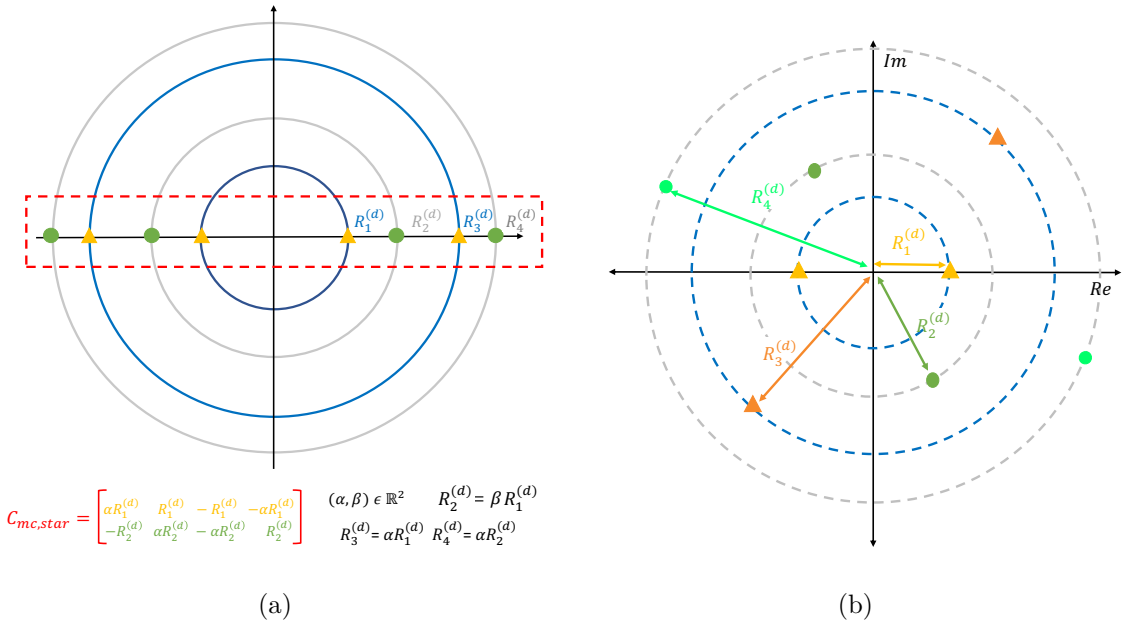
$$\ddot{E} = \frac{(R_1^{(d)})^2 \left(1 + |\alpha|^2\right) \left(3 + 2|\beta|^2 + |\beta|^4\right)}{6} \tag{3.10}$$

Hence, assuming that average energy of $\boldsymbol{C}_{mc}$ is fixed to $\ddot{E} = 1$, the radii of the inner ring, $R_1^{(d)}$, is calculated as following,

$$R_1^{(d)} = \sqrt{\frac{6}{\left(1 + |\alpha|^2\right) \left(3 + 2|\beta|^2 + |\beta|^4\right)}} \tag{3.11}$$

Case 2 : Irregular SCMA With Different Constellation Sizes

Secondly, we study a SCMA system with three distinct groups of users as presented in Figure 3.2.1(b). That is, the idea that each group will be characterized with a different *codebook size* and all of them use the same sparsity degree. For instance, $M_1^{(g)} = 2, M_2^{(g)} = 4, M_3^{(g)} = 8, N = 2$ in the system to be studied in the following.



(a) (b)

Figure 3.3.1: Four-rings star-QAM mother constellation for regular (a) and irregular (b) SCMA codebook design of size $M = 4$ and sparsity degree $N = 2$ when α and β are respectively reel and complex.

For the case of adaptive SCMA system with different constellation sizes, the same approach is used but it must be adapted such that we have a distinct mother constellation for each value of M, these mother constellations are denoted $\breve{\boldsymbol{C}}_{m,1} \in \mathbb{C}^{N \times M_1^{(g)}}, \breve{\boldsymbol{C}}_{m,2} \in \mathbb{C}^{N \times M_2^{(g)}}, \breve{\boldsymbol{C}}_{m,3} \in \mathbb{C}^{N \times M_3^{(g)}}$. These matrices are given by,

$$\breve{\boldsymbol{C}}_{mc,1} = \begin{bmatrix} R_1^{(d)} & -R_1^{(d)} \\ -R_3^{(d)} & R_3^{(d)} \end{bmatrix}$$

$$\breve{\boldsymbol{C}}_{mc,2} = \begin{bmatrix} \alpha R_1^{(d)} & R_1^{(d)} & -R_1^{(d)} & -\alpha R_1^{(d)} \\ -R_2^{(d)} & \alpha R_2^{(d)} & -\alpha R_2^{(d)} & R_2^{(d)} \end{bmatrix}$$

$$\breve{\boldsymbol{C}}_{mc,3} = \begin{bmatrix} \alpha R_3^{(d)} & \alpha R_1^{(d)} & R_3^{(d)} & R_1^{(d)} & -R_1^{(d)} & -R_3^{(d)} & -\alpha R_1^{(d)} & -\alpha R_3^{(d)} \\ -R_4^{(d)} & -R_2^{(d)} & \alpha R_4^{(d)} & \alpha R_2^{(d)} & -\alpha R_2^{(d)} & -\alpha R_4^{(d)} & R_2^{(d)} & R_4^{(d)} \end{bmatrix} \tag{3.12}$$

where $R_2^{(d)} = \beta R_1^{(d)}, R_3^{(d)} = \beta R_2^{(d)}$ and $R_4^{(d)} = \beta R_3^{(d)}$. In this case, the factor graph matrix $\mathbf{F}_{\text{ss}}$ as defined in (3.1) is used.

The average energy per symbol of employed 8-rings constellation of irregular SCMA when $M_1^{(g)} = 2, M_2^{(g)} = 4$ and $M_3^{(g)} = 8$ is expressed as,

$$\breve{E} = \frac{(R_1^{(d)})^2 [\,4\,(1 + |\beta|^4) + (1 + |\alpha|^2)\,(3 + 3|\beta|^2 + |\beta|^4 + |\beta|^6)]}{12} \tag{3.13}$$

Hence, assuming that average energy of $\boldsymbol{C}_{mc}$ is fixed to $\breve{E} = 1$, the radii of the inner ring, $R_1^{(d)}$, is calculated as following,

$$R_1^{(d)} = \sqrt{\frac{12}{4\left(1+|\beta|^4\right) + \left(1+|\alpha|^2\right)\left(3+3|\beta|^2+|\beta|^4+|\beta|^6\right)}} \tag{3.14}$$

Consequently, once the complex values of α and β are found, the mother constellation is completely calculated.
Hence, our optimization objective is reduced into finding the adequate α and β values that maximize the minimum square Euclidean distance of the combination set $\mathcal{CC}$. The Algorithm 2 proposed bellow in the two aforementioned cases. Obviously, codebooks for all users must be calculated before we can calculate $\mathcal{CC}$.

3.3.2 Proposed Transformation Operators Design

The last step of the codebook design is to find the optimal transformation operators. Here, we consider that scaling and interleaving operations are already included in the mother constellation design. Therefore, the only transformation that we will apply on the mother constellation, to generate the codebook of each user, is the rotation. Most existing works in the literature employ the typical rotation angles in [84], [130].
As for our proposition, an optimized set of user-specific rotation matrices is adopted. The n^{th} entry of the $N_j \times N_j$ diagonal rotation matrix for user j is defined as,

$$[\mathbf{T}_j]_{n,n} = e^{j\theta_{j,n}}, \tag{3.15}$$

where $\theta_{j,n} \in [0, \pi[$ since the different dimensions of each mother constellation are symmetric. In order to simplify the optimization, quantized angles can be employed to represent the semi-circle, i.e. $\theta_{j,n}$ can be obtained from a uniform grid as following,

$$\theta \in \left\{ \frac{i\pi}{N_\theta}; 0 \leqslant i \leqslant N_\theta - 1 \right\} \tag{3.16}$$

when N_θ is a design parameter.
It is worth mentioning that for the system proposed in Fig.3.2.1(c), a total number of distinct rotation angles of $2\left(N_1^{(g)} + N_2^{(g)} + N_3^{(g)}\right)$ is needed. However, for the system with different constellation sizes, we assign two rotation angles for each dimension with eight constellation points which is considered as a combination of two vectors with four constellation points. This makes the total number of distinct rotation angles of $2\left(M_1^{(g)} + M_2^{(g)} + 2M_3^{(g)}\right)$.
In Algorithm 3, we propose a numerical search algorithm to assign the optimal rotation

Algorithm 2: Optimization of mother constellation design

Input: SCMA system parameters: $J, M, N, K, \mathbf{F}$
Initialize: $\alpha = r_\alpha e^{j\theta_\alpha} = 1, \beta = r_\beta e^{j\theta_\beta} = 1$
$\mathbf{T}_j$ are first assigned successively
$\mathbf{V}_j$ are calculated from factor graph matrices as expressed in Equation (3.1)
while $\theta_\alpha \leq \pi$ **do**
  while $r_\alpha \leq \alpha_{max}$ **do**
    while $\theta_\beta \leq \pi$ **do**
      while $r_\beta \leq \beta_{max}$ **do**
        $\boldsymbol{C}_{mc}$ in given in Equations (3.9) and (3.12) in case 1 and case 2 respectively
        For $j = 1 \cdots J, \boldsymbol{C}_j = \mathbf{V}_j \mathbf{T}_j \boldsymbol{C}_{mc}$
        Compose the codewords combination set, $\mathcal{CC}$ as defined in 3.4
        Compute, d_{min}, the minimal distance between the points of $\mathcal{CC}$
        if *it is currently the maximum value of D in (3.6)* **then**
          save α and β
        end
        set the modulus value of r_β:
        $r_\beta = r_\beta + \Delta r$
      end
      set the phase value of θ_β:
      $\theta_\beta = \theta_\beta + \Delta\theta$
    end
    set the modulus value of r_α:
    $r_\alpha = r_\alpha + \Delta r$
  end
  set the phase value of θ_α:
  $\theta_\alpha = \theta_\alpha + \Delta\theta$
end
PS: $\alpha_{\max}$ and $\beta_{\max}$ are the maximum modulus values within the optimization process. $\Delta\theta$ and Δr are the search steps for θ and r, they are set to $\frac{\pi}{24}$ and 0.1 respectively.
Output: $\breve{\boldsymbol{C}}_{\boldsymbol{m},\{i|1\leq i\leq 3\}}$ or $\ddot{\boldsymbol{C}}_{\boldsymbol{m}}$ based on the optimized values of α and β

angle for each dimension of each user. This requires to find the optimal values of α and β first which itself requires to know the rotation matrices. To solve this problem the rotation angles are assigned successively in the first step, α and β are optimized before optimizing the rotation angles in the second step as shown in Figure 3.3.2.

3.4 Adaptation of Irregular SCMA

We have proposed in the previous section a new method to design irregular SCMA codebooks. In this section, we will present how we could adapt the structure of irregular SCMA in order to match the user requirement. Unlike the regular SCMA scenarios, for each need, a different structure of adaptive SCMA can be employed. For instance, it is not ideal to equally treat the users when they have different channel states, this may lead

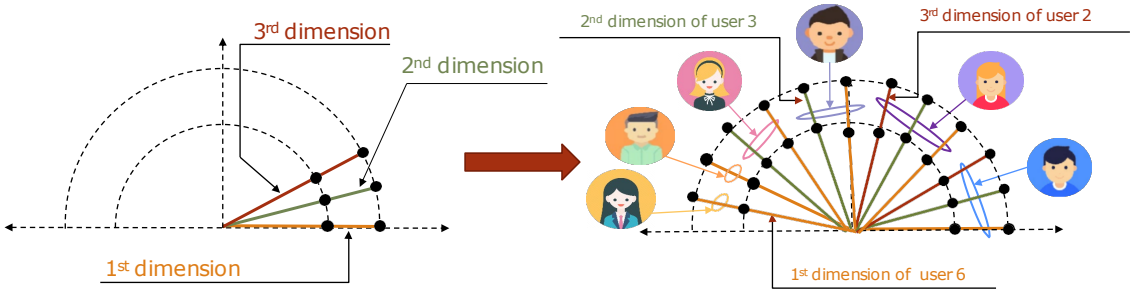


Figure 3.3.2: Rotation angles of different users are initialized to be assigned successively as described in Step 1 of Algorithm 3. Here, in order to simplify the illustration we considered a special case of two rings with no phase interleaving ($\alpha = 1$) and β of modulus 1.)

Algorithm 3: Optimization of the transformation operations

Input: J users, N_θ rotation angles

$$\begin{cases} N_\theta = 2\left(N_1^{(g)} + N_2^{(g)} + N_3^{(g)}\right) & \text{(in Case 1)} \\ N_\theta = 2\left(M_1^{(g)} + M_2^{(g)} + 2M_3^{(g)}\right) & \text{(in Case 2)} \end{cases}$$

Step 1: Phase assignment initialization The angles of the dimensions of the J users are initialized successively as presented in Figure 3.3.2, i.e.,

$\theta_{j,n}^{(\text{init})} - \theta_{j,n-1}^{(\text{init})} = \frac{\pi}{N_\theta}, 2 \leq n \leq N_j, 1 \leq j \leq J$, where
$\theta_{1,1}^{(\text{init})} = 0, \theta_{j,1}^{(\text{init})} = \theta_{j-1,N_j}^{(\text{init})} + \frac{\pi}{N_\theta}, 2 \leq j \leq J.$

Step 2: Phase assignment optimization

1. Optimize α and β based on Algorithm 2, these values are used in the following.
2. A numerical search algorithm is used to find Optimized $\theta_{j,n}$ of each dimension n, $1 \leq n \leq N_j$, of each user j, $1 \leq j \leq J$.
3. Compute the optimized rotation matrix $[\mathbf{T}_j]_{n,n} = e^{j\theta_{j,n}}$

Output: $\mathbf{T}_j$, $1 \leq j \leq J$

to poor QoS for users with bad channel conditions. In our work [134], differently from [129]–[133], the knowledge of the channel state of each user was exploited to improve the total performance of the system by adjusting either the number of REs to be allocated or the size of the codebook of each user.

Hence, the proposed adaptive SCMA design is more realistic, the idea is to divided users into different groups such that users belonging to the same group have the same needs and will be associated to the same resources. In this section three different scenarios will be studied:

(a) **Adaptation scenario I**

Description : all users have the same channel states but different needs in terms of delay, bandwidth and packet-delivery,

Objective : adjust the QoS according to user's needs.

(b) **Adaptation scenario II**:
Description : users have different channel states,
Objective : achieve more fairness among users.

(c) **Adaptation scenario III**:
Description : users have different channel states,
Objective : achieve higher overall throughput.

3.4.1 Adaptation Scenario I : Different Quality of Service Levels

In this scenario, we assume that all users have the same channel states but different needs in terms of delay, bandwidth and packet-delivery, our aim is to adapt irregular SCMA to reply to these needs.
Referred as QoS, the quality of service is a fundamental key performance indicator for wireless service over the 5G and beyond networks. Herein, we try to study how this indicator can be employed to design our SCMA codebooks in association of a variety of needs of SCMA users. The idea is that the QoS can be related directly to some of the system parameters. When considering for instance the amount of data in the same frame, if the QoS is important, the user is receiving and/or sending more of his data in an effective manner. It is to say that when using a simultaneous connection, one user can be proven against other ones, so we have to figure out how to let the data of this user dominate the rest of the data exchanged in a single frame as shown in Figure 3.4.1.
For 5G network, the QoS is enforced based on flows of QoS [136]. Where each QoS flow packets are classified and marked using QoS flow identifier (QFI). Then, the 5G QoS flows are mapped in the access network to the data radio bearers.
We consider in this thesis some of the QoS parameters defined by 3GPP as part of QoS implementation for 5G NR in 2018 [137]. We are interested in (i) the end-to-end delay, (ii) bandwidth and (iii) and packet-delivery ratio. We denote $P_{QoS}^{(j)}$ the QoS parameters

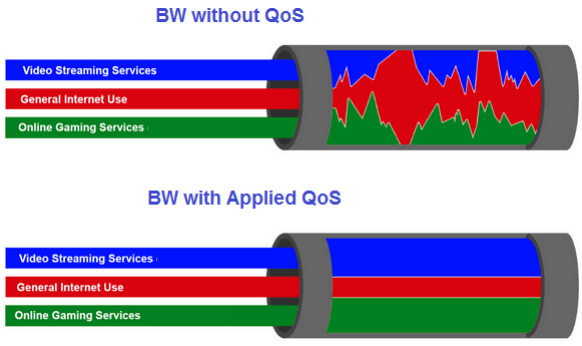


Figure 3.4.1: Role of QoS in de-cluttering flows [135]

Algorithm 4: Sparsity degree assignment according to the QoS parameters

Input: $P_j^{(QoS)}, \quad \forall 1 \leq j \leq J$

$P_j^{(QoS)} = \{\frac{1}{\text{Delay}_j}, \text{Bandwidth}_j, \text{Packet-Delivery}_j\}$

$N_{\text{QoS}} = |P_j^{(QoS)}| = 3$; $|\cdot|$ denote the cardinality of the set.

Step 1: Assign weights for QoS parameters

$w_j^{(i)}$ is the weight or the relative importance of each parameter $q_j^{(i)} \in P_j^{(QoS)}$ associated with each user $j, 1 \leq j \leq J$.

$\text{Sum}_i = \sum_{j=1}^{J} q_j^{(i)}, i, 1 \leq i \leq N_{\text{QoS}}$

Compute $w_j^{(i)} = \frac{q_j^{(i)}}{\text{Sum}_i}, j, 1 \leq j \leq J, i, 1 \leq i \leq N_{\text{QoS}}$

Step 2: Calculate QoS importance for each user

The relative QI of each user,

$\text{QI}_j = \sum_{i=1}^{N_{\text{QoS}}} w_j^{(i)}, j, 1 \leq j \leq J$

Sort users in descending order according to their relative QoS importance, the sorted users set is denoted U_{QoS}

Step 3: Sparsity degree assignment

Divide U_{QoS} into $N_G = 3$ groups, each group is of cardinally $\lfloor \frac{J}{N_G} \rceil$ (where $\lfloor . \rceil$ denoted the nearest integer)

$$N_j = \begin{cases} 3; \ ; & \text{for the group formed of the first } \lfloor \frac{J}{N_G} \rceil \text{ entries in } U_{\text{QoS}} \\ 1; \ ; & \text{for the group formed of the last } \lfloor \frac{J}{N_G} \rceil \text{ entries in } U_{\text{QoS}} \\ 2; \ ; & \text{for the group formed of the resting } J - 2\lfloor \frac{J}{N_G} \rceil \end{cases}$$

Output: $N_j, j, 1 \leq j \leq J$

set of each user $j, 1 \leq j \leq J$.

The QoS parameters must allow us to decide if a given user must be served with better QoS relatively to the other ones. Hence, the objective of our new proposed approach is to calculate the relative QoS importance of each user such that the SCMA scheme can be adapted accordingly. The idea is to first compute a specific weight to each parameter in order to measure its relative importance for each user, in other words, this weight judge which importance must be accorded to each user when compared to the other ones on the basis of this parameter only. Thereafter, the relative QoS importance (QI) of each user, QI_j, is calculated by accumulating its relative importance on each parameter. Sorting the users based on their QI_j allows to divide them into three groups. The first group represents the users that we want to emphasize since they have the higher relative QoS importance, while the third group is the one in which users have lesser requirements in terms of delay,

bandwidth and/or packet-delivery ratio. As for the second group, we represent the users with no-specific needs. Finally, our proposed strategy is to assign a specific *sparsity degree* $\left(N_1^{(g)}, N_2^{(g)} \text{ or } N_3^{(g)}\right)$ to each group. The idea behind this strategy is that more sparse is the codebook of a given user, less the data of this user are presented in the combined codeword to be transmitted, and consequently less its communication link is efficient. Hence, it is reasonable to think that this choice is more suitable to users accepting more end-to-end delay and/or less packet-delivery ratio. On the other hand, more resources (i.e. higher sparsity degree) is reserved for users with higher relative QoS importance. This method is explained in Algorithm 4.

3.4.2 Adaptation Scenario II : Fairness Among Users

In this scenario, we assume users have different channel states, therefore, treating them in the same manner means that users with bad channel conditions will have lesser QoS when compared to users with better channel conditions. Our objective is to restore some fairness among users by allocating more resources to users with bad channel conditions. The users are divided into three groups, as shown in Figure 3.4.2, such that the users is the same group have a similar channel state namely bad, good or excellent. To reach our aim, we propose to assign a specific *sparsity degree* to each group, i.e. $N_1^{(g)} = 3$ for bad channels and $N_3^{(g)} = 1$ for excellent ones. This approach was inspired form the power-domain NOMA where power allocation (i.e. resources allocation) among users is based on the near-far propriety as was shown in Figures 1.3.4 and 3.4.3.
However, this method requires the knowledge of the CSI, hence, in the following paragraph, we will rapidly review some existing methods before choosing the one to be em-

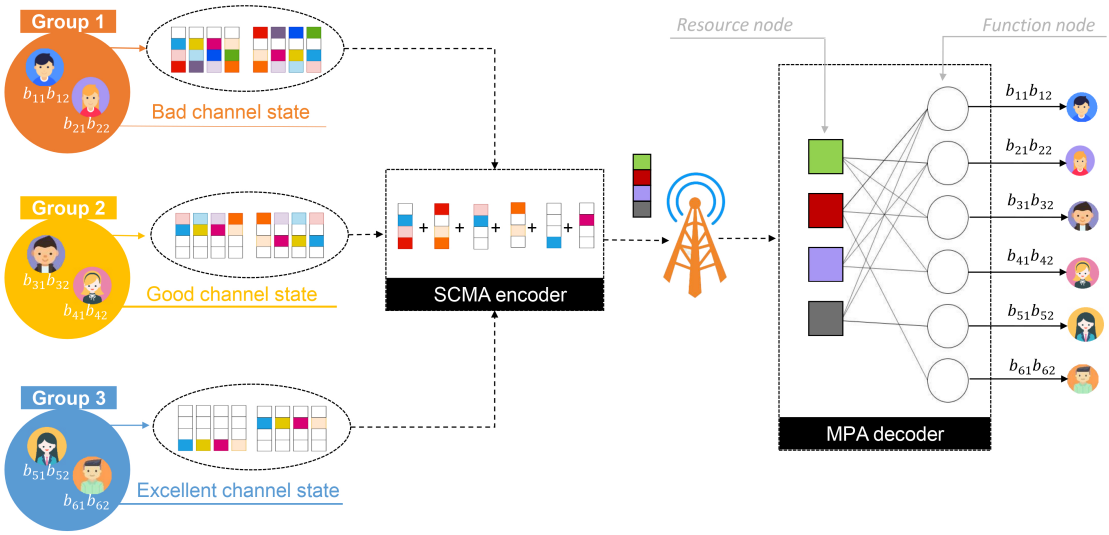


Figure 3.4.2: Users are divided into three groups based on their channel conditions. A specific sparsity degree is assigned to each group, i.e. $N_1^{(g)} = 3$ for bad channels and $N_3^{(g)} = 1$ for excellent ones.

