

Université du Québec
Institut national de la recherche scientifique
Centre Énergie, Matériaux et Télécommunications

**ALGORITHMES DE GESTION DES RESSOURCES RADIO POUR LA
COMMUNICATION PÉRIPHÉRIQUE À PÉRIPHÉRIQUE (D2D) DANS
LES RÉSEAUX CELLULAIRES SANS FIL**

Par

Tuong Duc Hoang

Thèse présentée pour l'obtention du grade de
Doctorat en philosophie, Ph.D.
en télécommunications

Jury d'évaluation

Examineur externe	Prof. Wessam Ajib <i>Université du Québec à Montréal</i>
	Prof. Chadi Assi <i>Concordia University</i>
Examineur interne	Prof. André Girard <i>INRS-ÉMT</i>
Directeur de recherche	Prof. Long Bao Le <i>INRS-ÉMT</i>

Acknowledgments

I would like to gratefully acknowledge and express a sincere thank to my supervisor, Professor Long Bao Le for giving me the opportunity to pursue doctoral study at INRS-ÉMT, University of Québec. From the very first day, he has always pointed me in good research directions and encouraged me to pursue them to concrete results. Without his constant support, encouragement, and availability, this Ph.D. dissertation could not have been finished. As well, I would like to sincerely thank my co-supervisor, Prof. Tho Le-Ngoc from McGill University, who provided me with useful guidance and advice for my work. Without his precious support, I would not have been possible to conduct this study.

I would like to express my gratitude to other members of my Ph.D. committee – Professor André Girard of INRS-ÉMT, University of Québec who has regularly reviewed and constructively commented on the progress of my doctoral study. I would also like to thank Professor Wessam Ajib of Université du Québec à Montréal and Professor Chadi Assi of Concordia University for serving as the external examiner to my Ph.D. dissertation.

I would like to thank the NSERC-PERSWADE program and Dr. Benoit Pelletier for providing me a research internship at InterDigital, Ltée, Canada. His valuable guidance in 3GPP standards and vehicle-to-vehicle (V2V) system level simulation has helped me to supplement the practical aspects of my research. My gratitude is also delivered to all lab-mates for the magnificent and unforgettable time at the Networks and Cyber Physical System Lab (NECPHY-Lab). Special thanks to Dr. Hieu Nguyen, Dr. Vu Ha, and Duong Nguyen for your helpful discussions and advice. Additionally, many thanks to Achref for helping me with the French translation to my Ph.D. dissertation.

Most importantly, my thesis could not have been completed without the endless encouragement and support from my family and my in-laws. Their love has always been a source of motivation for me in difficult times. Lastly, I am forever grateful to my wife, Nguyễn Thanh Huyền, whose unconditional love, encouragement and sacrifice have allowed me to finish the long journey of the Ph.D. program. I hope that this accomplishment makes her proud.

Abstract

Device-to-device (D2D) communication has been recently proposed as an important technology for the next-generation wireless cellular system, which promises to significantly improve the system spectrum-efficiency, and energy-efficiency by exploiting the advantages of proximity communication. However, many challenges must be resolved to enable efficient integration of D2D communication into the cellular networks. The overall objective of this doctoral research is to develop novel and efficient resource allocation algorithms for D2D communications.

Toward this end, we investigate three key design issues to support the harmonious coexistence of D2D and existing cellular communications, namely spectrum and energy-efficient resource allocation for single hop D2D communication, mode selection and resource allocation for relay-based D2D communication, and joint scheduling and resource allocation for D2D communication. These designs have resulted in several novel contributions, which can be summarized as follows.

First, we propose the spectrum-efficient resource allocation design for single hop D2D communication in the cellular networks, which is presented in Chapter 5. In particular, we present a resource allocation model which allows dynamic power allocation and subchannel assignment for both cellular and D2D links. It is then demonstrated that the proposed algorithm can improve the system spectrum-efficiency significantly in comparison with existing D2D resource allocation algorithms.

Second, we develop a general energy-efficient resource allocation framework for single hop D2D communication in cellular networks which targets to maximize the minimum weighted energy-efficiency (EE) of D2D links while maintaining the minimum required data rates of the cellular links. The research outcomes of this study are presented in Chapter 6. Particularly, we propose a low-complexity power control and subchannel allocation algorithm, which can approach the optimal solution of the underlying resource allocation problem. We also present the distributed implementation for the proposed algorithm, which helps reduce the computation burden for the BS.

Third, we study the resource allocation problem for relay-based D2D communications, which is covered in Chapter 7. The proposed design allows D2D links to dynamically choose either the direct or relay mode. We then propose an efficient mode selection and resource allocation algorithm which optimizes the system spectrum-efficiency. We show that the proposed algorithm can dramatically outperform the conventional resource allocation schemes.

Finally, we consider the joint scheduling and resource allocation design for D2D communication in the cellular networks, which is described in Chapter 8. The proposed design framework allows to dynamically select the set of scheduled D2D links and optimize the system spectrum-efficiency. Toward this end, we develop a monotonic-based algorithm which asymptotically achieves the optimal solution. We then propose a low-complexity algorithm, which can perform much better than the conventional ones and approach the optimal solution.

Contents

Acknowledgments	iii
Abstract	v
Contents	vii
List of Figures	ix
List of Tables	xi
List of Algorithms	xiii
List of Abbreviations	xv
1 Extended Summary	1
1.1 Background and Motivation	1
1.2 Research Contributions	4
1.2.1 Resource Allocation for D2D Communication Underlaid Cellular Networks Using Graph-based Approach	4
1.2.1.1 System Model	4
1.2.1.2 Optimal Power Allocation Algorithm	6
1.2.1.3 Graph-based Iterative Rounding Algorithm	7
1.2.1.4 Numerical Results	8
1.2.2 Energy-Efficient Resource Allocation for D2D Communications in Cellular Networks	10
1.2.2.1 System Model	11
1.2.2.2 Problem Transformation	12
1.2.2.3 Dual-Based Algorithm	13
1.2.2.4 Numerical Results	16
1.2.3 Joint Mode Selection and Resource Allocation for Relay-based D2D Communications	17
1.2.3.1 System Model	17
1.2.3.2 Problem Formulation	19
1.2.3.3 Power Allocation	20
1.2.3.4 Joint Mode Selection and RG Allocation	21
1.2.3.5 Numerical Results	22

1.2.4	Joint Prioritized Scheduling and Resource Allocation for OFDMA-based Wireless Networks	23
1.2.4.1	System Model	23
1.2.4.2	Monotonic Based Optimal Approaching (MBOA) Algorithm	25
1.2.4.3	Numerical Results	27
1.2.5	Concluding Remarks	28
2	Résumé Long	29
2.1	Contexte et motivation	29
2.2	Contributions de la recherche	32
2.2.1	Affectation des ressources pour la communication D2D Réseaux cellulaires sous-appliqués à l'aide de l'approche graphique	32
2.2.1.1	Modèle du système	33
2.2.1.2	Algorithme optimal d'allocation de puissance	34
2.2.1.3	Algorithme d'arrondi itératif basé sur les graphes	36
2.2.1.4	Résultats numériques	38
2.2.2	Allocation de ressources éconergétiques pour les communications D2D dans les réseaux cellulaires	40
2.2.2.1	Modèle du système	40
2.2.2.2	Transformation du Problème	42
2.2.2.3	Algorithme basé sur le dual	43
2.2.2.4	Résultats numériques	46
2.2.3	Sélection du mode et allocation de la ressource conjointes pour les communications D2D relais	47
2.2.3.1	Modèle du système	48
2.2.3.2	Formulation du problème	49
2.2.3.3	Allocation de puissance	51
2.2.3.4	Conjointe Sélection du mode et allocation RG	52
2.2.3.5	Résultats numériques	52
2.2.4	La conjointe Planification partagée et allocation de ressources pour les réseaux sans fil basés sur OFDMA	53
2.2.4.1	Modèle du système	54
2.2.4.2	Algorithme d'approche optimale monotonique (MBOA)	56
2.2.4.3	Résultats numériques	58
2.2.5	Remarques finales	58
3	Introduction	61
3.1	Background and Motivation	61
3.2	Research Challenges	64
3.2.1	Challenges in Resource Allocation for Single Hop D2D Communications	65
3.2.2	Challenges in Relay-based D2D Communication	66
3.2.3	Challenges in Scheduling and Resource Allocation for D2D Communication	67
3.3	Literature Review	68
3.3.1	Resource Allocation for Single Hop D2D Communication	68
3.3.2	Resource Allocation for Relay-based D2D Communication in Cellular Networks	69

