RESOURCE ALLOCATION FOR DOWNLINK NON-ORTHOGONAL MULTIPLE ACCESS (NOMA) SYSTEM

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy in the Faculty of Science and Engineering

2017

By
Ziad Qais Abdulkareem Al-Abbasi
School of Electrical and Electronic Engineering
## Contents

Abstract 10

Declaration 11

Copyright 12

Acknowledgements 13

Dedication 14

List of Abbreviations 15

List of Mathematical Notations 19

List of Variables 20

1 Introduction 23
   1.1 Motivation 24
   1.2 Aims and Objectives 25
   1.3 Key Contributions 26
   1.4 Thesis Organization 28
   1.5 List of Publications 30

2 Non-Orthogonal Multiple Access (NOMA) for 5G Networks 31
   2.1 Introduction 31
   2.2 Wireless Channel Models 31
      2.2.1 Large Scale Fading Effect 32
      2.2.2 Small Scale Fading Effect 33
   2.3 Overview of Orthogonal Frequency Division Multiplexing (OFDM) 35
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.4</td>
<td>Overview of Non-Orthogonal Multiple Access (NOMA) System</td>
<td>37</td>
</tr>
<tr>
<td>2.4.1</td>
<td>Superposition Coding</td>
<td>39</td>
</tr>
<tr>
<td>2.4.2</td>
<td>Successive Interference Cancellation</td>
<td>40</td>
</tr>
<tr>
<td>2.4.3</td>
<td>Related Works on Single-Antenna NOMA System</td>
<td>44</td>
</tr>
<tr>
<td>2.4.4</td>
<td>Related works on Multiple-Antenna NOMA System</td>
<td>48</td>
</tr>
<tr>
<td>2.5</td>
<td>NOMA for Small Cells Dense Networks</td>
<td>51</td>
</tr>
<tr>
<td>2.5.1</td>
<td>Challenges Against Dense Cell-Deployment</td>
<td>52</td>
</tr>
<tr>
<td>2.5.2</td>
<td>Modeling Cell-Deployment and Performance Metrics</td>
<td>53</td>
</tr>
<tr>
<td>2.6</td>
<td>Chapter Summary</td>
<td>57</td>
</tr>
<tr>
<td>3</td>
<td>Sum Rate Maximization for NOMA System</td>
<td>58</td>
</tr>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>58</td>
</tr>
<tr>
<td>3.2</td>
<td>Problem Formulation and Solution for Two-User Scenario</td>
<td>59</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Problem Solution using Lagrange Dual Function</td>
<td>61</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Equal per RB Power Allocation (ERPA)</td>
<td>63</td>
</tr>
<tr>
<td>3.2.3</td>
<td>Average Channel based Power Allocation (ACPA)</td>
<td>65</td>
</tr>
<tr>
<td>3.3</td>
<td>Multiuser NOMA with Hierarchical Pairing Concept</td>
<td>66</td>
</tr>
<tr>
<td>3.3.1</td>
<td>Hierarchical Pairing Power Allocation (HPPA)</td>
<td>70</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Equal-per-Pair Power Allocation</td>
<td>71</td>
</tr>
<tr>
<td>3.3.3</td>
<td>Proportional-per-Pair Power Allocation</td>
<td>72</td>
</tr>
<tr>
<td>3.3.4</td>
<td>Subband-based ERPA</td>
<td>73</td>
</tr>
<tr>
<td>3.4</td>
<td>The Proposed Hybrid System</td>
<td>73</td>
</tr>
<tr>
<td>3.4.1</td>
<td>RB Allocation and Classification</td>
<td>75</td>
</tr>
<tr>
<td>3.4.2</td>
<td>Power Allocation for the Hybrid System</td>
<td>76</td>
</tr>
<tr>
<td>3.5</td>
<td>Numerical Results</td>
<td>78</td>
</tr>
<tr>
<td>3.5.1</td>
<td>Two-user Scenario</td>
<td>79</td>
</tr>
<tr>
<td>3.5.2</td>
<td>Multiuser Scenario</td>
<td>80</td>
</tr>
<tr>
<td>3.5.3</td>
<td>Hybrid Multiple Access Against NOMA System</td>
<td>84</td>
</tr>
<tr>
<td>3.6</td>
<td>Chapter Summary</td>
<td>86</td>
</tr>
<tr>
<td>4</td>
<td>EE-SE Trade-off for NOMA System</td>
<td>88</td>
</tr>
<tr>
<td>4.1</td>
<td>Introduction</td>
<td>88</td>
</tr>
<tr>
<td>4.2</td>
<td>System Model</td>
<td>88</td>
</tr>
<tr>
<td>4.3</td>
<td>EE-SE Optimization</td>
<td>89</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Problem Formulation</td>
<td>89</td>
</tr>
<tr>
<td>4.3.2</td>
<td>The Proposed Subgradient based Solution</td>
<td>92</td>
</tr>
</tbody>
</table>
5 MIMO-NOMA Systems

5.1 System Model ............................................. 110
5.2 Linear Precoder for MISO-NOMA System ................. 112
  5.2.1 Linear Precoding .................................... 112
  5.2.2 Precoder Optimization .............................. 114
5.3 IA based MIMO-NOMA System ............................ 115
  5.3.1 Designing the IA precoders .......................... 115
  5.3.2 Singular Value Decomposition (SVD) .............. 116
  5.3.3 Designing MIMO Precoders using SVD ........... 117
  5.3.4 Designing MIMO Precoders using DPC ........... 117
5.4 Problem Formulation and Proposed Solutions ........... 118
  5.4.1 Low Complexity Closed-form Solution ............ 118
  5.4.2 Subgradient based Power Allocation ............. 123
5.5 Numerical Results ....................................... 126
  5.5.1 MISO-NOMA System ................................ 127
  5.5.2 MIMO-NOMA System with IA ...................... 129
5.6 Chapter Summary .................................... 133

6 NOMA based Homogeneous Small Cells Dense Network (Hom-SCDN)

6.1 Introduction ............................................ 135
6.2 System Model of NOMA based Hom-SCDN ................ 136
  6.2.1 Problem Formulation and Solution ................ 136
  6.2.2 Approximated Model and EE Optimization Problem .. 140
    6.2.2.1 Approximated Model .......................... 140
    6.2.2.2 Power Minimization Problem formulation and
             Solution ....................................... 144
6.3 Numerical Results .................................... 147
  6.3.1 Numerical results for the sum rate optimization schemes 148
  6.3.2 Numerical results for the approximated model and EE
          optimization .................................... 151
6.4 Chapter Summary .................................... 153
7 Conclusions and Future Work 154
  7.1 Conclusions ................................................. 154
  7.2 Future Work ................................................... 156

References 158

A A proof that EE is continuously differentiable in terms of SE 173

B The Grid Structure of the LTE Resource Block 174
  B.1 Introduction .................................................... 174

C Overview of Radio Resource Allocation 177
  C.1 Introduction ...................................................... 177
     C.1.1 RRA Classification in terms of Decision Control .... 178
     C.1.2 RRA Classification in terms of Allocation Mechanism 178
     C.1.3 RRA Optimization Techniques ......................... 178

D Constrained Optimization 180
  D.1 A General Problem Formulation ............................ 180
  D.2 Problem Solution Using Lagrange Dual Method .......... 182
List of Tables

2.1 ITU Pedestrian model [1] ........................................ 34
2.2 Summary of the related works to single antennas - NOMA system. ......................................................... 47
2.3 Summary of the related works to multiple antennas - NOMA system. ........................................................ 50
3.1 Simulation parameters [2–4] ........................................ 79
4.1 Simulation Parameters. [2, 5]. ................................................ 98
6.1 Simulation parameters [2–4] ........................................ 148
B.1 Bandwidth divisions according to LTE ................................. 175
# List of Figures

2.1 Classification of wireless channel models [6]. .............................. 32
2.2 Orthogonality of OFDM spectrum with eight subcarriers, the subcarriers do not interfere with each other because at the peak of each subcarrier; the signals from other subcarriers are zero [7]. .................................................................................................................................................................................................................................................. 35
2.3 OFDM Transceiver [8,9]. ................................................................. 36
2.4 Power spectrum illustration in terms of RBs for: (a). OFDMA (b). NOMA ................................................................................................................................. 36
2.5 SIC receiver of two users NOMA system with $\sqrt{P_{1,s}} > \sqrt{P_{2,s}}$ (i.e. $|\tilde{h}_{2,s}|^2 > |\tilde{h}_{1,s}|^2$) [10,11] ................................................................. 40
2.6 Rate region boundaries achieved by two users in the downlink of orthogonal and non-orthogonal cases [12,13]. ................................. 44
2.7 Grid Cells Pattern ............................................................................ 55
2.8 Voronoi diagram for Matern Distribution and Poisson Distribu- tion ........................................................................................................... 56
3.1 NOMA structure with HP concept. ..................................................... 67
3.2 Illustration of the SIC process with perfect decoding in hierarchically paired NOMA with four users (two pairs: $Z = 2$), where the users within the pairs are arranged in an increasing order. ....................................................................................................................... 69
3.3 Illustration of the power allocation stages in HPPA method for 8 users. ......................................................................................................................... 72
3.4 Structure of the hybrid orthogonal - non orthogonal scheme. ........... 74
3.5 Sum rate against different values of $P_t$. ........................................... 80
3.6 Sum rate against different cell diameters with $P_t = 30 \text{ dBm}$. ....... 81
3.7 CDF against the achievable rate of each single user using two different $P_t$ levels. .......................................................... 81
3.8 Sum rate against different values of $P_t$ for $K = 16$ .................. 83
3.9 Coverage probability of 10 users and $P_t = 30 \text{ dBm}$ against different target rates .......................................................... 83
3.10 Sum rate against different number of users with $P_t = 20 \text{ dBm}$. 84
3.11 Sum rate of the hybrid scheme against HPPA based NOMA system with $P_t = 20 \text{ dBm}$, and a cell diameter of: (a). diameter=300m. (b). diameter=1000m. ................................. 85
3.12 Sum rate of the hybrid scheme against HPPA based NOMA system with $P_t = 20 \text{ dBm}$, $K = 10$, and a cell diameter of: (a). diameter=300m. (b). diameter=1000m. ................................. 86
4.1 Convergence performance of the proposed subgradient-based NOMA system where $P_t = 30\text{dBm}$. ................................. 99
4.2 Achievable sum rate against increasing transmission power for:
(a) $K = 30$. (b) $K = 10$. ......................................................... 100
4.3 EE comparison against increasing transmission power for $K = 15$ with $P_c = 10W$. .............................................................. 101
4.4 EE versus SE comparison as a function of various minimum rate requirements (i.e., $\Phi_i = [1.5 \, 1.8 \, 2.0 \, 2.4 \, 2.8 \, 3.0] \text{Mbps}$ and $\Phi_j = 0.5\Phi_i$) with $K = 15$ and $P_t = 40\text{dBm}$, for: (a) $P_c = 5W$. (b) $P_c = 10W$. .............................................................. 102
4.5 EE comparison against increasing number of users with $P_t = 30\text{dBm}$ and: (a) $P_c = 5W$. (b) $P_c = 10W$. ......................................................... 104
4.6 EE comparison against increasing number of users with $P_t = 30\text{dBm}$ and $\zeta = 0.1\nu_j/b$, for: (a) $P_c = 5W$. (b) $P_c = 10W$. ......................................................... 105
4.7 EE-SE trade-off comparison as a function of increasing number of RBs (i.e., varying the number of RBs as $[50, 55, 60, 65, \ldots, 100]$ m) with $P_t = 30\text{dBm}$, $K = 15$, for: (a) $P_c = 5W$. (b) $P_c = 10W$. ......................................................... 106
4.8 EE-SE trade-off behavior as a function of different cell diameters (i.e., varying the diameter as $[200, 300, 400, \ldots, 1000]$ m) with $P_t = 30\text{dBm}$, $K = 15$, for: (a) $P_c = 5W$. (b) $P_c = 10W$. ......................................................... 107
5.1 Sum rate versus various transmission power levels with $K = 8$, $D = 8$ and $M = 8$. .............................................................. 128
5.2 Sum rate versus increasing number of antennas with $P_t = 30dBm$, $D = 4$ and $K = 4$.
5.3 Sum rate vs different numbers of users per each group with $P_t = 30dBm$ and $K = 24$ with SVD and DPC precoding.
5.4 Sum rate versus various transmission power levels with IA and SVD precoding with $D = 3$ and $K = 6$.
5.5 Sum rate versus various transmission power levels with IA and DPC precoding with $D = 3$ and $K = 6$.
5.6 Sum rate versus increasing number of users with IA and SVD precoding with $D = 6$ and $P_t = 30dBm$.
5.7 Sum rate versus increasing number of users with IA and DPC precoding with $D = 6$ and $P_t = 30dBm$.

6.1 SCDN with UFR planning for $\rho_K = 2.5 \times 10^{-4} m^{-2}$ and $\rho_{BS} = 0.5 \times 10^{-4} m^{-2}$.
6.2 SCDN with 3FR planning for $\rho_K = 2.5 \times 10^{-4} m^{-2}$ and $\rho_{BS} = 0.5 \times 10^{-4} m^{-2}$. BSs with the same color use the same RBs.
6.3 Sum rate against different BSs densities for NOMA based SCDN with the users at a density of $\rho_K = 2.5 \times 10^{-4} m^{-2}$.
6.4 The average rate achieved by each individual cell for various BS densities for NOMA based SCDN.
6.5 The overall sum rate achieved by multicell for various BS densities for NOMA based SCDN.
6.6 Overall EE for NOMA based SCDN using different BS densities ($\rho_{BS}$) with $P_c = 2W$, $\Phi_{min}^{BS} = 5Mbps$, $\rho_K = 3 \times \rho_{BS}$ and $P_t = 25dBm$.

B.1 RB Grid Structure (a) Uplink (b) Downlink. [3]
Abstract

In wireless networks, the exponentially increasing demands for wireless services are encountered by the scarcity of the available radio resources. More bandwidth is required for not only accommodating the increasing number of users, but also to meet the requirements of the new services such as TV on demand, wireless gaming, and mobile Internet. Non-orthogonal multiple access (NOMA) has attracted a great attention recently due to its superior spectral efficiency (SE) over orthogonal multiple access and could play a vital role in improving the capacity of future networks. In particular, power based NOMA multiplexes the users in power domain via superposition coding (SC) and allows them to access the whole spectrum simultaneously while using successive interference cancellation (SIC) at the receiver side for signal detection. Since NOMA exploits the power domain for multiple access, power allocation is vital to achieve superior SE with NOMA. Resource allocation and its optimization are general methods used to further improve the NOMA based networks performance. In this thesis, the resource allocation in the downlink NOMA system is considered and optimized for different objective functions such as the sum rate and the energy efficiency (EE). In addition, the combination of NOMA and multiple antenna is considered using linear and non-linear precoders. In all the considered cases, suboptimal power allocation schemes are proposed and compared to the numerically obtained optimal one. Results confirm that NOMA outperforms OFDMA. It also support the effectiveness of the proposed schemes as compared to the existing ones and to the optimal one. The results also reveal that using multiple antennas with NOMA can significantly enhance the overall performance. Furthermore, a NOMA-multicell scenario is considered to test the proposed schemes under the effect of intercell interference (ICI). The results prove that the proposed methods effective as compared to the optimal one at a much lower complexity.
Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.
Copyright

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://www.campus.manchester.ac.uk/medialibrary/policies/intellectual-property.pdf), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University’s policy on presentation of Theses.
Acknowledgements

I would like to express my utmost gratitude to my Ph.D. supervisor Dr. Daniel K. C. So for his patience, constant encouragement, guidance and support in directing and guiding my research over the years. This work would not have been possible without his rich expertise, excellent judgment, enthusiasm and dedication to top-quality research. I will forever be grateful.

Many thanks to my friends in the Middle Technical University (Baqubah Technical Institute), for their support during my study. In particular, I would like to express my gratitude to my friends and all the staff in the department of Electrical Techniques for their kind words in time of need.

I would like to express my thanks and gratitude to the Iraqi government and especially to the Higher Committee for Education Development (HCED) in Iraq for awarding me this scholarship and for continuously supporting me throughout my study.

I would also like to acknowledge the generous financial support from the HCED-Iraq, which enabled me to pursue this PhD research without any financial difficulties.

Last but not least, I would like to thank my family, my in-laws, and my parents, for their endless love, patience and support.
Dedication

With deepest gratitude,
I dedicate this thesis to my father,

Mr. Qais Abdulkareem Al-Abbasi,

who has been my continuous source of support, knowledge and inspiration,
Thank you for everything,
# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1G</td>
<td>First Generation mobile network</td>
</tr>
<tr>
<td>2G</td>
<td>Second Generation mobile network</td>
</tr>
<tr>
<td>3FR</td>
<td>Frequency Reuse with a reuse factor of 3</td>
</tr>
<tr>
<td>3G</td>
<td>Third Generation mobile network</td>
</tr>
<tr>
<td>3GPP</td>
<td>Third Generation Partnership Project</td>
</tr>
<tr>
<td>4G</td>
<td>Fourth Generation mobile network</td>
</tr>
<tr>
<td>5G</td>
<td>Fifth Generation mobile network</td>
</tr>
<tr>
<td>ACPA</td>
<td>Average Channel based Power Allocation</td>
</tr>
<tr>
<td>AWGN</td>
<td>Additive White Gaussian Noise</td>
</tr>
<tr>
<td>BC</td>
<td>Broadcast Channel</td>
</tr>
<tr>
<td>BER</td>
<td>Bit Error Rate</td>
</tr>
<tr>
<td>BS</td>
<td>Base Station</td>
</tr>
<tr>
<td>CCI</td>
<td>Co-Channel Interference</td>
</tr>
<tr>
<td>CDF</td>
<td>Cumulative Distribution Function</td>
</tr>
<tr>
<td>CDMA</td>
<td>Code Division Multiple Access</td>
</tr>
<tr>
<td>CFO</td>
<td>Carrier Frequency Offset</td>
</tr>
<tr>
<td>CR</td>
<td>Cognitive Radio</td>
</tr>
<tr>
<td>D2D</td>
<td>Device to Device</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>DPC</td>
<td>Dirty Paper Coding</td>
</tr>
<tr>
<td>EE</td>
<td>Energy Efficiency</td>
</tr>
<tr>
<td>EPA</td>
<td>Equal per stream Power Allocation</td>
</tr>
<tr>
<td>ERPA</td>
<td>Equal per RB Power Allocation</td>
</tr>
<tr>
<td>FDMA</td>
<td>Frequency Division Multiple Access</td>
</tr>
<tr>
<td>FPA</td>
<td>Fixed Power Allocation</td>
</tr>
<tr>
<td>FRA</td>
<td>Future Radio Access</td>
</tr>
<tr>
<td>FSPA</td>
<td>Full Search Power Allocation</td>
</tr>
<tr>
<td>FTPA</td>
<td>Fractional Transmit Power Allocation</td>
</tr>
<tr>
<td>HetNet</td>
<td>Heterogeneous Network</td>
</tr>
<tr>
<td>Hom-SCDN</td>
<td>Homogeneous Small Cells Dense Network</td>
</tr>
<tr>
<td>HP</td>
<td>Hierarchical Pairing</td>
</tr>
<tr>
<td>HPPA</td>
<td>Hierarchical Pairing based Power Allocation</td>
</tr>
<tr>
<td>IA</td>
<td>Interference Alignment</td>
</tr>
<tr>
<td>ICI</td>
<td>Inter-Cell Interference</td>
</tr>
<tr>
<td>ISI</td>
<td>Inter-Symbol Interference</td>
</tr>
<tr>
<td>LDD</td>
<td>Lagrange Dual Decomposition</td>
</tr>
<tr>
<td>LOS</td>
<td>Line of Sight</td>
</tr>
<tr>
<td>LTE</td>
<td>Long Term Evolution</td>
</tr>
<tr>
<td>M2M</td>
<td>Machine to Machine</td>
</tr>
<tr>
<td>MA</td>
<td>Margin Adaptive</td>
</tr>
<tr>
<td>MAC</td>
<td>Media Access Control</td>
</tr>
<tr>
<td>MAI</td>
<td>Multiple Access Interference</td>
</tr>
<tr>
<td>Acronym</td>
<td>Definition</td>
</tr>
<tr>
<td>-----------</td>
<td>----------------------------------------------------</td>
</tr>
<tr>
<td>MBB</td>
<td>Mobile BroadBand</td>
</tr>
<tr>
<td>MHP</td>
<td>Matern Hard Process</td>
</tr>
<tr>
<td>MIMO</td>
<td>Multiple-Input Multiple-Output</td>
</tr>
<tr>
<td>MISO</td>
<td>Multiple-Input Single Output</td>
</tr>
<tr>
<td>MMSE</td>
<td>Minimum Mean Square Error</td>
</tr>
<tr>
<td>MPA</td>
<td>Message Passing Algorithm</td>
</tr>
<tr>
<td>MPSK</td>
<td>Multilevel Phase Shift Keying</td>
</tr>
<tr>
<td>MQAM</td>
<td>Multilevel Quadrature Amplitude Modulation</td>
</tr>
<tr>
<td>MUSA</td>
<td>Multi-User Shared Access</td>
</tr>
<tr>
<td>NLOS</td>
<td>Non-Line of Sight</td>
</tr>
<tr>
<td>NOMA</td>
<td>Non-Orthogonal Multiple Access</td>
</tr>
<tr>
<td>OBF</td>
<td>Opportunistic Beamforming</td>
</tr>
<tr>
<td>OFDMA</td>
<td>Orthogonal Frequency Division Multiple Access</td>
</tr>
<tr>
<td>OMA</td>
<td>Orthogonal Multiple Access</td>
</tr>
<tr>
<td>PDMA</td>
<td>Pattern Division Multiple Access</td>
</tr>
<tr>
<td>PPP</td>
<td>Poisson Point Process</td>
</tr>
<tr>
<td>QoS</td>
<td>Quality of Service</td>
</tr>
<tr>
<td>RA</td>
<td>Rate Adaptive</td>
</tr>
<tr>
<td>RB</td>
<td>Resource Block</td>
</tr>
<tr>
<td>RRA</td>
<td>Radio Resource Allocation</td>
</tr>
<tr>
<td>SA</td>
<td>Spatially Average</td>
</tr>
<tr>
<td>SC</td>
<td>Superposition Coding</td>
</tr>
<tr>
<td>SC-FDMA</td>
<td>Single Carrier-Frequency Division Multiple Access</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>-------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>SCDN</td>
<td>Small Cells Dense Network</td>
</tr>
<tr>
<td>SCMA</td>
<td>Sparse Code Multiple Access</td>
</tr>
<tr>
<td>SE</td>
<td>Spectral Efficiency</td>
</tr>
<tr>
<td>SIC</td>
<td>Successive Interference Cancellation</td>
</tr>
<tr>
<td>SINR</td>
<td>Signal to Interference plus Noise Ratio</td>
</tr>
<tr>
<td>SISO</td>
<td>Single-Input Single-Output</td>
</tr>
<tr>
<td>SON</td>
<td>Self Originizing Network</td>
</tr>
<tr>
<td>SVD</td>
<td>Singular Value Decomposition</td>
</tr>
<tr>
<td>SWIPT</td>
<td>Simultaneous Wireless Information and Power Transfer</td>
</tr>
<tr>
<td>TDMA</td>
<td>Time Division Multiple Access</td>
</tr>
<tr>
<td>UFPA</td>
<td>User’s index based Fixed Power Allocation</td>
</tr>
<tr>
<td>UFR</td>
<td>Universal Frequency Reuse</td>
</tr>
<tr>
<td>VLC</td>
<td>Visible Light Communication</td>
</tr>
<tr>
<td>ZF</td>
<td>Zero Forcing</td>
</tr>
</tbody>
</table>
List of Mathematical Notations

\[ \prod \] Product symbol
\[ \exp(x) \] Exponential function \( e^x \)
\[ \sum \] Summation symbol
\[ j \] Imaginary unit \( j = \sqrt{-1} \)
\[ I_N \] Identity matrix of size \( N \times N \)
\[ \text{tr}(X) \] Trace of matrix \( X \)
\[ X^* \] Hermitian transpose of matrix \( X \)
\[ \log_x(.) \] Logarithmic function to base \( x \)
\[ \ln(.) \] Natural logarithm
\[ |.| \] Magnitude of a complex number
\[ \text{min} \] Argument of the minimum
\[ \text{max} \] Argument of the maximum
\[ O(.) \] Complexity order
List of Variables

\[ \alpha \] The weight factor of Dinkelbach function.

\[ B_s \] The resource block bandwidth.

\[ \beta \] The power allocation factor.

\[ d \] The distance between the BS and a user.

\[ d_0 \] The reference distance.

\[ D \] Number of data streams.

\[ \Delta \] The convergence tolerance.

\[ G \] The fading effect of MIMO channel gain for each user.

\[ \bar{G} \] The average channel gain for each user.

\[ \gamma_s \] The received SINR over the s-th RB.

\[ \hat{h}_{k,s} \] The channel between the BS and the user.

\[ \delta \] The beamforming normalizing factor.

\[ (H) \] Strong or good channel conditions.

\[ \mathbf{H}_k \] Channel matrix between the serving BS and user k.

\[ I_{\text{max}} \] The maximum number of subgradient iterations.

\[ I_s \] The power allocated to the stronger user pairs over the s-th RB.
$K$ Total number of users.

$(L)$ Weak or low channel conditions.

$L$ Total number of small cells.

$\lambda$ Lagrangian multiplier.

$L$ DPC LQ-decomposed lower triangular matrix.

$M$ Total number of transmitting antennas at the BS.

$\bar{M}$ Sum of the channel powers of each user pair.

$\mu$ Lagrangian multiplier.

$N$ Total number of receiving antennas at each user.

$N_c$ Total number of subcarriers.

$N_0$ The noise power spectral density.

$n_{k,s}$ The AWGN variance.

$O$ The sets of the users indices.

$\Omega$ The sets of the RBs indices.

$P_{k,s}$ The transmitted power allocated to the k-th user over the s-th RB.

$PL$ The path loss.

$PL_0$ The path loss at reference distance.

$P_{tx,i}$ Transmitted power.

$P_t$ Total power transmitted by the BS.

$\Phi_{min}$ The minimum rate requirement.

$\psi$ Lagrangian multiplier.
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Q$</td>
<td>DPC LQ-decomposed matrix.</td>
</tr>
<tr>
<td>$\mathcal{Q}$</td>
<td>The feasible range of the optimized variable.</td>
</tr>
<tr>
<td>$R_{k,s}$</td>
<td>The achievable rate of the $k$-th user over the $s$-th RB.</td>
</tr>
<tr>
<td>$\rho_K$</td>
<td>The PPP users density.</td>
</tr>
<tr>
<td>$\rho_{BS}$</td>
<td>The PPP BS density.</td>
</tr>
<tr>
<td>$S$</td>
<td>Total number of RBs.</td>
</tr>
<tr>
<td>$\tau$</td>
<td>Lagrangian multiplier.</td>
</tr>
<tr>
<td>$\mathbf{U}_k$</td>
<td>The receiver precoding matrix of user $k$.</td>
</tr>
<tr>
<td>$\nu$</td>
<td>The path loss exponent.</td>
</tr>
<tr>
<td>$\Sigma$</td>
<td>SVD singular values matrix.</td>
</tr>
<tr>
<td>$\varsigma$</td>
<td>The power split factor.</td>
</tr>
<tr>
<td>$\mathbf{V}_k$</td>
<td>The transmitter precoding matrix of user $k$.</td>
</tr>
<tr>
<td>$W_{ZF}$</td>
<td>ZF Precoding matrix.</td>
</tr>
<tr>
<td>$W_{MMSE}$</td>
<td>MMSE Precoding matrix.</td>
</tr>
<tr>
<td>$W_T$</td>
<td>The total bandwidth.</td>
</tr>
<tr>
<td>$X_{k,s}$</td>
<td>The transmitted signal to the $k$-th user over the $s$-th RB.</td>
</tr>
<tr>
<td>$Y_k$</td>
<td>Received signal from user $k$.</td>
</tr>
<tr>
<td>$Z$</td>
<td>The total number of users pairs.</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

The evolution of wireless cellular systems has always been driven by the need for a higher throughput. In fact, the continuous need for higher data rates has been the fuel that has led us from 2G systems to 4G systems, with data-rates growth from tens of kbit/s up to the current level of about tens of Mbit/s [13,14]. One key property that distinguishes the generations of different wireless systems from each other is the multiple access technologies [13,15]. For instance, 1G used frequency division multiple access (FDMA), while 2G mostly used time division multiple access (TDMA). On the other hand, 3G used code division multiple access (CDMA). CDMA system is based on direct-sequence spread-spectrum, where the users' information bits are being coded at a very low rate then modulated by pseudo noise sequences. Then at the receiver, a spreading code is used to separate the information of different users [12,16]. This system suffers interference due to the simultaneous transmissions, intra-cell and inter-cell interference. CDMA is an example of multiple accesses where multiple transmitters are allowed to send information altogether over a single communication channel. This facilitates using a number of frequency bands that is the difference between the upper and lower frequencies in a continuous frequency set. In 3G networks, CDMA allows its users to access a wide variety of contents via their cellular devices and terminals [12,16]. In addition, 4G is based on orthogonal frequency division multiple access (OFDMA). Most of these multiple access schemes are of orthogonal nature, particularly for the case of downlink transmission. In other words, to suppress cross user interference, these schemes assign orthogonal resources to different users either in frequency, time, or
code domain [13–15]. However, in terms of sum rate and spectral efficiency (SE), orthogonal systems are never the achieving schemes [15]. Due to the fast growth of mobile Internet, it is expected that by 2020 the data traffic increase to about 1000 fold [17]. This means that, unlike its previous generations, the future 5G systems will have to provide a variety of mobile broadband (MBB) services such as video and virtual reality as well as emerging new applications [13].

1.1 Motivation

Mobile communications are developing quickly due to the continual improvement of living standards, and it is expected that people will be surrounded by smart objects in smart homes, streets, offices, and cities; as a result, a smart world will be created [18]. This smart world will bring substantial increase in the number of active connections, the data traffic demand and the variety of supported services [18]. All of this require significant amount of information to be exchanged between the users and the base stations (BSs). Therefore, the traffic demand in the downlink direction will be huge and future generations of mobile networks are expected to guarantee high quality of service (QoS) and provide ubiquitous services for multiple users. However, the radio resources represented by the wireless spectrum and transmission power are scarce which make it not easy for the current wireless communication technologies to meet such challenges with orthogonal multiple access (OMA) as a backbone air interface. This is because they are limited by orthogonal resource allocation and the number of simultaneously transmitting users. To satisfy these challenges, it is necessary to design and develop advanced technologies that are capable of supporting dense user-connections and efficiently using the available spectrum.