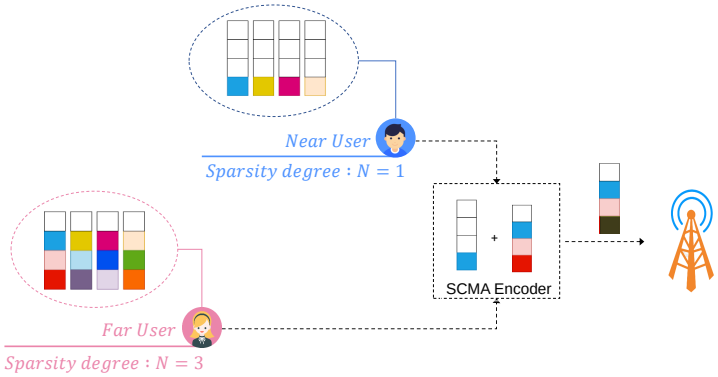


Figure 3.4.3: Irregular uplink SCMA system with near and far users using different sparsity degrees, $N = 1$ and $N = 3$ respectively, based on their channel states [134].

ployed in the rest of the current Chapter.

Existing SCMA Channel Estimation Methods

Existing channel estimation methods can be classified into three major categories [138]–[144]: pilot-aided estimation, data-aided estimation pilot, so known as semi-blind estimation, and blind estimation.
The pilot-estimation combines the knowledge of the transmitted and received signals to estimate the channel matrix $\mathbf{H}$ as defined in subsection 2.2.2. Then, the channel estimation is conducted either based on the *Least-square estimation* technique when the channel and noise distributions are unknown, or based on *MMSE estimator* when the channel and noise distributions are known. Moreover, data-aided estimation methods use and transmit data symbols in addition to some pilot sequences to estimate the channel.
It is proven that using data-aided algorithms increase the spectral efficiency more than the pilot-aided techniques when using orthogonal or non-orthogonal sequences for uncoded and turbo-coded channels. In [143], two channel estimation techniques were investigated: the pilot-estimation and the data-feedback-aided estimation algorithms, with the aim to enhance the spectral efficiency in uplink SCMA scenario.
As for the blind channel estimation, despite that it saves time-frequency resources, due to the lack of pilot signals, they have significant computational complexity and are less effective than classical approaches especially for SCMA systems as demonstrated in [144].
In the following, we will employ the data-aided estimation method proposed in [143] which is also described in Algorithm 5 of Appendix C.

Simulation Results

Here, we are assuming that users have not the same channel conditions. Thus, we are studying the SCMA performances when users have different channel states. We suppose that the $J = 6$ users can be classified into three groups such as the average SNR varies

from one group to another. That is why, we introduce the SNR level fluctuation (in dB) among the groups, which is denoted by δ, such that when the average SNR for all users is γ then the average SNR per groups is either $\gamma - \delta, \gamma$ or $\gamma + \delta$.

A performance evaluation of regular SCMA for different values of SNR level fluctuation

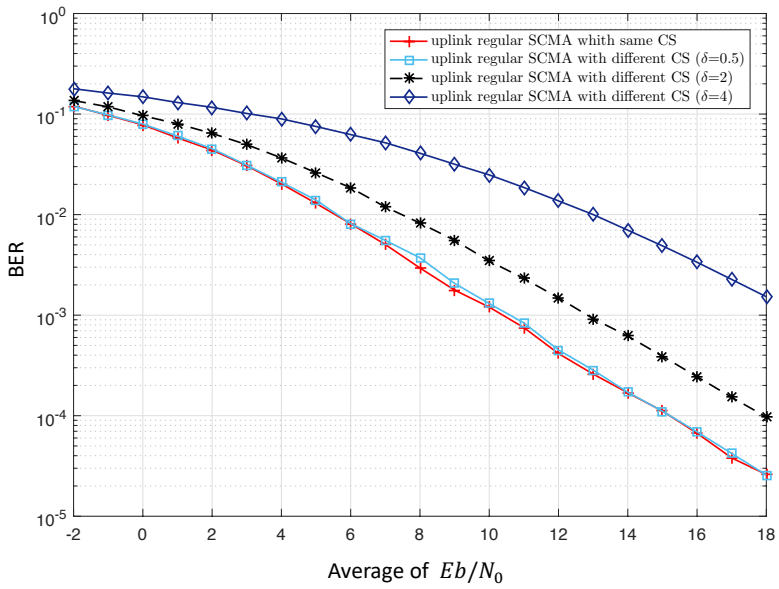


Figure 3.4.4: Performance evaluation of uplink SCMA in [84] when users have different channel states (CSs) for different values of SNR level fluctuation $\delta = 0.5, \delta = 2, \delta = 4$

is shown in Figure 3.4.4. Obviously, the BER increases when δ increases, for instance the BER is 5 and 50 times higher for $Eb/N_0 = 15$dB when $\delta = 2$ and $\delta = 4$. Hence, it is recommended to design an adaptive SCMA codebooks which take into consideration the variable nature of the state of channel for each user.

For the rest of simulations in this Chapter, $\delta = 2$ is considered.

The idea here is to assign a different sparsity degree to each group depending on its average channel condition which could be practically proportional to the average distance between users and the base station as shown in Figure 3.4.3.

The BER performance of our proposed adaptive SCMA, as described in Figure 3.2.1(c), is compared with the regular SCMA, as described in Figure 3.2.1(a), where the adaptive sparsity degrees are $N_1^{(g)} = 3, N_2^{(g)} = 2, N_3^{(g)} = 1$ and that of regular SCMA is N=2 and both codebook sizes are equal to M= 4. The design of irregular SCMA codebooks to be employed in this scenario was introduced earlier in 3.3.1. In Figure 3.4.5, we compare the performance of regular SCMA with SCMA based on the optimized codebooks as proposed in this thesis. Despite the fact that regular uplink SCMA (red solid line) outperforms the proposed adaptive SCMA (blue solid line) when all users have the same channel state, it is the opposite when different channel states are taken into consideration, that is the performance of regular SCMA degrades and that of the adaptive SCMA becomes better. Clearly in Figure 3.4.5, the BER performance of the optimized adaptive SCMA codebook

(green dashed line) is better than that of regular uplink SCMA (black dashed line) through Rayleigh fading channels, e.g. for a BER of 10^{-4}, a gain of almost 3 dB is achieved. This can be explained by the fact that we allocate more REs ($N_1^{(g)} = 3$) for users with the worst channel conditions.

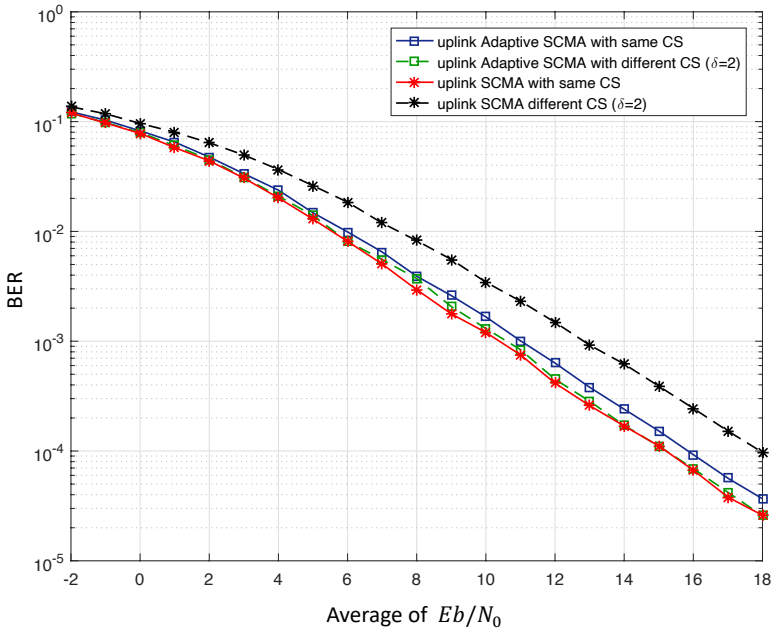


Figure 3.4.5: Performance comparison between regular uplink SCMA and adaptive uplink SCMA with different sparsity degrees ($N_1^{(g)} = 3, N_2^{(g)} = 2, N_3^{(g)} = 1$) when the different channel states are taken into consideration.

3.4.3 Adaptation Scenario III : Higher Overall Throughput

The idea of codebooks with different constellation sizes was inspired from the adaptive digital modulation where the size of the constellation increases when the channel is better and vice versa. Different order modulations allow you to send more bits per symbol and thus achieve higher throughputs or better spectral efficiencies. Following the same reasoning, we propose to adapt the codebook size of each user according to its channel state. However, it must also be noted that when using a larger size of codebook, SNRs are needed to overcome any inter-codeword interference and maintain a BER. The use of adaptive SCMA allows a wireless system to choose the highest codebook size depending on the channel conditions.

Here, we consider having three groups of users with three different channel states, hence the two codebooks with 8 codewords are assigned to the group with the better channel conditions, and the two codebooks with 2 codewords are assigned to the group with the worst ones. BER performance of the proposed adaptive SCMA with different constellation sizes is compared to the regular SCMA is Figure 3.4.7. Our proposal provides better

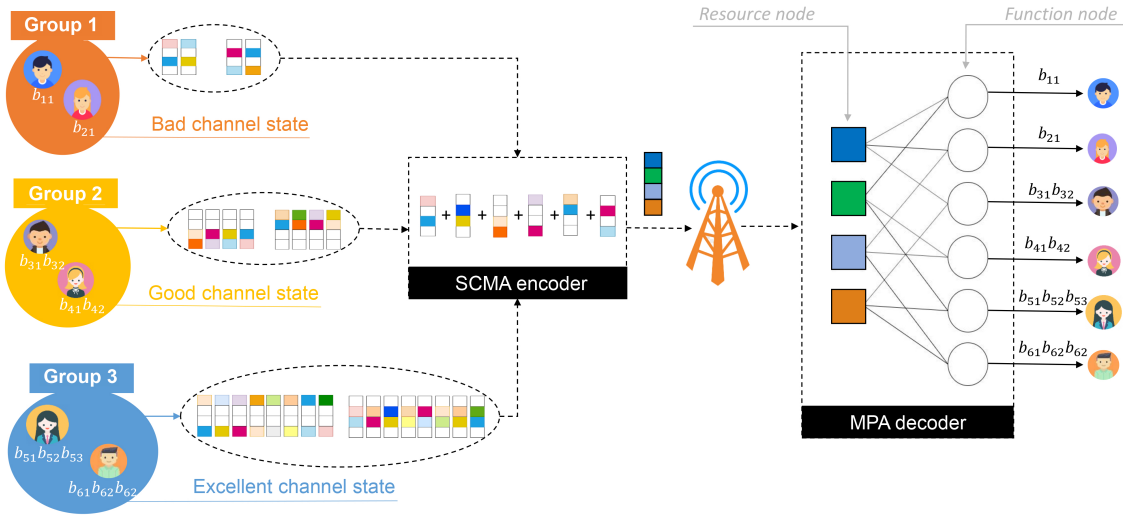


Figure 3.4.6: Users are divided into three groups based on their channel conditions. A specific codebook size is assigned to each group, i.e. $M_1^{(g)} = 2$ for bad channels and $M_3^{(g)} = 8$ for excellent ones.

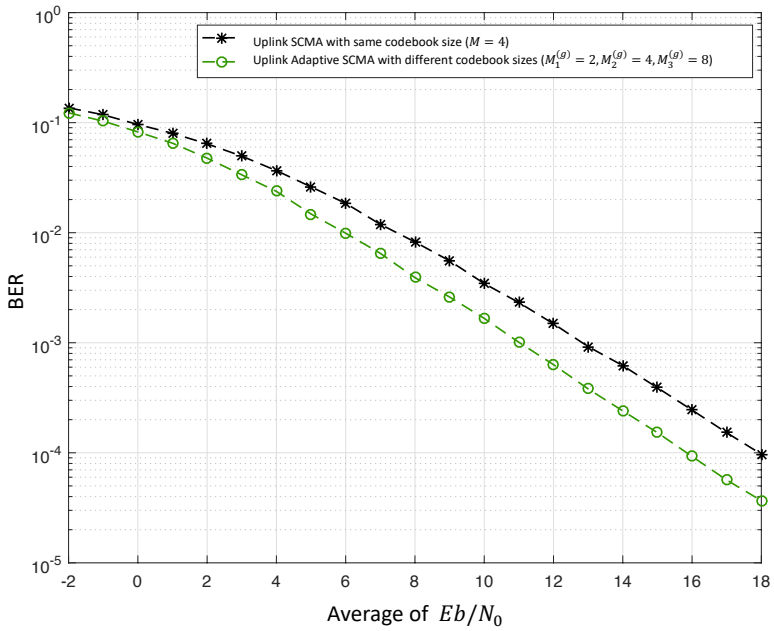


Figure 3.4.7: Performance comparison between regular uplink SCMA and adaptive uplink SCMA with different constellation sizes ($M_1^{(g)} = 2, M_2^{(g)} = 4, M_3^{(g)} = 8$) when the different channel states are taken into consideration.

performance, for instance a gain of 2 dB is achieved when BER is of 10^{-4}.

3.5 Conclusions

This Chapter was dealing with propositions on the irregular SCMA architecture. First, we proposed a new design of irregular SCMA in two cases: with different sparsity degrees

and with different codebook sizes. Then, the designed codebook were employed to adapt SCMA in three scenarios: (I) all users have the same channel states but different needs in terms of delay, bandwidth and packet-delivery, (II) users have different channel states, more fairness among users was achieved, and (III) users have different channel states, higher overall throughput was achieved. Simulation results confirms that the different objectives were achieved.

Chapter 4

Machine Learning for SCMA Detector and Channel Denoiser

Contents

4.1 Introduction **94**

4.2 Motivation and Objectives **94**

4.3 Machine Learning Configuration **95**

4.3.1 Employed ML's Techniques 95

4.3.2 Training and Hyper-Parameters 96

4.4 Joint Denoiser and Detector of DL-based SCMA (DAE-DNN-SCMA) **99**

4.4.1 Motivation behind the Proposed Solution 99

4.4.2 Denoising SCMA Signals over AWGN Channel 100

4.4.3 Deep Neural Network based SCMA Decoder 104

4.5 Fair SCMA Detector based Deep Learning (Fair-DNN-SCMA) **107**

4.5.1 Proposed Solution 108

4.5.2 Pre-processing 109

4.6 Analysis of Results **109**

4.6.1 Impact of Training Data with Different Noise Intensity Levels 110

4.6.2 Impact of Hyper-Parameters 112

4.6.3 Performance Evaluation of Proposed Detectors 113

4.6.4 Complexity Study of the Proposed Models 115

4.7 Conclusions **115**

4.1 Introduction

Field of study in artificial intelligence, ML is devoted to data analysis with an aim to create knowledge, in an automatic way from data, by training a model. This model can then be exploited on new data to make decisions.
We present in this Chapter our strategy to apply and use machine learning, mainly the deep learning techniques, in order to enhance the performance of the SCMA detector in an effective manner. We start by presenting the tools and mechanisms that we employed in the remainder of this Chapter. Later, we explain our proposed architectures and we analyze the several results comparing our systems with a conventional SCMA detector, such as the MPA one.

4.2 Motivation and Objectives

Detecting SCMA signals, with the help of deep learning methods, is still very challenging because of the presence of noise and fading channels. In addition, available works on SCMA based on machine learning do not take into consideration the variation of channel in realistic scenarios [75], [76], [78]–[80], [145], [146].
Motivated by this, we will try in this Chapter to help advancing this domain. The major contributions of this Chapter can be summarized as follows,

1. A joint denoising and decoding approach for designing SCMA detector is proposed. First, a denoising autoencoder is trained to mitigate the effect of AWGN by trying to remove it from received signal. The input of the DAE is the noisy SCMA signal, while the output is an estimation of the original noise-free one, i.e. the DAE aim is to reconstruct the transmitted data by learning their representation. Then, a DNN is adopted to learn how to detect the transmitted bits from the denoised SCMA signal. The two blocks of the DAE-DNN detector are jointly trained using an end-to-end objective function.

2. In fact, codebook and channel state information are supposed to be known at MPA-based receiver. Hence, previous comparisons of DL-based SCMA detectors with MPA are not fair. Here, the knowledge of SCMA codebook is integrated in the detection process by exploiting the distances between the received vector and each one among possible superimposed combination codewords. This will allow firstly to estimate the nearest superimposed codeword vector, the chosen vector will be the input of a DNN which objective is to decode the transmitted users' bits. The idea is that the knowledge of the codebook results in the knowledge of the set of vectors to which belongs the noise-free received vector, such that the corrupted received superimposed codeword is replaced by an estimated noise-free one.

4.3 Machine Learning Configuration

To fit the objectives mentioned in the previous section, in this PhD work, specific machine learning architectures and techniques are employed. In this section, we start by presenting the several used techniques, then, our strategy of adjusting the hyper-parameters of the neural network is explained.

4.3.1 Employed ML's Techniques

In the following parts of this Chapter, we propose two methods to detect a SCMA signal based on deep learning: (i) a joint denoiser and detector, denoted DAE-DNN-SCMA, will be illustrated in section 4.4, and (ii) a fair SCMA detector based deep learning, denoted as Fair-DNN-SCMA, will be highlighted in section 4.5 of the current Chapter. For each system we proposed and train different models: a denoising autoencoder and deep neural network .

Denoising Autoencoder

Generally, an AE is defined as a neural network trained in an unsupervised way in order to attempt copying inputs to outputs [48]–[51]. The key feature of this bottleneck, as presented in Figure 4.3.1, is that they teach themselves how to compress data from the input layer into a shorter representation, and then uncompress that representation into whatever format that best matches the original input. When using the back-propagation technique [147], the unsupervised algorithm used by an AE, continuously trains itself by setting the target output values, so-called labels, to equal the inputs. In order to help AEs to better learn data and to avoid overfitting, some noise is artificially added to the original data which introduces a new variant of AEs, this variant is called denoising autoencoder. Two approaches can be employed to inject noise into data: either by adding a Gaussain noise or by randomly dropping out some inputs as shown in figure 4.3.1.
Hence, the small-sized hidden encoding layers are forced to use dimensional reduction to eliminate noise and reconstruct the inputs. Thus, the data denoising and the reduction of dimension for data visualization are two interesting practical applications of AEs. In this PhD work, we are especially interested in the firstly presented use of AE. Hence, a DAE is incorporated into our studied models.

Deep Neural Network

Existing DL based SCMA detecting methods generally use fully connected DNNs as proposed in [75], [76], [78]–[80], [145], [146]. Here, the first DAE based pre-processing (i.e. denoising) step is followed also by a structure of fully connected layers which form a DNN.

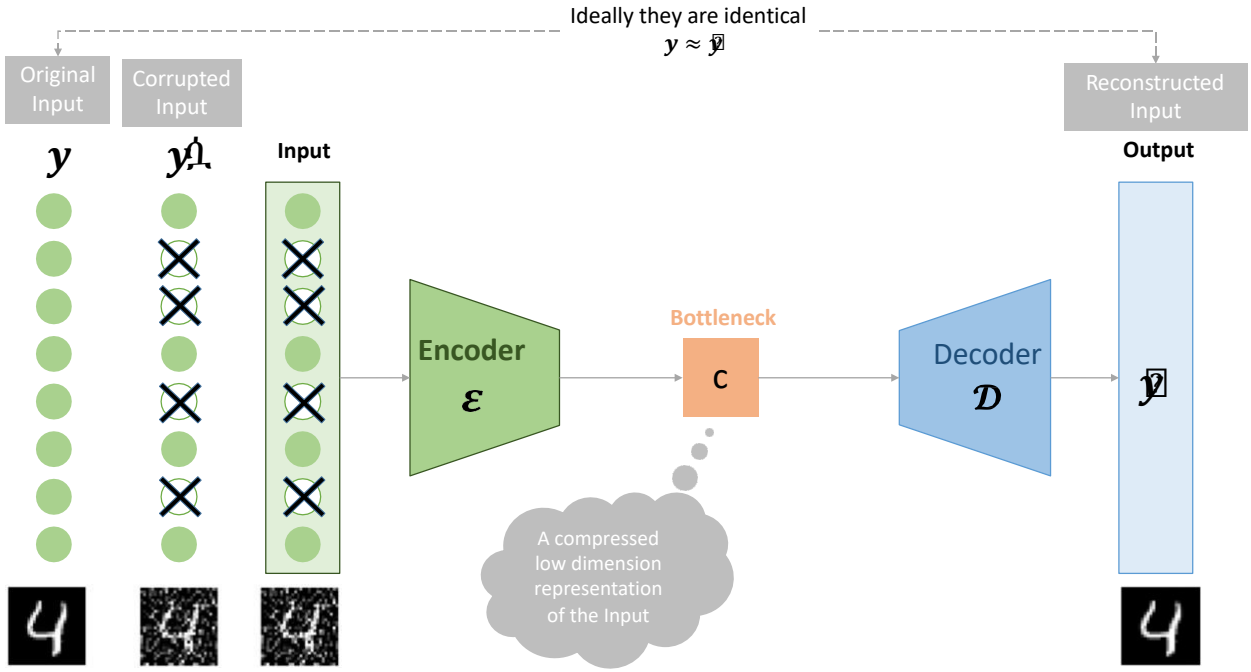


Figure 4.3.1: Denoising autoencoder architecture: Example of a DAE applied on images from the modified national institute of standards and technology (MNIST) handwritten digit database [148].