3.3.3	Scheduling and Resource Allocation for D2D Communication in Cellular Networks	70
3.4	Research Objectives and Contributions	71
3.5	Dissertation Outline	73
4	Background	75
4.1	Mathematical Optimization	75
4.1.1	Basic Concepts	75
4.1.2	Convex Optimization Problem	76
4.1.2.1	Definition	76
4.1.2.2	Lagrange Dual Function and Lagrange Dual Problem	77
4.1.2.3	Karush-Kuhn-Tucker Conditions	77
4.1.3	Monotonic Optimization	78
4.1.3.1	Polyblock Approximation Algorithm	80
4.1.4	Branch and Bound Algorithm	80
4.2	Resource Allocation for OFDMA-based System	82
4.2.1	System Model	83
4.2.2	System Utilities	83
4.2.3	Resource Allocation in Interference-free OFDMA System	84
4.2.4	Resource Allocation for Interference-Limited OFDMA-Based System	85
4.3	Summary	86
5	Resource Allocation for D2D Communication Underlaid Cellular Networks Using Graph-based Approach	87
5.1	Abstract	87
5.2	Introduction	88
5.2.1	Related Works	91
5.2.2	Contributions and Novelty of the Current Work	93
5.3	System Model and Problem Formulation	94
5.3.1	System Model	94
5.3.2	Problem Formulation	95
5.4	Optimal Power Allocation Algorithm	97
5.5	Subband Assignment Algorithms	100
5.5.1	Graph-based Resource Allocation Formulation	100
5.5.1.1	Graph-based Model	100
5.5.1.2	Subband Assignment Problem	101
5.5.2	Iterative Rounding Algorithm	103
5.5.2.1	Outline of Design Principals	103
5.5.2.2	Resource Allocation Algorithm	104
5.5.2.3	Properties of Iterative Rounding Algorithm	107
5.5.3	Branch and Bound Algorithm	108
5.5.3.1	Branching	108
5.5.3.2	Bounding	109
5.6	Complexity Analysis and Extensions	111
5.6.1	Complexity Analysis	111
5.6.2	Further Extensions	111
5.7	Numerical Results	113

5.7.1	Simulation Setting	113
5.7.2	Numerical Results for Scenario I	113
5.7.3	Numerical Results for Scenario II	117
5.8	Conclusion	120
5.9	Appendices	122
5.9.1	Proof of Proposition 5.1	122
5.9.2	Proof of Proposition 5.3	123
5.9.3	Proof of proposition 5.4	123
5.9.4	Proof of proposition 5.5	126
6	Energy-Efficient Resource Allocation for D2D Communications in Cellular Networks	127
6.1	Abstract	127
6.2	Introduction	128
6.3	System Model and Problem Formulation	131
6.3.1	System Model	131
6.3.2	Problem Formulation	133
6.4	Problem Transformation	134
6.5	Dual-Based Algorithm	136
6.5.1	Algorithm Development	136
6.5.2	Distributed Implementation with Limited Message Passing	141
6.6	Relaxation-Based Algorithms	143
6.6.1	Optimal Branch-and-Bound Algorithm	144
6.6.1.1	Upper-bound Calculation	145
6.6.1.2	Lower-bound Calculation	147
6.6.2	Relaxation-Based Rounding Algorithm	147
6.7	Complexity Analysis	148
6.8	Numerical Results	149
6.9	Conclusions	156
6.10	Appendices	157
6.10.1	Proof of Proposition 6.1	157
6.10.2	Proof of Theorem 6.1	158
6.10.3	Proof of Proposition 6.2	159
6.10.4	Proof of Proposition 6.3	160
6.10.5	Proof of Proposition 6.4	162
7	Joint Mode Selection and Resource Allocation for Relay-based D2D Communications	165
7.1	Abstract	165
7.2	Introduction	166
7.3	System Model and Problem Formulation	167
7.3.1	System Model	167
7.3.2	D2D Communication Modes	168
7.3.2.1	Direct Mode	168
7.3.2.2	Relay Mode	168
7.3.3	Problem Formulation	169
7.4	Solution Approach	170

7.4.1	Power allocation	170
7.4.2	Joint Mode Selection and RG Allocation	171
7.5	Power Allocation for Relay Mode	172
7.5.1	Case 1: $p_c = P_c^m$	173
7.5.2	Case 2: $p_d = P_d^m$ or $p_r = P_r^m$	174
7.6	Numerical Results	176
7.7	Conclusion	179
8	Joint Prioritized Scheduling and Resource Allocation for OFDMA-based Wireless Networks	181
8.1	Abstract	181
8.2	Introduction	182
8.2.1	Related Works	183
8.2.2	Contributions and Novelty of the Current Work	185
8.3	System Model	186
8.4	Problem Formulation and Transformation	187
8.4.1	Problem Formulation	187
8.4.2	Problem Transformation	189
8.5	Monotonic Based Optimal Approaching (MBOA) Algorithm	190
8.5.1	Introduction to Monotonic Optimization	192
8.5.2	Problem Transformation	193
8.5.3	Algorithm Design	194
8.5.4	Analysis of the MBOA Algorithm	196
8.6	Iterative Convex Approximation Algorithm	198
8.6.1	Power Allocation	199
8.6.2	Link Removal Procedure and Computational Complexity Analysis	201
8.7	Practical Implementation	201
8.7.1	Channel state information estimation and feedback phase	202
8.7.2	Algorithm execution phase	202
8.8	Numerical Results	204
8.8.1	Numerical Studies in the Imperfect CSI scenarios	205
8.8.2	Numerical Studies in the Perfect CSI scenarios	209
8.9	Conclusion	213
8.10	Appendices	214
8.10.1	Proof of Proposition 8.1	214
8.10.2	Proof of Proposition 8.2	214
8.10.3	Proof of Proposition 8.5	215
8.10.4	Proof of Theorem 8.1	215
9	Conclusions and Future Research Directions	217
9.1	Major Research Contributions	217
9.2	Further Research Directions	218
9.2.1	Multi-cell D2D Communication	219
9.2.2	MIMO based D2D Communication	219
9.2.3	Full-Duplex based D2D Communication	219
9.2.4	Robust Resource Allocation for D2D Communication in Imperfect CSI Scenarios	220

9.2.5	Scheduling and Resource Allocation for Vehicular-to-Vehicular (V2V) Communication	220
9.3	List of Publications	221
9.3.1	Journals	221
9.3.2	Conferences	221
	Références	223