A considerable number of literature works (e.g., in [10, 11, 19–25]) have nominated power based non-orthogonal multiple access (NOMA) as a promising solution that offers high system capacity and better SE as compared to its orthogonal predecessors. According to NOMA, a signal from multiple users can be multiplexed over the same resource blocks (RBs) at the transmitter, while a non-linear algorithm known as successive interference cancellation (SIC) is applied at the receiver to separate the multiplexed signal of different
users. In theory, NOMA is optimal in achieving the boundaries of multiuser capacity [26]. This is because it allows all the users to be multiplexed in the power domain using superposition coding (SC) and are all allowed to use the whole bandwidth at the same time; hence, it is more spectral efficient than OFDMA. Extra high data-rate communications can be achieved through extra performance gains obtained by combining several technologies working together in harmony. NOMA offers this advantage because some recent works proved it can work altogether with various networking platforms like SCDNs [26] and it can also be combined with different capacity-enhancing and air interface technologies like OMA and multiple-input multiple-output (MIMO) [24, 25, 27–29]. In particular, it is expected that applying NOMA as the air interface in small cell environments will offer further performance improvement due to the advantage of the additional frequency reuse gain [26]. Inspired by this fact, this work investigates new ways to propose efficient resource allocation schemes for the downlink NOMA system to enhance its performance and to overcome the limitations of its orthogonal counterparts.

1.2 Aims and Objectives

The main aim of the research project was to investigate the resource allocation in the downlink of NOMA system. The research covered several performance objectives for NOMA system, for instance, sum rate maximization, energy efficiency (EE), EE and SE trade-off. This research also considered several scenarios such as single antenna, multiple antennas, single cell, and multicell networks. The objectives of the research are:

1. To perform a deep literature review on NOMA and to understand the concept of several technologies to be applied in the research such as air interface technologies, MIMO precoding techniques, optimization techniques, and multicell modeling approaches.

2. To propose resource allocation mechanisms that can improve the overall performance of NOMA system, especially in terms of sum rate and EE. In addition, another aim of this research is to simplify the application of NOMA for a large number of users. Moreover, to also make the proposed schemes behave close to the optimal one and outperform existing
techniques in terms of system performance and offer less complexity.

3. To investigate combining NOMA with MIMO and MISO technologies and optimize their performance. In addition, to study simplifying the application of SIC of MIMO-NOMA by applying interference alignment (IA).

4. The final objective in this work is to investigate NOMA to underlay small cells dense networks (SCDN). As both NOMA and SCDN promise capacity gains independently, their combination could form a promising prototype that can further improve the capacity of 5G networks.

1.3 Key Contributions

During the course of this research, the major contributions of this thesis are listed below.

C.1 Proposing low complexity power allocation schemes with closed-form solutions to maximize the overall system sum rate under the proportional rate constraint. Those schemes have shown to achieve performance that is close to the optimal numerical solution. Considering the proportional rate constraint is a key contribution of this research and differentiate it from existing resource allocation methods for NOMA. Not only does this constraint ensure fairness between users, it is crucial in NOMA. Firstly for NOMA, the weaker users will have to detect their signals by treating the stronger users as interferers. The stronger users will also need to detect the weaker users’ signals first and remove them before they detect their own signals. In practice, this requires sufficient power allocated to the weaker users for such detection to be successful. This can be achieved by a proportional rate constraint. Secondly, according to the rate boundary of NOMA [12], NOMA achieves the highest performance gain over OMA when the weaker users achieve a good rate. Therefore, simply achieving a high rate for the stronger users as in conventional minimum rate constraints will not fully utilize the potential of NOMA and may also be impractical. Thus, we considered proportional fairness constraint in this research.
C.2 Deriving two closed-form suboptimal solutions for a two-user NOMA scenario. The closed-form solution is shown to achieve performance that is close to the optimal one and outperforms all existing techniques.

C.3 The proposed solution in C.2 is restricted to two users only. We then extend the obtained solution for a larger number of users by proposing a subband-based approach whereby two users are multiplexed into each subband. However, splitting the whole bandwidth into subbands cannot fully utilize the potential of NOMA, where the entire bandwidth can be occupied by all users. Therefore, we propose a hierarchical power allocation process whereby the users are divided into two groups such that a closed-form power allocation solution (the obtained two-users suboptimal one) can be applied. This divide-and-allocate approach is repeated until all users are allocated with a transmission power. The proposed architecture simplifies the sum rate optimization for a large number of NOMA users. In addition, proposing a low complexity power allocation scheme that allocates power to each RB in proportion to the sum of the channel gains of all multiplexed users.

C.4 Proposing the idea of hybrid multiple access, which represents a combination between NOMA and OFDMA, as a candidate for the next generation wireless networks. The hybrid scheme is effective in scenarios where there is a significant gap between the channel gains of the cell-edge and the cell-center users. In such case it is more efficient to apply the hybrid scheme than solely applying NOMA or OFDMA. Results confirm the superiority of the proposed hybrid multiple access scheme over conventional NOMA.

C.5 Investigating the trade-off between the EE and the SE in the downlink NOMA system. An optimization problem is formulated to maximize the EE under the total power and proportional minimum rate constraints. Then two solutions are proposed to solve the formulated optimization problem to establish an energy efficient design of NOMA system and achieve efficient power allocation. The first one is to simplify the formulated problem using Dinkelbach approach and then solved using an iterative subgradient based approach. The second solution is to allocate the power for each user based on hierarchical user-pairing. Moreover,
we demonstrate several aspects of the EE-SE trade-off for NOMA system. For instance, the impact of the dynamic transmitted power, circuit power, and the number of RBs.

C.6 Proposing the application of IA for a downlink, multiuser MIMO-NOMA. By grouping a number of users based on NOMA, the other users are treated as interference and suppressed by using IA. In so doing, there will be less users per NOMA multiplexed stream, and thus will simplify the SIC procedure. We also investigate the use of singular value decomposition (SVD) and dirty paper coding (DPC) for spatial dimension exploitation. The power allocation for both multiple-input single output (MISO) and MIMO-NOMA are investigated. A low complexity suboptimal solution for a two-user scenario is obtained and another subgradient based power allocation approach is also proposed. We also demonstrated the choosing of the best number of users to be multiplexed per each group before applying IA. While IA is usually applied to align the ICI, this work represent the first effort to consider data stream alignment in MIMO-NOMA.

C.7 Investigating the performance of NOMA based homogeneous SCDN (Hom-SCDN). An optimized power allocation is applied with the existence of inter-cell interference (ICI). In addition, a power minimization problem is formulated to minimize the power transmitted by each BS to control the ICI; then a suboptimal solution is proposed based on the target rate required to be achieved by each BS. Both frequency reuse with a reuse factor of 3 (3FR ) and universal frequency reuse (UFR) are considered in this thesis.

1.4 Thesis Organization

This thesis is organized in seven chapters as follows. Chapter 1 introduces the overview of this research, the predicted future challenges facing the next generations of wireless networks such as increasing traffic demand, huge capacity requirements, NOMA as a promising solution, and the importance of managing the radio resources. It also discusses the aims, the objectives and the main contributions of the thesis.
Chapter 2 provides theoretical background of capacity enhancement techniques of 5G networks. This chapter also includes a brief overview of NOMA, introduces the concept of NOMA and discusses its implementation. It also addresses the wireless channel modeling, multiple antennas systems, and small cells dense networks. Moreover, this chapter also presents the resent works done about several aspects of NOMA system.

Chapter 3 proposes a resource allocation scheme aiming to maximize the overall sum rate of a downlink NOMA system. This chapter also presents the proposed hierarchical pairing (HP) concept and the proposed power allocation techniques that are applied in combination with this concept. A hybrid air interface technique is also proposed in this chapter along with its resource allocation schemes.

Chapter 4 proposes energy efficient resource allocation schemes with subgradient based power allocation technique. In addition, it investigates the trade-off between the EE and the SE of NOMA system.

Chapter 5 proposes resource allocation technique for a downlink MIMO-NOMA system. This chapter also proposes the implementation of IA for MIMO-NOMA system. Moreover, both SVD and DPC were used as MIMO precoders in this chapter.

Chapter 6 proposes the implementation of NOMA based Hom-SCDN. Both sum rate and EE optimization are considered in this chapter. The sum rate is optimized using similar approaches to those proposed in Chapter 3. In addition, to optimize EE, a minimum target rate based power allocation scheme is proposed to minimize the effect of ICI and optimize EE.

Chapter 7 concludes the thesis and outlines prospects for the future extension of the work. This chapter is followed by the list of references and appendices. Appendix A of this thesis proves that EE is continuously differentiable in terms of SE as will be discussed in Chapter 4 of this thesis. Appendix B presents the RB structure according to long term evolution (LTE) standards. Appendix C provides an overview of the radio resource allocation (RRA) concept. Appendix D provides the basic steps of applying Lagrange approach that is used in this thesis to solve the formulated optimization problems.
1.5 List of Publications

The following papers have been extracted from this thesis.


Chapter 2

Non-Orthogonal Multiple Access (NOMA) for 5G Networks

2.1 Introduction

To meet the expected traffic demand of 5G mobile communications, the required capacity growth can be achieved through a number of techniques that include evolving the air interface which has led to the idea of adopting non-orthogonal air interface [15]. Other capacity enhancement techniques include deploying a large number of small cells (i.e., SCDN), and also using a large numbers of antennas at both the transmitter and/or the receiver sides (i.e., MIMO technology) [14]. Moreover, the combination of these techniques is also expected to hugely improve the overall system capacity and SE. This chapter explains the concept of NOMA and presents a literature review of NOMA related technologies that could be combined with NOMA to improve the achievable capacity. Before discussing the concept of NOMA and its related works, the wireless channel models used in this thesis is first presented.

2.2 Wireless Channel Models

The nature of the wireless channel plays a vital role in the resource allocation and optimization process. The main obstacle against the reliability of wireless communication is the difficulty of modeling the fluctuating nature of the wireless channel. In general, wireless channels could be classified as of line
of sight (LOS) nature, where there is a direct signal between the transmitter and the receiver, and as of non-line of sight (NLOS) nature where the signal between the transmitter and the receiver faces obstacles like buildings or terrains [12, 30]. Modeling a wireless channel for cellular networks in urban environment is mostly done with a wireless channel of NLOS nature. This is because the terrain in such areas has tall height that can act as serious obstacles in the way of the transmitted signal. In general, the wireless channel models can be classified as illustrated in Figure 2.1. As this figure illustrates, the fading channel can be mainly classified as of large scale and small scale fading nature, and each of these types has further divisions to be clarified in the next subsections.

![Figure 2.1: Classification of wireless channel models [6].](image)

### 2.2.1 Large Scale Fading Effect

Mainly, large scale models focus on determining the path loss levels as the others parameters could be decided based on known standards. Since the associated path loss is a function of the distance between the transmitter and the receiver, both empirical and analytical models predict that the received signal power declines accordingly as the distance increases. For a given distance \(d\), the average path loss is given by [1, 31]

\[
PL = -PL_0 - 10\nu \log_{10} \left( \frac{d}{d_0} \right) + \xi \ (dB)
\] (2.1)

where \(PL_0\) represents the path loss at reference distance \((d_0)\) which can be determined from field measurements, and \(d\) stands for the distance between the user and the serving BS with \(\nu\) being the path loss exponent which varies
from 2 to around 6 depending on the propagation environment, for example $v = 2$ models the free space effects. Without this factor, the path loss model assumes that the average path loss at a distance $d$ from the transmitter is the same in all directions. However, the received power at any two points located the same distance from a transmitter may differ vastly in reality, which contradicts this assumption. Hence, measurements shown that the path loss at a distance $d$ from the transmitter. Finally, $\xi$ represents the shadowing effect which is of random nature and has a log-normal distribution.

### 2.2.2 Small Scale Fading Effect

Small scale fading effects encompass the rapid fluctuations in the received signal strength over a very short period of time or a short distance [30]. As depicted in Figure 2.1 the small scale fading channels can be categorized as time variant channels and multipath fading channels. The time variant fading channels depend on the Doppler frequency shift and the speed of the mobile terminal to have fast or slow fading nature [32]. Multipath fading, on the other hand, occurs mostly in an urban environment as a result of the tall buildings that enhance the probability of NLOS propagation by providing many surfaces from which the signal might get reflected [30]. Therefore, in such environment, the transmitted signal might reach the destination in the form of several copies with each copy has its own channel response, phase, arrival time, and amplitude. The instant summation of these copies could have destructive or constructive nature which makes it fluctuates with time [12, 30]. This occurs due to the relative motion of the receiving terminal which would results in multipath fading. The frequency of each multipath component is shifted due to this motion causing what is called the Doppler shift. Doppler shift is a function of the direction of motion and the speed with respect to the direction of arrival of each signal copy. The summation of the envelopes of two or more Gaussian signals forms a Rayleigh distribution, and thus the statistical time varying nature of the received envelope is modeled using Rayleigh distribution [30].

Depending on the transmitted signal bandwidth and channel delay spread, multipath-small scale fading channels can be classified into flat fading channels (narrowband channels) or selective fading channels (wideband channels) as shown in Figure 2.1. Flat fading occurs when the channel has a
coherence bandwidth that is greater than the transmitted signal bandwidth, where the coherence bandwidth is the bandwidth over which the channel has a linear phase response and a constant envelop [30]. On the other hand, frequency selective fading occurs when the channel has a coherence bandwidth that is smaller than the transmitted signal bandwidth. This results in time dispersion of the information symbols through the channel that distorts the received signal and leads to inter-symbol interference (ISI). At the receiver side, solving the frequency selective fading problem often requires a complex receiver. Because different components of the transmitted signal are affected by the channel with various gains [30]. On the other hand, it is easier to reverse the effects of the flat fading using simple error correction and equalization techniques as opposed to frequency selective fading. Thus, in OFDM, the wideband channel is splitted into many narrowband subcarriers so that each subcarrier suffers flat fading effect [30]. In this thesis, the wireless channel is modeled as a six-path frequency selective fading channel using the ITU pedestrian - B model where the average power and the relative delays of the multipath are listed in Table 2.1 [1, 31].

Table 2.1: ITU Pedestrian model [1]

<table>
<thead>
<tr>
<th>Tap</th>
<th>Channel A</th>
<th>Channel B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Relative delays (ns)</td>
<td>Average power (dB)</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>110</td>
<td>-9.7</td>
</tr>
<tr>
<td>3</td>
<td>190</td>
<td>-19.2</td>
</tr>
<tr>
<td>4</td>
<td>410</td>
<td>-22.8</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2.3 Overview of Orthogonal Frequency Division Multiplexing (OFDM)

In general, multiple access techniques can be categorized into orthogonal and non-orthogonal techniques. The signals in the first category are made to be orthogonal to their counterparts that leads to no cross correlation between these signals, such as OFDMA which is the multiuser extension of OFDM and it is widely used in 4G networks. The advantage of orthogonality is that it allows the simultaneous transmission over the subcarriers through a restricted frequency space with no interference [7]. It is achieved among the OFDM subcarriers by a careful selection of the subcarrier spacing as depicted in Figure 2.2, which in some cases set to be equal to the symbol rate [7].

![Orthogonality of OFDM spectrum with eight subcarriers](image)

Figure 2.2: Orthogonality of OFDM spectrum with eight subcarriers, the subcarriers do not interfere with each other because at the peak of each subcarrier; the signals from other subcarriers are zero [7].

OFDM (and OFDMA) is a physical layer technology that divides a high rate data stream into a number of low rate data substreams to reduce the
severity of fading. In other words, it converts a wide frequency-selective fading channel into a set of parallel, narrow, flat-fading channels. This will simplify the application of channel equalization where a one-tap multipliers can be sufficient for each subcarrier [7]. Recall that OFDM allocates all subcarriers to one user at the same time, and hence the data carried by all the subcarriers belong to that user only. If more than one user intend to transmit by OFDM, they have to queue for their turns in time. OFDMA solved this problem by directly allocating the subcarriers in frequency domain to different users [7]. OFDMA robustness against ISI made it suitable to be depended as the air interface of 4G communication systems.

The idea of this technique was to use parallel data transmission and OFDM presented the idea of adapting orthogonal subchannels to provide robustness against multipath distortion and simplify the required equalization. Figure 2.3 shows the structure of OFDM transceiver. At the transmitter, the multi-level modulation stage includes applying quadrature amplitude modulation (MQAM) or phase shift keying modulation (MPSK). This stage maps the input data into complex symbols [8, 9]. The IFFT stage represents the modulation stage in OFDM, where the complex symbols are mapped into OFDM symbols [8, 9]. After that, a cyclic prefix is inserted to battle the effect of ISI. The cyclic prefix represents a copy of part of the OFDM symbol that is taken from the symbol beginning and added to its tail. In some cases the cyclic prefix length represents up to 25% of the symbol length [8, 9]. At the receiver side, the inverse of the same operations that are applied at the transmitter are repeated in reversed order to extract the intended data, as illustrated in Figure 2.3 [8, 9].

![Figure 2.3: OFDM Transceiver [8,9]](image-url)
2.4 Overview of Non-Orthogonal Multiple Access (NOMA) System

Recent reports suggest that the smart terminals popularity and the requirements of new mobile services are continuously growing in unprecedented manner [15]. This means it is hard to meet the expected future demand by the current transmission rates of OFDM based wireless communications. In particular, OFDMA (4G) already has some weak points which might become more pronounced in the future if it would be adopted in 5G networks. For instance, OFDM requires the inclusion of the cyclic prefix to combat multipath fading, which waste part of the radio resources. In addition, OFDM is highly sensitive to the carrier offset. Therefore, 5G system needs to adopt a new multiple access technology to increase the capacity and the access capability as well as to provide a higher SE [15]. Driven by this goal, several researchers have proposed a number of new multiple access schemes as 5G standard air interface candidates. The common factor among those proposed schemes is the idea of using non-orthogonal transmission. This is because non-orthogonality facilitates serving multiple users over the same frequency and time resources via power domain and/or code domain multiplexing that would enhances the system access performance.

The first example from the proposed schemes is NOMA which is a multiplexing scheme that was proposed as a promising candidate to represent a novel cellular multiple access scheme for future radio access (FRA) [15, 33]. NOMA introduces non-orthogonality via power-domain user multiplexing. Thus, NOMA exploits the power domain which is not utilized sufficiently in previous systems [13]. To ensure user demultiplexing, NOMA implies allocating a large power difference between the paired users at the transmitter and the application of SIC at the receiver [15]. In particular, NOMA differs from traditional orthogonal transmission schemes in that, on the transmitter side, it would intentionally introduce intra-cell interference so it can apply non-orthogonal transmission. On the receiver side, on the other hand, NOMA utilizes interference cancellation to enhance the SE by applying SIC to conduct multiuser signal separation. In other words, NOMA users are multiplexed in the power domain at the transmitter side using SC and are being separated by SIC at the receiver side [15]. In addition, NOMA differs from usual
OMA schemes in that, the OMA users are being allocated a narrow bandwidth which means that OMA cannot maintain the sum-rate capacity of a wireless system and makes OMA based systems struggle to deal with a large number of active connections. Recall that, for OFDMA, the set of subcarriers allocated to one user will not be used by another. The advantage of such procedure is the robustness against multiple access interference (MAI) [7,34]. However, the drawbacks here are the inefficient use of the radio spectrum, and the sensitivity to frequency asynchronism, i.e., the carrier frequency offset (CFO) problem that could affect the orthogonality [7]. This is because CFO not only degrades the performance of the OFDMA user who has it, but also causes MAI to other users, hence, OFDMA with CFO is not robust against MAI [34]. With MC-CDMA systems, orthogonality is maintained by spreading the symbols by orthogonal codes. This orthogonality might be destroyed due to frequency selective fading effect, thus leading to MAI [34]. Increasing the transmit power is one solution to the MAI problem but not an efficient one, because increasing the transmission power for one user will also increase the interference for other users [34]. In case of NOMA, using SIC detection is expected to make it resistant to MAI. Moreover, since all NOMA users are allowed to use the same RB, there will be no drawbacks from CFO. Furthermore, NOMA allows all the users to occupy all RBs, which means it allows multiple users to be served simultaneously using the whole spectrum and this makes it provide better SE and throughput than OFDMA [35,36]. Thus, there is no need for subcarrier or RB allocation with NOMA. As compared to CDMA and OFDMA, NOMA has no clear problems related to the near-far effect which is the obsession in 3G [15].

Other examples of the proposed non-orthogonal air interface candidates include:

- Sparse code multiple access (SCMA) that is proposed by Huawei [15]. SCMA is a frequency domain NOMA technique that uses sparse code book and code domain multiple access at the transmitter and requires the use of message passing algorithm (MPA) at the receiver. SCMA is expected to be applicable at both the uplink and downlink of 5G wireless networks. However, despite the well-defined code structure, it is still challenging to design and optimize the code.

- Pattern division multiple access (PDMA) is another non-orthogonal
multiple access candidate proposed by Datang [15]. PDMA is a scheme that uses non-orthogonal characteristic pattern in signal domains (e.g., power domain, space domain and code domain) at the transmitter to distinguish between multiple users and it applies SIC detection at the receiver [15]. However, there are some challenges that PDMA is still facing, such as how to combine PDMA with MIMO to design space domain coding pattern and how to design patterns at transmitter to differentiate between users more easily.

- Another candidate is the multiuser shared access (MUSA), which is a 5G multiple access scheme proposed by ZTE. It uses special spreading sequences to spread multiple user data sequentially, and then each user’s data spread are overlapped to be transmitted. At the receiver, MUSA uses advanced SIC receiver to demodulate and recover the data of each user [15].

It is worth mentioning that this thesis is mainly focused on power based NOMA and how to optimize its performance.

2.4.1 Superposition Coding

In NOMA system, SC is used for superimposing multiple users signals over the same RB in power domain [12, 37]. Hence, SC is effectively increasing the capacity of NOMA system without expanding the bandwidth.

Figure 2.4 illustrates the spectral occupancy of a two-user scenario of a) OFDMA and b) NOMA system. The horizontal axis denotes the bandwidth in terms of RBs, and the vertical axis is the power allocated for each user over each RB. For OFDMA, the RBs are exclusively allocated to one of the users and the power allocated to user 1 and 2 at the s-th RB is $P_s^{(1)}$ and $P_s^{(2)}$ respectively. For NOMA, both users occupy all the RBs and the user with a better channel at the s-th RB will be allocated the power $P_s^{(H)}$, and the weaker one with the power $P_s^{(L)}$.

Consider a single cell downlink NOMA scenario where the BS is transmitting the signal $X_{k,s}$ to the $k$-th user ($k = \{1, 2, \ldots K\}$) and at the s-th RB ($s = \{1, 2, \ldots S\}$) with transmission power $P_{k,s}$. Where the users are assumed to be arranged in an ascending order with user 1 as the weakest user (i.e., the user with the worst channel conditions) and the $K$-th user as the strongest
user (i.e., the user with the best channel conditions). The received signal by user $k$ at the $s$-th RB is given by [19]:

$$Y_{k,s} = \sum_{i=1}^{K} \sqrt{P_{i,s}} \tilde{h}_{i,s} X_{i,s} + n_{i,s}$$

(2.2)

where $\tilde{h}_{i,s}$ represents the channel coefficients between the BS and the user at the $s$-th RB and $n_{i,s}$ represents the additive white Gaussian noise (AWGN). From equation (2.2), it is clear that each user will receive a signal with its data and the data intended for the other users. Therefore, the power allocated to one NOMA user will have significant effect on the achievable rates of other users. This necessitates optimizing the allocated power to all users.

### 2.4.2 Successive Interference Cancellation

The SIC is a superposition scheme where the signal is the linear combination of the users signals [11, 22, 35]. It is applied at NOMA receiver so each user
could extract its respective data as shown in Figure 2.5. The mechanism behind SIC is that the user with weak channel conditions treats the signal of the user with the better channel as noise and decodes its own data from the received signal. On the other hand, the user with the better channel performs SIC, where it decodes the data of the weaker user and then proceeds to subtract it from the received signal and decode its own data [22]. Thus, NOMA could be considered as another method for interference cancellation and also it could be used to enhance the SE [19].

$$Y_{1,s} = \sqrt{P_{1,s}} X_{1,s} \tilde{h}_{1,s} + \sqrt{P_{2,s}} X_{2,s} \tilde{h}_{2,s} + n_{1,s}$$

Decoding the weak User Signal

$$Y_{2,s} = \sqrt{P_{1,s}} X_{1,s} \tilde{h}_{1,s} + \sqrt{P_{2,s}} X_{2,s} \tilde{h}_{2,s} + n_{2,s}$$

Cancelling the weak User Signal

$$\sqrt{P_{1,s}} X_{1,s} \tilde{h}_{2,s}$$

Decoding the Strong User Signal

$$\sqrt{P_{2,s}} X_{2,s} \tilde{h}_{2,s}$$

Figure 2.5: SIC receiver of two users NOMA system with $\sqrt{P_{1,s}} > \sqrt{P_{2,s}}$ (i.e. $|\tilde{h}_{2,s}|^2 > |\tilde{h}_{1,s}|^2$) [10, 11]

Using the two-user scenario as an example illustrated in Figure 2.5 and assuming that the users are arranged in an ascending order based on their channel gains, the user with the worst channel conditions (denoted as user 1) treats the signal of the user with the better channel as noise and decodes its data from the received signal. On the other hand, the user with the better channel (denoted as user 2) performs SIC, as it first decodes the data of the other user and then cancel it from the received signal. The interference-cancelled received signal is then used to decode the data for this user [22, 35].
This is possible because the user with the better channel conditions can decode any data that the weaker user can decode [12].

For two users multiplexed over the $s$-th RB using NOMA principles, the achievable rate by the user with the higher $(H)$ and lower $(L)$ channel gain is respectively given by

$$R_s^{(H)} = B_s \log_2 (1 + \gamma_s^{(H)}) \text{ bits/s}$$  \hspace{1cm} (2.3)$$

$$R_s^{(L)} = B_s \log_2 (1 + \gamma_s^{(L)}) \text{ bits/s}$$  \hspace{1cm} (2.4)$$

it is worth mentioning that, for successful SIC application, the success of the step mentioned in (2.3) is necessary for (2.4) to be attempted as explained earlier in Figure 2.5.

The two users sum rate over the $s$-th RB is given by

$$R_s = R_s^{(H)} + R_s^{(L)}. \hspace{1cm} (2.5)$$

It must be noted that the superscripts $(H)$ and $(L)$ are not fixed to user 1 and 2 and are assigned according to the channel gains of the two users in each RB. In other words, we do not assume that user 1 always have a better channel gain than user 2 over all RBs. In addition, the terms

$$\gamma_s^{(H)} = \frac{P_s^{(H)} |h_s^{(H)}|^2}{B_s N_0}$$  \hspace{1cm} (2.6)$$

$$\gamma_s^{(L)} = \frac{P_s^{(L)} |h_s^{(L)}|^2}{P^{(H)}_s |h_s^{(L)}|^2 + B_s N_0}$$  \hspace{1cm} (2.7)$$

represent the received SINR of the two users in the $s$-th RB, where $|h_s^{(H)}|^2 = \frac{\xi^{(H)} |\tilde{h}_s^{(H)}|^2}{PL_0^{(H)}}$ and $|h_s^{(L)}|^2 = \frac{\xi^{(L)} |\tilde{h}_s^{(L)}|^2}{PL_0^{(L)}}$ represent the channel gains which include the effect of fading, the log-normal shadowing factor $\xi$, and the path loss effect $PL$ [dB], and $d$ stands for the distance between the user and the serving BS with $\nu$ being the path loss exponent. In addition, $N_0$ denotes the noise power spectral density. On the other hand, the rate achieved by two users in the orthogonal case can be given by [12, 13]
$R_s^{(2)} = \varsigma B_s \log_2 \left( 1 + \frac{P_s^{(2)} |h_s^{(H)}|^2}{\varsigma B_s N_0} \right) \text{bits/s}$ \hspace{1cm} (2.8)

$R_s^{(1)} = (1 - \varsigma) B_s \log_2 \left( 1 + \frac{P_s^{(1)} |h_s^{(H)}|^2}{(1 - \varsigma) B_s N_0} \right) \text{bits/s}$ \hspace{1cm} (2.9)

where $\varsigma \in [0, 1]$ is the split for the degree of freedom that controls the bandwidth fraction being assigned to each user.

Figure 2.6 illustrates the rate-region boundaries achieved by the NOMA users given in (2.3) and (2.4) and that of the orthogonal users given in (2.8) and (2.9). It is obvious from this figure that the performance of the SC based NOMA scheme outperforms the orthogonal one except for the case where there is only one user being communicated to at the two corner points. This figure shows that SC-NOMA helps the weak user to maintain a performance that is close to the single-user bound and concurrently provide a reasonable rate to the strong user. It is also clear from this figure that as the asymmetry between the two users becomes deeper, the performance gap widens further.

For the weak user to achieve near single-user performance with the orthogonal scheme, it needs to be allocated a significant fraction of the degrees of freedom, which might result in a considerable performance degradation to the strong user. On the other hand, despite the small amount of the transmission power to be allocated to the strong user, its degree of freedom will be enhanced thanks to the SC which will allow this user to use the whole bandwidth. Being allocated a small amount of transmission power means that the strong user with SC-NOMA will cause low interference to its weaker counterpart.
Figure 2.6: Rate region boundaries achieved by two users in the downlink of orthogonal and non-orthogonal cases [12, 13].

2.4.3 Related Works on Single-Antenna NOMA System

Recently, NOMA attracted huge attention in the literature, e.g., power allocation techniques [10, 22, 35, 38], combination of NOMA with MIMO [27, 28], NOMA for multiple antenna relay network [39], ICI mitigation [21], SC for NOMA [40], etc. NOMA has been presented as a mean to enhance the SE for future networks and also confirmed to be superior to OFDMA. However, studies on the NOMA resource allocation aspects is still in its early stage. So far, resource allocation for OMA techniques have been studied intensively in literature. For instance, in [41], the dynamic resource allocation scheme based on maximizing the sum capacity with proportional fairness of OFDMA system is proposed. Moreover, [11, 19, 23, 42] presented a comparison between orthogonal and non-orthogonal multiple access techniques. In [19], the authors presented the concept and the practical considerations of NOMA with SIC at the receiver side. In addition, this work investigated a comparison between OFDMA and NOMA and the authors found that NOMA achieves about 30% better throughput than OFDMA. The author in [23] considered NOMA with SC to improve the SE in a coordinated system with two coordinated BSs using Alamouti code. The author proved that using SC could help to provide a
sufficient transmission rate to the cell edge users without affecting that of the cell center user. In addition, a number of power allocation techniques have been proposed in [10, 11, 22, 38], but none of the proposed techniques was in closed-form. For example, the authors in [22] analyzed NOMA with user’s index based fixed power allocation (UFPA) in a scenario with user target rate and another with opportunistic rate allocation. They proved analytically that NOMA is better than OFDMA in terms of the achievable sum rate and the coverage probability. Moreover, the authors in [10] and [11] presented how NOMA could improve the capacity and the throughput of the network over orthogonal access approaches. In particular, [10] presented several power allocation schemes: the full search power allocation (FSPA), the fractional transmit power allocation (FTPA), and the fixed power allocation (FPA) which was also presented in [38]. Among those schemes, the FSPA achieves the best performance but suffers from high computational complexity. The FPA has the lowest complexity but have poor performance. The FTPA is a balance between the two, but required a prior computer simulation to determine a specific parameter to obtain the best performance.