The hyper-parameters of our model will be largely studied in order to meet our requirements. Furthermore, this will allow us to compare our work with other existing ones et consequently to evaluate the effect of adding the first block of DAE.

4.3.2 Training and Hyper-Parameters

It is important to mention that artificial neural networks have two main hyper-parameters that control the topology or the architecture of the network: the number of layers and the number of nodes in each hidden layer. Further, the activation functions used for hidden, input and output layers should be specified when configuring the model as following.

(a) **Number of layers in each model**: It is defined also as the depth of the neural network. It is obvious that, the number of layers can influence the performances of the proposed model. The idea is that the performance of the deep neural model could suffer from the vanishing gradients [149] when large number of hidden layers is used. In addition, increasing the depth of the model, increases the accuracy until it reaches a maximum, and finally drops.
A theoretical study of the impact of the depth of neural networks, which is presented in [150], explains that while deep networks perform better than shallow ones in modeling complex non-linear functions, increasing the depth may not always produce the desired results. On the other hand, adding each layer is also expensive in

terms of the time required to train the model.

(b) **Number of nodes in each layer**: the average number of nodes per layer is called the width of the artificial model. It is a critical hyper-parameter to adjust until obtaining optimal accuracy of the trained model before the emergence of vanishing gradient problem.

(c) **The activation function of hidden layers**: the most commonly-used three activation functions that we can consider to be employed in hidden layers, are rectified linear activation, hyperbolic tangent and sigmoid, as presented respectively in Figure 4.3.2, Figure 4.3.3 and Figure 4.3.4.

(i) Rectified linear activation (ReLU) : if the input value I_v is negative, then the output value to be returned is $O_v = 0$, otherwise, the value $O_v = I_v$ is returned.

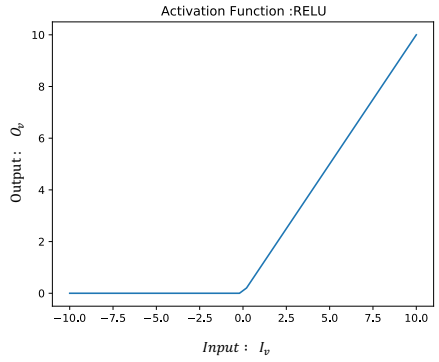


Figure 4.3.2: ReLU activation function.

(ii) Hyperbolic tangent (TanH) : for any real input value $I_v \in \mathbb{R}$, the function output value is given as $O_v \in [-1, 1]$. The larger the input is ($I_v >> 0$), the

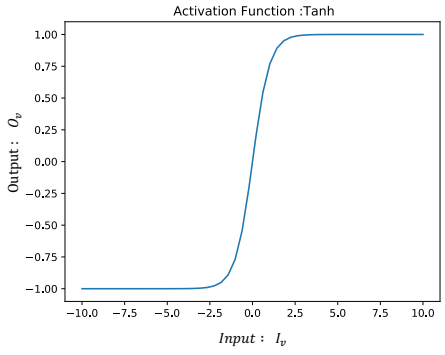


Figure 4.3.3: TanH activation function.

closer the output value will be to 1 ($O_v \approx 1$), whereas the smaller the input is ($I_v << 0$), the closer the output will be to -1 ($O_v \approx -1$).

(iii) Logistic (Sigmoid) : for any real input value $I_v \in \mathbb{R}$, the output value $O_v \in [0, 1]$. Such that, the larger is the input ($I_v >> 0$), the closer is the output to one,

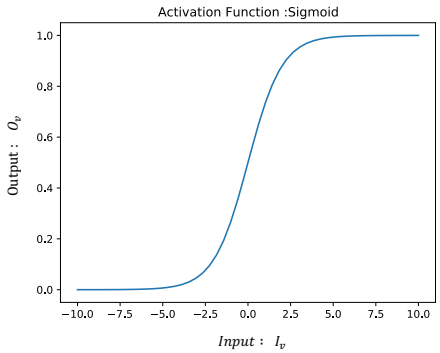


Figure 4.3.4: Sigmoid activation function.

$O_v \approx 1$, whereas when the input is very negative ($I_v << 0$), we have $O_v \approx 1$.

In order to choose the appropriate activation function for our hidden and output layers, it is important to acquire a prior knowledge on the input and output data structure. That so, the interpretation of data at each hidden layer is important for the whole model. The aim is that these layers learn how to treat similar data until arriving to the last layer, that is why employing the batch normalization technique is very essential.

(d) **Input and output layer configuration**: As their names indicate, the input and output layers of the model represent the very beginning of workflow and the last part of our artificial neural network, these two layers must be carefully configured. It is important to mention that the purpose behind the use of an artificial network, is to bring the initial data into the system for further processing by subsequent layers of artificial neurons until giving the final required analysis of these data.

(i) The dimension of the input layer, so-called shape, should be configured according to the dimension of the data to be treated. For example, when training a model to treat images, the dimension of first layer should consider the size of the image (height times width) multiplied by its number of color components, i.e. 1 or 3 when using respectively grey-scale or red-blue-green images (one for each of red, green and blue components). Moreover, for the case of complex vector as input, the input of the artificial model can be a vector of real values formed from the real and imaginary parts of each element in the complex vector.

(ii) For output layer, the important parameter to fix in addition to the activation function, is the number of output nodes. The idea is that the number of output

nodes could be 1 or more depending on the use cases : regression or classification. For example, in the case of multi-classification of input data into n different categories, the number of output nodes should be fixed to n and the value of output i refers to the probability of the actual inputs to belong to category i. The final decision is taken by finding the maximum probability. Here, we use either the label-encoding technique, or *one hot encoding* technique, to assign a unique binary code to specify each category [151].

4.4 Joint Denoiser and Detector of DL-based SCMA (DAE-DNN-SCMA)

In this section, we will introduce the first contribution of this Chapter, that is combining an unsupervised learning technique (namely DAE) and a supervised one (namely DNN) to propose a new SCMA detector. In the following, we will explain the motivation behind this combination, the structure of each one between the two learning blocks, and finally their real impact on the performance of SCMA detector.

4.4.1 Motivation behind the Proposed Solution

Here, we train the DAE-SCMA on a dataset of SCMA signals which are corrupted over an AWGN channel. The idea is that, the AE could learn the structure of transmitted SCMA signals, one way to help him in this task is to add some noise which is exactly the case when employing the received signals for learning, since the received signals are the transmitted ones added to the AWGN. In other words, there is no need to add noise to the data at the input of the DAE, as explained in Figure 4.3.1, they are already corrupted by the channel effect. Thus, a self-supervising process is employed by the DAE in order to estimate a reconstruction of the free-noise signal from the corrupted one.
Afterwards, once our estimated noise-free signal is ready, the proposed DNN is employed to detect the transmitted bits from estimated denoised SCMA signal. The architecture of the proposed SCMA detector is illustrated in Figure 4.4.1. The joint denoising and detection process requires to train the whole network with an end-to-end loss function. The idea is to improve the mean squared error of the the DAE-SCMA to give a better approximation of the denoised SCMA signal which will be in turn used as the input of the DNN detector. Regarding the second block, the final outputs are evaluated by measuring their accuracy via the categorical cross entropy loss. The real objective is to increase the overall SCMA performance which will be evaluated in terms of BER.
In the following subsections, we will present how the two blocks of the proposed DAE-DNN-SCMA detector are configured and trained.

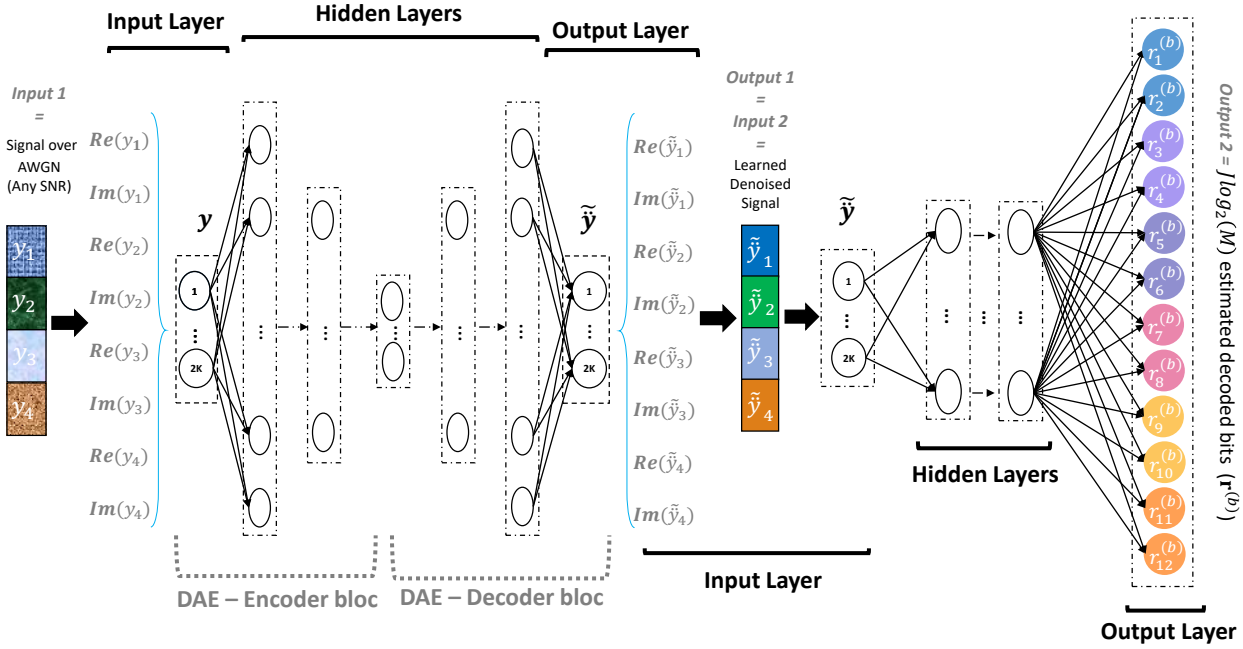


Figure 4.4.1: Block diagram of the proposed DAE-DNN-SCMA detector.

4.4.2 Denoising SCMA Signals over AWGN Channel

In the 1^{st} part of our proposed joint detector, we propose to configure and to train a DAE hoping to eliminate the noise that corrupted the transmitted signal over the AWGN channel.

Employed Model Structure

Our strategy is to to feed our DAE with SCMA signals corrupted with a Gaussian noise which is added to the input vector in a random manner. Then, the configured model is trained to predict the original uncorrupted SCMA signal as its output.
The autoencoder as presented in Figure 4.3.1 is composed of two phases: encoder and decoder. They are implemented using either fully-connected dense layers or convolutional layers, these implementations are respectively called deep autoencoder and convolutional autoencoder. That is, in convolutional autoencoder, we use a single weight associated with signals entering all neurons in a single convolution kernel, so-called filter. As for deep autoencoder, each neuron is considered as independent and a different weight is assigned to each incoming signal. In Table 4.4.1, it is obvious that using a convolutional autoencoder is less complex than using a deep autoencoder in term of number of parameters needed to train the full model in order to achieve equivalent performance. To conclude, we opted to use a convolutional DAE instead of fully connected layers.
The K-dimensional received vector over AWGN channel, $\mathbf{y}$, is given by,

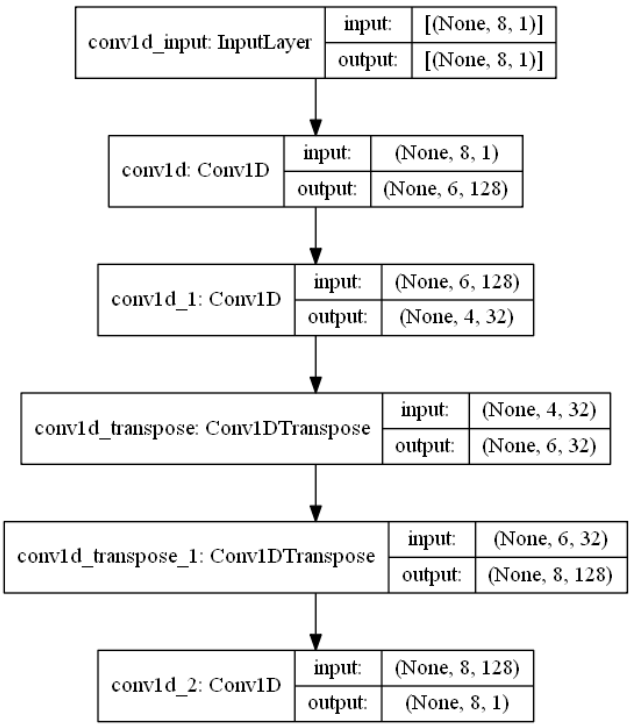


Figure 4.4.2: The graph of our configured DAE

$$\mathbf{y} = \sum_{j=1}^{J} \mathbf{x}_j + \mathbf{n} = \ddot{\mathbf{y}} + \mathbf{n}, \tag{4.1}$$

the received SCMA signal is a Gaussain noise added to a combination of the codewords of all users. Here, the input vector is arranged to form one column of real values.
The structure of our employed DAE is then, composed of an encoder which is based on deep convolutional layers, and a decoder which provides the reverse operation based on the transpose convolutional layers. The reduction of the spatial dimension of the inputs is enabled by the depth of the structure. Our model configuration is presented in Figure 4.4.2.

This procedure forces our model to learn important features from data [152] such that

Model	Depth	Total Width	Parameters Set Size	Training time	Accuracy
Deep AE	6	256	226144	$1,75$ (seconds)	$92,68\%$
Conv. AE	6	256	272	$1,42$ (seconds)	$92,85\%$

Table 4.4.1: Comparison between deep autoencoder and convolutional autoencoder in terms of complexity and training time for fixed number of nodes and fixed depth after 100 epochs of training.

the original uncorrupted version of the inputs is better recovered. Hence, the presence of the AWGN in equation (4.1) is not expected to hinder the capacity of DAE to reconstruct the original signal, $\ddot{\mathbf{y}}$, especially when the SNR ratio is not lower than certain level. In other words, we built and trained a DAE to estimate the clean (without noise) transmitted superimposed codeword from the SCMA received signal, $\mathbf{y}$. The estimated vector is denoted $\widetilde{\ddot{\mathbf{y}}} = \left(\widetilde{\ddot{y}}_1, \cdots, \widetilde{\ddot{y}}_K\right)^T$.

Training and Validation Phase

The input of the DAE-SCMA can be expressed as $\mathbf{y} = \mathbf{\Omega}(\ddot{\mathbf{y}})$ where $\mathbf{\Omega}(\cdot)$ is a corruption process. Whereas the output is expressed as $\widetilde{\ddot{\mathbf{y}}} = \widetilde{\mathbf{\Omega}}(\mathbf{y})$. So that, DAE-SCMA has to learn how does the inverse function, $\widetilde{\mathbf{\Omega}}(\cdot)$, work.

1. Encoder/decoder functions:
   Let us denote $\mathcal{E}(\mathbf{x}_e; \mathbf{W}_e, \mathbf{b}_e)$ and $\mathcal{D}(\mathbf{x}_d; \mathbf{W}_d, \mathbf{b}_d)$ as the encoder and decoder functions of the DAE, respectively, where $\mathbf{x}_e, \mathbf{x}_d$ are the inputs, $\mathbf{W}_e, \mathbf{W}_d$, $\mathbf{b}_e$ and $\mathbf{b}_d$ are the weights and biases of the encoder and the decoder respectively. The DAE-SCMA process is expressed as following,

$$\widetilde{\mathbf{\Omega}}(\mathbf{y}) = \widetilde{\ddot{\mathbf{y}}} = \mathcal{D}(\mathcal{E}(\mathbf{y}; \mathbf{W}_e, \mathbf{b}_e); \mathbf{W}_d, \mathbf{b}_d) \tag{4.2}$$

2. Loss function:
   The training process requires to choose a loss function, $\mathcal{L}$, to estimate the reconstruction loss of our DAE so that the weights and biases can be updated to reduce the loss on the next iteration. The loss function is computed and optimized as following,

$$\mathbf{W}_e^*, \mathbf{b}_e^*, \mathbf{W}_d^*, \mathbf{b}_d^* = \underset{\mathbf{W}_e, \mathbf{b}_e, \mathbf{W}_d, \mathbf{b}_d}{\arg\min} \mathcal{L}(\ddot{\mathbf{y}}, \widetilde{\ddot{\mathbf{y}}}) \tag{4.3}$$

   Training forces $\mathcal{E}$ and $\mathcal{D}$ to implicitly learn the structure of the superimposed codeword, $\ddot{\mathbf{y}}$, despite the fact that it is corrupted by the added Gaussian noise over the channel.

   Here, the popular mean squared error (MSE) is adopted as the training loss function of our DAE. The error between predictions and expected values is given by,

$$\mathcal{L}_{\text{MSE}}(\ddot{\mathbf{y}}, \widetilde{\ddot{\mathbf{y}}}) = \frac{1}{K} \sum_{i=1}^{K} \|\widetilde{\ddot{y}}_i - \ddot{y}_i\|^2 \tag{4.4}$$

   Thus, our proposed DAE will learn the structure of the received signal under different SNR regimes, that is, received signals are corrupted with noise with different levels of intensity as presented in Figure 4.4.3. Here, we adopt to evaluate the system performance as a function of a normalized SNR measure, E_b/N_0, so-called SNR per

bit.

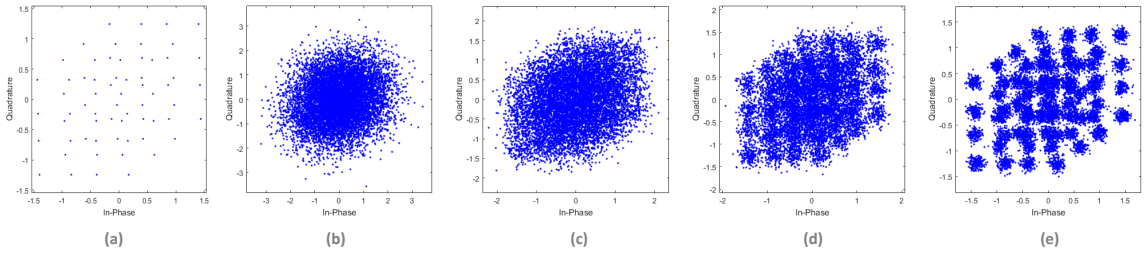


Figure 4.4.3: Illustration of the impact of the presence of noise on transmitted constellation for different E_b/N_0 values: (a) multiplexed constellations before transmission ($\ddot{\mathbf{y}}$), (b) $\mathbf{\Omega}(\ddot{\mathbf{y}})$ when $E_b/N_0 = 2$, (c) $\mathbf{\Omega}(\ddot{\mathbf{y}})$ when $E_b/N_0 = 9$, (d) $\mathbf{\Omega}(\ddot{\mathbf{y}})$ when $E_b/N_0 = 15$ and (e) $\mathbf{\Omega}(\ddot{\mathbf{y}})$ when $E_b/N_0 = 20$

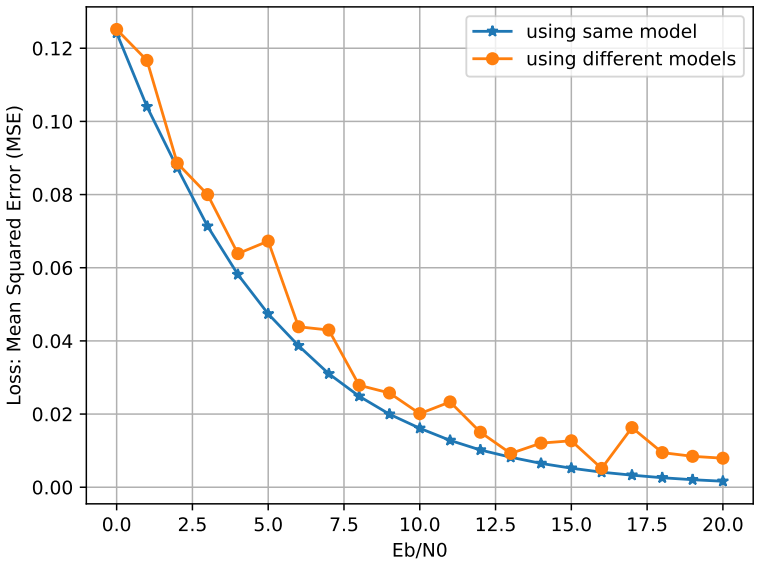


Figure 4.4.4: MSE as a function of E_b/N_0 values (a) the same DAE is trained with signals of different E_b/N_0 values, (b) different DAEs are trained separately with signals of fixed E_b/N_0 value per DAE.

Figure 4.4.4 shows the achieved MSE versus different E_b/N_0 values. Two cases were studied: (a) the same DAE is trained with signals of different E_b/N_0 values (the blue curve), (b) different DAEs are trained separately with signals of fixed E_b/N_0 value per DAE (the orange curve). A we can see clearly, it is better to allow the model to learn the structure of the signal by meeting examples with different levels of corruption.

After the learning phase, the curves for the training and validation sets can confirm that the model adopted in Figure 4.4.2 have learned well the reconstruction problem. In Figure 4.4.5, we can see that the model converged reasonably quickly, and both training and

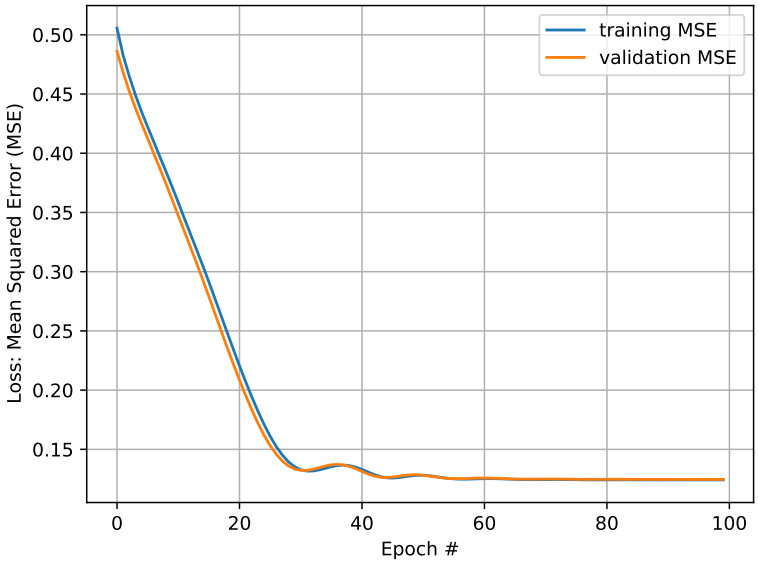


Figure 4.4.5: Mean squared error over 100 training epochs of DAE-SCMA when the used samples of SCMA received signal are corrupted over AWGN channel with $E_b/N_0 \in [6, 7, 8, 9]$.

validation performances remain equivalent. That is, the training and validation losses decrease to a point of stability with a minimal gap between the two final loss values. Hence, our model achieves a good reconstruction fit of input signal, holds a steady performance throughout training, and does not suffer from overfitting. The performance and convergence behavior of our model confirms also that MSE is a good match to DAE architecture when it comes to solving this problem.