List of Figures

1.1	D2D communication in cellular networks	3
1.2	The flowchart of Iterative Rounding Algorithm	9
1.3	Sum rate versus minimum rate of each cellular link	10
1.4	System sum rate versus K_c as $K_d = 30$	10
1.5	Minimum EE of D2D links versus D2D link distance	17
1.6	Minimum EE of D2D links versus number of D2D links	17
1.7	Sum rate versus d_{\max}	22
1.8	Sum rate versus R_c^{\min}	22
1.9	Number of admitted D2D links versus R_c^{\min}	27
1.10	Sum rate of D2D links versus minimum rate of cellular link R_c^{\min}	27
2.1	Communication D2D dans les réseaux cellulaires	31
2.2	L'organigramme de l'algorithme d'arrondi itératif	38
2.3	Taux de somme par rapport au taux minimum de chaque lien cellulaire	39
2.4	Taux de somme du système par rapport à K_c comme $K_d = 30$	39
2.5	EE minimum de liens D2D par rapport à la distance de liaison D2D	47
2.6	EE minimum des liens D2D par rapport au nombre de liens D2D	47
2.7	Taux de somme par rapport contre d_{\max}	53
2.8	Taux de somme par rapport contre R_c^{\min}	53
2.9	Nombre de liens D2D admis par rapport à R_c^{\min}	59
2.10	Taux de somme des liens D2D par rapport au taux minimum de liaison cellulaire R_c^{\min}	59
3.1	D2D communication in cellular networks	62
3.2	D2D communication in cellular networks	63
3.3	D2D communication in cellular networks	72
4.1	The procedure of polyblock approximation algorithm	79
5.1	Hyper-edge and vertex presentation of the subband assignment problem	101
5.2	The flowchart of Iterative Rounding Algorithm	103
5.3	Tree structure of BnB algorithm	109
5.4	Simulation setting	112
5.5	System sum rate versus P_d^{\max} as $K_c = 10$	113
5.6	System sum rate versus P_d^{\max} as $K_c = 20$	114
5.7	Sum rate versus minimum rate of each cellular link	115
5.8	System sum rate versus K_c as $K_d = 30$	116

5.9	System sum rate versus K_c as $K_d = 10$	117
5.10	Number of active D2D links versus K_c as $K_d = 30$	117
5.11	System sum rate versus K_d as $K_c = 10$	118
5.12	Number of active D2D links versus K_d as $K_c = 10$	118
5.13	Sum rate versus K_c (Comparing with [1])	120
5.14	Sum rate versus K_c (Comparing with [2])	120
5.15	Sum rate versus K_d (Comparing with [1])	121
5.16	Sum rate versus K_d (Comparing with [2])	121
6.1	Simulation setting	150
6.2	Minimum EE of D2D links versus d_{\max} for $L = 2$	152
6.3	Minimum EE of D2D links versus minimum rate of cellular links for $L = 2$	152
6.4	Convergence behavior of dual-based Algorithm 6.2	153
6.5	Convergence behavior of the relaxation phase of RBR Algorithm 6.4	153
6.6	Minimum EE of D2D links versus D2D link distance when $L = 4$	154
6.7	Minimum EE of D2D links versus minimum required cellular-link rate with $L = 4$	154
6.8	Minimum EE of D2D links versus circuit power	155
6.9	Minimum EE of D2D links versus the noise power	156
6.10	Minimum EE of D2D links versus number of D2D links	156
7.1	System configuration	176
7.2	Sum rate versus d_{\max}	177
7.3	Sum rate versus R_c^{\min}	178
7.4	Sum rate versus L	178
8.1	$q(x)$ function with different values of Q	191
8.2	Number of admitted D2D links versus number of cellular links K	206
8.3	Sum rate of D2D links versus number of cellular links K	206
8.4	Number of admitted D2D links versus minimum rate of D2D link R_d^{\min}	207
8.5	Sum rate of D2D links versus minimum rate of D2D link R_d^{\min}	207
8.6	Number of admitted D2D links versus K	208
8.7	Sum rate of D2D links versus number of cellular links K	208
8.8	Number of admitted D2D links versus R_c^{\min}	210
8.9	Sum rate of D2D links versus minimum rate of cellular link R_c^{\min}	210
8.10	Number of admitted D2D links versus d_{\max}	211
8.11	Sum rate of D2D links versus maximum distance of each D2D link d_{\max}	211
8.12	Number of admitted D2D links versus number of cellular links K	212
8.13	Sum rate of D2D links versus number of cellular links K	213

List of Tables

- 5.1 Summary of related works and our current work 90
- 5.2 Simulation parameters 112
- 6.1 Simulation parameters 150

List of Algorithms

1.1	Dual-Based Algorithm	14
2.1	Algorithme basé sur le dual	44
4.1	BRANCH AND BOUND ALGORITHM	82
5.1	Iterative Rounding Algorithm	105
5.2	LR ($E, \mathbf{w}(E)$)	105
6.1	General Algorithm	136
6.2	Dual-Based Algorithm	137
6.3	Upper-bound Calculation	146
6.4	Relaxation-Based Rounding Algorithm	149
7.1	Algorithm to solve problem (7.24)	174
8.1	The MBOA Algorithm	197
8.2	Calculate $\Pi_{\mathcal{G}_f}(\mathbf{x}^{(n)})$	197
8.3	ITERATIVE CONVEX APPROXIMATION ALGORITHM	199
8.4	Solve Problem $\mathcal{P}(\mathcal{L}_i)$	201

List of Abbreviations

3GPP	3rd Generation Partnership Project
5G	Fifth-Generation
AC	Admission Control
BnB	Branch and Bound
BS	Base Station
CQI	channel quality index
CRN	Cognitive Radio Network
CSI	Channel State Information
D-R	D2D transmitter to Relay receiver
D2D	Device-to-Device
DC	Different of two Convex functions
DCI	Downlink Control Information
DF	Decode and Forward
DIGA	Dynamic Interference Graph Algorithm
EE	Energy-Efficient/Energy-Efficiency
GP	Geometric Programming
ICA	Iterative Convex Approximation
IoT	Internet of Thing
IP	Integer Programming
KKT	Karush-Kuhn-Tucker
LP	Linear Programming
LRM	Local Ratio Method
LTE	Long Term Evolution

LTE-A	Long Term Evolution Advanced
MBOA	Monotonic Based Optimal Approaching
MIMO	Multiple Input Multiple Output
MINLP	Mixed Integer Nonlinear Programming
NP-Hard	Non-Deterministic Polynomial-time Hard
OFDM	Orthogonal Frequency-Division Multiplexing
OFDMA	Orthogonal Frequency-Division Multiple Access
P2P	Peer-to-Peer
PA	Power Allocation
PDCCH	Physical Downlink Control Channel
PDSCH	Physical Downlink Shared Channel
PSSCH	Physical Sidelink Control Channel
PU	Primary User
QoS	Quality of Service
R-D	Relay transmitter to D2D receiver
RA	Resource Allocation
RBR	Relaxation-based Rounding
RG	Resource Group
RHS	Right Hand Side
s.t.	subject to
SA	Subchannel Allocation
SCA	Successive Convex Approximation
SCI	Sidelink Control Information
SE	Spectrum-Efficient/Spectrum-Efficiency
SINR	Signal to Interference plus Noise Ratio
SNR	Signal to Noise Ratio
SU	Secondary User
UE	User Equipment
V2V	Vehicle to Vehicle

Chapter 1

Extended Summary

1.1 Background and Motivation

Wireless cellular operators have seen the ever increasing demand from high-speed applications and rapidly growing number of connected devices. Toward this end, future wireless networks are expected to deliver much larger capacity and support significantly higher communication rates. Specifically, it is predicted that the 5G wireless system should accommodate 1000-fold increase in the system capacity and 100-fold increase in the data rate of connected devices [3]. Device-to-device (D2D) communication has been recently proposed as an important technology toward achieving these objectives [4] by supporting local traffic through direct communications among mobile devices. In D2D communication, two nearby devices can establish a direct communication link under the control of the cellular base station (BS). Different design aspects of D2D communication such as hardware interface, D2D discovery, and resource allocation, have been investigated in both academic and industry communities to enable D2D communications and support future system scenarios and applications [5].

In general, D2D communication can help significantly improve the system capacity by exploiting the advantages of proximity communication. Thanks to the short communication distance among the nearby devices, robust communication with a high data rate can be established to meet stringent requirements of emerging broadband wireless applications such as video sharing and online gaming. In D2D communication, the proximity devices can communicate directly to each other

instead of bypassing through the BS; hence, transmission delay can be reduced significantly. D2D communication can support not only the traditional local voice and data services but also many emerging D2D based applications such as social-aware networking, video sharing, online gaming, and public safety applications.