Another important property that future system must possess is to be energy efficient. EE has been widely studied for OFDMA system [43, 44]. For example, the authors in [43] investigated the EE maximization of OFDMA system taking into account the data rates and bit error rate (BER) constraints to ensure proportional fairness among the users. The subcarrier allocation in both [41] and [43] was based on the access probability between the users and their serving units, which is calculated using each user’s channel to noise ratio. Both articles depicted the importance of fairness constraints to help users in achieving their target rates. On the other hand, the investigation in designing energy efficient NOMA is also very limited as previous works focused mainly on sum rate maximization. In [45, 46], the same authors optimized the subchannel assignment and power allocation to maximize EE for the downlink NOMA network. In terms of power allocation, the authors used DC programming to solve the formulated problem and allocate the power and they also used the FTPA scheme. In addition, the EE-SE trade off is an important feature for energy efficient system design as it allows the balancing between EE and SE in resource allocation. The EE-SE trade-off in OFDMA has been deeply studied in [47–54]. In [47], a framework of EE-SE trade-off
for a downlink OFDMA-network was presented and it shows that the EE is quasiconcave in SE. In [48], a joint EE and SE optimization problem is considered for OFDMA based multicell networks by optimizing both subchannels and power allocation. The authors of [51] focused on EE-SE trade-off in multicell cooperative OFDMA networks. The authors formulated an optimization problem to maximize EE under QoS constraints. Moreover, the authors in [52] addressed the EE–SE relationship in MIMO-OFDMA broadcast channel (BC) and they proposed a two-layer resource allocation algorithm that is based on the quasiconcavity of the EE–SE relationship. However, no works yet have considered the EE-SE trade-off in NOMA system. Chapter 4 of this thesis will provide a detailed study about EE and EE-SE trade-off in NOMA system.

Nevertheless, other aspects of NOMA system were also investigated by a number of works. For example, cooperative NOMA was considered in [55] and shows that it achieves maximum diversity gain for all multiplexed users. The importance of fairness for NOMA is also considered in [36], where the authors suggested that improving the performance of the worst-condition user will make NOMA significantly outperforms its orthogonal counterparts. In addition, the authors in [56] presented how NOMA could enhance the cell-edge user throughput while achieving fairness. The design of NOMA-receiver for a downlink NOMA system is investigated in [57]. A dynamic frequency reuse scheme for NOMA system was assessed in [21]. The authors also proposed intra-beam SC and intra-beam SIC for NOMA system. They suggested that the protection from ICI is necessary to improve the throughput of the cell-edge and the cell-center users. In [58, 59] the same authors investigated the gains that relay assisted NOMA transmission could offer over the conventional NOMA system. The authors in [60–62] applied NOMA with visible light communication (VLC) system to boost the throughput in VLC downlink networks. Unlike the aforementioned works, [63–65] investigated the performance of uplink-NOMA system.

All of these works highlight and confirm the significant performance gain that NOMA offers over OMA. They also show the importance of NOMA for 5G networks. This thesis is another effort to further investigate the outcomes of NOMA. Table 2.2 summarize the main aspects of these works and their limitations.
Table 2.2: Summary of the related works to single antennas - NOMA system.

<table>
<thead>
<tr>
<th>Authors [Reference]</th>
<th>The considered NOMA aspect</th>
<th>Outcome</th>
<th>Study limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>[19]</td>
<td>Practical considerations of NOMA with SIC</td>
<td>NOMA achieves about 30% better throughput than OFDMA</td>
<td>The allocated power was not optimized.</td>
</tr>
<tr>
<td>[10, 11, 22, 38]</td>
<td>Power allocation for NOMA system</td>
<td>Power allocation is necessary to properly perform SC and SIC</td>
<td>none of the proposed techniques was in closed-form.</td>
</tr>
<tr>
<td>[45, 46]</td>
<td>Designing energy efficient NOMA system</td>
<td>NOMA outperforms OMA in terms of both EE and SE</td>
<td>Considered two-user case only.</td>
</tr>
<tr>
<td>[55]</td>
<td>cooperative NOMA</td>
<td>NOMA achieves maximum diversity gain for all multiplexed users</td>
<td>Did not optimize the allocated power and used FPA.</td>
</tr>
<tr>
<td>[36]</td>
<td>Fairness for NOMA</td>
<td>Improving the performance of the worst-condition user makes NOMA outperforms OMA</td>
<td>No closed form solution as the formulated optimization problem was solved with standard solvers.</td>
</tr>
<tr>
<td>[56]</td>
<td>NOMA cell-edge user throughput</td>
<td>NOMA can enhance the cell-edge user throughput while achieving fairness</td>
<td>No closed form solution as the formulated optimization problem was solved by the standard interior-point method.</td>
</tr>
<tr>
<td>[60–62]</td>
<td>NOMA with VLC</td>
<td>NOMA can boost the throughput in VLC downlink networks</td>
<td>VLC is limited by the narrow bandwidth of the light sources, which forms a barrier to achieving high data rates.</td>
</tr>
</tbody>
</table>
2.4.4 Related works on Multiple-Antenna NOMA System

MIMO is one of the key technologies that were nominated to boost the capacity and the SE in LTE and LTE Advanced systems [13, 66]. This is because using multiple antennas at the BS means they could be exploited to boost the downlink beamforming which leads to better SINR or they could also be used to achieve downlink spatial multiplexing to increase the throughput [13, 66].

Recently, there are a number of works considering the combination of MIMO and NOMA, for instance, the authors in [67] addressed the impact of quasi-degradation on the downlink of MISO-NOMA system. They also proposed a Hybrid NOMA (H-NOMA) precoding algorithm based on the quasi-degradation approach to improve the system performance. In addition, a QoS based optimization problem for two-user MISO broadcast systems was considered in [68]. Using a known target interference levels, power minimization and precoder optimization were obtained using LDD problem and Newton’s iterative algorithm, respectively. However, their work considered only two users scenario. In [69], a transmission framework for MIMO-NOMA system is proposed using signals alignment. The authors also addressed fixed and cognitive power allocation to evaluate NOMA system performance. However, the allocated power was not optimized. The authors in [70] considered the performance analysis and the application of MIMO to a downlink NOMA system. The authors divided the users into clusters and designed the precoding and the detection matrices to remove the inter-cluster interference. They applied fixed and cognitive based power allocation; however, they did not optimize the power allocation according to the instantaneous channel conditions which we used in this work to further improve the performance of MIMO-NOMA. To simplify the design of the precoding matrix and use an identity matrix and to simplify the signal overheading, they assumed that the BS does not have the global CSI and it knows only the the order of the users’ effective channel gains to implement NOMA. However, to ensure a robust design of the precoding matrix, we will assume throughout this work that the BS has the global CSI. The authors in [25] focused on sum rate maximization for MIMO-NOMA system and showed that NOMA is much better than TDMA. In [27], the ergodic capacity of MIMO-NOMA is investigated. The author highlighted how NOMA significantly outperforms TDMA in terms of ergodic capacity. However, in both [25] and [27] the presented scenario also
considered only two users. The authors of [24] suggested that NOMA with opportunistic beamforming (OBF) offers promising results but the main issues are transmission power allocation and user-scheduling. The authors in [39] investigated the performance of NOMA in relay-MIMO system. A closed-form expression for the outage probability was derived to show the privilege of NOMA over OMA schemes. However, the allocated power was based on UFPA and not optimized. A downlink MIMO-NOMA system with random beam-forming was presented in [20] to improve the cell-edge user experience. In addition, the authors in [17] proposed user-clustering algorithm for a multiuser MIMO-NOMA system considering two users per cluster. The transmission power is allocated to the users based on FTPA scheme; the authors optimized the precoding matrix with an objective to maximize the achievable sum rate using the Majorization Minimization approach. However, it is unclear how the clustering will be applied for an odd number of users, or when the number of users is greater than double the number of the antennas. A summary of the related works to multiple antenna based NOMA systems is depicted in Table 2.3. Chapter 5 of this thesis will present resource allocation schemes for MISO-NOMA and MIMO-NOMA systems.
Table 2.3: Summary of the related works to multiple antennas - NOMA system.

<table>
<thead>
<tr>
<th>Authors [Reference]</th>
<th>The considered NOMA aspect</th>
<th>Outcome</th>
<th>Study limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>[67, 68]</td>
<td>Optimized the performance of MISO-NOMA system.</td>
<td>Using multiple antennas at the transmitter can further boost NOMA achievable rate</td>
<td>The authors in [68] considered only two-user scenario.</td>
</tr>
<tr>
<td>[17]</td>
<td>user-clustering algorithm for a multiuser MIMO-NOMA system</td>
<td>optimizing the precoding matrix effectively maximize the achievable sum rate</td>
<td>unclear how the clustering will be applied for an odd number of users, or when the number of users is greater than double the number of the antennas.</td>
</tr>
<tr>
<td>[69]</td>
<td>MIMO-NOMA transmission framework</td>
<td>Signals alignment enhances NOMA performance</td>
<td>Did not consider user-pairing and suggested it as future work.</td>
</tr>
<tr>
<td>[25, 27]</td>
<td>MIMO-NOMA sum rate maximization</td>
<td>NOMA is much better than TDMA</td>
<td>the presented scenario considered only two users</td>
</tr>
<tr>
<td>[70]</td>
<td>user-clustering in MIMO-NOMA system</td>
<td>Managed to remove the inter-cluster interference</td>
<td>Did not optimize the power allocation according to the instantaneous channel conditions.</td>
</tr>
<tr>
<td>[39]</td>
<td>the outage probability of relay MIMO-NOMA system</td>
<td>NOMA has better outage than OMA</td>
<td>the allocated power was based on UFPA and not optimized.</td>
</tr>
</tbody>
</table>
CHAPTER 2. NOMA FOR 5G NETWORKS

2.5 NOMA for Small Cells Dense Networks

The third approach that is expected to enhance the capacity of 5G mobile networks is to deploy a large number of cells by what is known as network densification [13, 19, 71, 72]. Network densification means that the environment is going to have large numbers of both users and broadcasting cells [19, 73]. The large number of users means a high traffic demand, whereas a huge number of cells means many overlapping areas will exist within the broadcasting range. Practical examples of such environment could be seen in urban areas, such as; big cities and also on crowded events and occasions, for instance, a NASCAR track, Football stadiums, and the Olympic games where a couple of thousands of users are expected to exist. For instance, the dense urban area is expected to have an average of 1000 users in real life [33, 73–76].

Small cell is a term refers to access nodes that are small in size. As compared to macrocell, who has a range of kilometers, small cells are characterized with short coverage range of few hundreds of meters [77]. Another contrast is that the height of the BS antenna in small cells networks is expected to be lower than that of the normal macrocell networks, which might make it difficult to establish a line of sight link between the transmitting BS and the users [78]. Moreover, macrocell networks usually use high powered transmitters to cover a wide geographic area; consequently, not always an energy efficient networks and connectivity to cell-edge users could be a liability [78]. SCDNs represent a good solution for this kind of problems, because it involves many small cells that are deployed close to the users and low powered cells, which make it an important step toward improving the capacity and establishing green networks [78]. In addition, different cell sizes along with different mobility scenarios are included in the studies that investigated heterogeneous network (HetNet) performance to model the variety of cells in each tier. For instance, Macro-cell radius is usually about $(500m < r)$, microcell $(100m < r < 500m)$, pico-cell $(r < 100m)$, and different mobility scenarios [31, 72].

Not many works have considered NOMA based multicell networks, except [79, 80] which investigated the use of coordinated beamforming techniques to deal with ICI. However, the considered scenario is based on two cells only and they used a cognitive radio (CR) inspired power allocation. In Chapter 6 of this thesis, we will investigate the performance of NOMA based small cells
dense networks with a model that includes more than two cells.

2.5.1 Challenges Against Dense Cell-Deployment

Placing many small cells is expected to return in a number of benefits, theoretically, the number of the deployed small cells scales up the overall capacity achieved in a unit area. In addition, it helps in reducing the distance between the BSs and the users which would enhance the power of the signal received by the user, improves connectivity, and increase EE. Secondly, as compared to the 4G networks, the small cells are more energy efficient structures as they use less transmission power and their operation is demand-dependent, and they could pass through a passive state called dozing or sleeping state, where this in turn could reduce power consumption, interference and doubles EE [71, 81]. Thirdly, being small in size means it could be placed anywhere it required to fulfill the coverage nulls, and economically, it has low expenses and hence the financial side will not be an obstacle for the users. Cell densification could lead to the ease of synchronization between neighboring cells and also each cell could help the users in the detection process of the nearby cells in case that the macrocell is absent [75, 81]. Therefore, it is expected that increasing the number of cells in a specific coverage area and shrinking the cell radius would enhance the capacity and offer more spectrum reuse.

However, in practice, the closer-cells deployment not only strengthen the desired signal but also boost the interference from other cells. Therefore the ultra-dense networks are challenged in several ways [71]. In addition, it is challenging to control the boundaries of the densification process and to choose the right number of cells. From EE point of view, it is important to decide that a certain level of cell density be enough to serve increasing user’s density. It is obvious that as the cell density increases, ICI will rise accordingly among nearby cells. As a result, a solution such as frequency reuse or users grouping might be necessary to tackle such problem [78, 82, 83]. Another challenge is to control the unplanned deployment of this large number of cells, i.e. to deploy the cells in the right positions based on users need; where, according to [83–85]; a dense small cells deployment could lead to “random topology network” which makes it difficult to control in practice. However, adopting NOMA as the air interface for these networks could ease this problem as it allows the cells to share the same spectrum. Chapter 6 of
CHAPTER 2. NOMA FOR 5G NETWORKS

this thesis will study NOMA based SCDN to assess some of these challenges.

2.5.2 Modeling Cell-Deployment and Performance Metrics

Some works have already been done in the field of dense networks to tackle the problem of dealing with a large number of cells. For example, a survey was presented in [86] about the problems and the solutions of network densification. According to the authors, cells densification could lead to improving the spatial reuse of cells and lead to a better network capacity.

The ideal hexagonal grid was mostly used to model the spatially deployed cellular networks. However, the statistical and the theoretical studies presented in [87–89] proved that the cellular networks follow irregular topologies that randomly change from one geographical location to another rather than the idealized hexagonal grid structure. In [87], the authors simulated actual BS locations and illustrated that the user-SINR has an upper bound represented by user-SINR experienced in an idealistic grid network and a lower bound represented by user-SINR experienced in a random network. These works beside [88,90] supported the use of stochastic geometry to model the random cell deployment.

Stochastic geometry was first used to design the transmission ranges and strategies and to characterize the aggregate interference coming from a Poisson field of interferers. This field reflects that the transmitters are independently, randomly, and uniformly scattered in the spatial domain where a Poisson point process (PPP) variable could be used to model the number of transmitters in any bounded area [91,92]. The pillar strength of the stochastic geometry based analysis is represented by its ability to reflect the spatially-random nature of the wireless networks. [90,93,94]. In terms of the performance metrics for dense networks, stochastic geometry can be used to build an analytical paradigm that is applicable on average for several realizations of the cellular networks. Stochastic geometry does not model the performance of a specific cellular network realization at a specific geographical location. Instead, it investigates the average performance of all users in all locations. Usually, such average performance metric defined as spatially averaged (SA) performance, which is the main profit behind using stochastic geometry analysis. Examples of the SA performance metrics of interest in cellular networks are the coverage probability which is defined as the probability that the user
achievable rate exceeds a specified threshold. Another metric is the ergodic capacity that measures the long-term achievable rate averaged over all the states of the network realization (i.e., channel and interference) \([95]\).

The authors in \([2, 96–99]\) presented EE calculations for different environments, for example, rural, urban, and suburban models. The presented schemes also implied a consumed power calculation model that could be implemented for different cell types. The power calculations included the static power dissipated by BS parts besides the used power dissipated through transmission. These models will be depended throughout this thesis.

The authors in \([100]\) proposed an analytical representation for multi-tier HetNets, where the Poisson distribution and the grid schemes were used to model each tier separately from other tiers. In particular, with Poisson model, the BSs numbers and position follow a Poisson distribution, whereas in Grid model, the number of BSs is constant and they follow a uniform distribution throughout the covered area. The authors used the coverage probability as a performance metric and they suggested that the signal to interference threshold has a direct effect on the coverage probability. In addition, the same authors presented in \([101]\) a homogeneous, multicell scheme based on PPP to consider different BSs placement strategies such as grid and Poisson. The comparison was in terms of the coverage probability. Despite the fact that the grid model outweigh the Poisson one in terms of performance, the latter one seems to be more realistic in modeling cellular networks inside a geographic area with variable radius as well as a variable number of cells.

Figures 2.7 and 2.8 depict part of what was modeled and the results that were found by the author in \([100, 101]\). However, using PPP as a deployment scheme faced some criticism as it, sometimes; deploy two or more base stations in a way that they are too close to each other which make it a more hypothetical than practical deployment scheme. In addition, close cells lead to improper interference representation as well as bad resource usage. Hence, some literature went toward proposing a new Poisson – modified scheme or what is called the Matern Point process (some references call it Hard core process), where this model encompass the clustering of any two or more cells depending on how they are close to each other maintaining a minimum distance to evade the problem encountered with the PPP model of so closed
cells. Where, [71] used Matern Hard Process (MHP) as another tool for modeling the cell deployment besides PPP, the main difference between the two approaches is that the MHP is more uniform than the PPP which is tending to be more irregular, see Figure 2.8 for comparison. The considered terrain is assumed to be circular with radius R and Poisson distributed cells of density ($\lambda$) inside this area.

In regard to the frequency planning in a multicell scenarios, the authors in [102] proved analytically that, for a given BS/user density, the UFR can maintain optimal SE. According to UFR, each BS has full access to the entire available RBs which contribute in enhancing the spectrum usage in both time and space domains. If UFR is applied, multicell systems would suffer from performance decline due to ICI which degrades both system SE and link quality especially the cell-edge users will be the ones who get affected the most by the interference. Those users would not be able to maintain the required QoS, especially if they stay on the cell boundaries for a long time. One solution is to increase the BS transmission power, however, this could make the interference problem get worse as more interference will be caused to the nearby cells, as
Figure 2.8: Voronoi diagram for Matern Distribution and Poisson Distribution
a result, applying NOMA seems to be a promising solution as it allows all users to use the whole available RBs of all BSs.

From the previous discussion it is clear that some authors suggested that using a technique like dense networks could have some pros; but, at the same time, they faced some challenges. In Chapter 6 of this thesis, we will investigate the performance of NOMA based SCDN and optimize its sum rate and its EE to meet the expected future capacity requirements. In addition, it is clear from the literature works that PPP provides the more realistic representation of both BSs and users distributions. Hence, PPP will be used in Chapter 6 to model the deployment of both the users and the BSs in SCDN.

2.6 Chapter Summary

This chapter presented NOMA as one of the key techniques that can enhance the capacity of 5G networks to fulfill the expected high traffic demand. It addressed the concept of NOMA along with a general view of the currently used techniques to be used in combination with NOMA such as MIMO technology and SCDN. This chapter also addressed the performance gain that could be obtained by combining MIMO with NOMA. It also presented a review of literature works about the most important aspects for the modeling of multicell networks besides some of the expected problems as preliminaries to study the combination of NOMA and SCDN. It is clear from those works that NOMA promises more SE than the OMA. It also presented the main distribution schemes that are used in carrying out both BSs and users deployment in practical situations. Finally, the chapter presented an overview of the resource allocation concept and also identified how the wireless channel modeling will be applied throughout this thesis.
Chapter 3

Sum Rate Maximization for NOMA System

3.1 Introduction

Developing the air interface techniques represents one of the ways to enhance the capacity of future networks and to meet the predicted high mobile data demand. To reach this goal, NOMA have been presented as a candidate multiple access to improve the capacity and the SE. This chapter investigates the sum rate maximization for a downlink NOMA system over a frequency selective fading channel. Since obtaining the optimal solution for NOMA requires high complexity numerical operations, two low complex closed-form suboptimal solutions are proposed for a two-user case. However, the solutions are restricted to two users only. It is extended for larger number of users using a subband based approach, whereby two users will transmit in each subband. Yet, dividing the total bandwidth into subbands is not spectrally efficient, where all users can transmit through the whole bandwidth. Thus the concept of HP is also proposed in this chapter, where users are grouped in pairs and allowed to occupy the whole bandwidth. Then the users are paired according to their channel gains and the pairs are multiplexed in the power domain, which is obtained from a modified solution to the two-user case. This will also facilitate optimizing the sum rate for a NOMA system with a large number of users.
3.2 Problem Formulation and Solution for Two-User Scenario

The RA optimization problem is first formulated to maximize the sum rate of the two-user NOMA system, and the solution is then generalized to the multiuser case. Recall that from Chapter 2, if two users are multiplexed over the s-th RB using NOMA principles; then the achievable rate by the user with the higher \((H)\) channel conditions is as given in (2.3) and that of the user with the low \((L)\) channel conditions is as given in (2.4), and their sum rate over that RB is given by (2.5). In order to guarantee that each user is able to achieve its target rate, the optimization will include nonlinear constraints of proportional rate.

Mathematically, the problem of the two-user scenario is formulated as

\[
\begin{align*}
\text{maximize} & \quad R \\
\text{Subject to} & \quad \sum_{s=1}^{S} (P_{s}^{(H)} + P_{s}^{(L)}) \leq P_t \\
& \quad P_{s}^{(H)}, P_{s}^{(L)} \geq 0, \forall s \\
& \quad \sum_{s=1}^{S} R_{s}^{(H)} : \sum_{s=1}^{S} R_{s}^{(L)} = \Phi_{\text{min}}^{(1)} : \Phi_{\text{min}}^{(2)}. \quad (3.4)
\end{align*}
\]

where \(R\) denotes the sum rate over all the RBs, and it is given by \(R = \sum_{s=1}^{S} R_{s}\), constraints (3.2) and (3.3) are to guarantee a positive allocated power and limited by the maximum allowable \(P_t\). In addition, the minimum rate proportional fairness constraint (3.4) is to control the achievable throughput by all users where \(\Phi_{\text{min}}^{(1)}\) and \(\Phi_{\text{min}}^{(2)}\) are the minimum rate requirements for user 1 and user 2, respectively. Constraint (3.4) helps to maintain proportionality between the minimum achievable rates of the users. In other words, the proportionality is guaranteed for the minimum achievable rates, but the rate obtained in the solution is not restricted to this ratio, as long as the minimum rates for all users are satisfied. It is important to point out that by using the proportional fairness constraint, once the minimum rates for all users are
satisfied, the remaining resources will also be allocated in a proportional manner. Such approach is important to maintain fairness in distributing the radio resources among these users and to ensure that the weak users have enough power to decode their own data from the received signal while treating the stronger users as noise, and to ensure that the stronger users have enough power to apply SIC and cancel the effect of the weak users and detect their own data. Thus, this constraint is vital especially for power domain NOMA system to help in maintaining power domain multiplexing of the users (i.e., to establish SC). Without this constraint, the maximum sum rate could simply be achieved by allocating all the bandwidth and power to one user or a few users who have the best channel conditions and not all users will be allowed to transmit, and thus no fairness would be established among the users. In addition, another important property of this constraint is that it can utilize the potential advantage of NOMA over OMA [12]. The minimum rate requirement is assigned to each user based on the large scale fading factor (the distance based path loss and the log-normal shadowing factor) experienced by that user in addition to the small scale fading effects. Since path loss and shadowing is more dominant and vary slowly, the proportionality constraint is therefore effectively more on a long term basis rather than short term basis. The following proof shows that the formulated problem is of concave nature.

Proof: The achievable rate by the user with the higher \((H)\) channel conditions and the user with the low \((L)\) channel conditions is as given in (2.3) and (2.4), respectively, and their sum rate over that RB is given by (2.5). For the \(s\)-th RB, their sum rate can be expressed as

\[
R = R_s^{(H)} + R_s^{(L)}
\]

where

\[
R = B_s \log_2 \left(1 + \frac{p_s^{(H)} |h_s^{(H)}|^2}{B_s N_0}ight) + B_s \log_2 \left(1 + \frac{p_s^{(L)} |h_s^{(L)}|^2}{P_s^{(H)} |h_s^{(L)}|^2 + B_s N_0}ight)
\]

which by using the logarithm properties can be expressed as

\[
R = B_s \log_2 \left[\left(1 + \frac{p_s^{(H)} |h_s^{(H)}|^2}{B_s N_0}\right) \ast \left(1 + \frac{p_s^{(L)} |h_s^{(L)}|^2}{P_s^{(H)} |h_s^{(L)}|^2 + B_s N_0}\right)\right].
\]

First, take the first derivative with respect to \(P_s^{(H)}\).
\[ \frac{dR}{dP_s^{(H)}} = B_s \left( \frac{\left| h_s^{(H)} \right|^2 \left( 1 + \frac{P_s^{(L)|h_s^{(L)}|^2}{B_s N_0} \right)}{\left( 1 + \frac{P_s^{(H)|h_s^{(H)}|^2}{B_s N_0} \right)} \right) - \frac{P_s^{(L)|h_s^{(L)}|^2}}{\left( 1 + \frac{P_s^{(L)|h_s^{(L)}|^2}{B_s N_0} \right)} \right). \]

After that, the second derivative with respect to \( P_s^{(H)} \) is found to be

\[ \frac{d^2 R}{dP_s^{(H)2}} = -B_s \left[ \left( \frac{\left| h_s^{(L)} \right|^2}{P_s^{(H)|h_s^{(L)}|^2} + B_s N_0} \right)^2 \left( \frac{\left( P_s^{(H)|h_s^{(L)}|^2} \right)^4}{\left( P_s^{(H)}|h_s^{(L)}|^2 + B_s N_0 \right)^2} \right) + 2 \left( \frac{\left| h_s^{(L)} \right|^2}{P_s^{(H)|h_s^{(L)}|^2} + B_s N_0} \right)^3 \left( P_s^{(H)} + 2P_s^{(H)} P_s^{(L)} + \left( P_s^{(L)} \right)^2 \right) \right] - \frac{\left( \frac{\left| h_s^{(L)} \right|^2}{P_s^{(H)|h_s^{(L)}|^2} + B_s N_0} \right)^2 \left( \left( \frac{\left( h_s^{(L)} \right)^2}{P_s^{(H)|h_s^{(L)}|^2} + B_s N_0} \right)^2 \left( \left( P_s^{(H)} + P_s^{(L)} \right) \left( h_s^{(L)} \right)^2 + B_s N_0 \right)^2 \right)}{\left( P_s^{(H)} |h_s^{(L)}|^2 + B_s N_0 \right)^4} \left( \left( P_s^{(H)} |h_s^{(L)}|^2 + B_s N_0 \right)^4 \right) \left( \left( P_s^{(H)} |h_s^{(L)}|^2 + B_s N_0 \right)^4 \right) \right]. \]

Secondly, take the first derivative with respect to \( P_s^{(L)} \)

\[ \frac{dR}{dP_s^{(L)}} = B_s \frac{\left| h_s^{(L)} \right|^2}{\left( \frac{\left( P_s^{(H)}|h_s^{(L)}|^2 + B_s N_0 \right)}{1 + \frac{P_s^{(L)|h_s^{(L)}|^2}{B_s N_0}} \right)} \]

From which the second derivative with respect to \( P_s^{(L)} \) is found as

\[ \frac{d^2 R}{dP_s^{(L)2}} = B_s \left( \frac{\left| h_s^{(L)} \right|^2}{\left( \frac{\left( P_s^{(H)} + P_s^{(L)} \right) |h_s^{(L)}|^2 + B_s N_0 \right)}{2} \right)^2. \]

It is clear that both (3.5) and (3.6) are negative, which proves that the objective function is concave.

### 3.2.1 Problem Solution using Lagrange Dual Function

Taking into account the objective function in (2.5) and using Lagrange Dual Decomposition (LDD) approach in [103] to solve the formulated problem in (3.1) to (3.4), the Lagrangian function of the optimization problem could be expressed as
\[ F = B_s \log_2 \left( (1 + \gamma_s^{(H)}) (1 + \gamma_s^{(L)}) \right) - \lambda \left( \sum_{s=1}^{S} (P_s^{(H)} + P_s^{(L)}) - P_t \right) - \]

\[
\mu \left( \frac{B_s \log_2 (1 + \gamma_s^{(H)})}{\Phi_{min}^{(1)}} - \frac{B_s \log_2 (1 + \gamma_s^{(L)})}{\Phi_{min}^{(2)}} \right) \quad (3.7)
\]

where \( \mu \) and \( \lambda \) represent the Lagrange multipliers. Differentiating against \( P_s^{(H)} \), \( P_s^{(L)} \), \( \lambda \), and \( \mu \), respectively, we obtain

\[
\frac{dF}{dP_s^{(H)}} = \frac{B_s \left( \frac{\gamma_s^{(H)} (1 + \gamma_s^{(L)})}{P_s^{(H)}} - \frac{(\gamma_s^{(L)})^2 (1 + \gamma_s^{(H)})}{P_s^{(L)}} \right)}{\left( 1 + \gamma_s^{(H)} \right) \left( 1 + \gamma_s^{(L)} \right)} - \lambda \mu \left( \frac{B_s \gamma_s^{(H)}}{\Phi_{min}^{(1)} P_s^{(H)}} \left( 1 + \gamma_s^{(H)} \right) + \frac{B_s (\gamma_s^{(L)})^2}{\Phi_{min}^{(2)} P_s^{(L)}} \left( 1 + \gamma_s^{(L)} \right) \right) \quad (3.8)
\]

\[
\frac{dF}{dP_s^{(L)}} = -\lambda + \frac{\gamma_s^{(L)} \mu B_s}{\Phi_{min}^{(1)} P_s^{(L)} \left( 1 + \gamma_s^{(L)} \right)} + \frac{\gamma_s^{(L)} B_s}{P_s^{(L)} \left( 1 + \gamma_s^{(L)} \right)} \quad (3.9)
\]

\[
\frac{dF}{d\lambda} = P_t - \sum_{s=1}^{S} (P_s^{(H)} + P_s^{(L)}) \quad (3.10)
\]

\[
\frac{dF}{d\mu} = \frac{B_s \log_2 (1 + \gamma_s^{(L)})}{\Phi_{min}^{(2)}} - \frac{B_s \log_2 (1 + \gamma_s^{(H)})}{\Phi_{min}^{(1)}} \quad (3.11)
\]

Setting each of these equations to zero and solving (3.8) for the Lagrange variable \( \lambda \) we obtain
which can be used to solve (3.9) for $P_s^{(H)}$ as

$$P_s^{(H)} = \frac{|h_s^{(H)}|^2 + |h_s^{(L)}|^2}{\Phi_{\min}^{(1)} + \Phi_{\min}^{(2)}} B_s N_0 - \left( |h_s^{(L)}|^2 \Phi_{\min}^{(1)} + |h_s^{(H)}|^2 \Phi_{\min}^{(2)} \right) \mu B_s N_0 \left( \frac{|h_s^{(H)}|^2 + |h_s^{(L)}|^2}{\Phi_{\min}^{(1)} + \Phi_{\min}^{(2)}} \right).$$

At this point, the solution is optimal; however, solving for $P_s^{(L)}$ by using (3.10) would be

$$\sum_{s=1}^{S} P_s^{(L)} = P_t + \frac{B_s N_0}{\mu} \sum_{s=1}^{S} \left( \frac{|h_s^{(L)}|^2 \Phi_{\min}^{(1)} + |h_s^{(H)}|^2 \Phi_{\min}^{(2)}}{|h_s^{(H)}|^2 |h_s^{(L)}|^2} \right).$$

Solving this requires the use of complex numerical solutions.