4.4.3 Deep Neural Network based SCMA Decoder

For the 2^{nd} part of our proposed joint detector, we propose and train a DNN in order to decode the $J\log_2(M)$ bits of the J users from the estimated denoised received signal $\widetilde{\widetilde{\mathbf{y}}}$.

Employed Model Structure

The simulation parameters and the hyper-parameters of the employed DNN are given in Table 4.4.2. A more detailed evaluation of the effect of hyper-parameters on the overall performance will be given later. For the moment being, we chose the number of hidden layers as 6 and the number of nodes per each hidden layer as 48. This represents a good compromise between performance and complexity. On the other hand, this allow us to make a fair comparison with existing DL-based SCMA detectors as reviewed in subsection 1.4.5 of the first Chapter of the current manuscript. The fully connected DNN which is used in the rest of our proposition is presented in Figure 4.4.6 which illustrates its structure

graph with its dense layers and dimensions (width:Input/output).

Parameter	Value
Number of SCMA users J	6
Number of subcarriers K	4
Number of codewords for each user M	4
Number of nodes for input layer	8
Number of nodes for output layer	12
Number of nodes for DNN-SCMA's hidden layer	48
Number of hidden layers for DNN-SCMA	6
Size of dataset	500000
Data used to train the model	80% of SCMA samples
Data used to validate the training phase	10% of SCMA samples
Data used to test the model(Prediction phase)	10% of SCMA samples
E_b/N_0 values (dB) in the training set	$6, 7, 8, 9$
Training optimizer	ADAM
Learning rate	0.0001
Batch size for training	4096
Training epochs	3000

Table 4.4.2: Summary of the experiment configurations

Training and Validation Phase

During the training phase, we adopt the train-valid-test split procedure [149]. That is, the strategy is to split the dataset into train, and test data for validation: 80% of data is used to train the model and 10% of data is used to validate the training phase and to guide our efforts in possibly stopping training early. Once the model is trained without causing an under- or over-fitting issue [151], we predict the behavior of our final model on new data (10% of data), which is different from the one used to train and validate the model. As shown in Figure 4.4.1, our aim here is to estimate $J\log_2(M)$ transmitted bits from the received signal as defined in (4.1), after the denoising operation. That is, the learned denoised signal by the first DAE block is used as an input training data for the second DNN block. The $J\log_2(M)$ transmitted bits vector and the $J\log_2(M)$ estimated decoded bits vector are denoted as $\mathbf{t}^{(b)}$ and $\mathbf{r}^{(b)}$, respectively.

1. Activation function:
   The employed activation function for the DAE is the rectified linear one (Relu), while (TanH) was chosen as the activation function for the hidden layers of the fully connected DNN.

2. Loss function:
   Differently from the DAE, the learning of the DNN is supervised under the metric of *Accuracy*. The adopted loss function is the categorical cross entropy (CCE) and

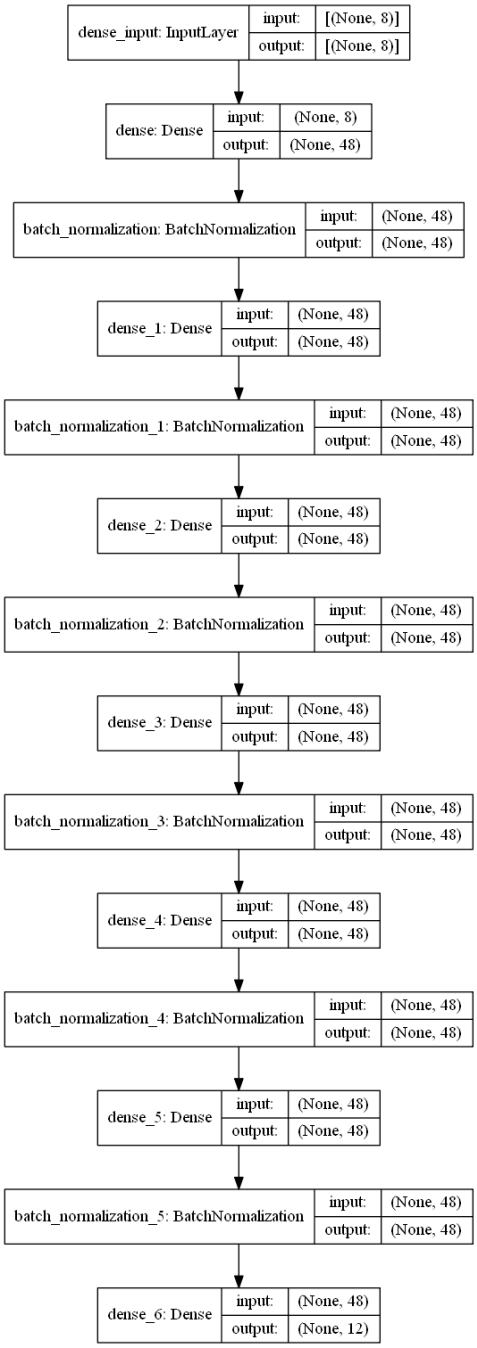


Figure 4.4.6: The graph of our configured DNN model

is presented as follows,

$$\mathcal{L}_{\text{CCE}}\left(\mathbf{r}^{(b)}, \mathbf{t}^{(b)}\right) = -\sum_{i=1}^{J\log_2(M)} t_i^{(b)} \log\left(\mathbb{P}\left(r_i^{(b)} = 1\right)\right) \tag{4.5}$$

where $\mathbb{P}(.)$ is the probability of an event, $\mathbf{t}^{(b)} = [t_1^{(b)}, ..., t_{J\log_2(M)}^{(b)}]^T$ and $\mathbf{r}^{(b)} = [r_1^{(b)}, ..., r_{J\log_2(M)}^{(b)}]^T$. The purpose behind training the proposed fully-connected DAE-DNN-SCMA detector is to estimate the optimum values of weights and biases, $\mathbf{W}_{\text{DNN}}$, $\mathbf{b}_{\text{DNN}}$, in order to minimize the loss between $\mathbf{r}^{(b)}$ and $\mathbf{t}^{(b)}$, which can be

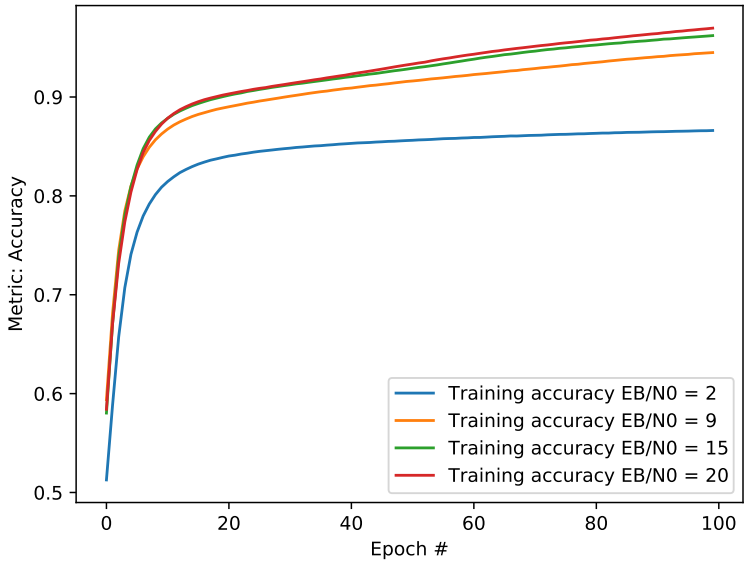


Figure 4.4.7: The Accuracy as a function of the number of epochs when the E_b/N_0 level is low, medium and high.

expressed as,

$$\min_{\mathbf{W}_{\text{DNN}},\mathbf{b}_{\text{DNN}}} \left(- \sum_{i=1}^{J\log_2(M)} t_b^{(i)} \log \left(\pi_i \left[\mathcal{R}\left(\mathbf{y}; \mathbf{W}_{\text{DNN}}, \mathbf{b}_{\text{DNN}} \right) \right] \right) \right) \tag{4.6}$$

where, $\mathcal{R}\left(\mathbf{y}; \mathbf{W}_{\text{DNN}}, \mathbf{b}_{\text{DNN}}\right)$ represents the DAE-DNN-SCMA detecting process which outputs are the estimated bits vector $\mathbf{r}^{(b)}$. $\pi_i[.]$ denotes the i^{th} element of a vector. The learning curves are presented in Figure 4.4.7.

3. Initialization of hyper-parameters:
In order to minimize the loss functions in (4.4) and (4.5), adaptive moment (ADAM) estimation [153] is used for the training of both neural networks, and the learning rate is set to 0.0001. During the training phase, to reduce the impact of the vanishing gradients problem for feed-forward neural networks, the Xavier variable initialization and the normalization via min-batches are adopted [154].

4.5 Fair SCMA Detector based Deep Learning (Fair-DNN-SCMA)

In [75], [76], [78]–[80], [146], authors proposed DL based SCMA detectors and they compared the performance of their propositions with that of the conventional MPA detector for SCMA systems over AWGN channels. None among these propositions was capable of

outperforming the MPA, the reason behind that is the fact that the MPA detector knows the codebook of each user, this is not the case for the detectors in [75], [76], [78]–[80], [146]. Hence the above comparison is not fair. Motivated by this, we propose, in this research work, a new method such that a fair comparison between the MPA detector and proposed DL based one is feasible.

4.5.1 Proposed Solution

The solution we propose here, lies in integrating the knowledge of the codebook in the learning process which is not the case for the method proposed in the subsection 4.4 of the current Chapter.

For a SCMA system as described in Chapter 2, the $K \times 1$ vector $\ddot{\mathbf{y}}$ as in equation (4.1) is a combination of J codewords, the j^{th} codeword is a $K \times 1$ vector which is chosen among the M codewords of, $\boldsymbol{C}_j$, the codebook of user j, as shown in Figure 2.2.1. Hence, $\ddot{\mathbf{y}} \in \mathcal{CC}$ where $\mathcal{CC}$ is the codewords combination set which is given by,

$$\mathcal{CC} = \left\{ \mathbf{z} \in \mathbb{C}^K \text{ where } \mathbf{z} = \sum_{j=1}^{J} \mathbf{c}_j, \mathbf{c}_j \in \boldsymbol{C}_j, 1 \leq j \leq J \right\}, \tag{4.7}$$

the cardinality of $\mathcal{CC}$ is M^J. The knowledge of codebooks $\boldsymbol{C}_j$ results in the knowledge of $\mathcal{CC}$.

In Figure 4.5.1, we present three examples of received SCMA signal constellation point on each subcarrier (the red diamante). The other colored circles represent the projection of the M^J members of $\mathcal{CC}$ on each RE (subcarrier).

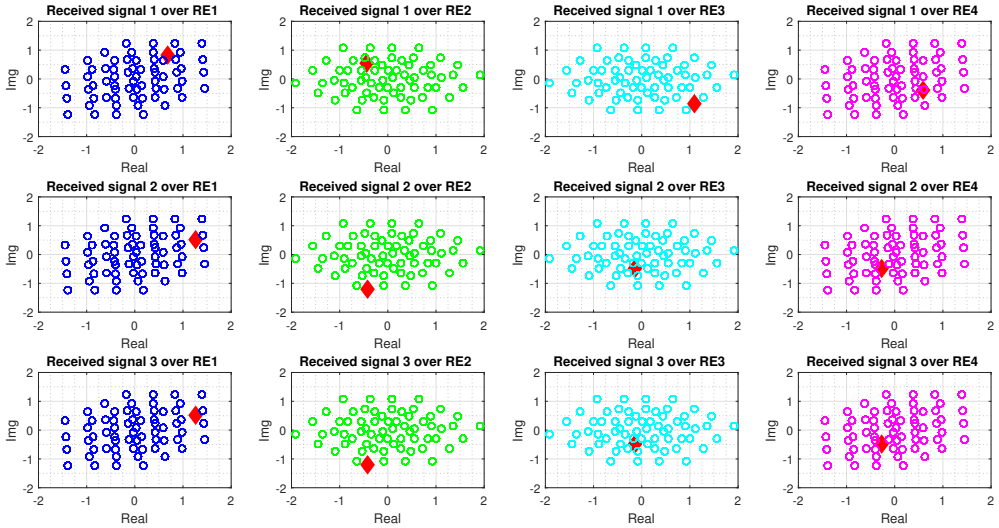


Figure 4.5.1: Example of noise-free received signal compared to the M^J possible combinations of codewords in $\mathcal{CC}$ over the $K = 4$ subcarriers; $M = 4$ and $J = 6$

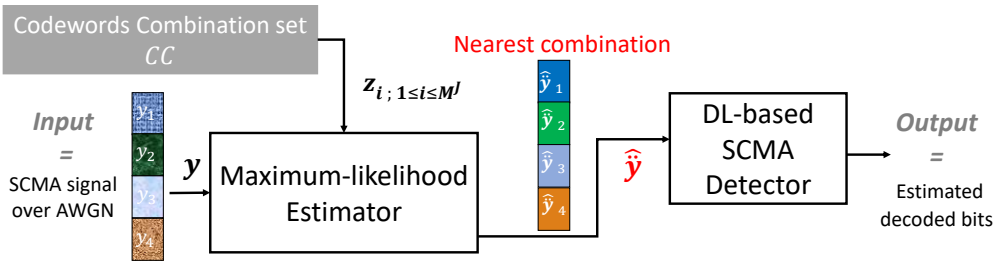


Figure 4.5.2: Block diagram of the proposed fair SCMA detector.

4.5.2 Pre-processing

Here, instead of sending, $\mathbf{y}$, the noise-corrupted received vector, as presented in equation (2.1), to the DNN inputs, we feed our model with an estimated value of $\ddot{\mathbf{y}}$, denoted as $\hat{\ddot{\mathbf{y}}}$. This estimation is facilitated by our prior knowledge of the research space. Since $\ddot{\mathbf{y}} \in \mathcal{CC}$, then its estimate is the most likely match among the M^J members of $\mathcal{CC}$. In the presence of Gaussian noise, the maximum-likelihood estimator of $\ddot{\mathbf{y}}$ is the minimum mean square error which is calculated based on the minimal square Euclidean distance between the received vector and each element of $\mathcal{CC}$, and it is expressed as,

$$\hat{\ddot{\mathbf{y}}} = \min_{\{\mathbf{z_i} \in \mathcal{CC}; 1 \leq i \leq M^J\}} \|\mathbf{y} - \mathbf{z_i}\|^2 \tag{4.8}$$

As shown in Figure 4.5.2, $\hat{\ddot{\mathbf{y}}}$ will replace the noisy received vector, $\mathbf{y}$, as the input of the DNN. We keep the same hyper-parameters as the DNN employed in subsection 4.4, we just need to re-train a new initialized model with the new inputs, the obtained detector is called Fair-DNN-SCMA.

As described in subsection 4.4, we can add a DAE block as the first step before estimating the transmitted superimposed codeword, i.e. the received vector $\mathbf{y}$ is replaced by its denoised version $\tilde{\ddot{\mathbf{y}}}$, then the equation (4.8) can be rewritten as,

$$\hat{\ddot{\mathbf{y}}} = \min_{\{\mathbf{z_i} \in \mathcal{CC}; 1 \leq i \leq M^J\}} \left\|\tilde{\ddot{\mathbf{y}}} - \mathbf{z_i}\right\|^2 \tag{4.9}$$

The obtained detector is called Fair-DAE-DNN-SCMA. The performance of different proposed detectors is compared in the next section.

4.6 Analysis of Results

In this section, the BER performances of our proposed methods are compared to conventional MPA detector [33] and the DL based one proposed in [80] for a SCMA transmission over AWGN channel. The codebook given in [1] in adopted for all the compared methods. The training data set is randomly simulated such that $M = 4$, $J = 6$ and $K = 4$ as described in Figure 2.2.1.

4.6.1 Impact of Training Data with Different Noise Intensity Levels

During the training phase for both of our proposed systems, data, with some E_b/N_0 (energy per bit to noise power spectral density ratio) values, are generated randomly. The choice of E_b/N_0 values may lead to different learning results and consequently to different performance levels of different proposed detectors. In order to avoid such situation, selecting appropriate E_b/N_0 values to generate our training data is investigated comprehensively. Thus, we can reduce the impact of weak signals due to too small SNR values and keep away our model from overfitting if only high SNR values were employed. Figure 4.6.1 presents the categorical cross entropy loss across training epochs as a function of E_b/N_0. Also, Figure 4.6.2 shows the detection accuracy of our DAE-DNN-SCMA detector for different E_b/N_0 values.

Hence, some further investigations had been conducted to study this problem, and the following scenarios were identified to be tested through intensive training of the DAE-DNN-SCMA detector,

- S1: we trained using 6 dB $\leq E_b/N_0 \leq$ 9 dB,
- S2: we chose some lower E_b/N_0 values, i.e. 2 dB $\leq E_b/N_0 \leq$ 5 dB,
- S3: we trained using all the tested values, i.e. $E_b/N_0 \in [0, 20]$ dB,
- S4: we proposed more complex architecture where a DAE-DNN model per E_b/N_0 is trained.

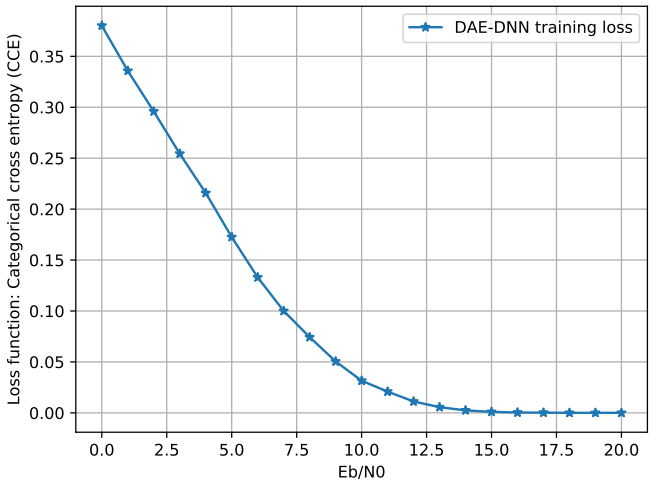


Figure 4.6.1: Categorical cross entropy loss versus E_b/N_0 for DAE-DNN-SCMA detector as introduced in section 4.4 after 100 epochs to train the DAE-SCMA first and 3000 to train the whole system DAE-DNN-SCMA using the estimated denoised SCMA signal samples.

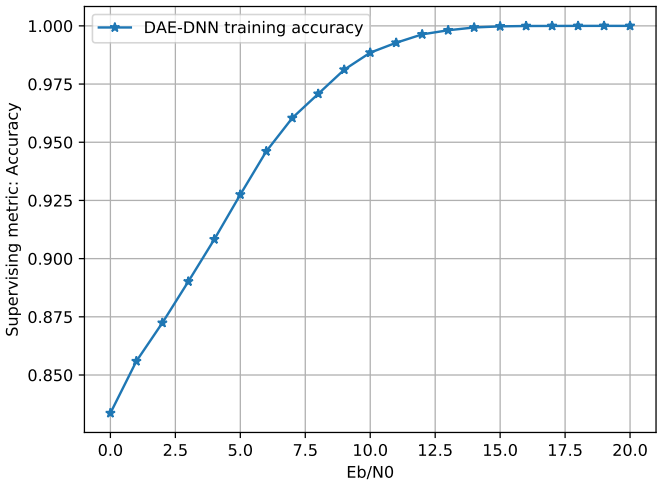


Figure 4.6.2: Accuracy versus E_b/N_0 for DAE-DNN-SCMA detector as introduced in section 4.4 after 100 epochs to train the DAE-SCMA first and 3000 to train the whole system DAE-DNN-SCMA using the estimated denoised SCMA signal samples.

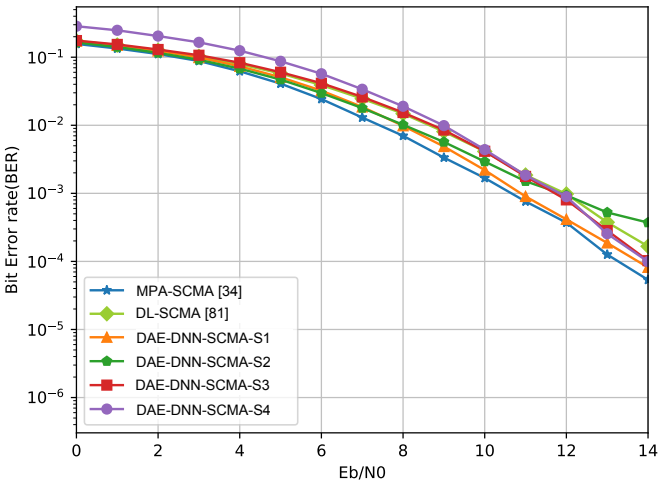


Figure 4.6.3: The impact of the choice of E_b/N_0 values on the training of DAE-DNN-SCMA detector when using the configuration of DNN-SCMA model presented in Table 4.4.2. The several DAE-DNN-SCMA scenarios' models, DAE-DNN-SCMA_S1, DAE-DNN-SCMA_S2, DAE-DNN-SCMA_S3 and DAE-DNN-SCMA_S4, are compared to MPA-SCMA detector in [33] and the proposed DL-SCMA in [80].