Third Generation Partnership Project (3GPP) considers D2D communication as an essential technology for the future of cellular system [6, 7]. The initial standardization of D2D communication was proposed in Release 12 of 3GPP for public safety applications [8]. Moreover, D2D communication plays more important roles in many different scenarios such as massive internet of things (IoT), extreme real-time communications, lifeline communications, ultra-reliable communications, and broadcast-like services [9]. In fact, potential scenarios and applications of D2D communication are still being discussed for 5G system, which is expected to be deployed in 2020.

One major challenge for enabling D2D communication in cellular networks is related to the development of efficient radio resource management techniques. In particular, it is desirable to employ a resource allocation scheme for D2D communication that can exploit the limited resources such as frequency and transmit power efficiently. Different general resource allocation approaches for D2D communication can be summarized as in Fig. 1.1. Specifically, D2D communication can be employed by using the frequency allocated to the cellular spectrum, i.e., in-band D2D, or by using unlicensed spectrum, i.e., out-band D2D [10]. However, in the out-band D2D, the quality of service (QoS) of D2D links would not be guaranteed due to uncontrollable interferences from other sources. In the in-band D2D, D2D links can be operated in either overlaying or underlaying scenarios. In the overlaying scenario, the frequency resource is orthogonally reserved for D2D and cellular links; nevertheless, operating in the overlaying scenario could result in inefficient resource utilization as some frequency resources might be unused due to the unavailability of either D2D or cellular links. Finally, in the underlaying scenario, frequency resources are shared between D2D and cellular links, which can potentially improve the system performance since the frequency resource and spatial diversity of the system can be exploited efficiently.

The potential benefits obtained from the underlaying scenario come with the price of more complicated resource allocation algorithms. In particular, these algorithms are required to efficiently allocate the frequency resources to each link and manage the interference between the links. However, to enable D2D communication in the cellular network, each operator may employ different

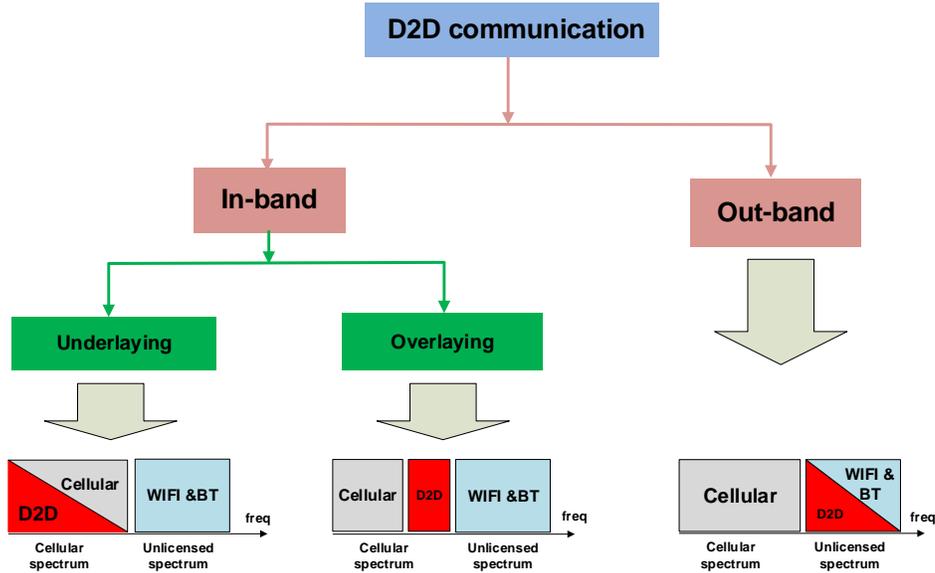


Figure 1.1 – D2D communication in cellular networks

approaches which can be chosen while considering different system models, set of optimization parameters, utility functions, and resource utilization schemes.

From the operator’s perspective, maximization of the system spectrum-efficiency (SE) can be more desirable since the revenue is proportional to the system throughput. Nevertheless, from the mobile users’ viewpoint, optimization of the system energy-efficiency (EE) can be more appropriate because this design can help prolong devices’ battery life. Toward this end, enabling direct D2D communication can be beneficial if the underlying pair of devices is close to each other. However, if two devices are far away from each other, they can establish the communication path with each other through the BS or a relay node. Therefore, resource allocation for relay-based D2D communication is an important design issue to enable D2D communication in cellular networks [11]. Besides resource allocation, scheduling control, which determines the set of links to transmit at a particular time, is also essential in achieving efficient resource utilization. Moreover, cellular and D2D can be treated as primary and secondary links, respectively, where the primary links are prioritized over the secondary links. Hence, joint scheduling and resource allocation design for this prioritized access model is an important design problem.

1.2 Research Contributions

The general objective of my Ph.D research is to develop efficient resource allocation algorithms for D2D communication which contribute to enable efficient integration of the D2D communication in cellular network while not causing severe performance degradation of the existing cellular links. Specifically, our main contributions can be described as follows.

1.2.1 Resource Allocation for D2D Communication Underlaid Cellular Networks Using Graph-based Approach

This contribution focuses on the spectrum-efficient radio resource allocation for D2D communications in cellular networks which simultaneously determines subband assignment and power control for both cellular and D2D links. Specifically, our work makes the following contributions.

- We formulate a resource allocation problem for joint D2D link selection, subband assignment, and power control problem. To solve this problem, we first derive the optimal power allocation for a given subband assignment for one pair of cellular and D2D links. Then the original subband assignment and power allocation is transformed to a subband assignment problem.
- Afterward, we develop a novel Iterative Rounding algorithm to solve the subband assignment problem by using graph theory. Numerical results demonstrate that the Iterative Rounding Algorithm achieves almost the same sum-rate as that attained by the optimal BnB algorithm, which is significantly outperforms the conventional algorithms.

1.2.1.1 System Model

We consider the spectrum sharing problem among multiple D2D and cellular links in the uplink direction. Let $\mathcal{N} = \{1, \dots, N\}$ be the set of subbands in the system.¹ We denote $\mathcal{K}_c = \{1, \dots, K_c\}$ as the set of cellular links, $\mathcal{K}_d = \{K_c + 1, \dots, K_c + K_d\}$ as the set of D2D links, and $\mathcal{K} = \mathcal{K}_c \cup \mathcal{K}_d$ as the set of all communications links.

¹Each subband can be a carrier and sub-channel in the multi-carrier wireless networks (e.g., LTE-based wireless networks).

Let h_{kl}^n be the channel gain from the transmitter of link l to the receiver of link k on subband n . We denote the power allocation and subband assignment vectors as $\mathbf{p} = [p_k^n]_{\forall k \in \mathcal{K}, \forall n \in \mathcal{N}}$ and $\boldsymbol{\rho} = [\rho_k^n]_{\forall k \in \mathcal{K}, \forall n \in \mathcal{N}}$ where p_k^n and ρ_k^n are, respectively, the power allocation and subband assignment of link k on subband n . For convenience, we adopt the following notations: $\mathcal{K}_k \equiv \mathcal{K}_c$ if $k \in \mathcal{K}_c$ and $\mathcal{K}_k \equiv \mathcal{K}_d$ if $k \in \mathcal{K}_d$. The signal to interference plus noise ratio (SINR) achieved by link $k \in \mathcal{K}$ on subband n can be expressed as $\Gamma_k^n(\mathbf{p}^n, \boldsymbol{\rho}^n) = \frac{\rho_k^n p_k^n h_{kk}^n}{\sigma_k^n + \sum_{l \in \mathcal{K} \setminus \mathcal{K}_k} \rho_l^n p_l^n h_{kl}^n}$. Therefore, the achievable rates of link $k \in \mathcal{K}$ on subband n and all the subbands can be expressed as $R_k^n(\mathbf{p}^n, \boldsymbol{\rho}^n) = \log_2(1 + \Gamma_k^n(\mathbf{p}^n, \boldsymbol{\rho}^n))$, and $R_k(\mathbf{p}, \boldsymbol{\rho}) = \sum_{n \in \mathcal{N}} R_k^n(\mathbf{p}^n, \boldsymbol{\rho}^n)$.