### 3.2.2 Equal per RB Power Allocation (ERPA)

To derive a simple closed-form solution, we first propose a low complexity suboptimal approach that allocates the power equally among all the RBs. In other words, we assume that the total transmit power of all RBs are equal and the total transmission power in each RB ($P_{RB}$) is obtained by simply dividing the total available power $P_t$ by the total number of RBs $S$ as follows
\[ P_{RB} = P_s^{(L)} + P_s^{(H)} = \frac{P_s}{S}. \] (3.15)

from which we can obtain that

\[ P_s^{(L)} = P_{RB} - P_s^{(H)}. \] (3.16)

Using (3.16) to solve (3.14) for \( P_s^{(L)} \) we obtain

\[ P_s^{(L)} = P_{RB} + \frac{\left( |h_s^{(L)}|^2 \Phi_{min}^{(1)} + |h_s^{(H)}|^2 \Phi_{min}^{(2)} \right) \mu B_s N_0 - \left( |h_s^{(H)}|^2 + |h_s^{(L)}|^2 \right) \Phi_{min}^{(1)} \Phi_{min}^{(2)} B_s N_0}{|h_s^{(H)}|^2 |h_s^{(L)}|^2 2 \mu \left( \Phi_{min}^{(1)} + \Phi_{min}^{(2)} \right)} . \] (3.17)

Since (3.13) and (3.17) still contain the Lagrangian variable \( \mu \), it has to be solved in order to allocate the transmission power for the strong and weak user. Using (3.17) to solve (3.11) for the second Lagrange variable \( \mu \) we obtain

\[ \mu = \frac{\psi_1 \left( \Phi_{min}^{(1)} - \Phi_{min}^{(2)} \right) + \left( \Phi_{min}^{(1)} + \Phi_{min}^{(2)} \right) \psi_3 \sqrt{\psi_1 \left( \Phi_{min}^{(1)} \Phi_{min}^{(2)} \right)}}{2 \left( \psi_2 B_s N_0 |h_s^{(L)}|^2 P_{RB} \left( \Phi_{min}^{(1)} + \Phi_{min}^{(2)} \right)^2 + \psi_4 \right)} \] (3.18)

where

\[
\begin{align*}
\psi_1 &= \left( |h_s^{(H)}|^2 - |h_s^{(L)}|^2 \right)^2 \Phi_{min}^{(1)} \Phi_{min}^{(2)} \Gamma_1 B_s N_0 \\
\psi_2 &= B_s N_0 |h_s^{(H)}|^2 |h_s^{(L)}|^2 \Gamma_2 \\
\psi_3 &= \sqrt{4 \frac{\psi_2 B_s N_0 |h_s^{(L)}|^2 P_{RB} + 4 \psi_2 + \psi_1}{\Phi_{min}^{(1)} \Phi_{min}^{(2)}}} \\
\psi_4 &= \psi_2 \left( \Phi_{min}^{(1)} + \Phi_{min}^{(2)} \right)^2 + \psi_4 \\
\Gamma_1 &= 2 \Phi_{min}^{(1)} \\
\Gamma_2 &= 2 \Phi_{min}^{(2)}
\end{align*}
\]

Finally, substituting (3.18) into (3.17), the suboptimal power for the weak user is found to be


\[ P_s^{(L)} = \frac{2|h_s^{(H)}|^2|h_s^{(L)}|^2 P_{RB} + \left(|h_s^{(H)}|^2 + |h_s^{(L)}|^2\right) B_s N_0}{2|h_s^{(H)}|^2|h_s^{(L)}|^2} - \frac{\psi_3 \sqrt{B_s N_0}}{2|h_s^{(H)}|^2|h_s^{(L)}|^2 \sqrt{\Gamma_1}} \]

while that of the stronger user is

\[ P_s^{(H)} = -\left(\frac{|h_s^{(H)}|^2 + |h_s^{(L)}|^2}{2|h_s^{(H)}|^2|h_s^{(L)}|^2}\right) B_s N_0 + \frac{\psi_3 \sqrt{B_s N_0}}{2|h_s^{(H)}|^2|h_s^{(L)}|^2 \sqrt{\Gamma_1}}. \] (3.20)

Despite the assumption made through the way to reach (3.19) and (3.20), it is important to highlight that the constraint (3.4) is still applicable for the obtained solution.

It is also worth mentioning that the superscripts \((H)\) and \((L)\) are included just to distinguish the parameters of the users with the better channel gain from those with weaker channel gains at the \(s\)-th RB and not over all RBs. It also does not necessarily mean that \(P_s^{(H)}\) is higher than \(P_s^{(L)}\), where it could be less than or equal to \(P_s^{(L)}\) depending on the final values from the proposed closed-form solutions.

### 3.2.3 Average Channel based Power Allocation (ACPA)

While the complexity of the proposed ERPA method is significantly lower than that of the optimal one, it still needs \(S\) times of calculations for each power allocation step. In order to achieve a further simplification, we propose the ACPA scheme that depends on the average channel gain of each user across the entire bandwidth for power allocation. In other words, the average channel gain of the strong and the weak user is determined by,

\[ \overline{G}_H = \frac{\sum_{s=1}^{S} |h_s^{(H)}|^2}{S} \quad \text{and} \quad \overline{G}_L = \frac{\sum_{s=1}^{S} |h_s^{(L)}|^2}{S}, \]

respectively, and these values will be used to determine the power to be allocated to the respective user. Applying this approach to (3.19) and (3.20), the suboptimal power for the strong user is given by

\[
\begin{align*}
P_s^{(H)} &= -\frac{\left(\overline{G}_H + \overline{G}_L\right) B_s N_0}{2\overline{G}_H \overline{G}_L} + \frac{\sqrt{B_s N_0} \sqrt{4 \sqrt{\psi_2} \overline{G}_L P_t + 4 \psi_2 + B_s N_0 \left(\overline{G}_H - \overline{G}_L\right)^2 \Gamma_1}}{2\overline{G}_H \overline{G}_L \sqrt{\Gamma_1}} \\
&= -\frac{\left(\overline{G}_H + \overline{G}_L\right) B_s N_0}{2\overline{G}_H \overline{G}_L} + \frac{\sqrt{B_s N_0} \sqrt{4 \sqrt{\psi_2} \overline{G}_L P_t + 4 \psi_2 + B_s N_0 \left(\overline{G}_H - \overline{G}_L\right)^2 \Gamma_1}}{2\overline{G}_H \overline{G}_L \sqrt{\Gamma_1}} \tag{3.21}
\end{align*}
\]
and that for the weak user is

\[ p_{(L)}^s = \frac{2G_H G_L P_t + (G_H + G_L) B_s N_0}{2G_H G_L} \sqrt{B_s N_0} \sqrt{\frac{4\psi_1 G_L P_t + 4\psi_1 + B_s N_0 (G_H - G_L)^2 \Gamma_1}{2G_H G_L \sqrt{\Gamma_1}}} \] (3.22)

This method offers simplicity over the ERPA method, and will also be compared to the optimal solution in Section 3.5.

### 3.3 Multiuser NOMA with Hierarchical Pairing Concept

The closed-form solution obtained in Section 3.2 allows low complexity implementation of power allocation for two-user NOMA. To extend the applicability of these closed-form solutions to a multiuser case, we propose the HP concept which groups the users in pairs as shown in Figure 3.1. We denote the total number of pairs to be \( Z \), and the pairs are arranged in an ascending order according to their channel gains from the bottom (first pair is the weakest) to the top (the \( Z \)-th pair is the strongest). The transmitted signal at the \( s \)-th RB is given by

\[ X_s = \sum_{j=1}^{Z} \left( \sqrt{p_{(H)}^{j,s}} X_{j,s}^{(H)} + \sqrt{p_{(L)}^{j,s}} X_{j,s}^{(L)} \right) \] (3.23)

which includes the information intended for all users whom will share the same time-frequency resource, where \( X_{j,s}^{(H)} \) and \( X_{j,s}^{(L)} \) are the superposition coded information bearing signal intended for the strong and weak user in the \( j \)-th pair respectively with \( p_{(H)}^{j,s} \) and \( p_{(L)}^{j,s} \) as the corresponding transmission power. The total power for this \( j \)-th pair at \( s \)-th RB is denoted as \( P_{j,s} = p_{(H)}^{j,s} + p_{(L)}^{j,s} \). It is worth mentioning that the transmitted signals with HP has exactly the same form as that of the conventional NOMA, as it is effectively a summation of all superposition coded signals from all users. The received signal by the stronger user of the \( z \)-th pair at the \( s \)-th RB is given by
where $h_{z,s}^{(H)}$ represents the channel gain between the BS and the strong user and $n_{z,s}^{(H)}$ represents the AWGN. The expression for the weaker user in this pair is similar to (3.24) but with the superscript $(H)$ as $(L)$. From (3.24), it is clear that each user will receive a signal with its data and those intended for other users.

Figure 3.1 illustrates NOMA structure with the HP concept. Starting from the bottom of Figure 3.1, the weakest user in the first pair will not perform SIC while the better user in this pair will perform SIC only to its partner in this pair.

![Figure 3.1: NOMA structure with HP concept.](image-url)

At the top of Figure 3.1, on the other hand, the weaker user of the $Z$-th pair (the strongest pair) will perform SIC to all of the previous pairs and consider its stronger partner as noise, while the strongest user at this pair will perform SIC to all of the previous pairs and its partner as well. Assuming
perfect decoding, Figure 3.2 depicts the SIC process for four users (two pairs: $Z = 2$) along with the pairing concept. It must be noted that the SIC detection process has the same principles as that in conventional NOMA. The key difference is in the power allocation procedure, which is described in the following subsection. It must be noted that the purpose and the procedure of pairing in our proposed approach is different to conventional pairing in NOMA \[10, 29\].

Driven by the high complexity of the power allocation algorithms, conventional NOMA pairing approaches allocate two users into a RB (or a sub-band) such that the algorithmic complexity can be lowered. However such a horizontal approach will not fully exploit the potential of NOMA, which benefits from using all bandwidth for all users. On the other hand for our HP approach, since the power allocation solution for two-user case is in closed-form, complexity is not an issue and we can expand to cases with more than two users to exploit the capacity improvements. Note that the proportional fairness among the users is still maintained in here for any arbitrary number of users; as the same constraint in (3.4) is still in use along with the derived suboptimal solutions in (3.19) and (3.20).

The proposed HP structure allows the use of the solutions in (3.19) and (3.20) to allocate power for the users within each pair. However, the power allocation across the pairs have to be determined. For the case of $K$ users, the sum rate of all possible pairs ($Z = \frac{K}{2}$) is given by

$$R = B_s \sum_{z=1}^{Z} \sum_{s=1}^{S} \left( \log_2 \left( 1 + \gamma_{z,s}^{(H)} \right) + \log_2 \left( 1 + \gamma_{z,s}^{(L)} \right) \right)$$  \hspace{1cm} (3.25)

where the terms

$$\gamma_{z,s}^{(H)} = \frac{P_{z,s}^{(H)} |h_{z,s}^{(H)}|^2}{I_s |h_{z,s}^{(H)}|^2 + B_s N_0}$$  \hspace{1cm} (3.26)

$$\gamma_{z,s}^{(L)} = \frac{P_{z,s}^{(L)} |h_{z,s}^{(L)}|^2}{\left( I_s + P_{z,s}^{(H)} \right) |h_{z,s}^{(L)}|^2 + B_s N_0}$$  \hspace{1cm} (3.27)

are the received SINR of the strong and weak user in the $z$-th pair at the $s$-th RB, respectively, and $I_s$ represents the power allocated to the preceding stronger pairs at the same $s$-th RB, and it is given by
Figure 3.2: Illustration of the SIC process with perfect decoding in hierarchically paired NOMA with four users (two pairs: $Z = 2$), where the users within the pairs are arranged in an increasing order.

$$I_s = \begin{cases} 
0 & z = Z \\
\sum_{k=z+1}^{Z} P_{k,s}^{(H)} + P_{k,s}^{(L)} & 1 \leq z \leq (Z - 1) 
\end{cases}$$

(3.28)

It is worth mentioning that the value of $I_s$ is not the same for all pairs. The optimal solution to maximize the sum rate in (3.25) with total power and minimum rate constraints requires complex numerical solutions. However, if the total transmission power for each pair is known, the optimization problem in (3.1)-(3.4) can be used to determine the optimal power allocation for the
can be repeated to each group as a second stage; i.e., two subgroups are
formed. In this scheme, the terms \(|h_s^{(H)}|^2\) and \(|h_s^{(L)}|^2\) will form the other. In this scheme, the terms \(|h_s^{(H)}|^2\) and \(|h_s^{(L)}|^2\) have only two candidates to allocate the power to. In here, the power will
be allocated to the hierarchically paired users in hierarchical manner based
on their channel gains. By splitting the users into two groups and combining
their channel gains, the above pair-based power allocation solution can be
modified to obtain the power allocated for each group. Following the concept
of NOMA, the stronger half of the users will form one group, and the weaker
half will form the other. In this scheme, the terms \(|h_s^{(H)}|^2\) and \(|h_s^{(L)}|^2\) in (3.29)
and (3.30) will become the sum of all channel gains in the weaker and stronger
group respectively.

Once the power for the two groups are determined, the same procedure
can be repeated to each group as a second stage; i.e., two subgroups are

\[
P_{z,s}^{(H)} = \frac{(B_s N_0 (|h_s^{(H)}|^2 + |h_s^{(L)}|^2) + 2|h_s^{(H)}|^2|h_s^{(L)}|^2 P_{z,s})}{2|h_s^{(H)}|^2|h_s^{(L)}|^2}
\]

\[
\sqrt{4 \left(\frac{|h_s^{(H)}|^2}{|h_s^{(L)}|^2}\right)^2 I_s P_{z,s} \Gamma_2 + 4\psi_2 (|h_s^{(H)}|^2 I_s + |h_s^{(L)}|^2 P_{z,s}) + B_s N_0 \left(\frac{\psi_1}{\frac{\psi_1}{\min} + 4\psi_2}\right)}
\]

\[
2|h_s^{(H)}|^2|h_s^{(L)}|^2 \sqrt{\Gamma_1}
\]

(3.29)

\[
P_{z,s}^{(L)} = \left(\frac{(|h_s^{(H)}|^2)^2 (|h_s^{(L)}|^2)^2 I_s \Gamma_2 (I_s \Gamma_2 + 4P_{s}^{(L)} \Gamma_1) + (B_s N_0)^2 \left(\left(|h_s^{(H)}|^2 - |h_s^{(L)}|^2\right) \Gamma_1 + |h_s^{(L)}|^2 \Gamma_2\right)^2}{2|h_s^{(H)}|^2|h_s^{(L)}|^2 \Gamma_1}
\]

\[
+ 2\psi_2 \left(|h_s^{(H)}|^2 - |h_s^{(L)}|^2\right) I_s \Gamma_2 + |h_s^{(L)}|^2 I_s \Gamma_1 + 2|h_s^{(L)}|^2 P_{s}^{(L)} \Gamma_1 \right)^{1/2} +
\]

\[
\frac{|h_s^{(H)}|^2 (|h_s^{(L)}|^2 I_s (\Gamma_2 - 2\Gamma_1) - 2\Gamma_1) B_s N_0 + |h_s^{(L)}|^2 B_s N_0 (\Gamma_2 - \Gamma_1)}{2|h_s^{(H)}|^2|h_s^{(L)}|^2 \Gamma_1}
\]

(3.30)

3.3.1 Hierarchical Pairing Power Allocation (HPPA)

The key to the low complexity closed-form solutions in (3.19) and (3.20) is to
have only two candidates to allocate the power to. In here, the power will
be allocated to the hierarchically paired users in hierarchical manner based
on their channel gains. By splitting the users into two groups and combining
their channel gains, the above pair-based power allocation solution can be
modified to obtain the power allocated for each group. Following the concept
of NOMA, the stronger half of the users will form one group, and the weaker
half will form the other. In this scheme, the terms \(|h_s^{(H)}|^2\) and \(|h_s^{(L)}|^2\) in (3.29)
and (3.30) will become the sum of all channel gains in the weaker and stronger
group respectively.

strong and weak user in each pair. Since there is now other pairs for each
RB, the interference will affect the optimal solution. The ones in (3.19) and
(3.20) will now be given by (3.29) and (3.30). In the following, we propose a
low complexity power allocation approach for this pairing-based scheme, and
also several simple power allocation schemes for comparisons.
formed in each group of the previous stage. This multiple stage approach is repeated until the subgroups become pairs of users, and the solution in (3.19) and (3.20) can be used directly.

To better illustrate the procedure, we consider an example of 8 users at the s-RB and the multistage process is shown in Figure 3.3. The application of this scheme is as follows:

- **First Stage** (2 groups each has 4 users): Divide the users into two groups with $G_{g1}$ and $G_{g2}$ as the sum of the users channel gains in each group, which will be used to replace $|h_s^{(H)}|^2$ and $|h_s^{(L)}|^2$, respectively, in (3.19) and (3.20). With the total available power being the same as the total transmission power per each RB ($P_{RB}$), apply the obtained suboptimal solutions in (3.19) and (3.20) to find the power across the two groups to be $P_{g1}$ and $P_{g2}$.

- **Second Stage** (4 subgroups of 2 users): By dividing each group of users further into 2 subgroups we will have four pairs of users, and each pair with a combined channel gain of $G_{g1,1}$, $G_{g1,2}$ $G_{g2,1}$, and $G_{g2,2}$. Using (3.19) and (3.20) with $P_{g1}$ and $P_{g2}$ as the total power for each subgroup, the power across each pair could be found easily as $P_1$, $P_2$, $P_3$, and $P_4$.

- **Third Stage** (8 users): This is the user level stage, where the suboptimal solution of the ERPA method in (3.19) and (3.20), along with $P_1$, $P_2$, $P_3$, and $P_4$ used to replace $P_{RB}$, are used to allocate the power for all users within the pairs.

It must be noted that the number of users in the proposed HPPA approach can be any positive integer and not mandatory to be in the power of two. In the case of an odd number of users, the number of users per each group can be unequal and one group can have one user more than the other group. The same steps are followed repeatedly until the power is allocated to all users.

### 3.3.2 Equal-per-Pair Power Allocation

In order to demonstrate the advantage of the HPPA method, a number of trivial approaches will be used to compare against its performance. The first approach is to arrange the users in a descending order of channel gain (the user with the highest channel gain is at the top and the one with the lowest
channel gain at the bottom as depicted in Figure 3.1) and pair every two consecutive users (i.e., users with similar channel gains are paired together). Then the power are allocated equally for all pairs so that the total transmission power is divided by the number of pairs in all RBs, and is given by

\[ P_{eq} = \frac{P_{RB}}{Z}. \] (3.31)

where \( P_{eq} \) represents the total power allocated for each pair at each RB. Next, the suboptimal approaches will be used based on (3.29) and (3.30) to obtain the power for each user within the pairs. The advantage of this approach over HPPA is simplicity because (3.29) and (3.30) are used only once per pair.

### 3.3.3 Proportional-per-Pair Power Allocation

The second approach is to determine the transmission power for each pair in proportion to the combined channel gain of the paired users. At the \( s \)-th RB, the sum of the channel gains at the \( z \)-th pair is denoted as \( \overline{M}_{z,s} = |h_{z,s}^{(H)}|^2 + |h_{z,s}^{(L)}|^2 \). The power allocated to each pair at the \( s \)-th RB is denoted as \( P_s \) and it could be obtained by

\[ P_s = \frac{P_{RB} \overline{M}_{z,s}}{Z} \sum_{z=1}^{Z} \overline{M}_{z,s}. \] (3.32)
The benefit of using (3.32) is that each pair will be allocated a power that is proportional to its channel gain, which would help in maximizing the user rates. After deciding the power across each pair, the power within the pairs will be allocated using (3.29) and (3.30).

### 3.3.4 Subband-based ERPA

This method is applied by dividing the whole spectrum into subbands and then NOMA is applied within each subband. A maximum of two users (one user pair) are multiplexed per each subband (horizontal pairing) and their power is allocated using the ERPA scheme. Due to the simplicity of this method, it is less spectral efficient than others as not all of the RBs will be shared among the users.

It must be noted that the last three approaches do not guarantee satisfying the minimum rate criterion, and so are only for comparative purposes.

### 3.4 The Proposed Hybrid System

In spite of the difference between OFDMA and NOMA in the sense of orthogonality between users, their combination could enhance the capacity. While orthogonal access assigns part of the spectrum to each user, multiplexing the users in a non-orthogonal manner offer fairness among these users in terms of the achievable throughput as they are all allowed to use the whole bandwidth regardless of their channel conditions [12, 56]. However, the difference between the channel gains of cell center and cell edge users could be significant and hence applying NOMA to the entire spectrum may not be beneficial. Thus we propose a hybrid multiple access scheme such that part of the spectrum are reserved for orthogonal access (dedicated) and the rest (shared) are for all users by NOMA. An illustration of this hybrid scheme in a two user case is depicted in Figure 3.4. As an example from this figure, the first RB is dedicated for user 1 while the S-th RB is dedicated for user 2; on the other hand, both users shared the second RB. This gives the hybrid method the advantage of, firstly, being less susceptible to interference and requiring less SIC process than NOMA, and secondly more spectral efficient than OFDMA since some users could share more spectrum as compared to the purely orthogonal
The hybrid scheme maximizes the overall sum rate and maintains the minimum rate requirements by adaptively allocating the RBs among the users in orthogonal and/or non-orthogonal manners. Whenever the received channel gain by a user is significantly higher than that received by other users, the hybrid scheme will exclusively allocate this RB to that user in an orthogonal manner rather than sharing it among all users by NOMA. Likewise, whenever the users receive comparable channel gains, NOMA is applied rather than OFDMA. In such a way, the hybrid scheme will have a smaller non-orthogonal part than that of NOMA scheme, which means that the hybrid scheme requires simpler SIC application than NOMA.

Figure 3.4: Structure of the hybrid orthogonal - non orthogonal scheme.

For this hybrid scheme, the optimization problem will have to determine which RB is allocated to the orthogonal and non-orthogonal counterparts. In addition, the RBs classified for orthogonal transmission will also have to be allocated to the users. Moreover, the power allocation will also have to be optimized. The global optimal solution will involve high complexity numerical computation and thus we propose a low complexity multistage suboptimal approach, where RB allocation is first performed assuming equal power allocation, followed by the power allocation approach that has been mentioned in the earlier sections.
CHAPTER 3. SUM RATE MAXIMIZATION FOR NOMA SYSTEM

3.4.1 RB Allocation and Classification

To classify whether a RB should be used for orthogonal or non-orthogonal transmission, the respective achievable rates are first computed and the best one will be selected. For the orthogonal case, the achievable rate of each user over the $s$-th RB is calculated using

$$R_{\text{orth},s} = B_s \log_2 \left( 1 + \frac{P_{RB}|h_{s}^{(k)}|^2}{B_sN_0} \right). \quad (3.33)$$

On the other hand, the non-orthogonal sum rate of all $K$ users over the $s$-th RB is given by

$$R_{\text{non},s} = B_s \log_2 \left( 1 + \frac{\beta_{k,s}P_{RB}|h_{s}^{(k)}|^2}{B_sN_0} \right) + B_s \sum_{k=1, k \neq \tilde{k}}^{K} \log_2 \left( 1 + \frac{\beta_{k,s}P_{RB}|h_{s}^{(k)}|^2}{B_sN_0 + \sum_{m=k+1}^{K} \beta_{m,s}P_{RB}|h_{s}^{(k)}|^2} \right) \quad (3.34)$$

where $\tilde{k} \in \{1, 2, \ldots, K\}$ represents the index of the users who has the strongest channel gain at the $s$-th RB, $\beta_{k,s} = \frac{|h_{s}^{(k)}|^2}{\sum_{j=1}^{K} |h_{s}^{(j)}|^2}$ refers to the power allocation factor which always has a positive quantity and of $\sum_{k=1}^{K} \beta_{k,s} \leq 1$, and $P_{RB}$ stands for the total power per each RB which will be assumed to be equal for all RBs during the allocation process and it is calculated as $P_{RB} = \frac{P_t}{S}$, and finally $|h_{s}^{(k)}|^2$ stands for the channel gain received by the $k$-th user over the $s$-th RB. Allocating the power in this way will guarantee that the RB classification process will be largely done based on the channel gain of each RB. It is worth mentioning that after the classification process, the power allocation for the shared RBs will be applied using HPPA and for the orthogonal part, a water-filling based optimal solution will be used to allocate the power.

Next, by examining the RBs one by one, the RB classification process will
be done by comparing the achievable $R_{orth,s}$ by each user (i.e., the individual user rate of all users over the $s$-th RB which is determined using (3.33)) against the $R_{non,s}$ (i.e., the sum rate of all users over the same $s$-th RB which is determined using (3.34)). After that, for the $s$-th RB, if $R_{non,s} \geq R_{orth,s}$ then the respective RB will be shared using NOMA. Otherwise, this RB will be classified as an orthogonal RB and will be allocated for dedicated use by the user who has it with the highest achievable rate $R_{orth,s}$ for the purpose of sum rate maximization. Algorithm 3.1 shows the RB classification and allocation steps for the proposed hybrid method.

### 3.4.2 Power Allocation for the Hybrid System

At the end of the classification process, all users will have non-orthogonal RBs but only some will also have the orthogonal ones. First, we will allocate the power to the dedicated and shared part proportional to the number of RBs allocated. That is, the amount of power allocated to the dedicated and shared parts are respectively $P_{orth} = \frac{P_t}{S_{orth}}$ and $P_{non} = \frac{P_t}{S_{non}}$, where $S_{non}$ and $S_{orth}$ are the corresponding number of RBs allocated to each part. For the shared part, since it is primarily a NOMA transmission, we use the proposed HPPA method for power allocation but with the total available power in the first stage as $P_{non}$ instead of $P_t$. On the other hand, the power allocation for the orthogonal part will be applied using an optimal water-filling based approach proposed in [41]. The sum rate for all users over the non-orthogonal part (i.e., the shared RBs) is given by

$$\begin{align*}
R_a &= BS \sum_{z=1}^{Z} \sum_{q \in \Omega_{non}} \left( \log_2 \left( 1 + \gamma_{z,q}^{(H)} \right) + \log_2 \left( 1 + \gamma_{z,q}^{(L)} \right) \right).
\end{align*}$$

(3.35)

On the other hand, the sum rate for the additional orthogonal part (i.e., the dedicated RBs) is given by

$$\begin{align*}
R_b &= BS \sum_{f \in \Omega_{orth}} \sum_{w \in \Omega_{orth}} \log_2 \left( 1 + \frac{P_w^{(f)} |h_w^{(f)}|^2}{BS N_0} \right).
\end{align*}$$

(3.36)
CHAPTER 3. SUM RATE MAXIMIZATION FOR NOMA SYSTEM

Algorithm 3.1 Steps of RBs allocation and classification algorithm

1. Initialize: \( K, S \)

2. \( P_{RB} = \frac{P}{S} \): Total power per RB which is allocated equally for the classification purpose.

3. Initialize: The sets of the indices of orthogonal and non-orthogonal RBs as \( \Omega_{orth} = \emptyset \) and \( \Omega_{non} = \emptyset \), respectively.

4. Initialize: The sets of the users indices as \( O_{orth} = \emptyset \).

   (a) Set \( S_{non} = 0, S_{orth} = 0 \) (RBs Counters), \( k = 1 \)

   i. for \( s = 1 \) to \( S \)

   - find \((u,s) = \arg \max |h^{(k)}_s|^2\)
   - Calculate \( R_{orth,s} \) from (3.33) for the user who has the best channel gain \( |h_s^{(H)}|^2 \) at this RB

   - Calculate \( R_{non,s} \) from (3.34) for all of the users over the same \( s \)-th RB.

   - If \( R_{non,s} > R_{orth,s} \) do

     - \( \{\Omega_{non}\} \leftarrow \{s\} \)
     - \( S_{non} = S_{non} + 1 \)

   - else do

     - \( \{\Omega_{orth}\} \leftarrow \{s\} \)
     - \( \{O_{orth}\} \leftarrow \{k\} \)
     - \( S_{orth} = S_{orth} + 1 \)

   - end if

   ii. end for
where $\Omega_{\text{non}}$ and $\Omega_{\text{orth}}$ are the set of non-orthogonal and orthogonal RB indices, respectively, and $O_{\text{orth}}$ represents the set of users with dedicated RBs. Finally, the total sum rate for the hybrid multiple access is the sum of (3.35) and (3.36).

The proposed hybrid scheme allows more than just two users to be multiplexed over the shared RBs using closed-form power allocation solution, and this is the main advantage over the MC-NOMA system that was investigated by [29] in which a maximum of two users were allowed to share a single subcarrier.