The performance of the DAE-DNN-SCMA detector, in terms of BER, when it is trained based on each one of the aforementioned scenarios is illustrated in Figure 4.6.3. Simulations results show that S1 is the best training strategy when compared to the other scenarios. We adopt S1, for the training of DAE-DNN-SCMA detector, in the rest of this work. Also, three among the 4 evaluated scenarios, namely S1, S2 and S3, outperform the DL based detector proposed in [80], which confirms that the denoising step had

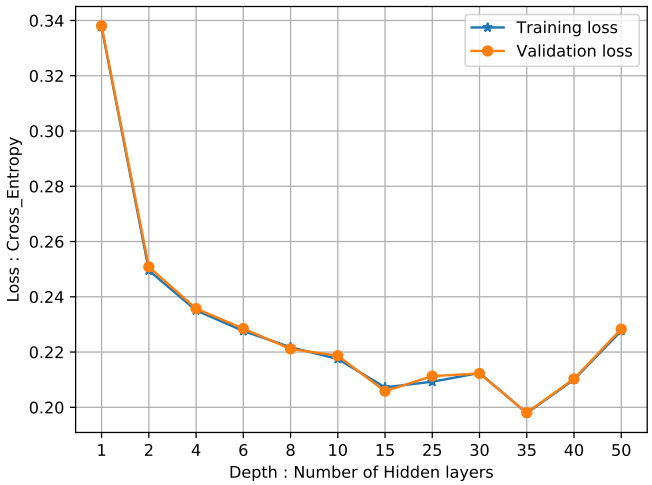


Figure 4.6.4: The loss function versus the depth after only 100 epochs of training when the SCMA signal is transmitted over AWGN channel using $E_b/N_0 \in [6, 7, 8, 9]$.

helped the DNN to better learn the SCMA signal structure. Finally, the performance of DAE-DNN-SCMA detector is slightly lower than that of MPA while still offering a lesser complexity.

4.6.2 Impact of Hyper-Parameters

In this subsection, we will study the effect of hyper-parameters of our neural network on the performance of the proposed detectors, we focus especially on two hyper-parameters, namely the depth (the number of hidden layers) and the width (the number of nodes in each hidden layer).

Figure 4.6.4 and Figure 4.6.5 show, respectively, the loss and the accuracy as a function of the depth. It is clear that the number of layers does influence the two performances metrics of the whole model, namely the loss and the accuracy. The performance will improve when the depth increases until a certain value of depth (depth = 35 layers) and then drops down. This can be explained by the vanishing gradients phenomenon.

Figure 4.6.6 illustrates the training time as a function of the depth, it can be approximated by an exponential function. However, the training time increases considerably when the depth is more that 10 layers. So, the depth must be chosen wisely to provide a good compromise among training time, complexity and performance. We will opt to a depth of 6 in this Chapter.

The other hyper-parameter to be optimized is the width, Figure 4.6.7 shows the accuracy versus the width. It is obvious that increasing the width when the other hyper-parameters are fixed will improve the performance while the risk of overfitting is minimized. The number of nodes per layer will is chosen to be 48 which is sufficient to have good performances

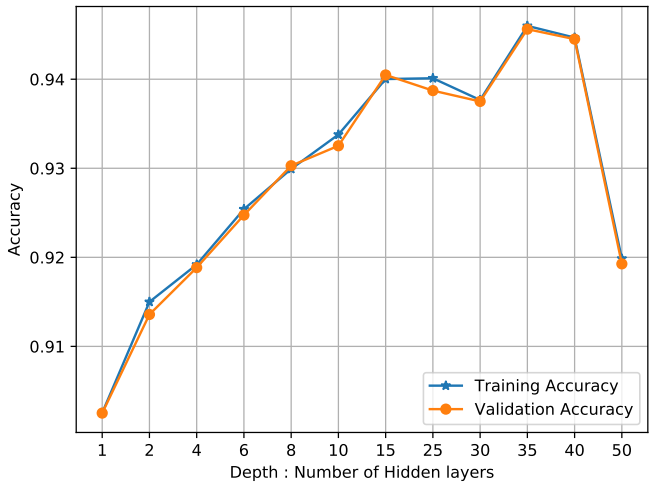


Figure 4.6.5: The accuracy versus the depth after only 100 epochs of training when the SCMA signal is transmitted over AWGN channel using $E_b/N_0 \in [6, 7, 8, 9]$.

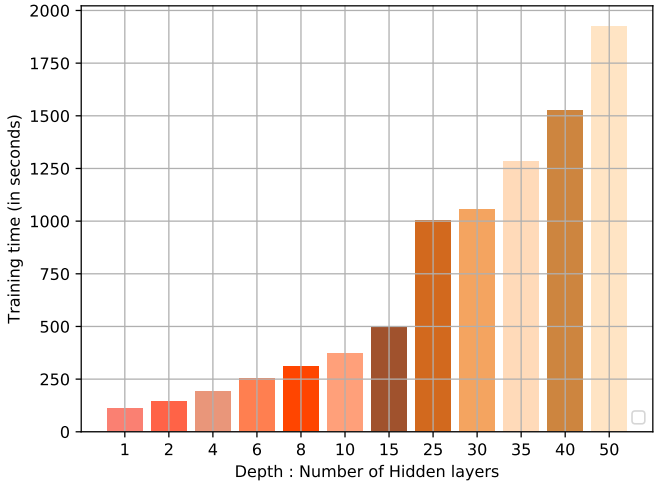


Figure 4.6.6: The required training time versus the depth after only 100 epochs of training when the SCMA signal is transmitted over AWGN channel using $E_b/N_0 \in [6, 7, 8, 9]$.

while having an acceptable complexity levels.

Once the values of these two main hyper-parameters are chosen, the other hyper-parameters are adjusted such as the performance can be optimized as illustrated by the simulation results presented in this section.

4.6.3 Performance Evaluation of Proposed Detectors

Figure 4.6.8 shows a performance comparison of the a SCMA system though AWGN channel when the following detectors are employed: traditional MPA [33], DL-based detector in [80], the DAE-DNN detector (see subsection 4.4), and the Fair-DNN and Fair-DAE-

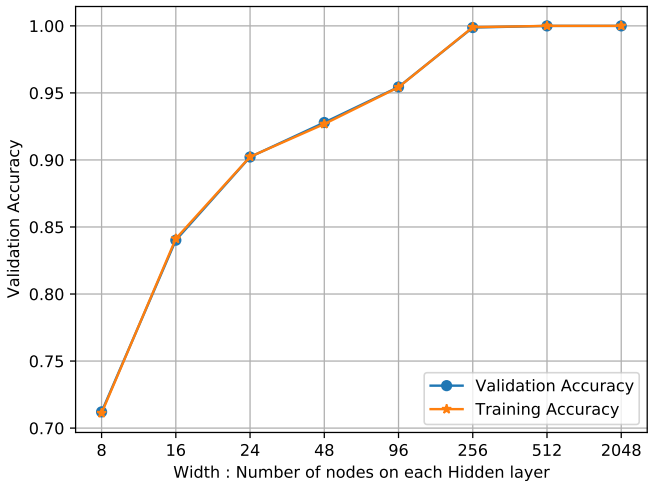


Figure 4.6.7: Training and validation accuracy for different values of width, when depth = 6, after only 100 epochs of training when the SCMA signal is transmitted over AWGN channel using $E_b/N_0 \in [6, 7, 8, 9]$.

DNN detectors (see subsection 4.5). Based on Figure 4.6.4 and Figure 4.6.5, we compare the DL-based detector in [80] using respectively depth = 6 and depth = 15, for only 100 training epochs. Simulations show that DL-based methods can outperform MPA when they are fairly compared to it. For example, comparing Fair-DNN (red solid line with square symbols) with MPA when BER $= 10^{-3}$ shows that Fair-DNN offers a gain of 0.5 dB. In addition, Fair-DNN is of lesser complexity. Furthermore, it was shown in the previous section that adding a denoiser to the detection process leads to performance improvement. However, we can see in Figure 4.6.8 that the two curves of Fair-DNN (red solid line with square symbols) and Fair-DAE-DNN (orange dashed line with circle symbols) are almost superimposed. This can be explained by the fact that the input of the DNN is the best match to received vector in the combination set $\mathcal{CC}$. Nevertheless, the denoising step can be useful if another channel model is adopted, for instance DAE-DNN-SCMA detector over a Rayleigh channel is actually under investigation. When DL-based detector in [80] is trained using different value of depth, for only 100 training epochs, simulations show that for instance when BER= 10^{-3}, using more layers (purple solid line with diamond symbols and orange solid line with diamond symbols), a gain of almost 1dB is achieved, which confirms what we proposed in the last subsection. However, when training the model with larger number of epochs, fixing the depth to 6 offered a gain of 0.25 dB when BER= 10^{-3}, compared to the case when depth is fixed to 15. Hence, we can assume that the performance of our proposed DL-based decoders depends not only on hyper-parameters that avoid the overfitting of the system, but also on the SNR and the complexity of the whole system which are two issues that should be respected.

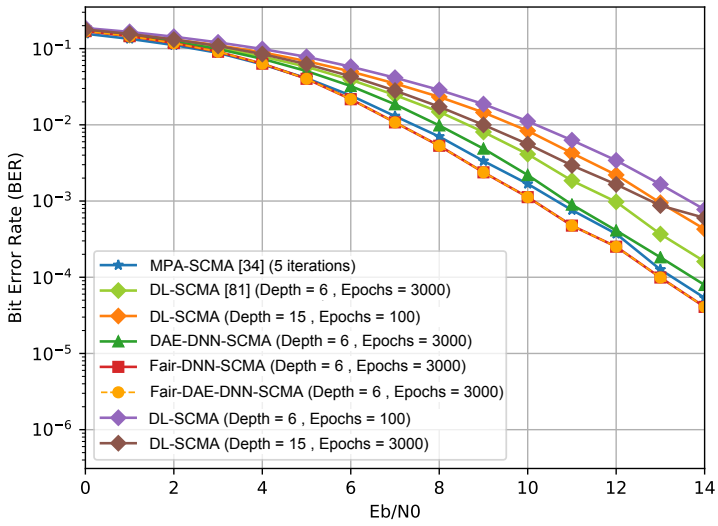


Figure 4.6.8: Performance comparison of MPA-SCMA detector [33], DL-SCMA [80], our proposed joint DAE-DNN-SCMA detector, our proposed Fair-DNN and Fair-DAE-DNN detectors over AWGN channel when $J = 6, M = 4$ and $K = 4$.

4.6.4 Complexity Study of the Proposed Models

Here we look into the complexity of proposed detectors in terms of arithmetic operations. In fact, DL-SCMA based detectors present less complex architecture compared to the iterative MPA in terms of addition and logarithmic or exponential operations but with more multiplication operators [80]. For instance, we use 14784 multiplication operations when training DL-SCMA detector against 9456 after 5 iterations of MPA. On the other hand, few logarithmic or exponential operations are used for offline training of DL-SCMA. However, the advantage of DL-based detectors is that their complexity goes only during the training process. That is, once the model is trained, the detection is performed through one shot operation instead of iterative process as in the case of MPA. In this proposition, adding the DAE does not increase the complexity a lot, moreover, the proposed fair detector does not require the DAE when the channel is AWGN.

4.7 Conclusions

In this Chapter, we studied the application of deep learning to SCMA detection over AWGN channels. Firstly, we highlighted our motivation and the objectives behind this application. Then, we presented the strategy and techniques that we employed to train our models. Here, we have suggested a joint denoising and decoding method based on a combination of denoising autoencoder as the first step and a deep neural network as the second one. The proposed DAE-DNN-SCMA detector outperforms existing DL-based SCMA detectors but still slightly less effective than the conventional MPA detector. Secondly, a

new DL-based detection method that can be fairly compared to MPA was introduced. The proposed Fair-DNN-SCMA detector assumes the knowledge of the SCMA codebook which enables a distance-based selection of the codewords combination to be fed into the DNN. Our proposed solution maintains lower computational complexity while achieving better BER performance. One perspective of this Chapter is to extend the proposed methods to the case of Raleigh fading channel.

General Conclusions and Future Works

This research work tries to enhance the performances of future wireless communication generations, as Advanced 5G, B5G and even 6G, by contributing to providing suitable multiple access schemes that are applicable in more realistic communications systems for several wireless services. In this thesis, we focus on optimizing encoding and decoding process of sparse code multiple access scheme.
In the following, the most important topics discussed in each Chapter are summarized, and the main conclusions and contributions of this thesis are highlighted. Thereafter, we present a list of possible near future work and perspectives that could lead to the continuation of some of our propositions.

Summary and conclusions

In Chapter 1, we discussed the problematic, the context and the motivations of this work. First, we will highlight the challenges imposed by the future generations when it comes to multiple access techniques, especially the need of massive connected devices over wireless networks, by introducing an overview of 5G and beyond technologies based on the several 5G scenarios, namely enhanced mobile broadBand, massive machine type communication and highly reliable low latency communication. Afterwards, the NOMA concept had been explained and reviewed in both power domain and code domain through existing publications and standardization works of 3GPP, IETF and ITU. These discussions were concluded by illustrating why SCMA is prominent. To end with this Chapter, we have reviewed the application of machine learning for communication systems especially the existing contributions which are related to NOMA techniques.
Then, in Chapter 2, the SCMA architecture, the design of SCMA codebooks and existing detectors of SCMA had been highlighted. The idea behind this Chapter is to further explain why SCMA is considered as a promising massive multiple access candidate for future generations of wireless communications systems. Simulations and results presented in this Chapter highlighted several comparisons of SCMA codebook designs and the impact of the choice of SCMA detector on the system efficiency in term of BER for different SNR levels.
Motivated by Chapter 2, we proposed an adaptive design for SCMA codebooks based on an irregular architecture of SCMA in Chapter 3 of this Ph.D. work. The channel state

information was taken into consideration under the assumption that different users could suffer from different channel conditions. Thus, we explored three cases of study to better satisfy the several demands of more realistic scenarios. First, we supposed that all users have the same channel states but they have different needs in terms of delay, bandwidth and packet-delivery and we adjusted the QoS according to user's needs. Then, we analyzed the system when users have different channel states. The purpose here was to achieve more fairness among users in terms of their QoS. And finally, we assigned irregular codebooks for each group of users in order to achieve higher overall throughput when assuming that users have different channel states. Simulations confirmed that using an adaptive SCMA design can achieve better performances compared to a conventional regular SCMA design. In Chapter 4, the design of an SCMA detector is conducted. Our proposition employ a branch of machine leaning, namely deep learning, with the purpose to enhance the performance of SCMA detector in an effective manner. We proposed first a joint denoising and decoding approach for designing SCMA detector. We trained a DAE to mitigate the effect of AWGN by trying to remove it from received signal. The estimated denoised signal at the output of our proposed DAE is used as an input of a DNN whose output is a binary sequence that represents an estimation of the transmitted data. Our proposed model outperform existing deep-learning based ones in terms of bit error rate and gives a performance which is slightly lesser than that of MPA while still having a reduced complexity. Inspired by a fair comparison of DL based SCMA detectors and the convectional SCMA one, MPA, we proposed a fair DNN detector exploiting the distances between the received vector and each one among possible superimposed combination codewords. This will allow firstly to estimate the nearest superimposed codeword vector, the chosen vector will be the input of a DNN which objective is to decode the transmitted users' bits. The proposed fair-DNN-SCMA detector assumes the knowledge of the SCMA codebook which enables a distance-based selection of the codewords combination to be fed into the DNN. Our proposed solution maintained lower computational complexity while achieving better BER performance when compared to MPA.

Perspectives

Beyond the contributions presented in this thesis, some questions remain open issues and need further investigations. Below, we list some topics that may be viewed as the natural continuity of our work:

- In this research work, the channel fading model is assumed to be Rayleigh or AWGN. However, this assumption is not always realist especially when a line-of-sight between the mobile terminal and the base station exists in some mobile environments, in this case, the fading is no longer of Rayleigh type. Thus, other channel fading

models must be considered when studying multiple access techniques in general and especially SCMA, for instance Rician or m-Nakagami fading.

- Further, as SCMA is a proposal for future cellular systems, the presence of co-cell interference (CCI) is unavoidable due to the channel reuse employed in cellular systems. Thus, it is interesting to know the sensitivity of our irregular SCMA decoder when CCI is present over different channel fading models.

- Our propositions in the whole manuscript consider SISO systems. An important perspective of our research work is to extend it to MIMO SCMA detectors using deep learning when different channel models are assumed. Therefore, implementing an analysis framework to investigate two fusion methods: early fusion to combine features before classification and late fusion to combine the outputs after classification could be an interesting continuity to our work, in order to perform the joint detection and decoding of multi-dimensional constellation.

- CNN can be used to reduce the number of parameters which are needed to train the deep learning model without sacrificing the performance. However, CNN training is a bit slower than that of a DNN. Also, RNN models such as LSTM, requires more parameters than CNN, but only about half of DNN, while being the slowest to train. Nevertheless, their advantage comes from being able to look at long sequences of inputs without increasing the network size. Thus, proposing a detector by employing CNN or LSTM could be a good perspective to evaluate the impact of the both models on the performance of our proposed fair SCMA decoder.

- An important issue in signal processing and image processing is the choice of an optimal representation of our signals in order to optimize the performance of a dedicated task (inference, classification or compression). The idea could be to train a deep learning model on new representation of received SCMA signal. Thus, the input of the proposed model will be a 2D-image resulting from transforming the received SCMA multi-dimensional constellation, and the task of the neural network is to classify these images into binary categorical sequences that define the output of the model.

- Inspired by the last proposed perspective and our proposed fair distance-based SCMA detector, an idea is exploit a set of vectors composed of the difference between the received vector and each one among the possible combinations of codewords. This gives a matrix of dimension $M \times M^J$ instead of a vector of M entries, this will enrich the inputs of the deep learning model. An DAE structure to reduce the noise before training a CNN using low-dimensional images could be be considered.

- SCMA is a more suitable technique for multiple access with small user overloading as it gives lower BER compared to MUSA. When increasing user overloading, the performance of SCMA degrades and thus MUSA gives better performance than it. Thus, exploring the exploitation of irregular SCMA structure for transmission with higher user overloading may enhance the performance of SCMA system compared to MUSA.

APPENDICES OF THE THESIS

A Codebook Example

To better illustrates SCMA mapping, a numerical example of complete SCMA codebooks, as described in [1] and depicted in Figure 2.3.4 and Figure 2.3.6, is provided in the following.

$$
\begin{aligned}
\mathrm{CB}_1 &= \left(\begin{bmatrix} 0 \\ -0.182 - 0.132i \\ 0 \\ +0.785 \end{bmatrix} \begin{bmatrix} 0 \\ -0.635 - 0.462i \\ 0 \\ -0.224 \end{bmatrix} \begin{bmatrix} 0 \\ +0.635 + 0.462i \\ 0 \\ +0.224 \end{bmatrix} \begin{bmatrix} 0 \\ +0.182 + 0.132i \\ 0 \\ -0.785 \end{bmatrix} \right) \\
\mathrm{CB}_2 &= \left(\begin{bmatrix} +0.785 \\ 0 \\ -0.182 - 0.132i \\ 0 \end{bmatrix} \begin{bmatrix} -0.224 \\ 0 \\ -0.635 - 0.462i \\ 0 \end{bmatrix} \begin{bmatrix} +0.224 \\ 0 \\ +0.635 + 0.462i \\ 0 \end{bmatrix} \begin{bmatrix} -0.785 \\ 0 \\ +0.182 + 0.132i \\ 0 \end{bmatrix} \right) \\
\mathrm{CB}_3 &= \left(\begin{bmatrix} -0.635 + 0.462i \\ +0.139 - 0.176i \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} +0.182 - 0.132i \\ +0.487 - 0.616i \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} -0.182 + 0.132i \\ -0.487 + 0.616i \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} +0.635 - 0.462i \\ -0.139 + 0.176i \\ 0 \\ 0 \end{bmatrix} \right) \\
\mathrm{CB}_4 &= \left(\begin{bmatrix} 0 \\ 0 \\ +0.785 \\ -0.006 - 0.224i \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ -0.224 \\ -0.019 - 0.785i \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ +0.224 \\ +0.019 + 0.785i \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ -0.785 \\ +0.006 + 0.224i \end{bmatrix} \right) \\
\mathrm{CB}_5 &= \left(\begin{bmatrix} -0.006 - 0.224i \\ 0 \\ 0 \\ -0.635 + 0.462i \end{bmatrix} \begin{bmatrix} -0.019 - 0.785i \\ 0 \\ 0 \\ +0.182 - 0.132i \end{bmatrix} \begin{bmatrix} +0.019 + 0.785i \\ 0 \\ 0 \\ -0.182 + 0.132i \end{bmatrix} \begin{bmatrix} +0.006 + 0.224i \\ 0 \\ 0 \\ +0.635 - 0.462i \end{bmatrix} \right) \\
\mathrm{CB}_6 &= \left(\begin{bmatrix} 0 \\ +0.785 \\ +0.139 - 0.176i \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ -0.224 \\ +0.487 - 0.616i \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ +0.224 \\ -0.487 + 0.616i \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ -0.785 \\ -0.139 + 0.176i \\ 0 \end{bmatrix} \right)
\end{aligned}
$$

Figure A.1: The 2-dimensional codebooks with 4-codewords, generated for $J = 6$ users, as described in [1].