We assume that each cellular link or active D2D link is allocated one subband, which is suitable for uplink communications [13]. Our design objective is to maximize the weighted sum-rate of all selected D2D links and cellular links while satisfying the minimum required rates of cellular links and active D2D links. The considered resource allocation problem can now be formulated as

$$\max_{\mathbf{p}, \boldsymbol{\rho}} \mathcal{R} = \sum_{k \in \mathcal{K}_c} \alpha R_k(\mathbf{p}, \boldsymbol{\rho}) + \sum_{k \in \mathcal{K}_d} (1 - \alpha) R_k(\mathbf{p}, \boldsymbol{\rho}) \quad (1.1a)$$

$$\text{s.t. } R_k(\mathbf{p}, \boldsymbol{\rho}) \geq R_k^{\min} \quad \forall k \in \mathcal{K}_c \quad (1.1b)$$

$$R_k(\mathbf{p}, \boldsymbol{\rho}) \geq \mathbf{I}\{\sum_{n \in \mathcal{N}} \rho_k^n = 1\} R_k^{\min} \quad \forall k \in \mathcal{K}_d, \quad (1.1c)$$

$$\sum_{n \in \mathcal{N}} p_k^n \leq P_k^{\max} \quad \forall k \in \mathcal{K} \quad (1.1d)$$

$$\sum_{k \in \mathcal{K}_c} \rho_k^n \leq 1, \quad \sum_{k \in \mathcal{K}_d} \rho_k^n \leq 1 \quad \forall n \in \mathcal{N} \quad (1.1e)$$

$$\sum_{n \in \mathcal{N}} \rho_k^n = 1 \quad \forall k \in \mathcal{K}_c, \quad \sum_{n \in \mathcal{N}} \rho_k^n \leq 1 \quad \forall k \in \mathcal{K}_d \quad (1.1f)$$

$$\rho_k^n \in \{0, 1\} \quad \forall k \in \mathcal{K} \quad \forall n \in \mathcal{N}, \quad (1.1g)$$

where α is a weight parameter that controls spectrum sharing of cellular and D2D links. In problem (1.1), constraints (1.1b) and (1.1c) are to ensure the minimum data rate of cellular and D2D links, and $\mathbf{I}\{A\}$ denotes the indicator function, which equals to 1 if A is true and equals 0, otherwise. Moreover, constraints (1.1e), (1.1f), and (1.1g) are imposed to guarantee that each subband is assigned to one cellular and one D2D link, and each link can exploit one subband. To solve this problem, we first investigate the optimal power allocation solution for a given subband assignment.

Then, the original problem is transformed to a subband assignment problem, which can be solved by the proposed Iterative Rounding Algorithm.

1.2.1.2 Optimal Power Allocation Algorithm

Note that we allow each cellular and active D2D link to use only one subband in the problem formulation (1.1). Therefore, if link $m \in \mathcal{K}$ is allocated subband n exclusively then the optimal power for this link is P_m^{\max} and the corresponding contribution of this link to the objective value is

$$w_m^n \triangleq \begin{cases} \alpha \log_2 \left(1 + \frac{P_m^{\max} h_{mm}^n}{\sigma_m^n} \right) & \text{if } m \in \mathcal{K}_c \\ (1 - \alpha) \log_2 \left(1 + \frac{P_m^{\max} h_{mm}^n}{\sigma_m^n} \right) & \text{if } m \in \mathcal{K}_d. \end{cases} \quad (1.2)$$

However, if cellular link k and D2D link l share subband n then the optimal power allocation must be determined from the following optimization problem

$$\begin{aligned} \max_{p_{Ck}^n, p_{Dl}^n} \quad & w_{kl}^n \triangleq \alpha R_{Ck}^n + (1 - \alpha) R_{Dl}^n \\ \text{s.t.} \quad & R_{Ck}^n \geq R_k^{\min}, R_{Dl}^n \geq R_l^{\min} \\ & p_{Ck}^n \in [0, P_k^{\max}], p_{Dl}^n \in [0, P_l^{\max}], \end{aligned} \quad (1.3)$$

where $R_{Ck}^n = \log_2 \left(1 + \frac{p_{Ck}^n h_{kk}^n}{\sigma_k^n + p_{Dl}^n h_{kl}^n} \right)$ and $R_{Dl}^n = \log_2 \left(1 + \frac{p_{Dl}^n h_{ll}^n}{\sigma_l^n + p_{Ck}^n h_{lk}^n} \right)$.

For this problem, it has been proved in [14] that if the problem is feasible then the optimal transmit powers $\mathbf{P} = (p_{Ck}, p_{Dl})$ have the form $\mathbf{P} \in \{(P_k^{\max}, p_{Dl}), (p_{Ck}, P_l^{\max})\}$. Let us now define the following quantities:

$$P_{Dl}^{(1)} \triangleq \max \left\{ \frac{(2^{R_l^{\min}} - 1)(P_k^{\max} h_{lk}^n + \sigma_l^n)}{h_{ll}^n}, 0 \right\}, P_{Ck}^{(1)} \triangleq \max \left\{ \frac{(2^{R_k^{\min}} - 1)(P_l^{\max} h_{kl}^n + \sigma_k^n)}{h_{kk}^n}, 0 \right\} \quad (1.4)$$

$$P_{Dl}^{(2)} \triangleq \min \left\{ \frac{1}{h_{kl}^n} \left(\frac{P_k^{\max} h_{kk}^n}{2^{R_k^{\min}} - 1} - \sigma_k^n \right), P_l^{\max} \right\}, P_{Ck}^{(2)} \triangleq \min \left\{ \frac{1}{h_{lk}^n} \left(\frac{P_l^{\max} h_{ll}^n}{2^{R_l^{\min}} - 1} - \sigma_l^n \right), P_k^{\max} \right\} \quad (1.5)$$

$$P_{Dl}^{(3)} \triangleq (-B_{Dl} + \sqrt{\Delta_{Dl}})/A_{Dl}, P_{Ck}^{(3)} \triangleq (-B_{Ck} + \sqrt{\Delta_{Ck}})/A_{Ck}, \quad (1.6)$$

Then the optimal power allocation of problem (1.3) is characterized in the following theorem.

Theorem 1.1. *If the problem (1.3) is feasible then its optimal power allocation solution belongs to the set $\mathcal{S}^* \triangleq \mathcal{S}_1 \cup \mathcal{S}_2$, where*

$$\mathcal{S}_1 \triangleq \begin{cases} \{(P_k^{\max}, P_{Dl}^{(1)}), (P_k^{\max}, P_{Dl}^{(2)}), (P_k^{\max}, P_{Dl}^{(3)})\}, & \text{if } P_{Dl}^{(3)} \in [0, P_l^{\max}] \\ \{(P_k^{\max}, P_{Dl}^{(1)}), (P_k^{\max}, P_{Dl}^{(2)})\}, & \text{otherwise} \end{cases} \quad (1.7)$$

$$\mathcal{S}_2 \triangleq \begin{cases} \{(P_{Ck}^{(1)}, P_l^{\max}), (P_{Ck}^{(2)}, P_l^{\max}), (P_{Ck}^{(3)}, P_l^{\max})\}, & \text{if } P_{Ck}^{(3)} \in [0, P_k^{\max}] \\ \{(P_{Ck}^{(1)}, P_l^{\max}), (P_{Ck}^{(2)}, P_l^{\max})\}, & \text{otherwise.} \end{cases} \quad (1.8)$$

As \mathcal{S}^* contains at most 6 possible power allocation solutions, we can determine the optimal solution by examining all potential solutions in \mathcal{S}^* easily.