### 3.5 Numerical Results

The considered scenario consists of a downlink of a subcarrier based NOMA system in a single cell that has $K$ users uniformly distributed within a circular coverage area with a BS located at the center. As illustrated in Figure 2.4, the total available bandwidth $W_T$ is divided into $S$ RBs; each occupying a bandwidth of $B_s$ and has $N_c$ subcarriers. The total transmission power is set to $P_t$. At the transmitter side of NOMA, the users are multiplexed in the power domain and are being separated by SIC at the receiver side. The wireless channel is modeled as a six-path frequency selective fading channel using the ITU pedestrian - B model that is depicted in Table 2.1. The channel will also be considered as a quasi-static channel where the channel conditions within one RB is fixed, but it varies from one RB to another. Unless stated otherwise, Table 3.1 depicts the simulation parameters [2-4] that are used in all of the simulation scenarios. Unless otherwise mentioned, the optimal power allocation for NOMA are numerically solved and the performance is compared to the proposed low complexity ERPA, ACPA, HPPA, equal per pair, proportional per pair, subband based NOMA (with a total of 16 subbands) schemes, and other existing schemes such as FTPA in [10], and UFPA in [22]. In addition, The OFDMA system will also be compared to show the advantage of NOMA.
### Table 3.1: Simulation parameters [2–4]

<table>
<thead>
<tr>
<th>Parameter Name</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transmitted power ($P_t$)</td>
<td>1 W (30 dBm)</td>
</tr>
<tr>
<td>Cell diameter</td>
<td>300 m</td>
</tr>
<tr>
<td>Path loss exponent ($\nu$)</td>
<td>3.76</td>
</tr>
<tr>
<td>Noise power density ($N_0$)</td>
<td>-174 dBm / Hz</td>
</tr>
<tr>
<td>Total bandwidth ($W_T$)</td>
<td>10 MHz</td>
</tr>
<tr>
<td>No. of RBs ($S$)</td>
<td>50</td>
</tr>
<tr>
<td>Bandwidth per RB ($B_s$)</td>
<td>200 kHz</td>
</tr>
<tr>
<td>No. of subcarriers per RB ($N_c$)</td>
<td>12</td>
</tr>
<tr>
<td>Shadowing standard deviation</td>
<td>8 dB</td>
</tr>
<tr>
<td>$PL_0$ at 2 GHz band</td>
<td>$15.3 + 10\nu \log_{10}(d_0)$</td>
</tr>
<tr>
<td>$\Phi_{\min}^{(1)}$</td>
<td>1 Mbps</td>
</tr>
<tr>
<td>$\Phi_{\min}^{(2)}$</td>
<td>0.5 Mbps</td>
</tr>
</tbody>
</table>

### 3.5.1 Two-user Scenario

In Figure 3.5, a sum rate comparison for the two-user scenario is made among the proposed methods against the optimal NOMA, OFDMA, and existing schemes from literature. It shows that as the maximum transmission power $P_t$ is increased, the sum rates of these schemes increase accordingly. Moreover, the sum rate for the two-user scenario is also evaluated for different BS to user separation as illustrated in Figure 3.6. From this figure, the sum rate is decreasing because the attenuation increases correspondingly with the distance.

More importantly, both of these figures show that NOMA with optimal power allocation performs significantly better than OFDMA as it achieves higher sum rate. In addition, the performance of the proposed suboptimal
methods only have small degradation from the optimal scheme but with much less complexity. The results in both figures show that the proposed methods are better than the other existing methods. In particular, ERPA provides better and closer performance to the optimal one than ACPA. This is because ERPA allocates power on a per-RB basis, while ACPA does so based on the average channel gain of all RBs.

Next, to demonstrate the fairness among the two users, Figure 3.7 shows the cumulative distribution function (CDF) against the achievable rate by the strong and the weak user. From this figure, it is clear that the trends for the two users are almost identical which shows that the proportional fairness constraint helps in maintaining a significant level of fairness between the users in spite of their channel differences.

### 3.5.2 Multiuser Scenario

In the multiuser case, the simulations involve a comparison among the numerically optimized conventional NOMA (i.e., all users transmit in the same band), the proposed HPPA, the three simplified schemes in Section
Figure 3.6: Sum rate against different cell diameters with $P_t = 30 \text{ dBm}$.

Figure 3.7: CDF against the achievable rate of each single user using two different $P_t$ levels.
3.3.2-4 (Equal-per-pair, Proportional-per-pair, and Subband based ERPA), and OFDMA. Figure 3.8 shows that the proposed HPPA is the best technique as compared to other schemes and is the closest to the optimal one. It is also better than the three simplified schemes, with equal-per-pair being the best amongst them. This also shows that putting all users in the same band has better performance than the subband based approach, and is made feasible by the low complexity closed-form solutions in our work.

Figure 3.9 compares the proposed methods against the optimal one in terms of the coverage probability. The coverage probability could be defined as the probability that a user achieves a rate that is higher than or equal to a predefined target rate. It also measure the coverage quality provided by the air interface scheme in regards to the applied power allocation scheme. It shows that the number of users who achieve the target rate using the proposed schemes is comparable to the optimal scheme. Again, this figure shows that HPPA is the best method and closest to the optimal one which verifies the effectiveness of the HP concept for NOMA system with a large number of users. The gap between the HPPA and the optimal scheme is due to fairness achieved in the power allocation. In order to maintain both rate requirement satisfaction and overall sum rate maximization, the optimal scheme is allocating power among the users in a slightly fairer manner than the HPPA scheme. Moreover, this figure also shows that the equal based scheme is the closest to the HPPA and offers better performance than the proportional and the subband based approaches. However, since all of these three schemes do not guarantee the satisfaction of the minimum rate requirements for all users they perform poorly comparing to the HPPA scheme.

Finally, Figure 3.10 illustrates the comparison in terms of the sum rate against increasing number of users. This figure shows that, as the number of users increases, the achievable sum rate of NOMA based schemes (except the subband based NOMA) also increases accordingly. This highlights the multiuser diversity gain that NOMA offers by multiplexing the users in power domain and allowing them all to share the whole available bandwidth. In addition, the performance of the subband based NOMA declines due to the fact that having only two users multiplexed per subband means that as the number of users increases, less RBs will be available to allocate for each sub-band which in turn affects the users’ achievable rate. Similar trend is also
CHAPTER 3. SUM RATE MAXIMIZATION FOR NOMA SYSTEM

Figure 3.8: Sum rate against different values of $P_t$ for $K = 16$

Figure 3.9: Coverage probability of 10 users and $P_t = 30\, dBm$ against different target rates
obtained for the case of OFDMA under the same optimization problem setup and constraints as that formulated for NOMA in (3.1) to (3.4), this is because as the number of users increases, more competition occurs among the available resources because each user attempts to acquire as many RBs as it can to maintain the minimum rate requirements.

![Figure 3.10: Sum rate against different number of users with $P_t = 20\ dBm$.](image)

### 3.5.3 Hybrid Multiple Access Against NOMA System

In Figure 3.11, the proposed hybrid scheme is compared against the HPPA-based NOMA system in terms of the sum rate for increasing number of users using different cell diameters (different user-BS separations) to evaluate the performance under different channel conditions. This figure shows that, in general, as the cell diameter increases, the user-BS separation increases accordingly which makes the channel conditions to become worse and this cause the achievable sum rate to decline for both of NOMA and the hybrid scheme. From this figure, despite the poor channel conditions, it is clear that the sum rate of the two schemes increases in proportion to the number of users. This is because of the multiuser diversity gain that is obtained
as the number of the multiplexed users increases. The advantage of the hybrid scheme over NOMA is the adaptability of the transmission scheme to the channel conditions. If the received channel gain by a certain user is significantly higher than that received by other users, the hybrid scheme will allocate this RB exclusively to this user rather than sharing it among all users using NOMA. On the other hand, if the users have comparable channel gain, it is better to apply NOMA than OFDMA. In this way, the overall sum rate is further increased while the minimum rate requirements are satisfied.

Figure 3.11: Sum rate of the hybrid scheme against HPPA based NOMA system with $P_t = 20 \, dBm$, and a cell diameter of: (a). diameter=300m. (b). diameter=1000m.

Figure 3.12 shows the sum rate comparison for increasing numbers of RBs using different cell diameters. It is intuitive that as the number of RBs increases, the overall sum rate increases due to the larger bandwidth. This figure also shows that the proposed hybrid method is again better than NOMA in terms of sum rate. In particular, the difference is larger when there are more RBs. This is due to the improved frequency diversity and that the hybrid scheme can exploit it better by optimizing the transmission methods that
maximizes the rate based on the channel quality. It should also be noted that by increasing the user-BS separation, the performance difference is larger because with poorer channel conditions, the users are better off in having more orthogonal transmissions.

![Figure 3.12: Sum rate of the hybrid scheme against HPPA based NOMA system with $P_t = 20\ dBm$, $K = 10$, and a cell diameter of: (a). diameter=300m. (b). diameter=1000m.](image)

### 3.6 Chapter Summary

NOMA has attracted a lot of attention recently due to its superior SE. In this chapter, a resource allocation scheme is developed for a downlink of a multiuser NOMA system. An optimization problem is formulated to maximize the sum rate under the total power and proportional minimum rate constraints. Due to the complexity of computing the optimal solution, we developed a low complexity suboptimal solution for two-user scenario and then extend it to the multiuser case by proposing a user-pairing approach as well as a number of power allocation techniques that facilitate dealing with
a large number of users in NOMA system. Numerical results are provided to support the effectiveness of the proposed approaches and show their close performance to the optimal one. The proposed method with pairing approach showed a comparable behaviour to the optimal one and outperformed other approaches such as FTPA. In addition, a new hybrid multiple access technique is proposed in this chapter to combine the properties of both NOMA and OFDMA techniques. Simulation results showed that the proposed hybrid method provides better performance than NOMA in terms of the overall achievable sum rate and the coverage probability. The next issue is to investigate the energy efficient approaches to enhance the performance of NOMA system.
Chapter 4

EE-SE Trade-off for NOMA System

4.1 Introduction

5G system needs a new multiple-access scheme to provide higher SE and increase the system capacity as well as being energy efficient [7, 14, 42, 104]. The conventional definition of network EE is the ratio of the total network throughput to the total power consumption [2]. However, having in mind the idea that the goal for 5G networks to be energy efficient and transmitting at very high rates while maintaining high QoS; recent articles considered EE with per-user QoS requirements [43, 44]. This implies that the transmission techniques could be exploited to serve the users in an energy efficient way while maintaining a certain level of QoS [104, 105]. This chapter will be a further step in this direction as it investigates maximizing the EE and demonstrating the EE-SE trade-off in NOMA system. We propose an energy efficient resource allocation approach in multiuser downlink NOMA system that guarantees a QoS level for each user. This chapter also presents an iterative based suboptimal power allocation technique that achieves energy efficient power allocation in downlink NOMA system.

4.2 System Model

The system model considers the downlink of a single cell NOMA system with $K$ users uniformly distributed within the coverage area. The considered system has a total bandwidth of $W_T$ that is divided into $S$ RBs. Without loss
of generality, for each RB, we assume that the users are arranged in ascending order based on their channel gains, i.e., the $K$-th user has the best channel gain. In NOMA, the weakest user will simply detect its signal in the presence of the other users as interferers. On the other hand, the stronger user will detect those users that are weaker than itself, and then cancel them from the received signal before detecting its own. In this way, the users are multiplexed in the power domain while occupying the same bandwidth and same time for transmission.

### 4.3 EE-SE Optimization

#### 4.3.1 Problem Formulation

For a total of $K$-users multiplexed over $S$ RBs, the achievable sum rate (with no user pairing being considered) is expressed as

$$ R = B_s \sum_{k=1}^{K} \sum_{s=1}^{S} \log_2 (1 + \gamma_{k,s}) $$

where the received SINR of $k \in \{1, ..., K - 1\}$ user is given by

$$ \gamma_{k,s} = \frac{P_{k,s}|h_{k,s}|^2}{\sum_{m=k+1}^{K} P_{m,s}|h_{m,s}|^2 + B_s N_0} $$

and that of the $K$-th user is

$$ \gamma_{K,s} = \frac{P_{K,s}|h_{K,s}|^2}{B_s N_0} $$

The total power consumption including both the dynamic and the static power consumed by other circuit components is given by [105, 106]

$$ P_{tc} = (1 + \tau) \sum_{k=1}^{K} \sum_{s=1}^{S} P_{k,s} + P_c + \zeta R $$

where $\tau$ represents the drain efficiency of the power amplifier, $P_{k,s}$ represents the transmission power allocated to the $k$-th user on the $s$-th RB, $P_c$ denotes the static circuit power, and $\zeta$ accounts for the dynamic power consumption.
factor, where the transmission associated circuit consumption is modeled as
a linear fraction of throughput. EE is defined as the ratio between the total
throughput and the total power consumed [105]

\[ EE = \frac{R}{P_{tc}}. \] (4.5)

Given this definition, the objective function is to maximize EE and guaran-
tee a certain level of QoS for each user. The objective function given by (4.5)
is of quasiconcave nature in terms of SE, as it is proved below.

**Lemma.** To state the quasiconcavity of the objective function, the EE-SE
expression of the \( k \)-th user could be written as

\[ EE(R_k) = \frac{R_k}{C + \zeta R_k} \] (4.6)

where \( C \) stands for \( (1 + \tau) \sum_{s=1}^{S} P_{k,s} + P_c \). Let \( R_i \) and \( R_j \) denote the achievable
rates of user \( i \) and user \( j \), respectively, where \( \{ i, j \} \in \{ 1, 2, ..., K \} \). Assume
that \( R_i \geq R_j \) as user \( i \) has better channel conditions than user \( j \). Now, ac-
cording to [107–109], a function is quasiconcave if, and only if, the following
condition is satisfied

\[ EE(\lambda R_i + (1 - \lambda) R_j) \geq \lambda EE(R_i) + (1 - \lambda) EE(R_j) \] (4.7)

where \( \lambda = \frac{R_i - R_j}{R_i} \) and \( \lambda \in [0, 1] \). By applying (4.7) to (4.6), the left hand side
will be

\[ EE(\lambda R_i + (1 - \lambda) R_j) = \frac{\lambda R_i + (1 - \lambda) R_j}{C + \zeta (\lambda R_i + (1 - \lambda) R_j)}. \] (4.8)

Since \( \lambda = \frac{R_i - R_j}{R_i} \), (4.8) can be reduced to

\[ EE(\lambda R_i + (1 - \lambda) R_j) = \frac{R_i - R_j + (1 - \lambda) R_j}{C + \zeta (R_i - R_j + (1 - \lambda) R_j)} = \frac{R_i - \lambda R_j}{C + \zeta (R_i - \lambda R_j)}. \] (4.9)
On the other hand, the right hand side is

\[ \lambda EE(R_i) + (1 - \lambda) EE(R_j) = \]

\[ \frac{\lambda R_i}{C + \zeta R_i} + \frac{(1 - \lambda) R_j}{C + \zeta R_j} \] \hspace{1cm} (4.10)

\[ = \frac{R_i - R_j}{C + \zeta R_i} + \frac{(1 - \lambda) R_j}{C + \zeta R_j} \] \hspace{1cm} (4.11)

\[ = \frac{(R_i - \lambda R_j) C + (2 - \lambda) \zeta R_i R_j - \zeta R_j^2}{(C + \zeta R_i)(C + \zeta R_j)}. \] \hspace{1cm} (4.12)

Since (4.9) and (4.12) satisfy the condition (4.7), it shows that EE-SE is concave and have a global maximum. While every quasiconcave function is a concave one, the reverse is false \([109, 110]\). Next, we will prove that the function is quasiconcave. In general, quasiconcavity is used to describe functions with many stationary points and also discontinuous functions; therefore quasiconcavity is a generalization of concavity. Hence, a function is said to be quasiconcave if it satisfies the following condition \([107-109]\)

\[ EE(\lambda R_i + (1 - \lambda) R_j) \geq \min (EE(R_i), EE(R_j)). \] \hspace{1cm} (4.13)

In other words, the following expression should be satisfied

\[ \frac{R_i - \lambda R_j}{C + \zeta (R_i - \lambda R_j)} \geq \min \left( \frac{R_i}{C + \zeta R_i}, \frac{R_j}{C + \zeta R_j} \right) \] \hspace{1cm} (4.14)

which, by demonstrating the expressions (4.14), it is clear that this condition holds for \( \frac{R_i - \lambda R_j}{C + \zeta (R_i - \lambda R_j)} \geq \frac{R_j}{C + \zeta R_j} \) and the function is quasiconcave. This completes the proof. The objective function in (4.5) is also continuously differentiable in terms of SE, as it is proved in Appendix A.

The mathematical formulation of the EE maximization problem subject to power and proportional fairness allocation among users is given by
\[
\text{maximize } \frac{\text{EE}}{P_{k,s}} \quad \text{(4.15)}
\]

Subject to
\[
\sum_{k=1}^{K} \sum_{s=1}^{S} P_{k,s} \leq P_t \quad \text{(4.16)}
\]

\[
P_{k,s} \geq 0, \forall k, s \quad \text{(4.17)}
\]

\[
R_{2i-1} : R_{2i} \geq \Phi_{2i-1} : \Phi_{2i} \quad \text{(4.18)}
\]

where \( i \in \{1, 2, \ldots, \frac{K}{2}\} \)

where (4.16) represents the total power constraint which puts an upper limit on the total transmission power of all users, and the lower limit of this power is set by (4.17). In addition, (4.18) represents the proportional fairness constraint, where \( \Phi_i \) and \( \Phi_j \) are the minimum rate requirements for the user with the best channel conditions and the worst channel conditions, respectively. The purpose of the proportional fairness constraint is to control the achievable throughput by all users and helps to maintain the proportionality between the minimum achievable rates of the users. In here, the formulated optimization problem uses the EE as the objective function to be maximized rather than the sum rate that was used in Chapter 3.

### 4.3.2 The Proposed Subgradient based Solution

By examining the problem in (4.15) to (4.18), it is clear that the objective function is nonlinear and of fractional nature. Solving such problem is very complex mathematically. To make it more tractable, we transform the objective function into a subtractive form. This is possible using the Dinkelbach approach which encompasses that for any optimization problem with a fractional objective function, there is an equivalent optimization problem with a subtractive objective function and both of them have the same optimal solution [43, 103, 111–114]. Based on the Dinkelbach approach, the objective function will be transformed into a subtraction form [43, 103, 111–114] by introducing \( \alpha \) as a weight factor. The new (transformed) version of the objective function is given by
where $A$ is the numerator that represents the total throughput, and $O$ is the
denominator that denotes the total power consumed, of the original fractional
objective function and both are functions of $P_{k,s}$, and $\alpha$ is a weighting factor.
Given the new form of the objective function, the problem is formulated as

$$
\begin{align*}
\text{maximize} & \quad F(P_{k,s}, \alpha) \\
\text{Subject to} & \quad (4.16), (4.17), (4.18).
\end{align*}
$$

This problem could be solved using iterative method with linear conver-
gence, where the solution of the power allocation in each iteration is derived
by the subgradient method based on Lagrange multipliers. The interesting
property of (4.19) is that, the $A(P_{k,s})$ part represents the system profit that is
achieved by transmitting the information and the $\alpha O(P_{k,s})$ part refers to the
the system cost (i.e., the power consumed), where $\alpha$ acts as a negative weight
on the total power consumed within the system (i.e., a balance between the
cost and the profit). In addition, the optimal value of $\alpha$ balances between the
gained profit and the cost [44,110].

Next, assume that the optimal solution of this problem is $P^*_{k,s}$ and that
$\alpha = \frac{A(P^*_{k,s})}{O(P^*_{k,s})}$. Based on the Dinkelbach approach, the solution to the problem in
(4.20)-(4.21) is obtained by determining the roots of the equality $F(P^*_{k,s}, \alpha) = 0$. Then, if the value of $F(P^*_{k,s}, \alpha)$ is positive, it means that $\alpha$ is less than its
optimal value. On the other hand, negative $\alpha$ means that it is higher than
its optimal value. Finally, if the solution equals to zero it means that the
optimal value of $\alpha$ is reached. This is due to the fact the maximum energy
is maintained if and only if $A(P^*_{k,s}) - \alpha^* O(P^*_{k,s}) = 0$. It is worth mentioning
that $\alpha$ could be used as a constant.

Using LDD optimization techniques in [103], the Lagrangian function of
the optimization problem in (4.19)-(4.21) can be expressed as
CHAPTER 4. EE-SE TRADE-OFF FOR NOMA SYSTEM

\[ F(P_{k,s}, \alpha) = B_s \sum_{k=1}^{K} \sum_{s=1}^{S} \log_2 (1 + \gamma_{k,s}) - \]

\[ \alpha \left( (1 + \tau) \sum_{k=1}^{K} \sum_{s=1}^{S} P_{k,s} + P_c + \zeta B_s \sum_{k=1}^{K} \sum_{s=1}^{S} \log_2 (1 + \gamma_{k,s}) \right) - \]

\[ \lambda \left( \sum_{k=1}^{K} \sum_{s=1}^{S} P_{k,s} - P_t \right) - \varphi \sum_{k=1}^{K} \left( \begin{array}{c}
B_s \sum_{s=1}^{S} \log_2 (1 + \gamma_{2k,s}) \\
\Phi_{2k}
\end{array} \right) + \left( \begin{array}{c}
B_s \sum_{s=1}^{S} \log_2 (1 + \gamma_{2k-1,s}) \\
\Phi_{2k-1}
\end{array} \right) \] (4.22)

where \( \lambda \) and \( \varphi \) represent the Lagrange multipliers. This expression could be simplified as

\[ F(P_{k,s}, \alpha) = (1 - \alpha \zeta) B_s \sum_{k=1}^{K} \sum_{s=1}^{S} \log_2 (1 + \gamma_{k,s}) - \]

\[ \alpha \left( (1 + \tau) \sum_{k=1}^{K} \sum_{s=1}^{S} P_{k,s} + P_c \right) - \lambda \left( \sum_{k=1}^{K} \sum_{s=1}^{S} P_{k,s} - P_t \right) - \]

\[ \varphi \sum_{k=1}^{K} \left( \begin{array}{c}
B_s \sum_{s=1}^{S} \log_2 (1 + \gamma_{2k,s}) \\
\Phi_{2k}
\end{array} \right) - \left( \begin{array}{c}
B_s \sum_{s=1}^{S} \log_2 (1 + \gamma_{2k-1,s}) \\
\Phi_{2k-1}
\end{array} \right). \] (4.23)

By taking the derivative against \( P_{k,s} \) for all users except the one with the best channel conditions (i.e., the \( K \)-th user), we have

\[ \frac{dF}{dP_{k,s}} = \frac{B_s |h_{k,s}|^2}{\left( \sum_{m=k+1}^{K} P_{m,s} |h_{k,s}|^2 + B_s N_0 \right) (1 + \gamma_{k,s}) \ln(2) - \alpha (1 + \tau) - \lambda.} \] (4.24)

For the user with the best channel conditions, the derivative is
Finally, the required water-filling to obtain the optimal solution is found to be

\[
P_{k,s} = \left[ \frac{(\Phi_k (1 - \alpha \zeta) + \varphi_k)}{\Phi_k (\alpha (1 + \tau) + \lambda) \ln (2)} - \sum_{m=k+1}^{K} P_{m,s} \left| h_{k,s} \right|^2 + B_s N_0}{\Phi_K N_0 (1 + \gamma_{K,s}) \ln (2)} \right]^{+} \tag{4.26}
\]

\[
P_{K,s} = \left[ \frac{\Phi_K (1 - \alpha \zeta) - \sum_{k=1}^{K-1} \varphi_k}{\Phi_K (\alpha (1 + \tau) + \lambda) \ln (2)} - N_0}{\left| h_{K,s} \right|^2} \right]^{+}. \tag{4.27}
\]

Since the objective function given by (4.5) was proved to be of quasiconcave nature in terms of SE, the subgradient method is applicable to find the optimal solution of the formulated problem in (4.15)-(4.18). Using the subgradient method, the Lagrangian multipliers will be updated using

\[
\lambda_k^{(i+1)} = \left[ \lambda_k^{(i)} - \Theta^{(i)} \left( P_t - \sum_{k=1}^{K} \sum_{s=1}^{S} P_{k,s} \right) \right]^{+} \tag{4.28}
\]

\[
\varphi_k^{(i+1)} = \left[ \varphi_k^{(i)} - \Upsilon^{(i)} \left( \sum_{k=1}^{K} \frac{B_s}{\Phi_{2k}} \sum_{s=1}^{S} \log_2 (1 + \gamma_{2k,s}) \right) - \frac{B_s}{\Phi_{2k-1}} \sum_{s=1}^{S} \log_2 (1 + \gamma_{2k-1,s}) \right]^{+} \tag{4.29}
\]

where \( \Theta^{(i)} \) and \( \Upsilon^{(i)} \) are small step sizes to be updated at each iteration and
chosen to be $0.1/\sqrt{i}$ [115], which will make it easier to reach a value that balances between convergence speed and optimality. The problem solution is listed in Algorithm 4.1 which will be used in the EE calculations depicted in Algorithm 4.2.

**Algorithm 4.1** Subgradient based power allocation method.

- **Initialization** $P_{k,s} = 0, P_{K,s} = 0, \lambda_k^{(i)} = 0.01, \varphi_k^{(i)} = 1,

- **while** $\lambda_k$ and $\varphi_k$ are not convergent, **do**

  - **Calculate** $P_{k,s}$ and $P_{K,s}$, from (4.26) and (4.27), respectively.
  - **update** $\lambda_k$ and $\varphi_k$ using (5.48) and (5.49), respectively.

- **end while**

- **Return** $P_{k,s}^*$ and $P_{K,s}^*$
Algorithm 4.2 Energy efficient power allocation.

- **Initialization** the maximum tolerance \( \Delta \) and the maximum number of iterations \( I_{\text{max}} \)

- Set \( i = 0 \) (the iteration index) and \( \alpha_0 = 0 \)

- while \( |F(P_{k,s}, \alpha)| \geq \Delta \) or \( i \leq I_{\text{max}} \) do
  
  - For \( k = 1 \) to \( K \)
    
    * For \( s = 1 \) to \( S \)
      
      - Apply Algorithm 4.1 to find the optimal \( P_{k,s}^* \) and \( P_{K,s}^* \)
      
      - Solve (4.19) using known \( \alpha_i \) and the obtained power \( P_{k,s}^* \)
    
    * end
  
  - end

  * Calculate \( |F(P_{k,s}, \alpha)| \) and \( \alpha_{i+1} = \frac{A(P_{k,s}^*)}{\partial(P_{k,s}^*)} \).

  * \( i = i + 1 \)

- end while

- Return \( (P_{k,s}, \alpha) \)

Note that Algorithm 4.1 is being called and executed to allocate the power for all of the \( S \) RBs used by all of the \( K \) users as applied within each of the \( S \) and \( K \) loops in Algorithm 4.2.

### 4.4 Numerical Results

This section presents the simulation results where the proposed subgradient power allocation approach is simulated and compared to the HPPA scheme that was proposed in Chapter 3, and also the optimal NOMA that is obtained numerically. In addition a comparison between NOMA and OFDMA is also presented to verify the benefits of the energy efficient design of NOMA. This optimal OFDMA is obtained using an optimization problem set up that is similar to that of the optimal NOMA scheme and subject to the same power and
minimum rate constraints (i.e., as in (4.15) to (4.18)) with subcarrier allocation as in [41]. The system is simulated along with a six-path frequency selective fading channel using the ITU pedestrian - B model, where the average power of the multipath are taken from Table 2.1. The channel state information is assumed to be perfectly known at the BS. Unless otherwise mentioned, the simulation parameters are listed in Table 4.1, and these parameters are chosen based on the third generation partnership project (3GPP) LTE standard and related literature.

<table>
<thead>
<tr>
<th>Parameter Name</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total bandwidth ($W_T$)</td>
<td>10 MHz</td>
</tr>
<tr>
<td>No. of resource blocks ($S$)</td>
<td>50</td>
</tr>
<tr>
<td>No. of subcarriers per RB ($N_c$)</td>
<td>12</td>
</tr>
<tr>
<td>Shadowing standard deviation</td>
<td>8 dB</td>
</tr>
<tr>
<td>Noise power density ($N_0$)</td>
<td>-174 dBm / Hz</td>
</tr>
<tr>
<td>Path Loss Exponent ($\nu$)</td>
<td>3.76</td>
</tr>
<tr>
<td>$\Phi_j$</td>
<td>50 kbps</td>
</tr>
<tr>
<td>$\Phi_i$</td>
<td>100 kbps</td>
</tr>
<tr>
<td>Tolerance ($\Delta$)</td>
<td>$10^{-3}$</td>
</tr>
<tr>
<td>Maximum number of iterations ($I_{max}$)</td>
<td>25</td>
</tr>
<tr>
<td>Power amplifier drain efficiency ($\tau$)</td>
<td>0.38</td>
</tr>
<tr>
<td>Cell diameter</td>
<td>300 m</td>
</tr>
<tr>
<td>$PL_0$ at 2 GHz band</td>
<td>$15.3 + 10\nu \log_{10}(d_0)$</td>
</tr>
<tr>
<td>Dynamic power consumption factor($\zeta$)</td>
<td>$1^{\mu/\lambda}$</td>
</tr>
</tbody>
</table>

First of all, we would like to show the fast convergence of the proposed iterative algorithm, which is presented in Figure 4.1 for different number of
users. The algorithm stops when the difference between two successive values of $F(P_k,s,\alpha)$ is less than or equal to the tolerance $\Delta$. For the two cases considered in this figure, it can be seen that the proposed subgradient-based algorithm converges quickly, which demonstrate the feasibility of this iterative solution. The convergence speed of the subgradient method depends highly on the choice of the step size. In Algorithm 4.1 it was chosen to be $0.1/\sqrt{i}$ [115] which is of a diminishing nature, this makes it becomes a bit slower as $i$ value (i.e., the loop iterations) grows further but on the bright side the subgradient search becomes more accurate in finding the optimal set as compared to a fixed step size.

![Figure 4.1](image)

**Figure 4.1:** Convergence performance of the proposed subgradient-based NOMA system where $P_t = 30\text{dBm}$.

In Figure 4.2, it is clear that increasing the transmission power gradually can enhance the achievable sum rate (and SE accordingly) because with higher transmission power, the users will have sufficient coverage that enhance their ability to battle against the bad channel conditions. This figure also shows the close trends of the proposed suboptimal schemes as compared to the optimal one with the advantage of their simplicity over the optimal approach. In
particular, the subgradient based solution provides power allocation behavior that is very close to the optimal power allocation approach. On the other hand, the proposed HPPA scheme falls slightly far from the optimal approach but is still the simplest method as compared to both the subgradient approach and the optimal one, where the complexity of the proposed HPPA method is estimated to be about $(K - 1) \times S$ times of (3.29) and (3.30) computations, while that of the subgradient one is higher and is about $O\left(\frac{SK}{\sqrt{K}}\right)$ times of the water-filling computations in (4.26) and (4.27). In addition, Figure 4.2 shows that all NOMA schemes are superior to OFDMA system.

![Achievable sum rate against increasing transmission power](image)

**Figure 4.2:** Achievable sum rate against increasing transmission power for: (a) $K = 30$, (b) $K = 10$.

Figure 4.3 illustrates the EE performance using different levels of transmission power for the proposed NOMA schemes, the optimal numerical solution, and OFDMA. This figure depicts that all the simulated NOMA schemes generally outperform OFDMA. In addition, it shows that the proposed subgradient based NOMA offers the closest behaviour to the optimal one with the advantage of low computational complexity. On the other hand, HPPA based
NOMA behaviour has a gap below the optimal and the subgradient based NOMA approaches but with the least computational complexity as compared to them. This figure also shows that increasing the transmission power is not an effective solution to improve the EE especially after a certain power level. This is because although increasing the transmission power will increase the sum rate, it will also increase the power consumption together with the increased transmission power as in (4.4); thus EE is reduced.