B Review of Existing SCMA Codebook Designs

MQAM = M-sized constellation based on shuffling technique en Two N-dimensions
MLQAM = M-sized constellation based on shuffling technique with low number of projections
MHQAM = M-sized constellation optimized by a rotation angles of the hyper-cube
MCQAM = M-point circular constellation
M-Peng = Optimized M-sized mother constellation based on Euclidean distance
M-Beko = Optimized M-sized mother constellation based on the average alphabet energy
M-Bao = Optimized M-sized mother constellation based on criterion derived from cutoff rate of MIMO

Table B.1: Review of existing SCMA codebook designs

Reference	Year	Approach	Uplink / downlink	Optimization criterion	Description
Beko *et al.* [82]	2012	M-Beko	Uplink & downlink	Average alphabet energy for a given minimum Euclidean distance	M-Beko method formulates the constellation design as a non-convex optimization problem which objective is to minimize average alphabet energy between any two constellations points. This method can offer an optimal solution.
Boroujeni *et al.* [86]	2013	TMQAM	Uplink & downlink	Euclidean distance, product distance & average product distance	Using a unitary rotation to produce a multi-dimensional mother constellation. This document introduces also the first low-projection multi-dimensional constellation. USA Patent US9509379B2.
Taherzadeh *et al.* [84]	2014	Shuffling QAM	Uplink & downlink	Euclidean distance & product distance	A systematic approach based on lattice constellations is employed to design multi-dimensional constellation with a reasonable minimum Euclidean distance which is then rotated to reach a good product distance. Shuffling operation can be applied.
Yu *et al.* [98] [99]	2015 2018	Star-QAM	Uplink	Pairwise error probability	Codebook with large minimum Euclidean distance, using star-QAM improves SCMA BER gain without sacrificing the system implementation complexity.
Zhang *et al.* [129]	2015	IrSCMA	Uplink	Not specified	Irregular SCMA: Various sparsity degrees for various user needs. Authors introduced the notion of degree distribution of SCMA. The first proposition of an irregular SCMA.
Yu *et al.* [130]	2016	IrSCMA	Down-link	EXIT chart	Irregular SCMA: this paper extends the proposition in [129] via an optimization of the employed rotation angles. The degree of user superposition on a each RE is constant in contrast to [129].
Bao *et al.* [95]	2016	M-Bao	Uplink	Cutoff rate of MIMO system	Rotation matrices are obtained through exhaustive computer search over compact parameterizations of unitary matrices, a suboptimal solution was applied. Perfect channel knowledge is assumed.
Bao *et al.* [96]	2016	Spherical codes	Uplink	Squared Euclidean distance	Spherical codes design is formulated as a non-convex second-order cone programming problem. Hybrid lattice and spherical codes based constellations are also discussed. Low peak-to-average power ratio. -
Metkarunchit [87]	2017	Low-projection MC	Down-link	Cutoff rate of MIMO system	Circular QAM for SCMA mother constellation design, complexity of MPA depends on the reduced number of constellation points.
Peng *et al.* [83]	2017	M-Peng	Uplink & downlink	Euclidean distance	A joint design of codebook and mapping matrices. The proposed solution is semi-definite relaxation of non-convex quadratically constrained quadratic programming.
Yan *et al.* [102]	2017	Dimen-sion distance-based design	Uplink & downlink	Sum of distance between dimensions of interfering codewords	Codebook design is based on turbo trellis coded modulation. Phase rotation and interleaving employ an appropriate permutation set which is selected according to the introduced criterion.
Lai *et al.* [103]	2017	Dynamic codebook design	Down-link	Not specified	A transmission of SCMA codewords designed using random angles extracted from CSI wich will grant more secure transmission process. Upper-bound-aided codebook design was introduced.
Zhai *et al.* [131]	2017	IrSCMA, QAM	Uplink	Not specified	Flexible resources scheduling scheme according to user's features.
Bao *et al.* [100]	2018	Low-projection mother constella-tion	Uplink	Coded modulation capacity	Multi-stage optimization of multi-radius rings based multi-dimensional constellation by permuting an one-dimensional constellation. Introduction of a new design criterion. Application of bit-interleaved BICM with iterative multiuser detection.
Vameghesta-hbanati *et al.* [97]	2020	MHQAM	Uplink	Frame-error-rate of a LDPC-coded system	Construction of gray-labelled mother constellation based on hypercubes which is used along with BICM. Exhaustive search is employed.

C Data-aided-CSI Estimation Algorithm

Algorithm 5: Data-aided channel estimation and detection

Input : ($\mathbf{y}$,$\dot{\mathbf{y}}$)

Step 1 : Channel estimation by pilot-aided method: Estimate the channel $\widetilde{\mathbf{H}}^{(p)} \sim \mathcal{CN}(0,\, \sigma^{(p)})$ only using pilot sequences according to (10).

Step 2 : Improve the estimated channel
while *(the error in (11) in not minimized)* **do**

(a) Data detection: MPA detection is performed based on obtained estimated channel.

(b) Data re-construction: Reconstruction of transmitted data symbols after detection of received signal and the channel based only on these data symbols, denoted as $\widetilde{\mathbf{H}}^{(d)} \sim \mathcal{CN}(0,\, \sigma^{(d)})$ is estimated.

(c) Channel re-estimation: Re-estimate the channel $\widetilde{\mathbf{H}}$, using both, pilots and reconstructed data symbols:

$$\widetilde{\mathbf{H}} = \alpha_p \widetilde{\mathbf{H}}^{(p)} + \alpha_d \widetilde{\mathbf{H}}^{(d)} \tag{10}$$

where, α_p and α_d are the optimum coefficients to compute by satisfying the criterion of optimization which is the minimization of the mean square estimation error variance as following:

$$\min \left\{ |\alpha_p|^2 \sigma^{(p)} + |\alpha_d|^2 \sigma^{(d)} \right\}; \tag{11}$$

Step 3 : Final data detection: MPA detection is performed based on improved estimated channel.

Output : $J \log_2(M)$ decoded bits

D Review of Existing SCMA Detector Designs

Only dominant part of complexity order is considered, notations are defined in Table D.1,
$\surd \rightarrow$ assumed to be known, $\times \rightarrow$ assumed to be unknown.

D-L = Downlink
U-L = Uplink
SISO = single-input single-output
SIMO = single-input multiple-output
MISO = multiple-input single-output
MIMO = multiple-input multiple-output
BICM = Bit-Interleaved Coded Modulation
BP = Belief Propagation
CS-MPA = Compressed Sensing aided MPA
DMPA = Discretized MPA
EPA = Expectation Propagation
EPA = Expectation Propagation Algorithm
EXIT = Extrinsic Information Transfer
EML-MPA = Joint Extended Maximum Likelihood and MPA
FFT = Fast Fourier Transform
GA-MPA-B = Gaussian-Approximated MPA using Bits
GA-MPA-S = Gaussian-Approximated MPA using Symbols
GenA-MPA = Generalized Approximate MPA
ISD = Improved SD
LSD = List Sphere Decoding
MMSE = Minimum Mean Square Error
MMSE-PIC = MMSE with Parallel Interference Cancellation
MSD = Modified Sphere Decoding
MS-MPA = Multiple Scheduling MPA
MSTS = Modified Single Tree Search
PA-MPA = Partially Active Message Passing Algorithm
PD-MPA = Partial Decoding MPA
PSMA = Power-Domain Sparse Multiple Access
PM-MPA = Partial Marginalization Message Passing Algorithm
Q-MPA = Quantum-Assisted MPA
QSD-MPA = Quantum-Assisted Sphere Decoder based MPA
SC-EPA = Sparse-Channel based EPA
SD = Sphere Decoding
S-SCL = Soft-Successive Cancellation List
SS-MPA = Single Scheduling MPA

Table D.1: Key parameters in complexity analysis

Parameters	**Description**
N_t	Number of transmitter antennas
N_r	Number of receiver antennas
K	Spreading length (number of REs)
N	Codebook sparsity degree
N_{iter}	Total number of iterations
M	Codebook size
M_p	Number of projection points on the constellation
d_f	Degree of signal superposition on a given resource element
d_s	Maximum degree-of-freedom allowed in SIC-MPA receiver
J	Number of users
N_{OL}	Number of outer loops for MIMO-SCMA detection
N_{IL}	Number of inner loops for MIMO-SCMA detection
N_{v1}	Average number of visited layers for Type I (ISD method)
N_{v2}	Average number of visited layers for Type II (ISD method)
s	Sparsity level at the current iteration
PR	Pass rate (ratio of removed combinations to all combinations)

Table D.2: Review of existing SCMA detector designs

Reference		Year	Approach	Scenario	CSI	Add & Mult Complexity	Description
Type	Authors						
MPA based detectors	Zou *et al.* [155]	2015	SIC-MPA	SISO D-L	√	Neglected & $\mathcal{O}(N_{\text{iter}}N_r K M_p^{d_s})$	Joint use of SIC and MPA. Only active users per RE are considered. Specific constellation design is needed to be employed.
	Zhang *et al.* [156], [157]	2016 2018	DMPA	SISO	√	$\mathcal{O}(d_f^3 ln(d_f))$ & $\mathcal{O}(d_f^3 ln(d_f))$	Requires a small d_f which limits the choice of codebook. Based on PDF discretizing and FFT in VNs.
	Yang *et al.* [158]	2016	Threshold-based MPA	SISO U-L	√	$\mathcal{O}(N_{\text{iter}}JNM^{d_f-1}d_f)$ & $\mathcal{O}(2N_{\text{iter}}JNM^{d_f}d_f)$	Users are divided into two groups based on the codeword reliability which is integrated in the iterative process. Inconvenient : loss of posterior information of users using the same RE.
	Qi *et al.* [111]	2017	Max-log-MPA	SISO U-L	√	$\mathcal{O}(2KM^{d_f}d_f))$ & 0	Approximating the logarithm of a sum of exponential operations into a maximum operation.
	Tian *et al.* [159]	2017	Improved Log-MPA	SISO D-L	√	$\mathcal{O}(2JM(M^{d_f-1}-1))$ & 0	Log-MPA adds a correction term to MAX-log-MPA. The authors improved Log-MPA by updating the FNs only within a restricted search region.
	Wu *et al.* [123]	2017	PD-MPA	MIMO U-L	√	$\mathcal{O}(NN_{\text{iter}}(M-1)M^{d_f-1}d_f^2.PR)$ & Neglected	Removing redundant combinations while updating the VNs. Based on Log-MPA.
	Huang *et al.* [160]	2017	GenA-MPA	MIMO U-L	√	Neglected & $\mathcal{O}(JK^2 + KN_{\text{iter}}M_p^{d_f})$	Approximated priori probability transforms the vector estimation problem into a scalar one, it is suitable for hardware implementation.
	Huang *et al.* [160]	2017	SIC-GenA-MPA	MIMO U-L	√	Neglected & $\mathcal{O}(JK^2 + N_{\text{iter}}M_p^{d_f})$	Extends GenA-MPA by eliminating the calculated user layer, the complexity is reduced.
	Dai *et al.* [161]	2017	SS-MPA MS-MPA	SISO U-L	√	Not specified & Not specified	A lookup table is introduced to substitute the information calculation. SS-MPA jointly updates the message at FNs and VNs. MS-MPA executes different updates in parallel.
	Lai *et al.* [116]	2018	LDPC-MPA	SISO U-L	√	Significant complexity reduction	Joint factor graph for SCMA and LDPC channel decoding. MPA was simplified by using partial message passing.
	Gao *et al.* [162]	2018	CS-MPA	SISO	√	Not specified & $8K + 3KN + 2Ns^2 + s^3$	The detector is divided into two phases : MPA with few iterations followed by a sparse error correction. The main complexity lies in MPA.
	Han *et al.* [163]	2018	Serial MPA	SISO	√	Not specified & Not specified	Updates the probability of user output codeword immediately when the information of one of its corresponding FNs is updated. EXIT charts are used to analyze MPA performance.
	Dai *et al.* [164]	2018	PA-MPA	MIMO U-L	×	Neglected & $\mathcal{O}(N_{\text{OL}}N_{\text{IL}}\sum_{n=1}^{N_r}2^J)$	An outer-loop iterative process for joint MIMO and multi-user detection. A basic-switching sliding window to separate users into active and silent ones during each iteration.
	Jia *et al.* [165]	2018	PM-MPA	SISO U-L	Not specified	$\mathcal{O}(M^{d_f}NJ(N_{\text{iter}}-t))$ & $\mathcal{O}((N_{\text{iter}}(d_f-1)M^{d_f}NJ)$	The codewords of a given number of chosen users are determined after a predefined number of iterations. Only the message of the undetermined users are updated in the remaining iterations.
	Lai *et al.* [166]	2018	EML-MPA	SISO U-L	√	Neglected & $\mathcal{O}(M_p^{d_f})$	Trellis representation is introduced by mapping SCMA constellation to a Galois field. A truncated-messages based detector is based on selecting the most reliable messages.

Reference		Year	Approach	Scenario	CSI	Add & Mult Complexity	Description
Type	Authors						
	Ye *et al.* [167]	2019	Q-MPA	SISO U-L	√	The cost function is calculated $\mathcal{O}\left(\sqrt{M_p^{d_f+1}}\right)$ times	Algorithm-aided MPA is accelerated by exploiting the results of a quantum search which is applied at the FNs.
	Ma *et al.* [168]	2019	Edgewise serial MPA	SISO U-L	√	$\mathcal{O}(Kd_f[(d_f+1)M^{d_f}-M])$ & $\mathcal{O}(Kd_fM^{d_f}(2d_f+1))$	This method updates messages in an edge by edge way such as a more reliable extrinsic information is used in a balanced manner.
	Shi *et al.* [169]	2019	Log-MPA	SISO U-L	√	Not specified & Not specified	MPA with flexible number of iterations. The iterative process continues until the codeword convergence rate is higher than a threshold. A balance between BER and complexity can be achieved.
	Dai *et al.* [126]	2019	GA-MPA-S GA-MPA-B	MIMO U-L	×	Neglected & $\mathcal{O}\left(2N_{\text{iter}}d_fM\right)$	Gaussian-approximation based MPA. GA-MPA-S updates the symbol likelihood ratios and GA-MPA-B is performed on bit level. Complexity was pulled down from exponential to linear order.
	Peng *et al.* [170]	2020	EML-MPA	SISO U-L	√	$\mathcal{O}(KM^{d_f}d_f)$ & $\mathcal{O}(KM^{d_f}(d_f+2))$	Dynamic trellis based message passing algorithm is proposed such as the truncation is dynamically decreased as the iterations progress. A channel adaptive version of this algorithm is also introduced.
EPA based detectors	Meng *et al.* [171]	2017	EPA	SIMO U-L	√	$\mathcal{O}(N_{\text{iter}}N_rKMd_f)$	This method is based on approximate Bayesian interference. Instead of continuous exponential terms, message passing is reduced to mean and variance calculation of approximate Gaussian distributions.
	Yuan *et al.* [125]	2018	Stretch-BP-EP	MISO D-L	×	$\mathcal{O}(N_{\text{iter}}(N_t+J)$	Stretched factor graph representation that enables the design of a hybrid BP and EP receiver with channel decoder. It can be only applied to user terminals with single antenna.
	Chen *et al.* [172]	2018	Bayesian Interference	SISO U-L	√	$\mathcal{O}(N(d_f+M))$ & $\mathcal{O}(MNd_f)$	The authors proposed a Monte Carlo Markov chain based SCMA detection algorithm. A new joint update parallel probability sampler is proposed.
	Wang *et al.* [127]	2020	SC-EPA	MIMO U-L	√	Neglected & $\mathcal{O}(Nd_fN_{\text{iter}}(6N_r+5M_p))$	QR decomposition is combined with EPA which allows to exploit the channels sparsity. A high-parallelism message passing technique is proposed. The effect of channel estimation errors is studied.
SD based detectors	Yang *et al.* [173]	2017	SD	SISO U-L	√	$\mathcal{O}(d_fKM^{d_f})$ & $\mathcal{O}(KM^{d_f}(d_f+1))$	Space restriction of codeword research to a hyper-sphere centered around the received signal whose radius is adjusted according to the noise power at the receiver. Method with high complexity in the low SNR region.
	Wei *et al.* [88], [174]	2016 2017	LSD	SISO U-L	√	$\mathcal{O}(2N(d_f\log_2(M))^3)$ & Neglected	The search is limited to a list of all the candidate lattice points within a given radius around the received vector.
	Vameghestahbanati *et al.* [175]	2017	MSD	SISO U-L	√	$\mathcal{O}((4d_f+2)N_{v1}+2N_{v2})$ & $\mathcal{O}((4d_f+2)N_{v1}+2N_{v2})$	Based on a modified tree search method. Using the Tikhonov regularization to facilitate applying SD to SCMA systems.
	Li *et al.* [176]	2018	MSTS	SISO U-L	√	Lesser number of cost function calculations	Modified single tree search avoids the unnecessary node searching by sorting channel matrix in ascending order and by employing a non-zero low bound enumeration.
	Ye *et al.* [167]	2019	QSD-MPA	SISO U-L	√	The cost function is calculated less than $\mathcal{O}\left(\sqrt{sM_p^{d_f+1}}\right)$ times	A quantum search is used to identify all possible codeword combinations within a given hyper-sphere, thereafter MPA is applied on identified points only.
	Vameghestahbanati *et al.* [177]	2019	ISD	SISO U-L	√	$\mathcal{O}((4d_f+2)N_{v1}+2N_{v2})$ & $\mathcal{O}((4d_f+2)N_{v1}+2N_{v2})$	This method extends the MSD one to support any arbitrary regular or irregular constellation topology.
Other SCMA detectors	Sun *et al.* [117]	2019	LDPC-coded SCMA	SISO U-L	√	$\mathcal{O}(660N_{\text{IL}}N_{\text{OL}})$	The interface between the SCMA detector and the LDPC decoder is optimized such as a proof of concept was implemented using 40nm CMOS technology.
	Jiao *et al.* [120]	2020	Polar-coded SCMA	SISO U-L	×	$\mathcal{O}(N_{\text{iter}}Jd_fM^{d_f})$	Iterative joint channel estimation and decoding scheme is proposed where Max-Log-MPA and S-SCL algorithms are used, respectively, for SCMA and polar decoding.
	Lu *et al.* [76]	2018	Deep learning	SISO U-L	×	Not specified	A DNN was designed such as the propagation between two of its layers is calculated based on how messages are passed between FNs and VNs in traditional iterative MPA.
	Lin *et al.* [80]	2020	Deep learning	SISO U-L	×	Not specified	Autoencoders are employed to automatically design SCMA codebooks and construct the corresponding detetor through AWGN channels.

BIBLIOGRAPHY

[1] Altera Innovate Asia FPGA Design Contest. (2015) 5G Algorithm Innovation Competition.

[2] K. Routh and T. Pal, "A survey on technological, business and societal aspects of internet of things by q3, 2017", in *2018 3rd International Conference On Internet of Things: Smart Innovation and Usages (IoT-SIU)*, 2018, pp. 1–4.

[3] L. Dong, H. Zhao, Y. Chen, D. Chen, T. Wang, L. Lu, B. Zhang, L. Hu, L. Gu, B. Li, H. Yang, H. Shen, T. Tian, Z. Luo, and K. Wei, "Introduction on imt-2020 5g trials in china", *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 8, pp. 1849–1866, Aug. 2017.

[4] J.Rodriguez, "Fundamentals of 5g mobile networks", in. Wiley, 2014, ch. 10, pp. 221–240.

[5] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey", *Computer Networks*, vol. 54, no. 15, pp. 2787–2805, 2010.

[6] N. Abu-Ali, A. M. Taha, M. Salah, and H. Hassanein, "Uplink scheduling in lte and lte-advanced: Tutorial, survey and evaluation framework", *IEEE Communications Surveys Tutorials*, vol. 16, no. 3, pp. 1239–1265, 2014.

[7] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Non-orthogonal multiple access for 5g and beyond", *Proceedings of the IEEE*, vol. 105, no. 12, pp. 2347–2381, Dec. 2017.

[8] 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 Communications*, vol. 14, no. 4, pp. 17–25, 2016.

[9] S. M. R. Islam, N. Avazov, O. A. Dobre, and K.-s. Kwak, "Power-Domain Non-Orthogonal Multiple Access (NOMA) in 5G Systems: Potentials and Challenges", *IEEE Communications Surveys & Tutorials*, vol. 19, no. 2, pp. 721–742, 2017.

[10] M. S. Ali, H. Tabassum, and E. Hossain, "Dynamic user clustering and power allocation for uplink and downlink non-orthogonal multiple access (noma) systems", *IEEE Access*, vol. 4, pp. 6325–6343, 2016. DOI: 10.1109/ACCESS.2016.2604821.

[11] Z. Yang, Z. Ding, P. Fan, and N. Al-Dhahir, "A general power allocation scheme to guarantee quality of service in downlink and uplink noma systems", *IEEE Transactions on Wireless Communications*, vol. 15, no. 11, pp. 7244–7257, 2016.

[12] Y. Wu, C. Wang, Y. Chen, and A. Bayesteh, "Sparse code multiple access for 5g radio transmission", in *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*, Sep. 2017, pp. 1–6.

[13] A. Ghaffari, M. Leonardon, A. Cassagne, C. Leroux, and Y. Savaria, "Toward High-Performance Implementation of 5G SCMA Algorithms", *IEEE Access*, vol. 7, pp. 10 402–10 414, Jan. 2019.

[14] ITU-R, *Detailed specifications of the terrestrial radio interface of international mobile telecommunications-2020 (imt-2020)*, Recommendation ITU-R M.2150-0 (2021), Feb. 2021.

[15] C. Yeh, G. D. Jo, Y.-J. Ko, and H. K. Chung, "Perspectives on 6g wireless communications", *ICT Express*, 2022, ISSN: 2405-9595. DOI: https://doi.org/10.1016/j.icte.2021.12.017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S240595952100182X.

[16] M. Sauter, *From GSM to LTE-Advanced Pro and 5G: An Introduction to Mobile Networks and Mobile Broadband*. Wiley, 2017, ISBN: 9781119346937. [Online]. Available: https://books.google.fr/books?id=AEozDwAAQBAJ.

[17] D. Wang, W. Meng, and J. Han, *Security and Privacy in New Computing Environments: Third EAI International Conference, SPNCE 2020, Lyngby, Denmark, August 6-7, 2020, Proceedings*, ser. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering. Springer International Publishing, 2021, ISBN: 9783030669225. [Online]. Available: https://books.google.fr/books?id=O6kWEAAAQBAJ.

[18] Q. Qiu, S. Liu, S. Xu, and S. Yu, "Study on security and privacy in 5g-enabled applications", *Wireless Communications and Mobile Computing*, vol. 2020, p. 8 856 683, Dec. 2020, ISSN: 1530-8669. DOI: 10.1155/2020/8856683. [Online]. Available: https://doi.org/10.1155/2020/8856683.

[19] H. Holma, A. Toskala, and T. Nakamura, *5G Technology: 3GPP New Radio*. Wiley, 2020, ISBN: 9781119236313. [Online]. Available: https://books.google.fr/books?id=hkHADwAAQBAJ.