1.2.1.3 Graph-based Iterative Rounding Algorithm

Since optimal power allocation for a given subband assignment can be determined as in the previous section, problem (1.1) can be transformed to the subband assignment problem. We propose to solve the subband assignment problem by using the graph-based approach where each link or subband can be modeled as a vertex, and one subband assignment corresponds to one hyper-edge in the graph. Denote V^0 and E^0 as the set of vertexes and hyper-edges corresponding to the original subband assignment problem. For a set of edges E , let $V(E)$, $V_c(E)$, $V_d(E)$, and $V_n(E)$ be the sets of vertexes, actual cellular links, D2D links, and subbands associated with E , respectively. To describe the subband assignment decision, we introduce a binary variable x_e where $x_e=1$ means edge e is activated and $x_e=0$, otherwise. Moreover, let \mathbf{x} denote the vector whose elements are subband assignment variables x_e associated with all possible edges. The subband assignment problem, i.e., $\mathbf{IP}(V, E)$, is described as follows

$$\begin{aligned} \max_{x_e} \quad & \mathcal{R} = \sum_{e \in E} w_e x_e \\ \text{s.t.} \quad & C1: D(v, E) = 1 \quad \forall v \in V \cap K_C \\ & C2: D(v, E) \leq 1 \quad \forall v \in V_d(E) \cup V_n(E) \\ & C3: x_e \in \{0, 1\} \quad \forall e \in E, \end{aligned} \quad (1.9)$$

where $D(v, E) \triangleq \sum_{e \in E(v)} x_e$ denotes the degree of vertex v in the set of edges E associated with \mathbf{x} .

Note that $\mathbf{IP}(V^0, E^0)$ is corresponding to the original subband assignment problem. Denote $\mathbf{LP}(V, E)$ as the linear relaxation version of problem $\mathbf{IP}(V, E)$. It can be observed that problem $\mathbf{LP}(V^0, E^0)$ can be solved easily by standard optimization solutions [15]. However, solving problem $\mathbf{LP}(V^0, E^0)$ often results in fractional values for some edges e ($0 < x_e < 1$). To address this issue, we propose an Iterative Rounding Algorithm in which we solve a linear relaxation problem and perform suitable rounding for fractional variables in each iteration.

The Iterative Rounding Algorithm performs the following operations in three phases of each iteration t . In phase 1, it solves the linear relaxation problem for inactive links and available subbands corresponding to the graph with the set of vertexes $V^{(t)}$ and the set of edges $E^{(t)}$, which results in two sets of variables equal to fractional values ($0 < x_e < 1$) and one ($x_e = 1$), namely $E_a^{(t)}$ and $E_u^{(t)}$, respectively. We then arrange the edges in the set $E_u^{(t)}$ with fractional subband assignment variables in phase 2 based on which we employ the Local Ratio Method in phase 3 to determine the set of additional subband assignments $E_g^{(t)}$. Phases 2 and 3 have been indeed appropriately designed to minimize the performance loss due to rounding of the fractional subband assignment variables. The edges in $E_a^{(t)} \cup E_g^{(t)}$ will be used to perform the corresponding subband assignments for cellular and/or D2D links in each iteration. Finally, we update the set of available subbands and inactive links and go back to phase 1 of the next iteration until convergence. The main operations of the algorithm are illustrated in Fig. 1.2. We present the performance of Iterative Rounding Algorithm in the following theorem.

Theorem 1.2. *A feasible solution of the original subband allocation problem $\mathbf{IP}(V^0, E^0)$, offered by Iterative Rounding Algorithm, achieves at least half of the optimal objective value of the linear relaxation problem $\mathbf{LP}(V^0, E^0)$.*

This theorem implies that we can always guarantee that Iterative Rounding Algorithm achieves at least 1/2 optimal objective value. This is true even if the solution of problem $\mathbf{LP}(V^0, E^0)$ corresponds all fractional edges.

1.2.1.4 Numerical Results

We consider the simulation where there is a single BS with the coverage area of 500m serving $K_c = 20$ randomly distributed cellular users. Moreover, there are $N = 25$ subbands, which are

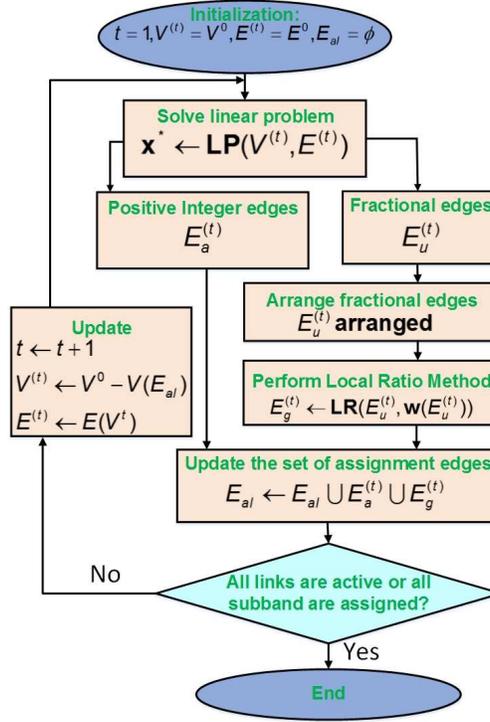


Figure 1.2 – The flowchart of Iterative Rounding Algorithm

shared by $K_c = 20$ cellular links and $K_d = 30$ D2D links unless stated otherwise. We compare the performance of our proposed algorithms with conventional algorithms developed for scenario I in [16] and [17]. The first conventional algorithm, referred to as the optimization-based conventional algorithm, in contrast the work [17] adopted a game-based approach.

In Fig. 1.3, we demonstrate the system sum-rate versus the minimum required rate of cellular links R_c^{\min} . It can be seen that the system sum-rate reaches the maximum value as $R_c^{\min} = 0$. This is because when as $R_c^{\min} = 0$, D2D links have more advantages than cellular links in accessing good subbands thanks to the short-range of D2D links. Hence, the rate of D2D links become higher for the smaller minimum required rate of each cellular link. It can also be observed that the system sum-rate decreases significantly as R_c^{\min} increases from zero before getting saturated at fixed value.

Fig. 1.4 shows the variations of the system sum-rate with the number of cellular links K_c as we fix $K_d = 30$. This figure demonstrates that the system sum-rate decreases with the number of cellular links. In fact, as K_c increases, the number of active D2D links is reduced, which results in the decrease in the system sum-rate. However, when K_c is sufficiently large, increasing K_c leads

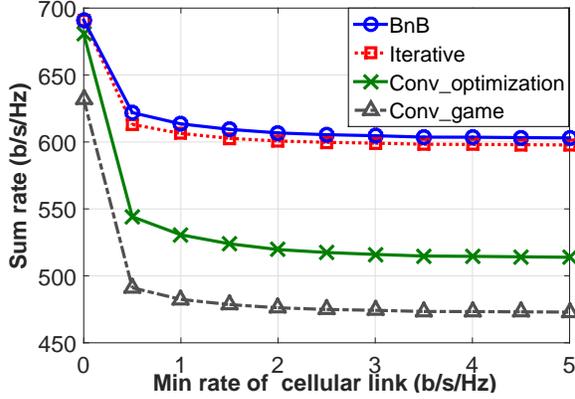


Figure 1.3 – Sum rate versus minimum rate of each cellular link

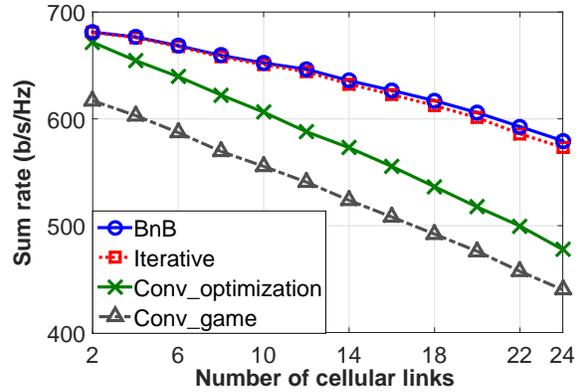


Figure 1.4 – System sum rate versus K_c as $K_d = 30$

to the scenario where active D2D links must share subbands with cellular links and the number of active D2D links decreases.