Figure 4.3: EE comparison against increasing transmission power for $K = 15$ with $P_c = 10$W.

In Figure 4.4, the EE against SE of all the considered schemes are plotted for 15 users at a transmission power of 40 dBm as a function of the minimum rate requirements using two different levels of circuit power. This figure shows the close performance of the proposed suboptimal schemes as compared to the optimal one. The EE-SE behavior in Figure 4.4 shows that in the lower SE range (i.e., starting from about $SE = 2.4 bps/Hz$ up to $SE = 14 bps/Hz$), it is possible to jointly increase the EE and SE at the same
time, and in this regime there is no trade-off. However, after a certain operating point (i.e., \( SE = 16 bps/Hz \)), EE starts to degrade as SE increases which depicts their trade-off. This can be explained by looking back at Figure 4.2 and Figure 4.3.

From Figure 4.2, it is obvious that increasing the SE is associated with increasing the transmission power, and Figure 4.3 shows that increasing the transmission power resulted in decreasing the EE. Therefore, this proves that there is a fundamental trade-off between the EE and the SE. This figure also proves that NOMA is more spectral efficient than OFDMA as it allows the users to occupy the whole bandwidth. In addition, this figure shows that EE decreases as the circuit power becomes higher which confirms the previous findings.

![Graph](image)

Figure 4.4: EE versus SE comparison as a function of various minimum rate requirements (i.e., \( \Phi_i = [1.5, 1.82, 0.2, 4.2, 83.0] Mbps \) and \( \Phi_j = 0.5 \Phi_i \)) with \( K = 15 \) and \( P_t = 40 \text{dBm} \), for: (a) \( P_c = 5 \text{W} \). (b) \( P_c = 10 \text{W} \).

Figure 4.5 and Figure 4.6 evaluate the EE for the proposed NOMA schemes, the optimal NOMA, and OFDMA against the increasing number
of users and each figure uses different dynamic power consumption factor \( \zeta \). From these figures, NOMA outperforms OFDMA for all of the compared cases. It is clear that EE monotonically increases with the number of users due to the multiuser-diversity gain that NOMA provides. The NOMA multiuser-diversity gain is obtained as a result of the flexibility of the power allocation schemes that efficiently exploit the differences in the users’ channel gains. However, this is not the case with OFDMA, where the EE improves first and then degrades when the number of users is close to the total number of RBs. The initial improvement is again due to multiuser diversity. However the subsequent degradation is because of the per user minimum rate constraint, which will ensure each user be allocated some RBs to meet the rate requirement. Hence some RBs are not allocated to the best possible users and thus the performance is degraded. In addition, both of the proposed NOMA schemes are very close to the optimal one. Despite being slightly degraded from the subgradient based NOMA, the HPPA scheme has the advantage of being the simplest scheme among them.

These figures also show that the circuit power has a significant effect on the performance of NOMA, where the EE noticeably declines with the increase of the circuit power. On the other hand, since NOMA achieves a higher sum rate than OFDMA, when the dynamic power consumption factor reduces, the power consumption for NOMA reduces more than that in OFDMA, resulting in a larger performance gap in EE between them. In Figure 4.5 and Figure 4.6, the trends nature is due to the existence of the sum rate \( \tilde{R} \) in (4.4) which is the denominator of (4.5). Where any increase in the achievable sum rate could have detrimental effect on the overall EE. Especially with the OFDMA behaviour where it first increases then declines as the number of users increases where more competition arise among the users over the available power and RB resources as each user tries to maintain the minimum rate requirements, where, for a fair comparison, OFDMA is applied in here under the same optimization problem set up as that of NOMA (i.e., the problem formulated in (4.15)-(4.18)).
Figure 4.5: EE comparison against increasing number of users with $P_t = 30$dBm and: (a) $P_c = 5W$. (b) $P_c = 10W$. 
CHAPTER 4. EE-SE TRADE-OFF FOR NOMA SYSTEM

Next, the performances of the compared schemes are evaluated in terms of the EE-SE trade-off using various numbers of RBs as illustrated in Figure 4.7 and for different levels of circuit power. From this figure, it is clear that increasing the number of RBs can improve the SE for all of the compared schemes. While the gain is clearly due to multiuser diversity in OFDMA, the same reason also applies to NOMA. Although NOMA uses all the RBs for each user and do not perform RB allocation as in OFDMA, the power allocation will exploit the channel variations of all users in the frequency domain and maximizes the EE-SE trade-off of the system. Therefore both SE and EE improves as the number of RB, which is effectively the bandwidth, increases.
Furthermore, the EE-SE trade-off performances is evaluated for different cell sizes as illustrated in Figure 4.8 and using different values of circuit power. This figure shows that the EE-SE of all schemes decrease monotonically as the distance increases. This is because of the larger BS to user separation which results in larger path loss and bad propagation conditions; hence, more power will be required to compensate for the channel conditions and to provide sufficient coverage to all users. As a result of this, both the EE and the SE decline. It can also be observed that the effect of increasing distance causes a more rapid decline in EE-SE trade-off for OFDMA than in NOMA. This again shows the superiority of NOMA over OFDMA.
Figure 4.8: EE-SE trade-off behavior as a function of different cell diameters (i.e., varying the diameter as \([200, 300, 400, ...., 1000]\) m) with \(P_t = 30\) dBm, \(K = 15\), for: (a) \(P_c = 5\) W. (b) \(P_c = 10\) W.

4.5 Chapter Summary

Establishing an energy efficient design for future networks is necessary in order to reduce the high energy consumption due to the predicted growth in traffic demand. This chapter investigated the EE and the relation between the EE and the SE of a downlink-multiuser NOMA system. We proposed a power allocation scheme with the goal of maximizing the EE under the overall power and proportional data rates constraints. Due to the complexity of the optimization objective function, the first step was to convert the problem into a simpler optimization problem with a subtractive form. After that, the converted problem was solved with an iterative approach to allocate the transmission power for each user to maintain a compromise between EE and SE. Another suboptimal scheme, the HPPA based NOMA, which we proposed in
Chapter 3 was involved in the comparison. Although the optimization programming is numerically stable, its computational complexity depends on the number of optimizing variables ($K$ and $S$), which can be large if the number of RBs and/or the number of UEs is large. In addition, despite its accuracy, the total complexity of the subgradient based approach also depends on the number of iterations in the outer-loop. In the worst case, it requires about $O(\frac{SK}{\Delta^2})$ times of water-filling computations in (4.26) and (4.27) to converge to the solution. The main advantage that the proposed HPPA scheme offers is the high simplicity as compared to the subgradient and the optimal ones. In particular, to allocate the power for $K$ users over $S$ RBs, the expressions in (3.29) and (3.30) need to be applied $(K - 1) \times S$ times only. Simulation results showed that the proposed NOMA schemes outperformed the optimal OFDMA. The results also supported the effectiveness of the proposed approaches and showed the close performance of these schemes to the optimal one.
Chapter 5

MIMO-NOMA Systems

MIMO is a technology based on using multiple antennas at both the transmitters and the receivers. As compared to single-antenna systems, MIMO transmission techniques exploit the spatial dimensions that are provided by multiple antennas at both ends of a wireless link to enhance the channel capacity by spatial multiplexing or spatial diversity without the need to add more bandwidth or power \([116,117]\). In other words, by assuming there are \(M\) transmit antennas, the concept of spatial multiplexing is to split each data stream into \(M\) separate substreams. Then those substreams are mapped into symbol streams to be transmitted simultaneously. Hence, as compared to a single antenna system, the transmission rate will be \(M\) times higher in this case. At the receiver side, after receiving all the substreams one of them will be regarded as the desired signal whilst the remaining substreams are regarded as spatial interference and removed using nullifying or linear filtering, this is also the case for each received substream \([116,117]\). Spatial diversity, on the other hand, implies using several transmitting and receiving antennas to transmit and receive several copies of the same data streams. In such case, each data stream copy will suffer independent fading path which will help to avoid signals detection in deep fading, and that how spatial diversity improves the system performance \([117]\). Since MIMO technology exploits the spatial domain and NOMA exploits the power domain, the combination of the two technologies can further improve the overall system SE \([13,25,69]\).

This chapter investigates the use of the spatial dimension in MIMO with NOMA, and develop low complexity suboptimal power allocation schemes for multiple users. We first consider multiuser MISO-NOMA and optimize
the precoder for sum rate maximization.

## 5.1 System Model

A BS with $M$ antennas serving $K$ users each equipped with $N$ antennas is used to model a downlink mult-user MIMO-NOMA system. Since NOMA allows the users to be multiplexed in the power domain, the transmitted signal from the BS to the $k$-th user will share the same time-frequency resource with the signals intended to other users, and is given by

$$
x^{(k)} = V^{(k)} \tilde{s}
$$

where $V^{(k)} \in \mathbb{C}^{M \times D}$ stands for the transmission precoding matrix used by the BS, $\tilde{s} \in \mathbb{C}^{D \times 1}$ is the information bearing vector that contains $D$ data streams.

Following the NOMA concept, each of these streams will consist of signals from all $K$-users; hence $\tilde{s}$ is given by

$$
\tilde{s} = \begin{bmatrix}
\hat{s}_1 \\
\hat{s}_2 \\
\vdots \\
\hat{s}_l \\
\vdots \\
\hat{s}_D
\end{bmatrix}
= \begin{bmatrix}
P^{(1)}_1 s^{(1)}_1 + P^{(2)}_1 s^{(2)}_1 + \cdots + P^{(K)}_1 s^{(K)}_1 \\
P^{(1)}_2 s^{(1)}_2 + P^{(2)}_2 s^{(2)}_2 + \cdots + P^{(K)}_2 s^{(K)}_2 \\
\vdots \\
P^{(1)}_l s^{(1)}_l + P^{(2)}_l s^{(2)}_l + \cdots + P^{(K)}_l s^{(K)}_l \\
\vdots \\
P^{(1)}_D s^{(1)}_D + P^{(2)}_D s^{(2)}_D + \cdots + P^{(K)}_D s^{(K)}_D
\end{bmatrix}
$$

where $P^{(k)}_l$ refers to the power allocated to the $k$-th user over the $l$-th data stream. It is worth mentioning that the system model here represents one symbol timing of the data stream. In addition, $s_l^{(k)}$ represents the data signal intended to the $k$-th user within the $l$-th data stream. On the other hand, the received signal by the $k$-th user can be expressed as
where \( y^{(k)} = H^{(k)}V^{(k)}\tilde{s}_k + \sum_{j=1,j\neq k}^{K} H^{(k)}V^{(j)}\tilde{s}_j + n^{(k)} \) (5.3)

where \( H^{(k)} \in \mathbb{C}^{N \times M} \) denotes the small scale fading effect between the \( k \)-th user and the BS, and \( n^{(k)} \in \mathbb{C}^{N \times 1} \) denotes the normalized AWGN vector observed by user \( k \). Note that each entry of the channel matrix \( H^{(k)} \) is generated using independently and identically distributed (i.i.d.) random variables according to \( CN(0,1) \). It is assumed that each channel is quasi-stationary and frequency flat fading. Each user decodes the desired signals by multiplying the received signal with a receiver beamforming matrix \( U^{(k)} \in \mathbb{C}^{N \times D} \), and thus, the post-processed signal can be modeled as

\[
\tilde{y}^{(k)} = U^{(k)*}y^{(k)} = U^{(k)*} \left( H^{(k)}V^{(k)}\tilde{s}_k + \sum_{j=1,j\neq k}^{K} H^{(k)}V^{(j)}\tilde{s}_j \right) + \tilde{n}^{(k)} \quad \text{(5.4)}
\]

where \( \tilde{n}^{(k)} = U^{(k)*}n^{(k)} \) stands for the effective noise component at the output of the beamformer.

Without loss of generality, we consider the users are ranked in a descending order based on their channels gains. The achievable sum rate by all \( K \)-users MIMO-NOMA system is given by [70]

\[
R_s = B_s \sum_{l=1}^{D} \log_2 \left( 1 + \frac{P_l^{(K)} G_l^{(K)}}{B_s N_0} \right) + B_s \sum_{k=1}^{K-1} \sum_{l=1}^{D} \log_2 \left( 1 + \frac{P_l^{(k)} G_l^{(k)}}{B_s N_0 + \sum_{j=k+1}^{K} P_l^{(j)} G_l^{(k)}} \right) \quad \text{(5.5)}
\]

where \( P_{L}^{(k)} \) accounts for the path loss that varies with the distance between the \( k \)-th user and the serving BS \( d_k \), and \( G_l^{(k)} \) is the effective channels of the \( k \)-th user, and it is given by

\[
G_l^{(k)} = \xi^{(k)} A \quad \text{(5.6)}
\]

where \( \xi \) stands for the log-normal shadowing, and the matrix \( A \) is calculated
as $|U^{(k)}H^{(k)}V^{(k)}|^2$, where $U^{(k)}$ and $V^{(k)}$ are the MIMO-NOMA receiver and the transmitter precoding matrices that are obtained based on the considered precoding techniques.

## 5.2 Linear Precoder for MISO-NOMA System

The performance of MISO-NOMA is first investigated as a special case with different precoding schemes. Precoding is the pre-processing that is applied at the transmitter and it requires the CSI to be known at the transmitter to maximize the throughput at the receiver. In addition, precoding is used to coordinate different data streams transmitted from different antennas using independent weights. In general, precoding schemes can be categorized into linear precoding and nonlinear precoding. Schemes from both types will be considered in this chapter. Precoding can benefit from the CSI to separate the data intended for a single user or multiple users; hence, it helps in simplifying multiuser detection which can be a burden for NOMA with high number of users. In this chapter, CSI is assumed to be known for the users and the BS while in practice channel estimation is required.

### 5.2.1 Linear Precoding

Linear precoding schemes are those that precode the data linearly and usually have simple structure than the non-linear techniques while offering good performance. In this chapter, two types of linear precoders are used with MISO-NOMA system as a comparison benchmark, namely, the zero forcing (ZF) and the minimum mean square error (MMSE). Moreover, unlike conventional MISO where precoding is applied to separate each user, here it is used to separate each NOMA stream, which contains more than one user. The ZF precoder is a channel inversion based precoding technique that converts the multiuser channel into independent single-streams [118, 119]. ZF mitigates the interference that degrades the transmitted signal by setting all the interference terms into zeros; hence, ZF is shown to achieve a large portion of DPC capacity [118]. According to ZF, the data to be transmitted is precoded with the pseudo-inverse of the channel matrix to achieve zero interference between the users. The precoding matrix for the ZF precoder can be expressed
where $H \in \mathbb{C}^{D \times M}$ denotes the small scale channel fading effect of $D$-data stream MISO-NOMA system, each stream has $K$-users multiplexed using NOMA and they detect their intended signals using SIC at the receiver. ZF and MMSE will be used to precode these data streams to null the effect of these streams from each other. In other words, each MISO-NOMA user has a channel vector as $h^{(k)} \in \mathbb{C}^{1 \times M}$ and all these $K$-users are multiplexed in one data stream, then each stream channel is effectively a combination of $K$-users channel (i.e., each $l$-stream has a channel of $H \in \mathbb{C}^{K \times M}$) combined altogether to form $\hat{H}$ by multiplying the users channel vectors as $\hat{H} = h^{(1)} \ast h^{(2)} \ast \ldots \ast h^{(K)}$. After that, the data stream channel is used to generate the precoding matrix $\hat{V} \in \mathbb{C}^{M \times D}$ by using ZF and MMSE, the normalized ZF-beamforming matrix is then given as $\hat{V}_{ZF} = \delta_{ZF} V_{ZF}$, where $\delta_{ZF} = \sqrt{P_t / tr(\hat{V}_{ZF} \hat{V}_{ZF}^*)}$ is the ZF normalizing factor with $P_t$ stands for the total transmission power. On the other hand, the MMSE precoder is also a channel inversion based precoder that is obtained by regularizing the pseudo inverse of the ZF precoder to improve its performance [120]. The precoding matrix for the MMSE precoder can be expressed as

$$V_{MMSE} = \left( \hat{H}^* \hat{H} + \frac{D}{P_t} I \right)^{-1} \hat{H}^*.$$ 

In addition, the normalized MMSE-beamforming matrix is then given as $\hat{V}_{MMSE} = \delta_{MMSE} V_{MMSE}$, where $\delta_{MMSE} = \sqrt{P_t / tr(\hat{V}_{MMSE} \hat{V}_{MMSE}^*)}$ is the MMSE normalizing factor. It is worth mentioning that the ZF precoder is effective only when the total number of receiving antennas is lower than or equal to the number of transmitting antennas [67]. In addition, at high SNR regions, MMSE behave in similar way to ZF. Both $\hat{V}_{MMSE}$ and $\hat{V}_{ZF}$ consists of $D$-precoding vectors ($v \in \mathbb{C}^{M \times 1}$) each single vector belongs to one data stream, hence, this vector is used by all users who are multiplexed within that stream. By re-writing the expression of (5.6) for MISO-NOMA system, the channel experienced by the $k$-th user multiplexed in the $l$-th stream is given as

$$G_{l}^{(k)} = \xi_{l}^{(k)} |h_{l}^{(k)} v_{l}^{(k)}|^2.$$ 

(5.9)
where \( h_{i}^{(k)} \) denotes the channel vector received by the \( k \)-th user over the \( l \)-th data stream, and \( v_{l}^{(k)} \) is the precoding vector that is obtained using ZF or MMSE and it used by this user as well as all users that are multiplexed over the \( l \)-th stream.

This term will be used as a replacement to that in (5.6) and all the equations that follows it including the proposed power allocation schemes as well as the achievable sum rate calculations for MISO-NOMA system.

### 5.2.2 Precoder Optimization

Apart from the standard linear precoders, we also investigate the precoder optimization to maximize the sum rate. The normalized precoder sum rate maximization problem is formulated as

\[
\begin{align*}
\text{maximize} & \quad R_s \\
\text{Subject to} & \quad \sum_{l=1}^{D} \text{tr} \left( v_{l}^{\text{opt}} v_{l}^{\text{opt}*} \right) \leq 1 \\
& \quad \text{rank} \left( v_{l}^{\text{opt}} v_{l}^{\text{opt}*} \right) = 1
\end{align*}
\]  

(5.10) \quad (5.11) \quad (5.12)

where constraint (5.11) is necessary to guarantee a normalized precoding matrix that will not amplify the transmission power allocated to each user. Constraint (5.12) refers to the dimension of the obtained optimal precoding vector for each user \( (v_{l}^{\text{opt}} \in \mathbb{C}_{M \times 1}) \). The solution of this problem is obtained numerically as it is impossible mathematically to reach a closed-form solution. Nonetheless, this numerically optimized precoder \( v_{l}^{\text{opt}} \) allows the comparison with ZF and MMSE precoders and indicates the potential improvement that can be achieved. In here, the channel experienced by each user is obtained through (5.9) by using the optimized precoding vector \( v_{l}^{\text{opt}} \). This optimization problem is different from the previously formulated ones in that it involves precoder constraints and multiple antennas scenario.
5.3 IA based MIMO-NOMA System

The second part of this chapter will investigate the performance of IA based MIMO-NOMA with nonlinear precoding. We propose the use of IA to suppress the interference from all other streams on the intended one. Then we consider applying SVD and DPC for precoding the MIMO-NOMA channels. After that, SIC will be used for each stream to decode the users that are multiplexed using NOMA. In other words, we are applying IA and using both SVD and DPC to deal with the MIMO channels, and then the standard SIC is used for NOMA decoding.

5.3.1 Designing the IA precoders

The basic idea of IA is to align the unwanted signals into an interference-subspace and suppress its projection to the desired signal subspace where the wanted signals are directed [121, 122]. In other words, it is an interference management process that encompasses aligning multiple interference signals in a signal subspace with dimension smaller than the number of interferers [123, 124]. This is possible by determining $V^{(k)} = \left[ v_1^{(k)}, v_2^{(k)}, \ldots, v_l^{(k)}, \ldots, v_D^{(k)} \right] \in \mathbb{C}^{M \times D}$ and $U^{(k)} = \left[ u_1^{(k)}, u_2^{(k)}, \ldots, u_l^{(k)}, \ldots, u_D^{(k)} \right] \in \mathbb{C}^{N \times D}$ which are the precoding matrix and the post processing matrix for the $k$-th user, respectively. These matrices shall meet the following conditions

$$\text{rank} \left( U^{*(k)} H^{(k)} V^{(k)} \right) = D$$  \hspace{1cm} (5.13)

$$U^{*(k)} H^{(k)} V^{(j)} = 0 \ \forall j \neq k$$  \hspace{1cm} (5.14)

where $D \leq \min(M, N)$ which denotes the number of data streams.

In the considered MIMO-NOMA system, two interference types exist, namely, the intra-stream interference and the inter-stream interference. It is necessary to cancel these interference effects to decode the useful signal efficiently. While the intra-stream interference is assumed to be canceled by SIC, the goal here is to design an interference mitigation technique that removes the inter-stream interference. The proposed IA scheme align the users of each data stream to a certain subspace and then the users in other
CHAPTER 5. MIMO-NOMA SYSTEM

streams are aligned to another unwanted subspace. In other words, the targeted user group uses the same alignment precoder to align other groups as interference. This is achieved by designing the receiver beamforming matrices to be orthogonal to the subspace where the inter-stream interference is aligned. Hence at each receiver, the unwanted streams space is constituted in \( \text{span} \left( \left[ U^*(k)H^{(k)}V^{(j)} \right]_{j=1,j\neq k}^K \right) \) while \( \text{span} \left( U^*(k)H^{(k)}V^{(k)} \right) \) would contain the space of the useful data stream. After identifying the effective inter-stream interfering channel, the transmit beamforming matrices are designed to achieve this alignment. Then, the transmit beamforming vectors for the \( k \)-th user in the \( l \)-data stream \( V^{(k)} \) is designed such that the symbols that are sent within the \( l \)-th data stream do not cause interference to other streams.

\[
V^{(k)} \subset \text{null} \left( \begin{bmatrix} U^*_c \times H^*_c \\ \text{effective inter-stream interference channel} \end{bmatrix} \right)_{c=1,...,K}^{j=1,...,D(j\neq l)}.
\]

(5.15)

It is worth mentioning that, with (5.15), the number of data streams \( D \) shall always be less than or equal to the total number of transmitting antennas \( M \) (i.e., \( D \leq M \)).

Accordingly, the effective channel is given by

\[
\tilde{H} = U^*(k)H^{(k)}V^{(k)}.
\]

(5.16)

After identifying the effective channel, it will be used to generate the transmit and the receive precoding matrices, where in this chapter, both SVD and DPC are applied for this purpose.

5.3.2 Singular Value Decomposition (SVD)

In here, SVD will be used to design the MIMO-NOMA channel precoders. Using the SVD, the MIMO channel matrix could be decomposed into \( D \) independent single-input single-output (SISO) subchannels over which different data streams can be transmitted separately and that is how spatial multiplexing is achieved. In this case, the achievable data rate will be scaled up by a factor that equals the rank of the decomposed matrix. The MIMO channel matrix \( H \) can be mathematically decomposed as [125]
\[ H = U \Sigma V^* \] (5.17)

where \( U \in \mathbb{C}^{N \times N} \) is a unitary matrix that contains the left singular vectors and it will act as the receiver post-processing matrix, \( \Sigma \in \mathbb{C}^{N \times M} \) is a diagonal matrix with its entries represent the singular values \((\nu_i)\) of the channel \( H \) that are nonzero, and each singular value satisfies the property \( \nu_i = \sqrt{\lambda_i} \), where \( \lambda_i \) is the \( i \)-th eigenvalue of \( HH^* \). \( V \in \mathbb{C}^{M \times M} \) is a unitary matrix that contains the right singular vectors and it will act as the transmitter precoding matrix.

### 5.3.3 Designing MIMO Precoders using SVD

In case of applying SVD precoding, both of \( \tilde{U}^{(k)} \) and \( \tilde{V}^{(k)} \) matrices will be used to obtain the matrix \( A \) in (5.6). In transmit precoding, the modulated symbol stream is precoded by applying linear transformation of the input signal through transmit precoding matrix as in (5.1). At the receiver side, the linear transformation is applied to the output signal through receiver shaping matrix \( U^* \) and the received signal is obtained as mentioned in (5.4).

Based on SVD, the effective channel can be decomposed as

\[ \tilde{H} = \tilde{U}^{(k)} \tilde{\Sigma}^{(k)} \tilde{V}^{(k)} \] (5.18)

where \( \tilde{U}^{(k)} \in \mathbb{C}^{D \times D} \) act as a unitary matrix that contains the left singular vectors, \( \tilde{\Sigma}^{(k)} \in \mathbb{C}^{D \times D} \) is a diagonal matrix with its entries represent the singular values of the effective channel, \( \tilde{V}^{(k)} \in \mathbb{C}^{D \times D} \) is a unitary matrix that contains the right singular vectors. Both \( \tilde{U}^{(k)} \) and \( \tilde{V}^{(k)} \) represent the transmitting and the receiving precoding matrices that will be used for MIMO-NOMA system.

### 5.3.4 Designing MIMO Precoders using DPC

DPC is used here to design the precoders of the MIMO-NOMA channels. One of the important non-linear precoding schemes for BCs is the DPC, which was first proposed by [126]. DPC act as an interference pre-cancellation scheme that removes the perfectly known interference to the transmitter to be pre-cancelled at the transmitter without increasing the transmit power [16, 127]. DPC can also be used to determine the achievable capacity region of MIMO-BC channels [16]. Another advantage of DPC is that if

\[ \tilde{H} = \tilde{U}^{(k)} \tilde{\Sigma}^{(k)} \tilde{V}^{(k)} \] (5.18)
CHAPTER 5. MIMO-NOMA SYSTEM

5.4 Problem Formulation and Proposed Solutions

Since MISO system can be regarded as a special case of MIMO system, the optimization problem will be formulated based on MIMO-NOMA and the proposed approaches can be applied to both MIMO and MISO.

5.4.1 Low Complexity Closed-form Solution

For a two user MIMO-NOMA case, the achievable rate by the user with poor channel conditions (weakest user) is given by

\[ R^{(L)} = B_s \sum_{l=1}^{D} \log_2 \left( 1 + \gamma_l^{(L)} \right) \]  \hspace{1cm} (5.19)

while that achieved by the user with the better channel conditions (strong user) is given by

\[ R^{(H)} = B_s \sum_{l=1}^{D} \log_2 \left( 1 + \gamma_l^{(H)} \right) \]  \hspace{1cm} (5.20)

where

\[ \gamma_l^{(H)} = \frac{P_l^{(H)} G_l^{(H)}}{B_s N_0} \]  \hspace{1cm} (5.21)
represent the received SINR of the stronger and weaker users in the $l$-th data stream respectively. Note that $G_i^{(H)}$ (and similarly $G_i^{(L)}$) is determined using (5.6). In (5.6), the matrix $A$ is calculated as $|\tilde{\mathbf{U}}^{(H)}\mathbf{H}^{(H)}\tilde{\mathbf{V}}^{(H)}|^2$ in case of SVD precoding, where $\tilde{\mathbf{U}}^{(H)}$ and $\tilde{\mathbf{V}}^{(H)}$ are the MIMO-NOMA receiver and the transmitter precoding matrices. In case of DPC precoding, on the other hand, the matrix $A$ in (5.6) is calculated as $|\tilde{\mathbf{L}}^{(H)}\mathbf{H}^{(H)}\tilde{\mathbf{Q}}^{(H)}|^2$.

The formulated sum rate maximization problem is expressed as

$$\text{maximize} \quad R_s$$

$$\text{Subject to} \quad \sum_{l=1}^{D} \left( P_i^{(L)} + P_i^{(H)} \right) \leq P_t$$

$$P_i^{(L)}, P_i^{(H)} \geq 0, \ \forall l, k$$

$$R_{2i-1}^{(H)} : R_{2i}^{(L)} \geq \Phi_{2i-1}^{(H)} : \Phi_{2i}^{(L)}$$

$$\text{where } i \in \left\{ 1, 2, ..., \frac{K}{2} \right\}$$

where constraints (5.24) and (5.25) are important to guarantee a positive allocated power and limited by the maximum transmission power $P_t$. Additionally, the proportional fairness constraint (5.26) is used to control the rate achieved by all users, where $\Phi_{2i-1}^{(H)}$ and $\Phi_{2i}^{(L)}$ are the minimum rate requirements for the user with the best channel conditions and the worst channel conditions, respectively. In this chapter, each user is assigned its minimum rate requirement based on the large scale fading factor (the distance based path loss and the log-normal shadowing factor) experienced by that user in addition to the small scale fading effects. This is to make the proportionality constraint more effective on a long term basis rather than short term, where the large scale fading factor constitutes the path loss and the shadowing who are more dominant than the small scale fading part and vary slowly. Unlike the formulated problems in Chapters 3 and 4 where a single antenna scenario is used, this problem is formulated for a multiple antennas scenario. Having multiple antennas represents the main challenge faced to solve this problem as it led to a channel with a matrix nature which makes it difficult to deal with.
especially to arrange the users in a certain order for SIC application. Thus, the minimum rate requirements and the users ordering were based on the large scale part of the channel (i.e., the path loss and the shadowing).