[20] P. Curwen and J. Whalley, *Understanding 5G Mobile Networks: A Multidisciplinary Primer*. Emerald Publishing Limited, 2021, ISBN: 9781800710368. [Online]. Available: https://books.google.fr/books?id=fKYgEAAAQBAJ.

[21] A. Imoize, O. Adedeji, N. Tandiya, and S. Shetty, "6g enabled smart infrastructure for sustainable society: Opportunities, challenges, and research roadmap", *Sensors*, vol. 21, pp. 1–57, Mar. 2021. DOI: 10.3390/s21051709.

[22] 3GPP, *Study on 3gpp ran sharing enhancements*, R1-163992, 3GPP TSG-RAN WG1 85, Jan 2015.

[23] ——, “5G;Study on scenarios and requirements for next generation access technologies ”, 3rd Generation Partnership Project (3GPP), Technical Report (TR) 38.913, Sep. 2018, Version 15.0.0. [Online]. Available: https://standards.iteh.ai/catalog/standards/etsi/08d3693a-e5df-4610-ac4e-bee031d00f2e/etsi-tr-138-913-v15-0-0-2018-09.

[24] M. Vaezi, Z. Ding, and H. V. Poor, *Multiple Access Techniques for 5G Wireless Networks and Beyond*. Springer, 2019, ISBN: 978-3-319-92089-4.

[25] Navita and Amandeep, “Performance analysis of ofdma, mimo and sc-fdma technology in 4g lte networks”, in *2016 6th International Conference - Cloud System and Big Data Engineering (Confluence)*, 2016, pp. 554–558. DOI: 10.1109/CONFLUENCE.2016.7508181.

[26] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, “A survey of non-orthogonal multiple access for 5g”, *IEEE Communications Surveys Tutorials*, vol. 20, no. 3, pp. 2294–2323, Jul. 2018.

[27] R. Hoshyar, F. P. Wathan, and R. Tafazolli, “Novel low-density signature for synchronous cdma systems over awgn channel”, *IEEE Transactions on Signal Processing*, vol. 56, no. 4, pp. 1616–1626, 2008.

[28] Y. Du, B. Dong, P. Gao, Z. Chen, J. Fang, and S. Wang, “Low-complexity lds-cdma detection based on dynamic factor graph”, in *2016 IEEE Globecom Workshops (GC Wkshps)*, vol. 67, Dec. 2016, pp. 1–6.

[29] L. Wen, “Non-orthogonal multiple access schemes for future cellular systems”, Ph.D. dissertation, University of Surrey, 2016.

[30] Z. Yuan, G. Yu, W. Li, Y. Yuan, X. Wang, and J. Xu, “Multi-user shared access for internet of things”, in *2016 IEEE 83rd Vehicular Technology Conference (VTC Spring)*, May 2016, pp. 1–5.

[31] X. Dai, S. Chen, S. Sun, S. Kang, Y. Wang, Z. Shen, and J. Xu, “Successive interference cancelation amenable multiple access (sama) for future wireless communications”, in *2014 IEEE International Conference on Communication Systems*, Nov. 2014, pp. 222–226.

[32] S. Chen, B. Ren, Q. Gao, S. Kang, S. Sun, and K. Niu, “Pattern division multiple access—a novel nonorthogonal multiple access for fifth-generation radio networks”, *IEEE Transactions on Vehicular Technology*, vol. 66, no. 4, pp. 3185–3196, 2017.

[33] H. Nikopour and H. Baligh, “Sparse code multiple access”, in *2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2013, pp. 332–336.

[34] B. Wang, K. Wang, Z. Lu, T. Xie, and J. Quan, "Comparison study of non-orthogonal multiple access schemes for 5g", in *2015 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, vol. 5, Jun. 2015, pp. 1–5.

[35] M. Moltafet, N. Mokari, M. R. Javan, H. Saeedi, and H. Pishro-Nik, "A New Multiple Access Technique for 5G: Power Domain Sparse Code Multiple Access (PSMA)", *IEEE Access*, vol. 6, pp. 747–759, 2018.

[36] A. Anwar, B.-C. Seet, M. A. Hasan, and X. J. Li, "A survey on application of non-orthogonal multiple access to different wireless networks", *Electronics*, vol. 8, no. 11, p. 1355, Nov. 2019.

[37] 3GPP-RP-163510, "Candidate NR multiple access schemes ", Qualcomm Incorporated, RP, Mar. 2017.

[38] L. Dai, B. Wang, Y. Yuan, S. Han, I. Chih-lin, and Z. Wang, "Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends", *IEEE Communications Magazine*, vol. 53, no. 9, pp. 74–81, 2015. DOI: 10.1109/MCOM.2015.7263349.

[39] 3GPP-RP-170681, " New SI: Study on 5G Non-Orthogonal Multiple Access ", ZTE Corporation, New WI/SI proposals for New Radio, Mar. 2017.

[40] Q. Wang, R. Zhang, L. Yang, and L. Hanzo, " Non-Orthogonal Multiple Access: A Unified Perspective", *IEEE Wireless Communications*, vol. 25, no. 2, pp. 10–16, 2018. DOI: 10.1109/MWC.2018.1700070.

[41] B. Makki, K. Chitti, A. Behravan, and M. S. Alouini, "A Survey of NOMA: Current Status and Open Research Challenges", *IEEE Open Journal of the Communications Society*, vol. 1, pp. 179–189, 2020. DOI: 10.1109/OJCOMS.2020.2969899.

[42] Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, J. Klingner, A. Shah, M. Johnson, X. Liu, Łukasz Kaiser, S. Gouws, Y. Kato, T. Kudo, H. Kazawa, K. Stevens, G. Kurian, N. Patil, W. Wang, C. Young, J. Smith, J. Riesa, A. Rudnick, O. Vinyals, G. Corrado, M. Hughes, and J. Dean, "Google's neural machine translation system: Bridging the gap between human and machine translation", *CoRR*, vol. abs/1609.08144, 2016. [Online]. Available: http://arxiv.org/abs/1609.08144.

[43] B. Ahn, "Real-time video object recognition using convolutional neural network", in *2015 International Joint Conference on Neural Networks (IJCNN)*, 2015, pp. 1–7. DOI: 10.1109/IJCNN.2015.7280718.

[44] B. Banerjee and V. Murino, "Efficient pooling of image based cnn features for action recognition in videos", in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 2637–2641. DOI: 10.1109/ICASSP.2017.7952634.

[45] W. Wang and J. Gang, "Application of convolutional neural network in natural language processing", in *2018 International Conference on Information Systems and Computer Aided Education (ICISCAE)*, 2018, pp. 64–70. DOI: 10.1109/ICISCAE.2018.8666928.

[46] F. Chollet, *Deep Learning with Python*. Manning, Nov. 2017, ISBN: 9781617294433.

[47] R. A. Viswambaran, G. Chen, B. Xue, and M. Nekooei, "Evolutionary design of long short term memory (lstm) ensemble", in *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2020, pp. 2692–2698. DOI: 10.1109/SSCI47803.2020.9308393.

[48] A. S. Shamsabadi, M. Babaie-Zadeh, S. Z. Seyyedsalehi, H. R. Rabiee, and C. Jutten, "A new algorithm for training sparse autoencoders", in *2017 25th European Signal Processing Conference (EUSIPCO)*, 2017, pp. 2141–2145. DOI: 10.23919/EUSIPCO.2017.8081588.

[49] W. Han, G. Wang, and K. Tu, "Latent variable autoencoder", *IEEE Access*, vol. 7, pp. 48 514–48 523, 2019. DOI: 10.1109/ACCESS.2019.2910152.

[50] L. Yasenko, Y. Klyatchenko, and O. Tarasenko-Klyatchenko, "Image noise reduction by denoising autoencoder", in *2020 IEEE 11th International Conference on Dependable Systems, Services and Technologies (DESSERT)*, 2020, pp. 351–355. DOI: 10.1109/DESSERT50317.2020.9125027.

[51] M. Kerner, K. Tammemäe, J. Raik, and T. Hollstein, "Novel architectures for contractive autoencoders with embedded learning", in *2020 17th Biennial Baltic Electronics Conference (BEC)*, 2020, pp. 1–6. DOI: 10.1109/BEC49624.2020.9277246.

[52] N. C. Luong, D. T. Hoang, S. Gong, D. Niyato, P. Wang, Y. Liang, and D. I. Kim, "Applications of deep reinforcement learning in communications and networking: A survey", *IEEE Communications Surveys Tutorials*, vol. 21, no. 4, pp. 3133–3174, 2019.

[53] T. Gruber, S. Cammerer, J. Hoydis, and S. t. Brink, "On deep learning-based channel decoding", in *2017 51st Annual Conference on Information Sciences and Systems (CISS)*, 2017, pp. 1–6. DOI: 10.1109/CISS.2017.7926071.

[54] N. Farsad, M. Rao, and A. Goldsmith, "Deep learning for joint source-channel coding of text", in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 2326–2330. DOI: 10.1109/ICASSP.2018.8461983.

[55] E. Nachmani, E. Marciano, L. Lugosch, W. J. Gross, D. Burshtein, and Y. Be'ery, "Deep learning methods for improved decoding of linear codes", *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 1, pp. 119–131, 2018. DOI: 10.1109/JSTSP.2017.2788405.

[56] N. Samuel, T. Diskin, and A. Wiesel, "Deep mimo detection", in *2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2017, pp. 1–5. DOI: 10.1109/SPAWC.2017.8227772.

[57] C.-K. Wen, W.-T. Shih, and S. Jin, "Deep learning for massive mimo csi feedback", *IEEE Wireless Communications Letters*, vol. 7, no. 5, pp. 748–751, 2018. DOI: 10.1109/LWC.2018.2818160.

[58] N. Ye, H. Han, L. Zhao, and A.-h. Wang, "Uplink nonorthogonal multiple access technologies toward 5g: A survey", *Wireless Communications and Mobile Computing*, vol. 2018, pp. 1–26, Article ID 6187580.

[59] C. Luo, J. Ji, Q. Wang, X. Chen, and P. Li, "Channel state information prediction for 5g wireless communications: A deep learning approach", *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 1, pp. 227–236, 2020. DOI: 10.1109/TNSE.2018.2848960.

[60] S. Dörner, S. Cammerer, J. Hoydis, and S. t. Brink, "Deep learning based communication over the air", *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 1, pp. 132–143, 2018. DOI: 10.1109/JSTSP.2017.2784180.

[61] E. S. Ali, M. K. Hasan, R. Hassan, R. A. Saeed, M. B. Hassan, S. Islam, N. S. Nafi, and S. Bevinakoppa, "Machine learning technologies for secure vehicular communication in internet of vehicles: Recent advances and applications", *Security and Communication Networks*, vol. 2021, p. 8 868 355, Mar. 2021, ISSN: 1939-0114. DOI: 10.1155/2021/8868355. [Online]. Available: https://doi.org/10.1155/2021/8868355.

[62] M. Eisen, C. Zhang, L. F. O. Chamon, D. D. Lee, and A. Ribeiro, "Learning optimal resource allocations in wireless systems", *IEEE Transactions on Signal Processing*, vol. 67, no. 10, pp. 2775–2790, 2019. DOI: 10.1109/TSP.2019.2908906.

[63] W. Cui, K. Shen, and W. Yu, "Spatial deep learning for wireless scheduling", *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 6, pp. 1248–1261, 2019. DOI: 10.1109/JSAC.2019.2904352.

[64] G. Gui, H. Huang, Y. Song, and H. Sari, "Deep Learning for an Effective Nonorthogonal Multiple Access Scheme", *IEEE Transactions on Vehicular Technology*, vol. 67, no. 9, pp. 8440–8450, 2018. DOI: 10.1109/TVT.2018.2848294.

[65] C. Lin, Q. Chang, and X. Li, " A Deep Learning Approach for MIMO-NOMA Downlink Signal Detection", *Sensors*, vol. 19, no. 11, 2019. DOI: 10.3390/s19112526.

[66] J. Luo, J. Tang, D. K. C. So, G. Chen, K. Cumanan, and J. A. Chambers, "A Deep Learning-Based Approach to Power Minimization in Multi-Carrier NOMA With SWIPT", *IEEE Access*, vol. 7, pp. 17 450–17 460, 2019. DOI: 10.1109/ACCESS.2019.2895201.

[67] M. Liu, T. Song, and G. Gui, "Deep Cognitive Perspective: Resource Allocation for NOMA-Based Heterogeneous IoT With Imperfect SIC", *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 2885–2894, 2019. DOI: 10.1109/JIOT.2018.2876152.

[68] N. Ye, X. Li, H. Yu, A. Wang, W. Liu, and X. Hou, "Deep Learning Aided Grant-Free NOMA Toward Reliable Low-Latency Access in Tactile Internet of Things", *IEEE Transactions on Industrial Informatics*, vol. 15, no. 5, pp. 2995–3005, 2019. DOI: 10.1109/TII.2019.2895086.

[69] M. K. Hasan, M. Shahjalal, M. M. Islam, M. M. Alam, M. F. Ahmed, and Y. M. Jang, "The Role of Deep Learning in NOMA for 5G and Beyond Communications", in *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, 2020, pp. 303–307. DOI: 10.1109/ICAIIC48513.2020.9065219.

[70] J. M. Kang, I. M. Kim, and C. J. Chun, "Deep Learning-Based MIMO-NOMA With Imperfect SIC Decoding", *IEEE Systems Journal*, vol. 14, no. 3, pp. 3414–3417, 2020. DOI: 10.1109/JSYST.2019.2937463.

[71] K. Chitti, J. Vieira, and M. Behrooz, "Deep-Learning based Multiuser Detection for NOMA", Nov. 2020. arXiv: 2011.11752.

[72] Narengerile and J. Thompson, "Deep learning for signal detection in non-orthogonal multiple access wireless systems", in *2019 UK/ China Emerging Technologies (UCET)*, 2019, pp. 1–4. DOI: 10.1109/UCET.2019.8881888.

[73] A. T. S and T. Raveendran, *Cnn based channel estimation using noma for mmwave massive mimo system*, 2021. arXiv: 2108.00367 [eess.SP].

[74] M. Kim, N.-I. Kim, W. Lee, and D.-H. Cho, "Deep learning-aided scma", *IEEE Communications Letters*, vol. 22, no. 4, pp. 720–723, 2018. DOI: 10.1109/LCOMM.2018.2792019.

[75] F. Sun, K. Niu, and C. Dong, "Deep learning based joint detection and decoding of non-orthogonal multiple access systems", in *2018 IEEE Globecom Workshops (GC Wkshps)*, 2018, pp. 1–5.

[76] C. Lu, W. Xu, H. Shen, H. Zhang, and X. You, "An enhanced SCMA detector enabled by deep neural network", *CoRR*, vol. abs/1808.08015, 2018.

[77] I. Abidi, M. Hizem, I. Ahriz, M. Cherif, and R. Bouallegue, "Convolutional neural networks for blind decoding in sparse code multiple access", in *2019 15th International Wireless Communications Mobile Computing Conference (IWCMC)*, 2019, pp. 2007–2012. DOI: 10.1109/IWCMC.2019.8766707.

[78] Y. Han, Z. Wang, Q. Guo, and W. Xiang, "Deep learning-based detection for moderate-density code multiple access in iot networks", *IEEE Communications Letters*, vol. 24, no. 1, pp. 122–125, 2020.

[79] S. Sharma and Y. Hong, "A hybrid multiple access scheme via deep learning-based detection", *IEEE Systems Journal*, pp. 1–4, 2020.

[80] J. Lin, S. Feng, Y. Zhang, Z. Yang, and Y. Zhang, "A novel deep neural network based approach for sparse code multiple access", *Neurocomputing*, vol. 382, pp. 52–63, 2020.

[81] M. Vameghestahbanati, I. D. Marsland, R. H. Gohary, and H. Yanikomeroglu, "Multidimensional constellations for uplink scma systems—a comparative study", *IEEE Communications Surveys Tutorials*, vol. 21, no. 3, pp. 2169–2194, 2019, ISSN: 2373-745X. DOI: 10.1109/COMST.2019.2910569.

[82] M. Beko and R. Dinis, "Designing good multi-dimensional constellations", *IEEE Wireless Communications Letters*, vol. 1, no. 3, pp. 221–224, 2012.

[83] J. Peng, W. Chen, B. Bai, X. Guo, and C. Sun, "Joint Optimization of Constellation With Mapping Matrix for SCMA Codebook Design", *IEEE Signal Processing Letters*, vol. 24, no. 3, pp. 264–268, Mar. 2017, ISSN: 1070-9908, 1558-2361. DOI: 10.1109/LSP.2017.2653845.

[84] M. Taherzadeh, H. Nikopour, A. Bayesteh, and H. Baligh, "Scma codebook design", in *2014 IEEE 80th Vehicular Technology Conference (VTC2014-Fall)*, 2014, pp. 1–5.

[85] A. Bayesteh, H. Nikopour, M. Taherzadeh, H. Baligh, and J. Ma, "Low complexity techniques for scma detection", in *2015 IEEE Globecom Workshops (GC Wkshps)*, 2015, pp. 1–6.

[86] M. Taherzadehboroujeni, H. Nikopour, A. Bayesteh, and A. Baligh, *System and method for designing and using multidimensional constellations*, U.S Patent 9,509,379, Nov. 2016.

[87] T. Metkarunchit, "Scma codebook design base on circular-qam", in *2017 Integrated Communications, Navigation and Surveillance Conference (ICNS)*, 2017, 3E1–1–3E1–8.

[88] F. Wei and W. Chen, "Low complexity iterative receiver design for sparse code multiple access", *IEEE Transactions on Communications*, vol. 65, no. 2, pp. 621–634, 2017.

[89] Yan Xin, Zhengdao Wang, and G. B. Giannakis, "Space-time diversity systems based on linear constellation precoding", *IEEE Transactions on Wireless Communications*, vol. 2, no. 2, pp. 294–309, 2003.

[90] C. D'Amours, "Parity bit selected spreading sequences: A block coding approach to spread spectrum", *IEEE Communications Letters*, vol. 9, no. 1, pp. 16–18, 2005.

[91] M. Shi, C. D'Amours, and A. Yongacoglu, "Design of spreading permutations for mimo-cdma based on space-time block codes", *IEEE Communications Letters*, vol. 14, no. 1, pp. 36–38, 2010.

[92] Z. Luo, W. Ma, A. M. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems", *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 20–34, 2010.

[93] S. P. Herath, N. H. Tran, and T. Le-Ngoc, "Rotated multi-d constellations in rayleigh fading: Mutual information improvement and pragmatic approach for near-capacity performance in high-rate regions", *IEEE Transactions on Communications*, vol. 60, no. 12, pp. 3694–3704, 2012.

[94] Q. Xie, Z. Yang, J. Song, and L. Hanzo, "Exit-chart-matching-aided near-capacity coded modulation design and a bicm-id design example for both gaussian and rayleigh channels", *IEEE Transactions on Vehicular Technology*, vol. 62, no. 3, pp. 1216–1227, 2013.

[95] J. Bao, Z. Ma, Z. Ding, G. K. Karagiannidis, and Z. Zhu, "On the Design of Multiuser Codebooks for Uplink SCMA Systems", *IEEE Communications Letters*, vol. 20, no. 10, pp. 1920–1923, Oct. 2016.

[96] J. Bao, Z. Ma, M. A. Mahamadu, Z. Zhu, and D. Chen, "Spherical codes for scma codebook", in *2016 IEEE 83rd Vehicular Technology Conference (VTC Spring)*, 2016, pp. 1–5.

[97] M. Vameghestahbanati, "Hypercube-based multidimensional constellation design for uplink scma systems", in *2020 IEEE International Conference on Communications (ICC)*, 2020.

[98] L. Yu, X. Lei, P. Fan, and D. Chen, "An optimized design of scma codebook based on star-qam signaling constellations", in *2015 International Conference on Wireless Communications Signal Processing (WCSP)*, 2015, pp. 1–5.

[99] L. Yu, P. Fan, D. Cai, and Z. Ma, "Design and analysis of scma codebook based on star-qam signaling constellations", *IEEE Transactions on Vehicular Technology*, vol. 67, no. 11, pp. 10 543–10 553, Nov. 2018, ISSN: 1939-9359. DOI: 10.1109/TVT.2018.2865920.

[100] J. Bao, Z. Ma, M. Xiao, T. A. Tsiftsis, and Z. Zhu, "Bit-interleaved coded scma with iterative multiuser detection: Multidimensional constellations design", *IEEE Transactions on Communications*, vol. 66, no. 11, pp. 5292–5304, 2018.

[101] M. Alam and Q. Zhang, "Performance study of scma codebook design", in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, 2017, pp. 1–5.

[102] C. Yan, G. Kang, and N. Zhang, "A dimension distance-based scma codebook design", *IEEE Access*, vol. 5, pp. 5471–5479, 2017.

[103] K. Lai, J. Lei, L. Wen, G. Chen, W. Li, and P. Xiao, "Secure Transmission With Randomized Constellation Rotation for Downlink Sparse Code Multiple Access System", *IEEE Access*, vol. 6, pp. 5049–5063, 2018.

[104] Z. Mheich, L. Wen, P. Xiao, and A. Maaref, "Design of scma codebooks based on golden angle modulation", *IEEE Transactions on Vehicular Technology*, vol. 68, no. 2, pp. 1501–1509, 2019.

[105] M. A. Imran, Y. A. Sambo, and Q. H. Abbasi, *Enabling 5G Communication Systems to Support Vertical Industries*. John Wiley & Sons Ltd, 2019.

[106] S. Zhang, K. Xiao, B. Xiao, Z. Chen, B. Xia, D. Chen, and S. Ma, "A capacity-based codebook design method for sparse code multiple access systems", in *2016 8th International Conference on Wireless Communications Signal Processing (WCSP)*, 2016, pp. 1–5. DOI: 10.1109/WCSP.2016.7752620.

[107] Y. Zhou, Q. Yu, W. Meng, and C. Li, "Scma codebook design based on constellation rotation", in *2017 IEEE International Conference on Communications (ICC)*, 2017, pp. 1–6. DOI: 10.1109/ICC.2017.7996395.