1.2.2 Energy-Efficient Resource Allocation for D2D Communications in Cellular Networks

In this contribution, we study the joint subchannel and power allocation that maximizes the minimum weighted EE of D2D links and guarantees the minimum data rates of cellular links. Specifically, we make the following contributions.

- We formulate a general energy-efficient resource allocation problem considering multiple cellular and D2D links where each D2D link can reuse the spectral resources of multiple cellular links. We first characterize the optimal power allocation solution for a cellular link as a function of the optimal power of the co-channel D2D link. Based on this result, we transform the original RA problem into the RA problem for only D2D links.
- We propose the dual-based algorithm that solves the resource allocation problem in the dual domain. Particularly, we adopt the max-min fractional programming technique to iteratively transform the resource allocation problem into a Mixed Integer Nonlinear Programming (MINLP) problem. Then, we solve the underlying MINLP problem by using the dual decomposition approach.
- Extensive numerical results are presented to evaluate the performance of the developed algorithms. Specifically, it is shown that the objective values achieved by the dual-based algorithm

is close to that of the optimal BnB algorithm and significantly higher than that of conventional algorithm and the spectrum-efficient resource allocation design.

1.2.2.1 System Model

We consider uplink resource allocation scenario where cellular links share the same spectrum with multiple D2D links in a single macro-cell system. We assume that K uplink cellular links in a set $\mathcal{K} = \{1, \dots, K\}$ occupying K orthogonal subchannels in the set $\mathcal{N} = \{1, \dots, K\}$ in the considered cell. Moreover, we assume that the set $\mathcal{L} = \{1, \dots, L\}$ of D2D links transmits data using the same set of subchannels.²

We present the allocated power vectors as $\mathbf{p} = [\mathbf{p}_C, \mathbf{p}_D]$ for all the links, where $\mathbf{p}_C = [p_{Ck}^k]_{\forall k \in \mathcal{K}}$, $\mathbf{p}_D = [p_{Dl}^k]_{\forall l \in \mathcal{L}, \forall k \in \mathcal{K}}$. p_{Ck}^k and p_{Dl}^k denote the allocated transmit powers on subchannel k of cellular link $k \in \mathcal{K}$ and D2D link $l \in \mathcal{L}$, respectively. Then, the SINR achieved by cellular link k and D2D link l on subchannel k can be expressed, respectively, as $\Gamma_{Ck}^k(\mathbf{p}, \boldsymbol{\rho}) = \frac{p_{Ck}^k h_{kk}^k}{\sigma_k^k + \sum_{l \in \mathcal{L}} \rho_l^k p_{Dl}^k h_{kl}^k}$, and $\Gamma_{Dl}^k(\mathbf{p}, \boldsymbol{\rho}) = \frac{\rho_l^k p_{Dl}^k h_{ll}^k}{\sigma_l^k + p_{Ck}^k h_{lk}^k}$. The data rates of cellular link $k \in \mathcal{K}$ on its subchannel k , D2D link $l \in \mathcal{L}$ on subchannel k , and D2D link l on all the subchannels can be calculated as

$$R_{Ck}^k(\mathbf{p}, \boldsymbol{\rho}) = \log_2 \left(1 + \Gamma_{Ck}^k(\mathbf{p}, \boldsymbol{\rho}) \right) \quad (1.10)$$

$$R_{Dl}^k(\mathbf{p}, \boldsymbol{\rho}) = \log_2 \left(1 + \Gamma_{Dl}^k(\mathbf{p}, \boldsymbol{\rho}) \right) \quad (1.11)$$

$$R_{Dl}(\mathbf{p}, \boldsymbol{\rho}) = \sum_{k \in \mathcal{K}} \rho_l^k R_{Dl}^k(\mathbf{p}, \boldsymbol{\rho}). \quad (1.12)$$

We assume that the total *consumed* power of D2D link l can be expressed as [18, 19]

$$P_{Dl}^{\text{total}} = 2P_0^l + \alpha_l \sum_{k \in \mathcal{K}} \rho_l^k p_{Dl}^k, \quad (1.13)$$

where $2P_0^l$ represents the fixed circuit power of both transmitter and receiver of D2D link l , and $\alpha_l > 1$ is a factor accounting for the transmit amplifier efficiency and feeder losses.

The objective of our resource allocation design is to maximize the minimum weighted EE of the D2D links while the minimum data rate of cellular links are guaranteed. Therefore, this design can

²The considered orthogonal subchannels can be sub-carriers or sub-channels in the OFDMA system or simply channels in the FDMA system.

be formulated as the following energy-efficient resource allocation problem to attain the max-min fairness in weighed EE for D2D links

$$\max_{\mathbf{p}, \boldsymbol{\rho}} \min_{l \in \mathcal{L}} \frac{w_l R_{Dl}(\mathbf{p}, \boldsymbol{\rho})}{P_{Dl}^{\text{total}}} \quad (1.14a)$$

$$\text{s.t. } R_{Ck}^k(\mathbf{p}, \boldsymbol{\rho}) \geq R_{Ck}^{\min}, \forall k \in \mathcal{K} \quad (1.14b)$$

$$p_{Ck}^k \leq P_{Ck}^{\max}, \forall k \in \mathcal{K}, \quad (1.14c)$$

$$\sum_{k \in \mathcal{K}} \rho_l^k p_{Dl}^k \leq P_{Dl}^{\max}, \forall l \in \mathcal{L}, \quad (1.14d)$$

$$\sum_{l \in \mathcal{L}} \rho_l^k \leq 1, \forall k \in \mathcal{N} \quad (1.14e)$$

$$\rho_l^k \in \{0, 1\}, \forall k \in \mathcal{K}, l \in \mathcal{L}. \quad (1.14f)$$

1.2.2.2 Problem Transformation

To solve problem (1.14), we first describe the optimal power allocation of D2D link $l \in \mathcal{L}$ on subchannel $k \in \mathcal{N}$ in the following proposition.

Proposition 1.1. *If D2D link $l \in \mathcal{L}$ is allowed to reuse subchannel $k \in \mathcal{N}$ of cellular link k , then its power on subchannel k , $p_{Dl}^k = \frac{1}{h_{kl}^k} \left(\frac{p_{Ck}^k h_{kk}^k}{2^{R_{Ck}^{\min}} - 1} - \sigma_k^k \right) \in [0, P_{Dlk}^{\max}]$, where p_{Ck}^k is the power of cellular link k , and $P_{Dlk}^{\max} = \min \left\{ P_{Dl}^{\max}, \frac{1}{h_{kl}^k} \left(\frac{P_{Ck}^{\max} h_{kk}^k}{2^{R_{Ck}^{\min}} - 1} - \sigma_k^k \right) \right\}$.*

From Proposition 1.1, the data rate of D2D link l on subchannel k and all subchannels can be re-written as

$$\hat{R}_{Dl}^k(\mathbf{p}_D, \boldsymbol{\rho}) = \rho_l^k \log_2 \left(1 + \frac{p_{Dl}^k}{a_{kl} + b_{kl} p_{Dl}^k} \right) \quad (1.15)$$

$$\hat{R}_{Dl}(\mathbf{p}_D, \boldsymbol{\rho}) = \sum_{k \in \mathcal{N}} \hat{R}_{Dl}^k(\mathbf{p}_D, \boldsymbol{\rho}), \quad (1.16)$$

where

$$a_{kl} \triangleq \frac{\sigma_l^k}{h_{ll}^k} + \frac{(2^{R_k^{\min}} - 1) h_{lk}^k \sigma_k^k}{h_{kk}^k h_{ll}^k} \quad (1.17)$$

$$b_{kl} \triangleq \frac{(2^{R_k^{\min}} - 1) h_{lk}^k h_{kl}^k}{h_{kk}^k h_{ll}^k}, \quad (1.18)$$

and the allocated transmit power must satisfy

$$p_{Dl}^k \leq P_{Dlk}^{\max}, \forall k \in \mathcal{N}, \forall l \in \mathcal{L}. \quad (1.19)$$