Using LDD approach \([103]\), the solution of (5.23)-(5.27) can be given as

\[
F = - \mu \left( \frac{B_s \sum_{i=1}^{D} \log_2 \left( 1 + \gamma_i^{(H)} \right)}{\phi_{2i-1}^{(H)}} - \frac{B_s \sum_{i=1}^{D} \log_2 \left( 1 + \gamma_i^{(L)} \right)}{\phi_{2i}^{(L)}} \right) + B_s \sum_{i=1}^{D} \log_2 \left( 1 + \gamma_i^{(H)} \right) + B_s \sum_{i=1}^{D} \log_2 \left( 1 + \gamma_i^{(L)} \right) - \psi \left( P_i^{(L)} + P_i^{(H)} - P_t \right)
\]

(5.28)

where \(\mu\) and \(\psi\) represent the Lagrangian multipliers. Differentiating with respect to \(P_i^{(H)}, P_i^{(L)}, \mu\) and \(\psi\) respectively, we obtain

\[
\frac{dF}{dP_i^{(H)}} = - \frac{B_s P_i^{(L)} \left( G_i^{(L)} \right)^2}{\left( B_s N_0 + P_i^{(H)} G_i^{(L)} \right)^2 \left( 1 + \gamma_i^{(L)} \right)} + \frac{G_i^{(H)}}{N_0 \left( 1 + \gamma_i^{(H)} \right)} - \psi - \\
\mu \left( \frac{G_i^{(H)}}{N_0 \phi_{2i-1}^{(H)}} \left( 1 + \gamma_i^{(H)} \right) + \frac{B_s P_i^{(L)} \left( G_i^{(L)} \right)^2}{\phi_{2i}^{(L)} \left( B_s N_0 + P_i^{(H)} G_i^{(L)} \right)^2 \left( 1 + \gamma_i^{(L)} \right)} \right)
\]

(5.29)

\[
\frac{dF}{dP_i^{(L)}} = \Phi_{2i}^{(L)} \left( B_s N_0 + P_i^{(H)} G_i^{(L)} \right) \left( 1 + \gamma_i^{(L)} \right) - \psi
\]

(5.30)

\[
\frac{dF}{d\mu} = \frac{B_s \sum_{i=1}^{D} \log_2 \left( 1 + \gamma_i^{(L)} \right) - B_s \sum_{i=1}^{D} \log_2 \left( 1 + \gamma_i^{(H)} \right)}{\phi_{2i-1}^{(H)}} + B_s \sum_{i=1}^{D} \log_2 \left( 1 + \gamma_i^{(L)} \right)
\]

(5.31)

\[
\frac{dF}{d\psi} = \sum_{i=1}^{D} \left( P_i^{(L)} + P_i^{(H)} \right) - P_t.
\]

(5.32)
By setting each of these equations to zero and solving (5.29) for the Lagrange variable $\psi$ we obtain

$$
\psi = \frac{B_s P_{l}^{(L)} \left( G_{l}^{(L)} \right)^2}{\left( B_s N_0 + P_{l}^{(H)} G_{l}^{(L)} \right)^2 \left( 1 + \gamma_{l}^{(L)} \right)} - \frac{G_{i}^{(H)}}{N_0 \left( 1 + \gamma_{i}^{(H)} \right)} + \mu \left( \frac{G_{i}^{(H)}}{N_0 \Phi_{2i-1}^{(H)} \left( 1 + \gamma_{i}^{(H)} \right)} + \frac{B_s P_{l}^{(L)} \left( G_{l}^{(L)} \right)^2}{\left( B_s N_0 + P_{l}^{(H)} G_{l}^{(L)} \right)^2 \left( 1 + \gamma_{l}^{(L)} \right)} \right).
$$

\[ (5.33) \]

Next, by substituting (5.33) in (5.30) and solving for $P_{l}^{(H)}$ we obtain

$$
P_{l}^{(H)} = - \frac{B_s N_0 \left( G_{l}^{(H)} \Phi_{2i}^{(L)} + G_{l}^{(L)} \Phi_{2i}^{(H)} \right) \mu + \left( G_{l}^{(L)} - G_{l}^{(H)} \right) \Phi_{2i-1}^{(H)} \Phi_{2i}^{(L)}}{\left( \Phi_{2i}^{(L)} + \Phi_{2i-1}^{(H)} \right) \mu G_{l}^{(L)} G_{l}^{(H)}}.
$$

\[ (5.34) \]

At this point, the solution is optimal; however, solving for $P_{l}^{(L)}$ by setting (5.32) to zero would be

$$
\sum_{l=1}^{D} P_{l}^{(L)} = P_t -
\sum_{l=1}^{D} \left( - \frac{B_s N_0 \left( G_{l}^{(H)} \Phi_{2i}^{(L)} + G_{l}^{(L)} \Phi_{2i}^{(H)} \right) \mu + \left( G_{l}^{(L)} - G_{l}^{(H)} \right) \Phi_{2i-1}^{(H)} \Phi_{2i}^{(L)}}{\left( \Phi_{2i}^{(L)} + \Phi_{2i-1}^{(H)} \right) \mu G_{l}^{(L)} G_{l}^{(H)}} \right).
$$

\[ (5.35) \]

Solving this expression requires the use of computationally intensive numerical solutions; hence, we simplify it by assuming that the total power ($P_t$) is equally divided among the $D$ data streams then the power is optimally allocated to the strong and weak users within each stream. The expression of (5.35) can be simplified as

$$
P_{l}^{(L)} = \frac{P_t}{D} - P_{l}^{(H)} = P_{TS} - P_{l}^{(H)}.
$$

\[ (5.36) \]

We denote this technique as the equal per stream power allocation (EPA)
technique. This approach is inspired by the scheme proposed in Chapter 3 but is used here based on the currently formulated problem. Finally, The closed-form solution is obtained by simply substituting (5.34) and (5.36) into (5.31). Accordingly, the suboptimal power for the strong user is found to be

\[ P^{(H)}_i = \left[ \frac{\sqrt{B_s N_0 \beta} - B_s N_0 \sqrt{T_H} \left( G^{(H)}_i + G^{(L)}_i \right)}{2G^{(H)}_i G^{(L)}_i \sqrt{T_H}} \right]^+ \] (5.37)

while that of the weaker user is found as

\[ P^{(L)}_i = \left[ \frac{2G^{(H)}_i G^{(L)}_i \sqrt{T_H} P_{TS} - \sqrt{B_s N_0 \beta}}{2G^{(H)}_i G^{(L)}_i \sqrt{T_H}} - \frac{B_s N_0 \sqrt{T_H} \left( G^{(H)}_i + G^{(L)}_i \right)}{2G^{(H)}_i G^{(L)}_i \sqrt{T_H}} \right]^+ \] (5.38)

where \( \beta = 4G^{(H)}_i G^{(L)}_i \Gamma_L \left( B_s N_0 + G^{(L)}_i P_{TS} \right) + B_s N_0 \Gamma_H \left( G^{(H)}_i - G^{(L)}_i \right)^2 \) and \( \Gamma_H = \frac{1}{2^{\frac{H}{2\lambda - 1}}} \), while \( \Gamma_L = \frac{1}{2^{\frac{L}{2\lambda}}} \).

These two equations form the proposed EPA technique that will allocate the user power when only two users are multiplexed per each data stream.

To apply the proposed solutions in (5.37) and (5.38) to a larger number of users, the concept of HP that was presented in Chapter 3 will be applied to form the proposed HP based power allocation scheme (EPA-HP). According to HP, the users are organized into \( Z \) pairs. At the \( Z \)-th pair (the strongest user-pair), the weaker user will consider its partner as noise and will perform SIC to all of the preceding pairs, while the strongest user at this pair will perform SIC to all of the previous pairs and its partner. Accordingly, the achievable rate by the weak user in the \( z \)-th pair is given by

\[ R^{(L)}_z = B_s \sum_{l=1}^{D} \log_2 \left( 1 + \frac{P^{(L)}_{z,l} G^{(L)}_{z,l}}{B_s N_0 + I_z + P^{(H)}_{z,l} G^{(H)}_{z,l}} \right) \] (5.39)

while that achieved by the better user in this pair is given by

\[ R^{(H)}_z = B_s \sum_{l=1}^{D} \log_2 \left( 1 + \frac{P^{(H)}_{z,l} G^{(H)}_{z,l}}{I_z + B_s N_0} \right) \] (5.40)
where $I_z$ represent the interference effect from other user pairs, and it is denoted by

$$I_z = \begin{cases} 
0 & z = Z \\
\sum_{c=z+1}^{Z} \left( P_{c,l}^{(L)} G_{z,l} \right) + \left( P_{c,l}^{(H)} G_{z,l} \right) & 1 \leq z \leq (Z - 1).
\end{cases} \quad (5.41)$$

The allocated power to the strong user using EPA-HP is given by

$$P_{z,l}^{(H)} = \left[ -\frac{B_s N_0 \sqrt{T_H} \left( G_{z,l}^{(H)} + G_{z,l}^{(L)} \right)}{2 G_{z,l}^{(H)} G_{z,l}^{(L)} \sqrt{T_H}} - 2 I_z G_{z,l}^{(H)} G_{z,l}^{(L)} \sqrt{T_H} + \sqrt{\Omega + \left( B_s N_0 \right)^2 \Gamma_H \left( G_{z,l}^{(H)} - G_{z,l}^{(L)} \right)^2} \right] + 2 G_{z,l}^{(H)} G_{z,l}^{(L)} \sqrt{T_H} \quad (5.42)$$

and that the power allocated to the weak user is given by

$$P_{z,l}^{(L)} = \left[ \frac{B_s N_0 \left( G_{z,l}^{(H)} + G_{z,l}^{(L)} \right) + 2 \left( P_{TS} + I_z \right) G_{z,l}^{(H)} G_{z,l}^{(L)}}{2 G_{z,l}^{(H)} G_{z,l}^{(L)}} \right] - \sqrt{\Omega + \left( B_s N_0 \right)^2 \Gamma_H \left( G_{z,l}^{(H)} - G_{z,l}^{(L)} \right)^2} \right] + 2 G_{z,l}^{(H)} G_{z,l}^{(L)} \sqrt{T_H} \quad (5.43)$$

where $\Omega = 4 G_{z,l}^{(H)} G_{z,l}^{(L)} \Gamma_L \left( B_s N_0 + G_{z,l}^{(H)} I_z \right) \left( B_s N_0 + G_{z,l}^{(L)} \left( I_z + P_{TS} \right) \right)$.

### 5.4.2 Subgradient based Power Allocation

The second proposed power allocation technique is obtained by setting (5.30) to zero and solving for $P_{l}^{(L)}$ as follows
\[
\frac{dF}{dP^{(L)}_l} = \frac{\mu B_s G^{(L)}_l}{\Phi^{(L)}_{2i} \left( B_s N_0 + P^{(H)}_l G^{(L)}_l \right) \left( 1 + \frac{P^{(L)}_l G^{(L)}_l}{B_s N_0 + P^{(H)}_l G^{(L)}_l} \right)} - \psi = 0
\]

\[
P^{(L)}_l = \left[ \frac{\mu + \Phi^{(L)}_{2i}}{\psi \Phi^{(L)}_{2i}} - \frac{B_s N_0 + P^{(H)}_l G^{(L)}_l}{B_s G^{(L)}_l} \right]^+ . \quad (5.44)
\]

In the same way, the expression of \(P^{(H)}_l\) is obtained as

\[
P^{(H)}_l = \left[ \frac{\mu + \Phi^{(H)}_{2i-1}}{\psi \Phi^{(H)}_{2i-1}} - \frac{N_0}{G^{(H)}_l} \right]^+ . \quad (5.45)
\]

The power allocation for multiple users can be extended from (5.44) and (5.45) to

\[
P_{k,l} = \left[ \frac{(\mu_k + \Phi^{(k)})}{\psi \Phi^{(k)}} - \frac{B_s N_0 + \sum_{m=k+1}^{K} P_{m,l} G^{(k)}_l}{B_s G^{(k)}_l} \right]^+ \quad (5.46)
\]

for all users except the strongest one, and

\[
P_{K,l} = \left[ \frac{\mu_K + \Phi^{(K)}}{\psi \Phi^{(K)}} - \frac{N_0}{G^{(K)}_l} \right]^+ \quad (5.47)
\]

for the strongest user.

The optimal power allocation can be obtained by following the steps listed in Algorithm 5.1. At each step of this algorithm, the Lagrangian multipliers will be updated using

\[
\psi^{(i+1)}_k = \left[ \psi^{(i)}_k - \Theta^{(i)} \left( P_t - \sum_{k=1}^{K} \sum_{l=1}^{D} P_{k,l} \right) \right]^+ \quad (5.48)
\]
\[ \mu_{k}^{(i+1)} = \left[ \mu_{k}^{(i)} - \Theta^{(i)} \left( \sum_{k=1}^{K/2} \left( \frac{B_s}{\Phi_{2k}} \sum_{l=1}^{D} \log_2 (1 + \gamma_{2k,l}) \right) - \frac{B_s}{\Phi_{2k-1}} \sum_{l=1}^{D} \log_2 (1 + \gamma_{2k-1,l}) \right) \right]^+ \] (5.49)

where \( \Theta^{(i)} \) and \( \Theta^{(i)} \) are small step sizes to be updated at each iteration and chosen to be \( 0.1 / \sqrt{i} \) [115].

It is worth mentioning that the power allocation schemes are the same for both MISO-NOMA and MIMO-NOMA systems.
Algorithm 5.1 Sum rate calculations using subgradient based power allocation method

- **Initialization** the maximum tolerance \( \Delta \) and the maximum number of iterations \( I_{\text{max}} \)
- **Set** \( i = 0 \) (the iteration index)
- **Initialization** \( P_{k,l} = 0, P_{K,l} = 0, \lambda^{(i)}_{k} = 0.01, \mu^{(i)}_{k} = 1, \)
- **while** \( i \leq I_{\text{max}} \) **do**
  - **For** \( k = 1 \) **to** \( K \)
    - **For** \( l = 1 \) **to** \( D \)
      - **while** \( \lambda_{k} \) and \( \mu_{k} \) are not convergent, **do**
        - **Calculate** \( P_{k,l} \) and \( P_{K,l} \), from (5.46) and (5.47), respectively.
        - **update** \( \lambda_{k} \) and \( \mu_{k} \) using (5.48) and (5.49), respectively.
      - **end while**
      - **Solve** (5.23) using the obtained power \( P_{k,l}^{*} \) and \( P_{K,l}^{*} \).
    - **end**
  - **end**
  - **Calculate** \( R_{s} \).
  - **i = i + 1**
- **end while**
- **Return** \( (P_{k,l}, R_{s}) \)

5.5 Numerical Results

A total of \( K \) users distributed uniformly in a circular area of diameter 300m is used to model a downlink of multiuser MISO and MIMO-NOMA systems. In addition, channel estimation is assumed to be perfectly applied and the CSI is assumed to be perfectly known at the BS. Table 3.1 lists the rest of the simulation parameters. For both MISO-NOMA and MIMO-NOMA systems, the obtained results present a comparison among the proposed EPA, EPA-HP,
the subgradient methods, the numerically obtained optimal NOMA and the FTPA based NOMA system [17].

5.5.1 MISO-NOMA System

First, we examine the performance of the proposed power allocation techniques for a multiuser MISO-NOMA system. The performance of this system will be investigated by using the proposed power allocation schemes along with the optimized precoder as compared to the ZF and MMSE precoders. It is known that DPC is optimal for the BC, however, it was not included in this comparison as the aim here is to examine the profit that could be obtained by optimizing the MISO precoder while ZF and MMSE are used as a comparison benchmark.

Figure 5.1 shows the achievable sum rate increasing in proportion to the total transmission power for all of the compared schemes. This figure depicts that the optimized precoder offers better performance than the ZF and the MMSE. In addition, for each of the compared precoding techniques, the proposed subgradient and EPA power allocation schemes offer better performance than FTPA and also show a close behavior to that of the optimal solution.
Next, Figure 5.2 displays the achievable sum rate over increasing the number of the BS antennas. It also indicates that the optimized precoder offers better trends than ZF and MMSE. Moreover, the proposed EPA and the subgradient based power allocation schemes offers better performance than FTPA from [17], and they are closer to the optimal solution.
Figure 5.2: Sum rate versus increasing number of antennas with $P_t = 30 \text{dBm}$, $D = 4$ and $K = 4$.

### 5.5.2 MIMO-NOMA System with IA

In here, the term EPA-HP refers to the multiuser version of EPA (i.e., the proposed EPA combined with HP from Chapter 3). Before we evaluate the sum rate performance against the transmission power for MIMO-NOMA, the optimal number of users per NOMA stream is first investigated. The total number of users is set to 24, and the sum rate performance is evaluated against different number of users per stream.

After the user-multiplexing step in each case, IA will be used to align the effect of each cluster into an interference subspace other than the respective one. It can be seen from Figure 5.3 that the SVD-IA approaches have comparable performance to the capacity-achieving DPC-IA ones. More importantly, as the multiplexed number of users per cluster increases, the achievable sum rate of NOMA system decreases accordingly. This is because as the cluster becomes more crowded by multiplexed users and more competition will occur over the available resources, the users will have less degree of freedom.
which will affect their achievable rates. Hence for the rest of this chapter we will consider the scenario where only two users are multiplexed per each data stream, and IA will be applied to align the effect of all other pairs into the interference subspace. This will leave each user only with the effect of its partner which means one user (the stronger user) only will apply SIC and the other user (the weaker one) will consider its partner as noise.

![Figure 5.3](image.png)

Figure 5.3: Sum rate vs different numbers of users per each group with $P_t = 30dBm$ and $K = 24$ with SVD and DPC precoding.

The next simulation is to compare the two-user based NOMA with IA approach to that of multiuser per stream without IA. Figure 5.4 and Figure 5.5 depict that the two-user per stream approach with IA offers better performance than the multiuser per stream approach. This is an obvious evidence as IA could boost the achievable sum rate of NOMA system further. In addition, IA application simplifies the complexity of SIC because each user will apply SIC to its partner only while all other users were already aligned by IA. Moreover, these figures show that DPC based schemes offer better sum rate than their SVD based counterparts.
Another comparison is made between the achievable sum rate for different numbers of users with two-user based IA-NOMA as illustrated in Figure 5.6 and Figure 5.7. These figures show that the sum rate is continuously increasing in proportion to the number of users due to the multiuser diversity gain offered by NOMA. These figures also confirm that IA can improve the achievable sum rate and reduce the SIC complexity for a large number of users. Once again, applying DPC precoding outperforms its SVD contender.
Figure 5.5: Sum rate versus various transmission power levels with IA and DPC precoding with $D = 3$ and $K = 6$.

Figure 5.6: Sum rate versus increasing number of users with IA and SVD precoding with $D = 6$ and $P_t = 30dBm$. 
### Figure 5.7: Sum rate versus increasing number of users with IA and DPC precoding with $D = 6$ and $P_t = 30dBm$.  

<table>
<thead>
<tr>
<th>No. of users</th>
<th>Sum rate (Mbps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>50</td>
</tr>
<tr>
<td>6</td>
<td>100</td>
</tr>
<tr>
<td>7</td>
<td>150</td>
</tr>
<tr>
<td>8</td>
<td>200</td>
</tr>
<tr>
<td>9</td>
<td>250</td>
</tr>
<tr>
<td>10</td>
<td>300</td>
</tr>
</tbody>
</table>

### 5.6 Chapter Summary

This chapter presented sum rate maximization for a downlink, multiuser MIMO-NOMA system with IA technique. Firstly, it investigated the optimal precoder design along with power allocation to maximize the sum rate of MISO-NOMA system. Then the multiuser MIMO-NOMA system is considered using SVD and DPC based IA, where a number of users are grouped and multiplexed together using NOMA while the others are regarded as interference and aligned to the null space. Two suboptimal power allocation schemes are proposed, namely, a low complexity closed-form solution for a two-user scenario, which is then extended to the multiuser case by the HP approach from Chapter 3, and a subgradient based solution. Simulation results showed that the optimized precoder offers better performance than ZF and MMSE precoders. It also verified that applying IA with NOMA could improve the achievable sum rate and offers simplicity in terms of SIC. In other words, the spatial dimension is better used for interference suppression than
for multiplexing gain in multiuser NOMA system. Moreover, the subgradient and closed-form power allocation solutions perform very close to the optimal solution, but at a much lower complexity.
Chapter 6

NOMA based Homogeneous Small Cells Dense Network (Hom-SCDN)

6.1 Introduction

All the previous chapters considered a single cell based NOMA system model, this chapter on the other hand, considers a multicell based NOMA scenario. As has been mentioned in Chapter 2, capacity enhancement can be obtained by deploying a large number of small cells which could be of different types and functionalities as in the case of HetNet or of the same type as in homogeneous networks [13, 71]. Meanwhile, NOMA is a highly promising scheme to enhance the bandwidth utilization [13]. This chapter will investigate the performance of NOMA based small cells dense networks by optimizing the achievable sum rate and the overall EE with a model that includes more than two cells. In addition, we propose a user pairing based power allocation to the users with the existence of ICI. Moreover, a new per cell power allocation scheme is also proposed to control ICI and to optimize EE of the whole network.
6.2 System Model of NOMA based Hom-SCDN

In this chapter, a homogeneous network of small cells is modeled with a total of \( L \) BSs spatially distributed following PPP with a density of \( \rho_{BS} \). In addition, the uniformly distributed users have a PPP density of \( \rho_K \) and they are allocated a total of \( S \) RBs that will be either fractionally or universally reused among all the BSs. To maximize the achievable network throughput in the considered scenario, each user connects to the BS that provides the best SINR. To reduce power consumption, improve user throughput and to mitigate ICI in a multi cellular network, one of the effective solutions that have been shown in 3GPP LTE Release 12 is to exploit the small cell on/off technique \([13,128]\). This technique opportunistically signals the BS to be in idle state when there is no need for its transmission (i.e., there are no active users within the cell boundaries of that BS) so that the unnecessary interference to its nearby cells will be reduced \([13,128]\). Throughout this work, the small cell on/off operation is taken into consideration. In this chapter, a problem similar to that formulated in Chapter 3 will be used to maximize the sum rate and allocate the power under the effect of ICI.

6.2.1 Problem Formulation and Solution

To maximize the sum rate for a multicell scenario, the same procedure that was used in Chapter 3 with a single cell NOMA could be generalized for multicell scenario in here. The optimization problem is formulated to maximize the sum rate of a two-user, multicell NOMA system, and then the solution is then generalized to the multi-user case. The optimization also included non-linear constraints of proportional fairness to guarantee that each user is able to achieve its target data rate. The overall network sum rate for a multicell, multiuser NOMA system is given as

\[
R_{net} = B_s \sum_{l=1}^{L} \sum_{k=1}^{K_l} \sum_{s=1}^{S_l} \log_2 (1 + \gamma_{l,k,s})
\]  

(6.1)

where \( L \) stands for the total number of BSs, \( K_l \) denotes the number of users being served by the \( l \)-th BS, and \( S_l \) refers to the total number of RBs allocated to the \( l \)-th BS in case of 3FR and it will be equal to the total number
of RBs \((S)\) in case of UFR. Finally, \(\gamma_{l,k,s}\) represents the SINR received at the \(s\)-th RB of user \(k \in \{1, \ldots, (K_l - 1)\}\) who is being served by the \(l\)-th BS, and it is given by

\[
\gamma_{l,k,s} = \frac{P_{l,k,s} |h_{l,k,s}|^2}{B_s N_0 + \sum_{j=k+1}^{K_l} P_{l,j,s} |h_{l,k,s}|^2 + \sum_{v=1}^{L} P_v |h_{l,k,s}|^2 \quad v = 1, (v \neq l) \tag{6.2}
\]

and that of the \(K_l\)-th user is

\[
\gamma_{l,K,s} = \frac{P_{l,K,s} |h_{l,K,s}|^2}{B_s N_0 + \sum_{v=1}^{L} P_v |h_{l,K,s}|^2 \quad v = 1, (v \neq l) \tag{6.3}
\]

where \(P_{l,K,s}\) denotes the transmission power at the \(s\)-th RB of the \(k\)-th user who is being served by the \(l\)-th BS. The term \(\sum_{v=1}^{L} P_v |h_{l,K,s}|^2 \) represents the ICI effect from other BSs on the \(l\)-th BS, where \(P_v\) represents the transmission power of each interfering BS.

By assuming that all the interfering BSs are transmitting using the maximum allowed limit \(P_t\) (i.e., \(P_v = P_t\)), the mathematical formulation of the optimization problem for a two-user scenario is given as

\[
\text{maximize} \quad R_{\text{net}} \tag{6.4}
\]

\[
\text{Subject to} \quad \sum_{k=1}^{K_l} \sum_{s=1}^{s_t} P_{l,k,s} \leq P_t \tag{6.5}
\]

\[
P_{l,k,s} \geq 0, \quad \forall l, k, s \tag{6.6}
\]

\[
R_{2i} : R_{2i-1} \geq \Phi_{2i} : \Phi_{2i-1} \tag{6.7}
\]

where \(i \in \left\{1, 2, \ldots, \frac{K_l}{2}\right\}\)

where (6.7) represents the proportional fairness constraint, and \(\Phi_{2i-1}\) and \(\Phi_{2i}\)
stand for the minimum rate requirements for the user with the best channel conditions and the worst channel conditions, respectively. Next, we will optimize the power allocated to all $K_l$ users that are served by the $l$-th BS while considering the effect from other BSs as ICI. This problem is formulated for a multicell scenario which is the main difference from the previously formulated problems in the preceding chapters. The Lagrangian function of the optimization problem in (6.4)-(6.7) for the $l$-th BS is represented as

$$F(P_{l,k,s}) = B_s \sum_{k=1}^{K_l} \sum_{s=1}^{S} \log_2 (1 + \gamma_{l,k,s}) - \mu \left( \sum_{k=1}^{K} \sum_{s=1}^{S} P_{l,k,s} - P_t \right) -$$

$$\varphi \sum_{k=1}^{K_l} \left( \frac{B_s \sum_{s=1}^{S} \log_2 (1 + \gamma_{l,k,s})}{\Phi_{2k}} - \frac{B_s \sum_{s=1}^{S} \log_2 (1 + \gamma_{l,2k-1,s})}{\Phi_{2k-1}} \right)$$

where $\varphi$ and $\mu$ represent the Lagrange multipliers. By taking the derivative against $P_{l,k,s}$ for all users except the one with the best channel conditions (i.e., the $K$-th user), we have

$$\frac{dF}{dP_{l,k,s}} = \frac{B_s |h_{l,k,s}|^2}{\left( \sum_{m=k+1}^{K_l} P_{l,m,s} |h_{l,k,s}|^2 + \sum_{v=1}^{L} P_v |h_{l,k,s}|^2 + B_s N_0 \right)} (1 + \gamma_{l,k,s}) \ln (2)$$

$$\mu + \frac{B_s |h_{l,k,s}|^2 \varphi_k}{\Phi_k \left( \sum_{m=k+1}^{K_l} P_{l,m,s} |h_{l,k,s}|^2 + \sum_{v=1}^{L} P_v |h_{l,k,s}|^2 + B_s N_0 \right)} (1 + \gamma_{l,k,s}) \ln (2)$$

For the user with the best channel conditions (the strong user), the derivative is
\[
\frac{dF}{dP_{l,K,s}} = \frac{B_s|h_{l,K,s}|^2}{\left( \sum_{v=1, (v \neq l)}^{L} P_v|h_{l,k,s}|^2 + B_sN_0 \right) (1 + \gamma_{l,K,s}) ln (2)} - \Phi_K \left( \sum_{v=1, (v \neq l)}^{L} P_v|h_{l,k,s}|^2 + B_sN_0 \right) (1 + \gamma_{l,K,s}) ln (2)
\]

Following the same steps that were mentioned in Chapter 4 to obtain the subgradient based solution, the power allocation approach of all users except the strong one is found to be

\[
P_{l,k,s} = \left( \Phi_k + \varphi_k \right) \left( \sum_{m=k+1}^{K} P_{l,m,s} + \sum_{v=1, (v \neq l)}^{L} P_v \right) \left| h_{l,k,s} \right|^2 + B_sN_0 \right)^+ \left( \frac{\Phi_k + \varphi_k}{\Phi_k \mu ln (2)} \right)
\]

On the other hand, the power allocation approach of the strong user is
$P_{l,K,s} = \left[ \left( \frac{\Phi_K - \sum_{k=1}^{K-1} \varphi_k}{\Phi_K \mu \ln(2)} \right) \left( N_0 B_s + \sum_{v=1, (v \neq l)}^{L} P_v |h_{l,K,s}|^2 \right) \right]^+ . \quad (6.12)$

Regarding the HPPA approach that was proposed in Chapter 3, the same steps can be repeated to obtain the solution for a multicell scenario with the one difference represented by including the ICI term in the SINR denominator. The final expressions for the user with weak channel conditions and the strong channel conditions are given in (6.13) and (6.14) respectively. The problem solution can be obtained by following the same procedure that is listed in Algorithm 5.1 which will be used to calculate the sum rate with the only difference is to replace (5.23) by (6.4).

### 6.2.2 Approximated Model and EE Optimization Problem

#### 6.2.2.1 Approximated Model

In SCDN, it is expected that as the BS density increases, the effect of ICI (i.e., $\sum_{v=1, (v \neq l)}^{L} P_v |h_{l,k}|^2$) becomes very high as compared to $B_s N_0$, in other words

$$\sum_{v=1, (v \neq l)}^{L} P_v |h_{l,k}|^2 \gg B_s N_0 \quad (6.15)$$

Thus, the expression in (6.1) can be approximated to be
\[ P_{l,p,s}^{(L)} = \left( \frac{\sqrt{\Gamma_H} \left( B_s N_0 \left( |h_{l,p,s}^{(H)}|^2 + |h_{l,p,s}^{(L)}|^2 \right) + 2|h_{l,p,s}^{(H)}|^2 |h_{l,p,s}^{(L)}|^2 \Lambda_5 \right) - 2\Lambda_4}{2\Lambda_4} \right)^{1/2} \]

\[ P_{l,p,s}^{(H)} = \left( \frac{(\Lambda_3 \sqrt{\Gamma_L} - 2\Lambda_4 \sqrt{\Gamma_H}) \Lambda_2 + B_s N_0 \left( |h_{l,p,s}^{(L)}|^2 (\Gamma_L - \Gamma_H) - |h_{l,p,s}^{(H)}|^2 \Gamma_H \right) + 2\Lambda_3 \sqrt{\Gamma_L} B_s N_0^*}{2\Lambda_4 \sqrt{\Gamma_H}} \right)^{1/2} \]

where

\[
\begin{aligned}
\Lambda_1 &= \left( 1 + \gamma_{l,s}^{(H)} \right) \left( 1 + \gamma_{l,s}^{(L)} \right) \\
\Lambda_2 &= I_s + \sum_v \frac{P_v}{L}, \quad v = 1, (v \neq l) \\
\Lambda_3 &= |h_{l,s}^{(H)}|^2 |h_{l,s}^{(L)}|^2 \sqrt{\Gamma_L} \\
\Lambda_4 &= |h_{l,s}^{(H)}|^2 |h_{l,s}^{(L)}|^2 \sqrt{\Gamma_H} \\
\Lambda_5 &= P_{RB} + \sum_v \frac{P_v}{L}, \quad v = 1, (v \neq l) \\
\Gamma_H &= 2 \frac{\phi^{(H)}}{\Phi_{\min}}^{-\frac{1}{2}} \\
\Gamma_L &= 2 \frac{\phi^{(L)}}{\Phi_{\min}}^{-\frac{1}{2}}
\end{aligned}
\]
where $P_l$ denotes the transmission power of the $l$-th BS, and its boundaries are $(0 \leq P_l \leq P_t)$.