[108] K. Xiao, B. Xia, Z. Chen, B. Xiao, D. Chen, and S. Ma, "On Capacity-Based Codebook Design and Advanced Decoding for Sparse Code Multiple Access Systems", *IEEE Transactions on Wireless Communications*, vol. 17, no. 6, pp. 3834–3849, Jun. 2018.

[109] M. Kulhandjian and C. D'Amours, "Design of permutation-based sparse code multiple access system", in *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2017, pp. 1–6.

[110] F. R. Kschischang, B. J. Frey, and H. Loeliger, "Factor graphs and the sum-product algorithm", *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 498–519, 2001.

[111] Y. Qi, G. Wu, S. Hu, and Y. Gao, "Parallel-implemented message passing algorithm for scma decoder based on gpgpu", in *2017 9th International Conference on Wireless Communications and Signal Processing (WCSP)*, 2017, pp. 1–6.

[112] W. B. Ameur, P. Mary, M. Dumay, J. Hélard, and J. Schwoerer, "Performance study of mpa, log-mpa and max-log-mpa for an uplink scma scenario", in *2019 26th International Conference on Telecommunications (ICT)*, 2019, pp. 411–416.

[113] K. Xiao, B. Xiao, S. Zhang, Z. Chen, and B. Xia, "Simplified multiuser detection for scma with sum-product algorithm", in *2015 International Conference on Wireless Communications Signal Processing (WCSP)*, 2015, pp. 1–5. DOI: 10.1109/WCSP.2015.7341328.

[114] D. Hui, S. Sandberg, Y. Blankenship, M. Andersson, and L. Grosjean, "Channel coding in 5g new radio: A tutorial overview and performance comparison with 4g lte", *IEEE Vehicular Technology Magazine*, vol. 13, no. 4, pp. 60–69, 2018.

[115] J. H. Bae, A. Abotabl, H.-P. Lin, K.-B. Song, and J. Lee, "An overview of channel coding for 5g nr cellular communications", *APSIPA Transactions on Signal and Information Processing*, vol. 8, e17, 2019.

[116] K. Lai, L. Wen, J. Lei, P. Xiao, A. Maaref, and M. A. Imran, "Sub-Graph Based Joint Sparse Graph for Sparse Code Multiple Access Systems", *IEEE Access*, vol. 6, pp. 25 066–25 080, 2018.

[117] W. Sun, Y. Su, Y. Ueng, and C. Yang, "An ldpc-coded scma receiver with multi-user iterative detection and decoding", *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 66, no. 9, pp. 3571–3584, 2019.

[118] H. Mu, Y. Tang, L. Li, Z. Ma, P. Fan, and W. Xu, "Polar coded iterative multiuser detection for sparse code multiple access system", *China Communications*, vol. 15, no. 11, pp. 51–61, 2018.

[119] Z. Pan, E. Li, L. Zhang, J. Lei, and C. Tang, "Design and optimization of joint iterative detection and decoding receiver for uplink polar coded scma system", *IEEE Access*, vol. 6, pp. 52 014–52 026, 2018.

[120] J. Jiao, K. Liang, B. Feng, Y. Wang, S. Wu, and Q. Zhang, "Joint channel estimation and decoding for polar coded scma system over fading channels", *IEEE Transactions on Cognitive Communications and Networking*, pp. 1–1, 2020.

[121] G. Kang, H. M. Kirn, Y. Shin, and O. Shin, "Multiple antenna mpa detection for uplink scma systems based on selective gaussian approximation", in *2017 23rd Asia-Pacific Conference on Communications (APCC)*, 2017, pp. 1–4.

[122] G. Kang, H. M. Kim, Y. Shin, and O. Shin, "Message passing algorithm based on qr decomposition for an scma system with multiple antennas", in *2017 International Conference on Information and Communication Technology Convergence (ICTC)*, 2017, pp. 941–944.

[123] Z. Wu, C. Zhang, X. Shen, and H. Jiao, "Low complexity uplink sfbc-based mimo-scma joint decoding algorithm", in *2017 3rd IEEE International Conference on Computer and Communications (ICCC)*, 2017, pp. 968–972.

[124] L. Tian, M. Zhao, J. Zhong, and L. Wen, "Resource-selection based low complexity detector for uplink scma systems with multiple antennas", *IEEE Wireless Communications Letters*, vol. 7, no. 3, pp. 316–319, 2018.

[125] W. Yuan, N. Wu, Q. Guo, Y. Li, C. Xing, and J. Kuang, "Iterative receivers for downlink mimo-scma: Message passing and distributed cooperative detection", *IEEE Transactions on Wireless Communications*, vol. 17, no. 5, pp. 3444–3458, 2018.

[126] J. Dai, K. Niu, and J. Lin, "Iterative Gaussian-Approximated Message Passing Receiver for MIMO-SCMA System", *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 3, pp. 753–765, Jun. 2019. (visited on 10/01/2019).

[127] P. Wang, L. Liu, S. Zhou, G. Peng, S. Yin, and S. Wei, "Near-optimal mimo-scma uplink detection with low-complexity expectation propagation", *IEEE Transactions on Wireless Communications*, vol. 19, no. 2, pp. 1025–1037, 2020.

[128] W. Abdessamad, Y. Nasser, K. Y. Kabalan, and O. Bazzi, "On the performance evaluation of mimo-scma systems", in *2016 8th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, 2016, pp. 135–140. DOI: 10.1109/ICUMT.2016.7765346.

[129] S. Zhang, B. Xiao, K. Xiao, Z. Chen, and B. Xia, "Design and analysis of irregular sparse code multiple access", in *2015 International Conference on Wireless Communications Signal Processing (WCSP)*, 2015, pp. 1–5.

[130] L. Yu, P. Fan, Z. Ma, X. Lei, and D. Chen, "An optimized design of irregular scma codebook based on rotated angles and exit chart", in *2016 IEEE 84th Vehicular Technology Conference (VTC-Fall)*, 2016, pp. 1–5.

[131] D. Zhai, "Adaptive codebook design and assignment for energy saving in scma networks", *IEEE Access*, vol. 5, pp. 23 550–23 562, 2017.

[132] S. Han, C. Guo, W. Meng, and C. Li, "A flexible resource scheduling scheme for an adaptive scma system", *Computer Networks*, vol. 129, pp. 384 –391, 2017, Special Issue on 5G Wireless Networks for IoT and Body Sensors.

[133] M. Zhao, S. Zhou, W. Zhou, and J. Zhu, "An Improved Uplink Sparse Coded Multiple Access", *IEEE Communications Letters*, vol. 21, no. 1, pp. 176–179, Jan. 2017.

[134] M. Rebhi, K. Hassan, K. Raoof, and P. Chargé, "An adaptive uplink scma scheme based on channel state information", in *Future Networks (5G and beyond) Workshop*, Paris, France, Mar. 2020, pp. 1–7.

[135] *5g nr qos architecture, qos attribute and qos flow*, Sep. 2020. [Online]. Available: https://www.techplayon.com/5g-nr-qos-architecture-qos-attribute-and-qos-flow/.

[136] Z. Bojkovic, D. Milovanovic, and T. Fowdur, *5G Multimedia Communication: Technology, Multiservices, and Deployment*. CRC Press, 2020, ISBN: 9781000197174. [Online]. Available: https://books.google.fr/books?id=8_T6DwAAQBAJ.

[137] 3GPP, "5G;System Architecture for the 5G System", 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 23.501, Sep. 2018, Version 15.3.0. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/123500_123599/123501/15.03.00_60/ts_123501v150300p.pdf.

[138] Y. Wang, S. Zhou, L. Xiao, X. Zhang, and J. Lian, "Sparse bayesian learning based user detection and channel estimation for scma uplink systems", in *2015 International Conference on Wireless Communications Signal Processing (WCSP)*, 2015, pp. 1–5.

[139] V. P. Klimentyev and A. B. Sergienko, "Detection of SCMA signal with channel estimation error", in *2016 18th Conference of Open Innovations Association and Seminar on Information Security and Protection of Information Technology (FRUCT-ISPIT)*, ISSN: 2305-7254, Apr. 2016, pp. 106–112. DOI: 10.1109/FRUCT-ISPIT.2016.7561515.

[140] A. B. Sergienko and V. P. Klimentyev, "Scma detection with channel estimation error and resource block diversity", in *2016 International Siberian Conference on Control and Communications (SIBCON)*, 2016, pp. 1–5. DOI: 10.1109/SIBCON.2016.7491765.

[141] Y. Wang, X. Zhang, S. Zhou, J. Lian, and L. Xiao, "User detection and channel estimation for scma uplink system in dispersive channel", in *2016 IEEE International Conference on Communication Systems (ICCS)*, 2016, pp. 1–5. DOI: 10.1109/ICCS.2016.7833605.

[142] K. Struminsky, S. Kruglik, D. Vetrov, and I. Oseledets, "A new approach for sparse bayesian channel estimation in scma uplink systems", in *2016 8th International Conference on Wireless Communications Signal Processing (WCSP)*, 2016, pp. 1–5. DOI: 10.1109/WCSP.2016.7752678.

[143] A. B. Sergienko and V. P. Klimentyev, "Spectral efficiency of uplink scma system with csi estimation", in *2017 20th Conference of Open Innovations Association (FRUCT)*, 2017, pp. 391–397. DOI: 10.23919/FRUCT.2017.8071339.

[144] D. Pokamestov, Y. Kryukov, E. Rogozhnikov, E. Mishchenko, and A. Brovkin, "Analysis of the downlink channel estimation methods in communication systems with scma", in *2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*, 2020, pp. 1–6. DOI: 10.1109/ICECCE49384.2020.9179266.

[145] M. Kim, N. Kim, W. Lee, and D. Cho, "Deep learning-aided scma", *IEEE Communications Letters*, vol. 22, no. 4, pp. 720–723, 2018.

[146] C.-P. Wei, H. Yang, C.-P. Li, and Y.-M. Chen, "Scma decoding via deep learning", *IEEE Wireless Communications Letters*, vol. 10, no. 4, pp. 878–881, 2021. DOI: 10.1109/LWC.2020.3048068.

[147] A. Subasi, "Chapter 3 - machine learning techniques", in *Practical Machine Learning for Data Analysis Using Python*, A. Subasi, Ed., Academic Press, 2020, pp. 91–202, ISBN: 978-0-12-821379-7. DOI: https://doi.org/10.1016/B978-0-12-821379-7.00003-5. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128213797000035.

[148] S. M. S. Heerin Yang and Jungmin So, "An ensemble of simple convolutional neural network models for MNIST digit recognition", *CoRR*, vol. abs/2008.10400, Aug. 2020.

[149] M. Moocarme, M. Abdolahnejad, and R. Bhagwat, *The Deep Learning with Keras Workshop: Learn how to define and train neural network models with just a few lines of code*. Packt Publishing, 2020, ISBN: 9781800564756. [Online]. Available: https://books.google.fr/books?id=XxL0DwAAQBAJ.

[150] S. Sun, W. Chen, L. Wang, X. Liu, and T.-Y. Liu, *On the depth of deep neural networks: A theoretical view*, 2015. arXiv: 1506.05232 [cs.LG].

[151] P. Gupta and N. Sehgal, *Introduction to Machine Learning in the Cloud with Python: Concepts and Practices*. Springer International Publishing, 2021, ISBN: 9783030712709. [Online]. Available: https://books.google.fr/books?id=t-grEAAAQBAJ.

[152] K. Janod, M. Morchid, R. Dufour, G. Linarès, and R. De Mori, "Denoised bottleneck features from deep autoencoders for telephone conversation analysis", *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 9, pp. 1809–1820, 2017. DOI: 10.1109/TASLP.2017.2718843.

[153] D. Kingma and J. Ba, "Adam: A method for stochastic optimization", *International Conference on Learning Representations*, Dec. 2014.

[154] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks", *Journal of Machine Learning Research - Proceedings Track*, vol. 9, pp. 249–256, Jan. 2010.

[155] Jia Zou, Hui Zhao, and Wenxiu Zhao, "Low-complexity interference cancellation receiver for sparse code multiple access", in *2015 IEEE 6th International Symposium on Microwave, Antenna, Propagation, and EMC Technologies (MAPE)*, 2015, pp. 277–282.

[156] C. Zhang, Y. Luo, and Y. Chen, "A Low Complexity Detection Algorithm for SCMA", en, Nov. 2016. (visited on 10/29/2019).

[157] C. Zhang, Y. Luo, and Y. Chen, "A low-complexity scma detector based on discretization", *IEEE Transactions on Wireless Communications*, vol. 17, no. 4, pp. 2333–2345, 2018.

[158] L. Yang, Y. Liu, and Y. Siu, "Low complexity message passing algorithm for scma system", *IEEE Communications Letters*, vol. 20, no. 12, pp. 2466–2469, 2016.

[159] L. Tian, M. Zhao, J. Zhong, P. Xiao, and L. Wen, "A low complexity detector for downlink scma systems", *IET Communications*, vol. 11, no. 16, pp. 2433–2439, 2017.

[160] Y. Huang, Y. Li, M. Zhao, X. Xu, and J. Wang, "Two Simplified Multiuser Detection Algorithms for Uplink SCMA Systems via Generalized Approximate Message Passing", in *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*, Sep. 2017, pp. 1–5. DOI: 10.1109/VTCFall.2017.8288416.

[161] J. Dai, K. Niu, C. Dong, and J. Lin, "Improved Message Passing Algorithms for Sparse Code Multiple Access", *IEEE Transactions on Vehicular Technology*, vol. 66, no. 11, pp. 9986–9999, Nov. 2017, ISSN: 0018-9545, 1939-9359. DOI: 10.1109/TVT.2017.2741525.

[162] P. Gao, Y. Du, B. Dong, W. Zhu, Z. Chen, and X. Wang, "Low-Complexity CS-Aided MPA Detector for SCMA Systems", *IEEE Communications Letters*, vol. 22, no. 4, pp. 784–787, Apr. 2018, ISSN: 1089-7798, 1558-2558, 2373-7891. DOI: 10.1109/LCOMM.2017.2779859.

[163] S. Han, Y. Huang, and B. Wang, "A method for analysing and improving the multi-user detection algorithm of scma", in *Wireless Internet*, Springer International Publishing, 2018, pp. 103–115.

[164] J. Dai, G. Chen, K. Niu, and J. Lin, "Partially Active Message Passing Receiver for MIMO-SCMA Systems", *IEEE Wireless Communications Letters*, vol. 7, no. 2, pp. 222–225, Apr. 2018, ISSN: 2162-2337, 2162-2345. DOI: 10.1109/LWC.2017.2767588.

[165] M. Jia, L. Wang, Q. Guo, X. Gu, and W. Xiang, "A low complexity detection algorithm for fixed up-link scma system in mission critical scenario", *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 3289–3297, 2018.

[166] K. Lai, L. Wen, J. Lei, J. Zhong, G. Chen, and X. Zhou, "Simplified sparse code multiple access receiver by using truncated messages", *IET Communications*, vol. 12, no. 16, pp. 1937–1945, 2018.

[167] W. Ye, W. Chen, X. Guo, C. Sun, and L. Hanzo, "Quantum search-aided multi-user detection for sparse code multiple access", *IEEE Access*, vol. 7, pp. 52 804–52 817, 2019.

[168] L. Ma, S. Tong, H. Zheng, B. Bai, and X. Dai, "Edgewise serial message passing detection of uplink scma systems for better user fairness and faster convergence rate", *IEEE Wireless Communications Letters*, vol. 8, no. 4, pp. 1285–1288, 2019.

[169] X. Shi, P. Ren, and D. Xu, "A flexible iterative log-mpa detector for uplink scma systems", in *IoT as a Service*, Springer International Publishing, 2019, pp. 292–302.

[170] X. Peng, Z. Pan, K. Lai, L. Wen, and J. Lei, "Low complexity receiver of sparse code multiple access based on dynamic trellis", *IET Communications*, vol. 14, no. 9, pp. 1420–1427, 2020.

[171] X. Meng, Y. Wu, Y. Chen, and M. Cheng, "Low Complexity Receiver for Uplink SCMA System via Expectation Propagation", in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, San Francisco, CA, USA: IEEE, Mar. 2017, pp. 1–5.

[172] J. Chen, Z. Zhang, S. Fu, and J. Hu, "A Joint Update Parallel MCMC-Method-Based Sparse Code Multiple Access Decoder", *IEEE Transactions on Vehicular Technology*, vol. 67, no. 2, pp. 1280–1291, Feb. 2018, ISSN: 0018-9545, 1939-9359. DOI: 10.1109/TVT.2017.2754552.

[173] L. Yang, X. Ma, and Y. Siu, "Low Complexity MPA Detector Based on Sphere Decoding for SCMA", *IEEE Communications Letters*, vol. 21, no. 8, pp. 1855–1858, Aug. 2017, ISSN: 1089-7798, 1558-2558, 2373-7891. DOI: 10.1109/LCOMM.2017.2697425.

[174] F. Wei and W. Chen, "A low complexity scma decoder based on list sphere decoding", in *2016 IEEE Global Communications Conference (GLOBECOM)*, 2016, pp. 1–6.

[175] M. Vameghestahbanati, E. Bedeer, I. Marsland, R. H. Gohary, and H. Yanikomeroglu, "Enabling Sphere Decoding for SCMA", *IEEE Communications Letters*, vol. 21, no. 12, pp. 2750–2753, Dec. 2017, ISSN: 1089-7798, 1558-2558, 2373-7891. DOI: 10.1109/LCOMM.2017.2747550.

[176] L. Li, J. Wen, X. Tang, and C. Tellambura, "Modified Sphere Decoding for Sparse Code Multiple Access", *IEEE Communications Letters*, vol. 22, no. 8, pp. 1544–1547, Aug. 2018, ISSN: 1089-7798, 1558-2558, 2373-7891. DOI: 10.1109/LCOMM.2018.2848273.

[177] M. Vameghestahbanati, I. Marsland, R. H. Gohary, and H. Yanikomeroglu, "A novel sd-based detection for generalized scma constellations", *IEEE Transactions on Vehicular Technology*, vol. 68, no. 10, pp. 10 278–10 282, 2019.

DOCTORAT BRETAGNE SCIENCES LOIRE POUR L'INGENIEUR

Le Mans Université

UNIVERSITÉ DE NANTES

Titre : Contributions sur les techniques d'accès multiple non-orthogonal pour les communications massives

Mot clés : Techniques d'accés multiple, Communication massive, SCMA, CD-NOMA, Apprentissage automatique

Résumé : Les techniques d'accès multiple présentent de nombreux défis et possibilités pour la conception de réseaux sans fil massifs. Par conséquent, d'importants efforts de recherche ont été consacrés au problème de la distribution égale et simultanée des ressources partagées (temps et/ou fréquence) entre des utilisateurs d'un même réseau. Ainsi, l'amélioration des techniques d'accès multiple des prochaines générations de communications mobiles mérite une étude approfondie, ce qui est l'objectif principal de cette thèse. Le travail de recherche présenté dans cette thèse se concentre sur l'accès multiple par code parcimonieux. Dans un premier temps, nous étudions l'adaptation du SCMA en fonction des besoins des utilisateurs en termes d'énergie, de bande passante et de qualité de service. L'architecture SCMA adaptative proposée non seulement prend en compte les différences entre les besoins des utilisateurs, mais permet également une utilisation plus réaliste de la connaissance des canaux de transmission en personnalisant le livre de codes de chaque groupe d'utilisateurs qui sont regroupés en fonction de leurs informations sur l'état du canal (CSI). La deuxième partie concerne l'application de techniques d'apprentissage profond dans le but de décoder efficacement le signal SCMA reçu. Ainsi, un détecteur SCMA en deux étapes basé sur l'apprentissage profond a été proposé dans l'hypothèse d'un bruit blanc Gaussien additif. La première étape consiste à débruiter le signal à l'aide d'un autoencodeur de débruitage avant de le décoder ensuite sur la base d'un réseau neuronal profond qui permet d'estimer simultanément tous les bits transmis en une seule fois. Les performances du détecteur SCMA en termes de taux d'erreur binaire et sa complexité ont été évaluées. Cependant, les performances de cette méthode sont légèrement inférieures à celles du détecteur SCMA conventionnel, étant plutôt itératif. Néanmoins, cette comparaison n'est pas juste, car le détecteur que nous avons proposé en premier lieu, contrairement au détecteur conventionnel, suppose que le livre de code SCMA est inconnu au niveau du récepteur. C'est pourquoi nous proposons un nouveau détecteur à réseau neuronal profond basé sur la distance en supposant que le livre de codes est connu. Le deuxième détecteur proposé peut être comparé équitablement au détecteur conventionnel de SCMA et surpasse ses performances.

Title: Contributions on non-orthogonal multiple access techniques for massive communications

Keywords: Multiple access techniques, Massive communication, SCMA, CD-NOMA, Machine learning

Abstract: Multiple access techniques present many challenges and opportunities for the design of massive wireless networks. Therefore, substantial research efforts were devoted to the problem of serving users of the same network equally and simultaneously with some shared resources (time and/or frequency). Thus, the improvement of multiple access techniques of the next generations of mobile communications deserves a thorough study, which is the main objective of this thesis. The research work presented in this thesis focuses on sparse code multiple access (SCMA) and it is organized into two main parts. First, we study the adaptation of SCMA according to the users' needs in terms of energy, bandwidth and quality of service. The proposed adaptive SCMA architecture not only rightfully addresses the differences in users requirements, but also allows a more realistic use of the knowledge of transmission channels by customizing the codebook of each group of users which are clustered based on their channel state information. The second part concerns the application of deep learning techniques in the aim of efficiently decoding the received SCMA signal. Thus, a two-stage deep learning based SCMA detector was proposed under the assumption of additive white Gaussian noise. The first stage consists in denoising the signal using a denoising autoencoder before decoding it afterwards based on a deep neural network which allows to simultaneously estimate all the transmitted bits in one-shot. The performance of SCMA detector in terms of bit error rate and its complexity were evaluated. However, the performance of this method is slightly worse than that of the conventional SCMA detector, being rather iterative. Nevertheless, this comparison is not fair, since our firstly proposed detector, contrary to the conventional one, assumes that the SCMA codebook is unknown at the receiver. That is why, we propose a new distance-based deep neural network detector under the assumption that the codebook is known. The second proposed detector can be fairly compared to the conventional detector of SCMA and outperforms its performance.