**Proof:** Consider the case where all the BSs have equal transmission power $P_l$ and consider $K = 3$ per the each cell as an example (i.e., each BS is serving three users). Assume that these users are arranged in an ascending order based on their channel gains. Then (6.1) will be expressed as

$$R_{net} = W_T \sum_{l=1}^{L} \log_2 \left( 1 + \frac{P_l}{\sum_{v=1}^{L} \sum_{v \neq l} P_v} \right)$$

(6.16)

the term $\sum_{s=1}^{S} B_s$ could be replaced with the total bandwidth $W_T$, and (6.17)
can be expanded as

\[
R_{\text{net}} = W_T \sum_{l=1}^{L} \log_2 \left( 1 + \frac{P_{l,1}}{(P_{l,2} + P_{l,3}) + \sum_{v=1, (v \neq l)}^{L} P_v} \right) * \left( 1 + \frac{P_{l,2}}{P_{l,3} + \sum_{v=1, (v \neq l)}^{L} P_v} \right) \left( 1 + \frac{P_{l,3}}{L \sum_{v=1, (v \neq l)}^{L} P_v} \right)
\]

(6.18)

\[
R_{\text{net}} = W_T \sum_{l=1}^{L} \log_2 \left( \frac{P_{l,1} + (P_{l,2} + P_{l,3}) + \sum_{v=1, (v \neq l)}^{L} P_v}{(P_{l,2} + P_{l,3}) + \sum_{v=1, (v \neq l)}^{L} P_v} \right) *
\]

\[
\left( \frac{P_{l,2}}{P_{l,3} + \sum_{v=1, (v \neq l)}^{L} P_v} \right) \left( \frac{P_{l,3}}{L \sum_{v=1, (v \neq l)}^{L} P_v} \right)
\]

(6.19)

By canceling the similar terms, we can have
\[ R_{net} = W_T \sum_{l=1}^{L} \log_2 \left( \frac{P_{l,1} + (P_{l,2} + P_{l,3}) + \sum_{v=1, (v \neq l)}^{L} P_v}{\sum_{v=1, (v \neq l)}^{L} P_v} \right) \]  

(6.20)

\[ R_{net} = W_T \sum_{l=1}^{L} \log_2 \left( 1 + \frac{P_{l,1} + (P_{l,2} + P_{l,3})}{\sum_{v=1, (v \neq l)}^{L} P_v} \right) \]  

(6.21)

This derivation hold for any arbitrary number of users and not just three. It could be noticed that the numerator of (6.21) is the total transmission power of the \( l \)-th BS, so (6.21) can be simplified as

\[ R_{net} = W_T \sum_{l=1}^{L} \log_2 \left( 1 + \frac{P_l}{\sum_{v=1, (v \neq l)}^{L} P_v} \right) \]  

(6.22)

### 6.2.2.2 Power Minimization Problem formulation and Solution

From the expressions (6.11)-(6.14), it is clear that the BS transmission power plays a vital rule in the achievable rate of its counterpart inside the network. Thus, it is necessary to optimize the level of the transmitted power by each BS to limit the ICI effect and balance it with a convenient level of QoS for the active users. This in turn will optimize the EE, minimize power consumption
and mitigate ICI. For this purpose, an optimized power allocation will be used in here to decide the transmission power for each BS while fulfilling a target rate requirements. In such case, the problem can be formulated as

$$\begin{align*}
\text{minimize} & \quad P_l \\
\text{Subject to} & \quad P_l \leq P_t \\
& \quad P_l \geq 0, \forall l \\
& \quad W_T \log_2 \left( 1 + \frac{P_l}{\sum_{v=1}^{L} P_v} \right) \geq \Phi_{BS_{\min}}
\end{align*}$$

where $\Phi_{BS_{\min}}$ stands for the minimum rate to be achieved by that BS, this rate could effectively represent the sum of the target rates requirements of the users that are connected to this BS. Constraint (6.24) is necessary to control the upper limit of the transmission power for each BS. On the other hand, constraint (6.25) guarantees that the BS power does not fall to a negative level. In particular, constraint (6.26) guarantees that each cell maintains a certain level of QoS for its users. The absence of this constraint might make the optimization problem to minimize the transmission power to a level that is below the reasonable limit where at least one user is requesting a service from that BS. In addition, this constraint is a minimum rate constraint that is mathematically represented in a subtractive form and not in a fractional form as in the previous cases that used proportional rate constraint. Moreover, the main difference between this problem and the previously formulated ones is that this a power minimization problem that aims at minimizing the objective function while all the other problems were formulated to maximize the objective function (i.e., the sum rate). The Lagrangian function of the formulated problem is expressed as
CHAPTER 6. NOMA BASED HOM-SCDN

\[ P_l - \mu \left( \sum_{v=1}^{L} P_l - P_l \right) \]

\[ -\psi \left( \Phi_{BS}^{min} - W_T \sum_{v=1}^{L} \log_2 \left( 1 + \frac{P_l}{\sum_{v=1}^{L} P_v} \right) \right) \]  \hspace{1cm} (6.27)

Since we are trying to minimize the transmission power, the power constraint in (6.24) may not be necessary as the condition would be \( \sum_{l=1}^{L} P_l < (L P_l) \), which means the Lagrangian function can be rewritten as

\[ J = P_l - \psi \left( \Phi_{BS}^{min} - W_T \sum_{v=1}^{L} \log_2 \left( 1 + \frac{P_l}{\sum_{v=1}^{L} P_v} \right) \right) \]  \hspace{1cm} (6.28)

The closed-form solution for the above formulated problem is found as

\[ P_l = \left( 2 \frac{\Phi_{BS}^{min}}{W_T} - 1 \right) \sum_{v=1}^{L} P_v \]  \hspace{1cm} (6.29)

which makes the transmission power to be used by each BS depends largely on the minimum rate requirement \( \Phi_{BS}^{min} \) of that BS. Since \( \Phi_{BS}^{min} \) is equal for all
of the BSs, the power will also be equal but not necessarily at the maximum level $P_t$. The expression in (6.29) can be used iteratively to decide each BS power, with constraint (6.24) can control the maximum limits of the power used by each BS as detailed in Algorithm 6.1.

**Algorithm 6.1** Iterative based BS power allocation.

- **Initialization** $P_l = 0, P_v = 0.0001$,
  - *while* $P_l$ is less than or equal to $P_t$, *do*
    * Calculate $P_l$ using (6.29).
    * update $P_v$ with the obtained $P_l$.
  - *end while*

- **Return** $P_l$.

6.3 Numerical Results

We consider the downlink of small cells network with cellular users spatially distributed within a $1 \times 1 \text{ km}^2$ coverage area. The wireless channel is modeled using path loss, shadowing, and fading, which follows the six-path frequency selective fading channel using the ITU pedestrian - B model that is depicted in Table 2.1. Unless stated otherwise, Table 6.1 depicts the simulation parameters [2–4] that are used in all of the simulation scenarios. The optimal power allocation for NOMA is numerically solved and the performance is compared to the proposed low complexity HPPA based NOMA schemes. The comparison is made between the straightforward NOMA and the HP based NOMA approaches because both are dealing with all users in one NOMA group, so they are equivalent to each other.
Table 6.1: Simulation parameters [2–4]

<table>
<thead>
<tr>
<th>Parameter Name</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum Transmitted power ( (P_t) )</td>
<td>0.1 W (20 dBm)</td>
</tr>
<tr>
<td>Dimensions of the Euclidean plane (Covered area)</td>
<td>( 1 \times 1 \text{km}^2 )</td>
</tr>
<tr>
<td>Path loss exponent ( (\nu) )</td>
<td>3.76</td>
</tr>
<tr>
<td>Noise power density ( (N_0) )</td>
<td>-174 dBm / Hz</td>
</tr>
<tr>
<td>Total system bandwidth ( (W_T) )</td>
<td>20 MHz</td>
</tr>
<tr>
<td>No. of RBs ( (S) )</td>
<td>100</td>
</tr>
<tr>
<td>Bandwidth per RB ( (B_s) )</td>
<td>200 kHz</td>
</tr>
<tr>
<td>No. of subcarriers per RB ( (N_c) )</td>
<td>12</td>
</tr>
<tr>
<td>Shadowing standard deviation</td>
<td>8 dB</td>
</tr>
<tr>
<td>( PL_0 ) at 2 GHz band</td>
<td>( 15.3 + 37.6 \log_{10}(d_0) )</td>
</tr>
<tr>
<td>( \Phi_{min}^{(H)} )</td>
<td>0.1 Mbps</td>
</tr>
<tr>
<td>( \Phi_{min}^{(L)} )</td>
<td>0.05 Mbps</td>
</tr>
<tr>
<td>Users density ( (\rho_K) )</td>
<td>( 2.5 \times 10^{-4} \text{m}^{-2} )</td>
</tr>
<tr>
<td>BS density ( (\rho_{BS}) )</td>
<td>( 0.2 \times 10^{-4} \text{m}^{-2} )</td>
</tr>
</tbody>
</table>

6.3.1 Numerical results for the sum rate optimization schemes

The optimal power allocation for NOMA is numerically solved and the performance is compared to the proposed low complexity HPPA based and sub-gradient based NOMA schemes. The frequency planning of UFR based SCDN and 3FR based SCDN is shown in Figure 6.1 and Figure 6.2, respectively. To guarantee equivalent comparison between them, the results are obtained using similar set up. It is worth mentioning that the main difference between 3FR and UFR is represented by the fact that, with 3FR, the total of \( S \) RBs will be equally divided into three groups to be reused within different clusters. In addition, with 3FR the frequency group is allocated to each cell based on the geographic position of that cell and how many active users in it. For example, if one cell has no users or few cell edge users; it will be allocated the same frequency group as that of its nearby cells, and then this cell is switched into idle state. If a cell has a different frequency group (different color) from its
nearby cells, it means this cell is in idle state and the users within its borders are being served by nearby cells as they offer better SINR.

Figure 6.3 displays the overall system sum rate achieved for different BS densities with both 3FR and UFR reuse schemes. From this figure, it is clear that the overall system sum rate increases in proportion to the BS density for both of the reuse schemes. In particular, the 3FR case offers better performance than the UFR one. This is because with 3FR, the ICI level is lower than that in the case of UFR.
Figure 6.2: SCDN with 3FR planning for $\rho_K = 2.5 \times 10^{-4} \text{ m}^{-2}$ and $\rho_{BS} = 0.5 \times 10^{-4} \text{ m}^{-2}$. BSs with the same color use the same RBs.

Figure 6.3: Sum rate against different BSs densities for NOMA based SCDN with the users at a density of $\rho_K = 2.5 \times 10^{-4} \text{ m}^{-2}$. 
6.3.2 Numerical results for the approximated model and EE optimization

Figure 6.4 and Figure 6.5 examine the equivalence of the derived (6.16) to that of (6.1) for a UFR case. It is clear that both expressions become equivalent when the average number of BSs reaches about 40 (at a density of $\rho_{BS} = 6 \times 10^{-4} \text{m}^{-2}$) or higher. In Figure 6.4 there is a small gap between the two compared schemes that add up and becomes noticeable in Figure 6.5.

![Graph showing average rate achieved per cell for various BS densities for NOMA based SCDN.](image)

Figure 6.4: The average rate achieved by each individual cell for various BS densities for NOMA based SCDN.
Figure 6.5: The overall sum rate achieved by multicell for various BS densities for NOMA based SCDN.

We also solved the formulated problem in (6.23) to (6.26) numerically as can be seen in Figure 6.6. This figure shows that minimizing the power improves the overall EE, this is because reducing the BS transmission power will cause less interference (i.e., less ICI) to the other BSs around it.
CHAPTER 6. NOMA BASED HOM-SCDN

Figure 6.6: Overall EE for NOMA based SCDN using different BS densities ($\rho_{BS}$) with $P_c = 2W$, $\Phi_{BS}^{min} = 5Mbps$, $\rho_K = 3 \times \rho_{BS}$ and $P_t = 25$dBm.

6.4 Chapter Summary

Unlike the previous chapters that considered single cell scenarios, this chapter presented the power allocation problem of NOMA based SCDN in the presence of co-channel interference (CCI). Both sum rate and EE are optimized in this chapter. For the sum rate, both the proposed HPPA-NOMA (Chapter 3) and subgradient based power allocation scheme (Chapter 4) are investigated with both UFR and frequency reuse with a reuse factor of 3 (3FR) planning. For EE optimization, a power minimization problem is formulated to reduce the effect of ICI and a simple solution based on the minimum rate requirement is proposed to allocate the power per each BS. Numerical results showed the close behavior of the proposed scheme as compared to the numerically obtained optimal scheme and how it is better than the upper bound one.
Chapter 7

Conclusions and Future Work

7.1 Conclusions

The aim of this project is to improve the performance of NOMA system by investigating the resource allocation process and its optimization in the downlink of NOMA system. Using LDD approach, the performance of NOMA is optimized with various objective functions such as the sum rate and EE. Various performance metrics, for instance the achievable sum rate, coverage probability, and EE-SE trade-off were applied to examine the effectiveness of the proposed schemes using Monte Carlo simulations.

In Chapter 3, an optimization problem is formulated using LDD to maximize the sum rate with total power and proportional minimum rate constraints. Two suboptimal power allocation methods have been proposed to allocate the transmission power to each user in a two-user scenario, then, the proposed techniques are extended to a multiuser scenario by the HP concept. The pairs are then multiplexed in the power domain, which is obtained from a modified solution to the obtained suboptimal ones in the two-user scenario. Power allocation is applied across and within all users pairs. Numerical results show that NOMA provides better performance than OFDMA. Moreover, the simulations confirm that the proposed ERPA and ACPA methods achieve comparable performance to the optimal solution with the advantage of lower complexity and also better than the other NOMA existing schemes such as FTPA. Among these two methods, ERPA method performs slightly better than ACPA at the cost of more complexity. For the multiuser scenario, the proposed HPPA showed the best performance among the compared schemes.
and the closest to the optimal one. This confirms the effectiveness of the HP concept applicability to NOMA system with a large number of users.

Furthermore, the idea of hybrid multiple access has been proposed as a combination between NOMA and OFDMA to utilize the transmission schemes for varying channel conditions. It has been shown that the hybrid scheme outperforms NOMA due to its flexibility in adapting the transmission approach according to the channel condition; especially when there is a significant gap between the channel gains of the served users, which makes it a favourable method to satisfy the future traffic demand.

Another important design property that NOMA must improve is the EE. Chapter 4 of this thesis addressed the EE-SE relationship for a single cell downlink NOMA system and also compared it to that of OFDMA system. The EE is proved to be quasiconcave in terms of SE. An optimization problem for maximizing the EE is formulated also using LDD approach. Due to the complexity of the objective function, the first approach simplifies the problem using the Dinkelbach concept. After that, an iterative based solution is obtained using the subgradient method to allocate the transmission power among the users. The proposed subgradient based NOMA system shown to offer simplicity over the numerically obtained optimal scheme. Simulations are used to model the impact of the transmission power, the circuit power, the number of users, and the SE variations against EE. The results confirm the energy efficient behaviour of NOMA and its superiority against OFDMA and also supports the validity of the proposed approaches and their effectiveness as compared to the optimal one.

Combining MIMO with NOMA is expected to further enhance the overall achievable capacity. Chapter 5 of this thesis also presented new resource allocation schemes for multiuser MIMO-NOMA system. Firstly, the performance of multiuser MISO-NOMA system is investigated along with precoder optimization approach, and compared to the existing ZF and MMSE precoders. The application of IA to multiuser MIMO-NOMA system is investigated by proposing SVD-IA and DPC-IA, in which a number of users are grouped and multiplexed together over each data stream using NOMA while the other streams are regarded as interference and are aligned to the null space.

Secondly, two power allocation schemes are proposed to maximize the
sum rate under the total transmit power constraint and the proportional minimum rate constraint. A low complexity closed-form solution for two-user scenario is obtained and then generalized to the multiuser case by the HP scheme that was proposed in Chapter 3. Also a subgradient based solution is also proposed to allocate the power iteratively to each user. For the multiuser MISO-NOMA case, simulation results showed that the optimized precoder outperforms both ZF and MMSE for all of the compared power allocation techniques. The results also showed that the proposed power allocation schemes outperform FTPA scheme from the literature and also showed their close behavior to the optimal one. Moreover, the two proposed power allocation schemes were also considered with multiuser MIMO-NOMA system. Simulations results showed that applying IA with NOMA offers better sum rate and requires simpler SIC than the case with no IA. All of the compared power allocation techniques offered better sum rate performance if the precoding is applied based on DPC rather than SVD. Moreover, the obtained results illustrate that the proposed EPA, EPA-HP and the subgradient method schemes outperformed FTPA and also showed close performance to the optimal scheme.

Finally, Chapter 6 of this thesis examined the performance of NOMA-SCDN with UFR and 3FR frequency planning to optimize both the sum rate and the EE. The proposed power allocation schemes from previous chapters were also applied and numerical results proved their effectiveness despite the existence of ICI. In addition, the major obstacle that seem to face NOMA-SCDN is the effect of ICI. To alleviate its effect and optimize EE, a power minimization problem is formulated using Lagrange approach. A minimum rate based power allocation solution is proposed and compared to the optimal solution and the upper bound solution. The proposed solution proved its effectiveness in addition to its simplicity.

### 7.2 Future Work

This thesis highlighted different resource allocation aspects for NOMA system. However, there are still several features related to NOMA that are not covered by this thesis and can therefore be considered for future works. Here we suggest possible research extensions to the work done in this thesis.
1. The performance of the hybrid multiple access presented in Chapter 3 can be investigated further by jointly optimizing the allocated RBs in the orthogonal part and the allocated power to both orthogonal and non-orthogonal parts.

2. This thesis considered NOMA as a stand-alone technique, however, it is interesting to consider the benefits of interacting NOMA with other technologies, such as, investigating the outcomes of applying the simultaneous wireless information and power transfer (SWIPT) and demonstrating how this combination can boost the energy harvesting process. In addition, it is interesting to consider the combination of NOMA with device to device communication (D2D), and machine to machine (M2M) and so forth.

3. The application of IA to MIMO-NOMA in Chapter 5 can be extended by considering other IA approaches that depend on, for example, the received SINR.

4. Chapter 6 of this thesis presented the performance examination of NOMA based SCDNs, in which we considered homogeneous network to avoid having a dominant large cell as the user connectivity to the BS was based on the received SINR. However, it is interesting to consider NOMA based HetNet along with small cell biasing. It is also important to investigate the power consumption of these dense networks. In addition, the BS in Chapter 6 are considered to be equipped with a single antenna, so it can be extended to consider a MIMO case. Moreover, the 3FR planning used in Chapter 6 did not optimize the number of RBs used in each division, therefore, in the future optimizing the number of RBs used per each cluster could be considered to enhance the overall system performance.

5. This thesis has extensively studied the downlink of NOMA system. It is important to consider the applicability of NOMA to the uplink as well.
References


REFERENCES


resource allocation for downlink non-orthogonal multiple access network,” *IEEE Transactions on Communications*, vol. 64, no. 9, pp. 3722–3732, Sept 2016.


REFERENCES


REFERENCES


direct-sequence spread-spectrum multihop packet radio network,” IEEE
Journal on Selected Areas in Communications, vol. 8, no. 5, pp. 762–771, Jun
1990.

a poisson field of interferers or scatterers,” IEEE Transactions on Signal

work interference and its applications,” Proceedings of the IEEE, vol. 97,

“Stochastic geometry and random graphs for the analysis and design of
wireless networks,” IEEE Journal on Selected Areas in Communications,

[95] L. Li and A. J. Goldsmith, “Capacity and optimal resource allocation
for fading broadcast channels I. ergodic capacity,” IEEE Transactions on

base station power model,” Communications Letters, IEEE, vol. 17, no. 11,

downlink base station power control strategy,” in Personal Indoor and
Mobile Radio Communications (PIMRC), 2011 IEEE 22nd International Sym-

power consumption,” Selected Areas in Communications, IEEE Journal on,

[99] A. De Domenico, R. Gupta, and E. Strinati, “Dynamic traffic manage-
ment for green open access femtocell networks,” in Vehicular Technology
REFERENCES


[120] D. H. Nguyen and T. Le-Ngoc, “MMSE precoding for multiuser MISO downlink transmission with non-homogeneous user SNR conditions,”
REFERENCES


Appendix A

A proof that EE is continuously differentiable in terms of SE

The objective function in (4.5) is also continuously differentiable in terms of SE, the following steps can be followed.

Proof. Since

\[
\lim_{\Delta R_k \to 0} \frac{EE(R_k + \Delta R_k) - EE(R_k)}{\Delta R_k}
\]  \hspace{1cm} (A.1)

where \( \Delta R_k \) is infinitely small increments of throughput, then (A.1) is simplified to be

\[
= \lim_{\Delta R_k \to 0} \frac{(R_k + \Delta R_k) (C + \zeta R_k) - R_k (C + \zeta (R_k + \Delta R_k))}{\Delta R_k (C + \zeta R_k) (C + \zeta (R_k + \Delta R_k))}
\]

\[
= \lim_{\Delta R_k \to 0} \frac{C}{(C + \zeta R_k) (C + \zeta (R_k + \Delta R_k))}
\]

\[
= \frac{C}{(C + \zeta R_k)^2}
\]

This completes the proof.
Appendix B

The Grid Structure of the LTE Resource Block

B.1 Introduction

The RB represents the smallest unit in the physical layer that the MAC scheduler can assign to a specific user. The channel gains represented by the combination of the path loss, shadowing, small scale fading and large-scale fading for each user are calculated over each RB. Therefore, the user achievable rate is calculated using the SINR of each RB allocated to that user. Due to the frequency and time selectivity nature of the wireless channels, the RBs channel gain values vary from one user to another [129, 130]. On the other hand, the resource element represents the smallest unit in the RB that contains a single subcarrier in the frequency domain and a single OFDM or single carrier (SC)-FDMA symbol of a complex value in the time domain. Figure B.1 clarifies how the basic LTE RB can be seen as a time-frequency grid [130]. In this figure, the terms $N_c^{UL}$ and $N_c^{DL}$ represent the number of subcarriers in the uplink and the downlink respectively, while $S^{UL}$ and $S^{DL}$ stand for the total number of RBs in the uplink and the downlink respectively.

According to LTE standards, the $10ms$ duration of the LTE frame is divided into 10 subframes, each subframe has a duration $1.0ms$ and is further splitted into two time slots. A single RB occupies one time slot of $0.5ms$ duration in the time domain and a bandwidth of about $180kHz$ in the frequency domain as shown in Figure B.1. The total number of RBs depends on the overall transmission bandwidth as listed in Table B.1.
APPENDIX B. THE GRID STRUCTURE OF THE LTE RESOURCE BLOCK

There are 12 subcarriers per each RB, and the number of symbols per each RB might vary in both the uplink and the downlink depending on the subcarrier bandwidth and the cyclic prefix length, where a normal cyclic prefix allows each time slot to occupy 7 OFDM symbols, while an extended cyclic
prefix would allow about 6 OFDM per each time slot [5, 129, 130]. The RB structure in both the uplink and downlink is very similar; the main difference between the downlink transmission and the uplink transmission is that the former transmission supports long symbol duration by using a subcarrier spacing of 7.5 kHz as compared to the 15 kHz of the later, such as the case for multimedia broadcasting over single frequency network. This means that the symbols duration are twice as long as the 15 kHz case, which makes it easier to use a longer cyclic prefix to combat the higher delay spread encountered especially in the large-sized cells [130].
Appendix C

Overview of Radio Resource Allocation

C.1 Introduction

Introducing NOMA with other techniques might not be enough to efficiently exploit the available power and bandwidth resources. The operators of cellular networks must provide satisfactory services to the users so they can maintain a large number of subscribers and attract new ones. From the user’s perspective, one of the most important features to guarantee user satisfaction and loyalty to the cellular operator is maintaining acceptable level of QoS. Therefore, maximizing the number of satisfied users is a reasonable objective to be pursued when designing cellular networks. The concept of user satisfaction is relative and depends on several aspects, in other words, technical aspects such as throughput, service type, and delay, as well as economic aspects such as subscription fees [131]. To encounter the challenges of cellular operators and to increase the number of satisfied users for various service types, RRA is vital. RRA is the process of distributing and managing the available scarce resources of the radio air interface to be effectively exploited by the active connections. Frequency resources, transmit power, and time slots, are the most important among the resources to be managed by RRA strategies. The nature of NOMA allows all the users to transmit using the whole bandwidth as well as occupying all the available time slots. Thus, the main issue would be allocating the transmit power in the most effective way possible.
C.1.1 RRA Classification in terms of Decision Control

RRA could be classified into, in terms of control, centralized and distributed techniques the former involves a central controller that establishes the resource sharing process based on the propagation conditions of the environment, and also to monitor which BS could provide highest SINR to a certain user [132,133]. On the other hand, the distributed techniques involve each cell being responsible for managing its own radio resources. The former is more efficient and more practically implemented, as it is more fluent in terms of system performance optimization. However, the latter is simpler and easier to be implemented [131–133]. Thirdly, self organizing network (SON) where the network becomes adaptive to its environment by performing self healing, self optimization, and self configuration functions through continuous communication among the network nodes [133,134].

C.1.2 RRA Classification in terms of Allocation Mechanism

In terms of allocation mechanism, there are two categories of resource allocation techniques: dynamic, and fixed resource allocation [41]. In dynamic resource allocation, the users get their resources based on their channel gains in an adaptive manner [41]. In fixed resource allocation, on the other hand, the users get a fixed share of the power or the bandwidth resources regardless of their channel gains [41]. As it depends on the instantaneous user channel gain to allocate resources, dynamic scheme offers better performance than the fixed one.

C.1.3 RRA Optimization Techniques

In terms of optimization techniques used, fixed resource allocation schemes represent a suboptimal approach because it does not take user conditions into account to assign its resources. On the other side, dynamic schemes are usually represented through two different optimization approaches; namely, margin adaptive (MA) and rate adaptive (RA). In MA, the objective function is to minimize the transmission power subject to the users rate constraints. On the other hand, with RA, the objective function is to maximize the total sum rate under the total power constraints [41]. The optimal solution for these
techniques could be found using numerical techniques which, in most times, is not easy to find, as a result the closed-form solutions are derived [41].

Mainly, two different approaches can be followed to solve the RRA design problems, namely the heuristic and the utility based approaches. Whilst the heuristic approach provides simplified and quick solutions to the RRA design problems, the utility based approach offers more flexibility and act as a general tool for RRA design [131]. Heuristic solutions include finding satisfactory answers to the studied problems through common sense and experience. Such solutions are suitable for the problems where the best possible solution is very complex or nearly impossible to reach [131]. On the other hand, utility-based RRA applied to evaluate the degree of satisfaction that communication network is able to satisfy the service requirements of its users with a certain criterion, such as in terms of throughput and delay; or to measure the usage benefit of certain resources, e.g., transmission power and/or bandwidth. Utility-based RRA is a flexible tool that is expected to be able to enhance user satisfaction in next generation of cellular networks [131].
Appendix D

Constrained Optimization

D.1 A General Problem Formulation

In most cases, the resource-allocation problems in wireless networks can be formulated as constrained optimization problems, that can be optimized according to both, the network or the individual point of view. The general formulation can be represented as [103, 135]

$$\begin{align*}
\text{minimize} & \quad f(x) \\
\text{subject to} & \quad g_i(x) \leq 0, \quad \text{for } i = 1, \ldots, m \\
& \quad h_j(x) = 0, \quad \text{for } j = 1, \ldots, l
\end{align*}$$

(D.1)

(D.2)

where $f(x)$ is the optimization objective function or utility function that represents the performance criteria or cost. $x$ stands for the parameter vector for optimizing the resource allocation and $Q$ represent its feasible range. Here $g_i(x)$ and $h_j(x)$ denote the inequality and equality constraints, respectively. Then the optimization problem is processed for finding the solution $\bar{x} \in Q$ which achieves all equality and inequality constraints. Where, for an optimal solution, this condition shall hold, $f(\bar{x}) \leq f(x), \forall x \in Q [135]$.

When the objective function and the constraints are all linear functions of the optimizing variable $x$, then the formulated problem in (D.1) - (D.2) is known as a linear optimization problem or linear program. One important property of such problem is that it has a global optimal point that is easy
to find using linear programming. However, most of the practical scenarios in wireless networking and resource allocation are formulated as nonlinear problems. As a result, it is hard to model and deal with these practical problems using linear programming approaches. If either the objective function or the constraint functions are nonlinear, then the optimization problem formulated in (D.1) - (D.2) is said to be a nonlinear optimization problem or nonlinear program. Generally speaking, nonlinear problems have multiple local optima which makes it a hard task to find the global optimum. In addition, if there are integer sets in the feasible set $\mathcal{Q}$, then the formulated problem in (D.1) - (D.2) is known to be an integer program [103, 135].

The physical meanings of the optimizing parameters, the objective function and the constraints in (D.1) - (D.2) varies from one layer to another. For instance, the optimizing variables in the physical layer could represent the transmitted power, modulation level, channel-coding rate, and channel/code selection. While the objective functions in this layer could represent the minimization of the overall power, maximizing the throughput, maximizing the EE (rate per joule), or minimizing the BER. On the other hand, the constraint in this layer can be applied to restrict the maximum transmitted power, available modulation constellation, available channel-coding rate, and limited energy.

In the media access control (MAC) layer, the optimizing variables might represent the transmission time/frequency, service rate, priorities for transmission. The objective function in the MAC layer might stand for maximizing the overall throughput, minimizing the buffer overflow probability, minimizing the delay. The constraint of this layer could be applied to contentions, limited time/frequency slot, limited information about other mobile users.

In the network layer, on the other hand, the optimizing variables could represent the route selection, routing cost. The objective function of the optimization problems of this layer can represent the minimum cost, maximum profit. Moreover, the constraints in this layer might set to restrict the maximum number of hops, security concerns.

Finally, in the application layer, the optimization variable might represent the source-coding rate, buffer priority, packet arrival rate. The objective function in this layer can denote the distortion minimization, delay minimization. The constraints in this layer can be applied to restrict the transmission of the base-layer, limited source rate, strict delay requirement.
Once the constrained resource allocation optimization problem is formulated, we need to reach the solutions. One of the most effective methods to find a closed-form solution for constrained optimization is the Lagrangian method. Next, we explain briefly the steps to get the closed-form solution using this method [135].

D.2 Problem Solution Using Lagrange Dual Method

Represent the formulated problem in (D.1) - (D.2) as a Lagrangian multiplier function $J$ as follows [135]

$$J = f(x) + \sum_{i=1}^{m} \lambda_{i}g_{i}(x) + \sum_{j=1}^{l} \mu_{j}h_{j}(x)$$  \hspace{1cm} (D.3)

where $\lambda_{i}$ and $\mu_{j}$ represent the Lagrangian multipliers. Next, differentiate the obtained form $J$ with respect to the optimizing variable $x$, then set the resulted expression to zero as

$$\frac{dJ}{dx} = 0$$  \hspace{1cm} (D.4)

After that, solve the expression in (D.4) with respect to $\lambda_{i}$ and $\mu_{j}$. Then use the obtained equivalent expressions of $\lambda_{i}$ and $\mu_{j}$ to replace them in the constraints to get the optimal value of $x$. It is worth mentioning that the main difficulty in this procedure lies where the closed-form solution are obtained for the Lagrangian multipliers. In some cases, mathematical approximations and tricks might help to obtain the closed-form solutions.