Therefore, problem (1.14) is equivalent to the following

$$\begin{aligned} \max_{(\mathbf{p}_D, \boldsymbol{\rho})} \min_{l \in \mathcal{L}} \frac{w_l \hat{R}_{Dl}(\mathbf{p}_{Dl}, \boldsymbol{\rho})}{P_{Dl}^{\text{total}}(\mathbf{p}_{Dl}, \boldsymbol{\rho})} \\ \text{s.t. } (1.14d), (1.14f), (1.14e), (1.19). \end{aligned} \quad (1.20)$$

In order to solve problem (1.20), we consider the following optimization problem

$$\begin{aligned} \max_{\mathbf{p}_D, \boldsymbol{\rho}} \eta(\zeta, \mathbf{p}_D, \boldsymbol{\rho}) \triangleq \min_{l \in \mathcal{L}} \left[w_l \hat{R}_{Dl}(\mathbf{p}_D, \boldsymbol{\rho}) - \zeta P_{Dl}^{\text{total}}(\mathbf{p}_D, \boldsymbol{\rho}) \right] \\ \text{s.t. } (1.14d), (1.14e), (1.14f), (1.19). \end{aligned} \quad (1.21)$$

Suppose that $\eta^*(\zeta) = \eta(\zeta, \mathbf{p}_D^*, \boldsymbol{\rho}^*)$ where $(\mathbf{p}_D^*, \boldsymbol{\rho}^*)$ is the optimal solution of problem (1.21), and \mathcal{D} denotes the set of feasible solutions of problem (1.20). Then, we can characterize the optimal solution of problem (1.21) in the following theorem, which is adopted from [20].

Theorem 1.3. $\eta^*(\zeta)$ is a decreasing function of ζ . In addition, if we have

$$\begin{aligned} \max_{(\mathbf{p}_D, \boldsymbol{\rho}) \in \mathcal{D}} \min_{l \in \mathcal{L}} [w_l \hat{R}_{Dl}(\mathbf{p}_D, \boldsymbol{\rho}) - \zeta^* P_{Dl}^{\text{total}}(\mathbf{p}_D, \boldsymbol{\rho})] \\ = \min_{l \in \mathcal{L}} [w_l \hat{R}_{Dl}(\mathbf{p}_D^*, \boldsymbol{\rho}^*) - \zeta^* P_{Dl}^{\text{total}}(\mathbf{p}_D^*, \boldsymbol{\rho}^*)] = 0 \end{aligned} \quad (1.22)$$

then $\zeta^* = \min_{l \in \mathcal{L}} \frac{w_l \hat{R}_{Dl}(\mathbf{p}_D^*, \boldsymbol{\rho}^*)}{P_{Dl}^{\text{total}}(\mathbf{p}_D^*, \boldsymbol{\rho}^*)}$ is the optimal solution of (1.20).

Theorem 1.3 allows us to transform a general max-min fractional problem (1.20) to a non-fractional optimization problem with the parameter ζ . In addition, the optimal solution of problem (1.20), ζ^* , can be found if $\eta^*(\zeta^*) = 0$. Since $\eta^*(\zeta)$ is a decreasing function of ζ , it can be seen that ζ^* can be indeed determined by bisection method.

1.2.2.3 Dual-Based Algorithm

Algorithm 1.1. Dual-Based Algorithm

- 1: Initialization: $\zeta^{\max}, \zeta^{\min}$
 - 2: **repeat**
 - 3: Initialization: Choose $\zeta = \frac{1}{2}(\zeta^{\min} + \zeta^{\max})$, $\boldsymbol{\lambda}^{(0)}, \mu_l^{(0)} = \frac{1}{L}$, step size $\theta^{(0)}$, and $\kappa^{(0)}$
 - 4: **repeat**
 - 5: Step 1: For all $k \in \mathcal{K}, l \in \mathcal{L}$, calculate p_{Dl}^{k*} according to (1.29)
 - 6: Step 2: For all $k \in \mathcal{K}$, perform subchannel allocation following (1.30)
 - 7: Step 3: Update dual variables $\boldsymbol{\lambda}$, and $\boldsymbol{\mu}$ by subgradient method.
 - 8: **until** Convergence
 - 9: Output $z^* = \min_{l \in \mathcal{L}} \left[w_l \hat{R}_{Dl}(\mathbf{p}_D^*, \boldsymbol{\rho}^*) - \zeta P_{Dl}^{\text{total}}(\mathbf{p}_D^*, \boldsymbol{\rho}^*) \right]$
 - 10: If $z^* > 0$, $\zeta^{\min} = \zeta$; otherwise $\zeta^{\max} = \zeta$
 - 11: **until** Convergence of ζ
 - 12: Output $\mathbf{p}_{Dl}^*, \boldsymbol{\rho}^*$, and $\zeta^* = \min_{l \in \mathcal{L}} \frac{w_l \hat{R}_{Dl}(\mathbf{p}_{Dl}^*, \boldsymbol{\rho}^*)}{P_{Dl}^{\text{total}}(\mathbf{p}_{Dl}^*, \boldsymbol{\rho}^*)}$
-

In this section, we propose a dual-based algorithm to solve problem (1.20), which is summarized in Algorithm 1.1. The algorithm comprises two iterative loops. In the outer loop, we adopt the max-min fractional programming technique investigated in Theorem 1.3 to attain the optimal value of ζ for problem (1.20). In the inner loop (lines 4-8), we solve problem (1.21) for a given ζ by employing the dual decomposition method.

In the following, we show how to solve problem (1.21) for a given value of ζ . First, it can be observed that problem (1.21) is equivalent to the following problem

$$\begin{aligned}
 & \max_{z, \mathbf{p}_D, \boldsymbol{\rho}} \quad z \\
 & \text{s.t.} \quad w_l \hat{R}_{Dl}(\mathbf{p}_{Dl}, \boldsymbol{\rho}) - \zeta P_{Dl}^{\text{total}}(\mathbf{p}_D, \boldsymbol{\rho}) \geq z, \forall l \in \mathcal{L} \\
 & \quad (1.14d), (1.14e), (1.14f), (1.19).
 \end{aligned} \tag{1.23}$$

To tackle problem (1.23), we consider its Lagrangian as $L_D(\mathbf{p}_D, \boldsymbol{\rho}, z, \zeta, \boldsymbol{\lambda}, \boldsymbol{\mu})$

$= z(1 - \sum_{l \in \mathcal{L}} \mu_l) + \sum_{l \in \mathcal{L}} \mu_l \left[w_l \hat{R}_{Dl}(\mathbf{p}_D, \boldsymbol{\rho}) - \zeta P_{Dl}^{\text{total}}(\mathbf{p}_D, \boldsymbol{\rho}) \right] + \sum_{l \in \mathcal{L}} \lambda_l (P_{Dl}^{\max} - \sum_{k \in \mathcal{N}} \rho_l^k p_{Dl}^k)$, where $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_L]^T$ and $\boldsymbol{\mu} = [\mu_1, \dots, \mu_L]^T$ represent the Lagrange multipliers. Then, the dual function can be written as

$$\bar{L}_D(\zeta, \boldsymbol{\lambda}, \boldsymbol{\mu}) \triangleq \max_{z, \mathbf{p}_D \in \mathcal{X}, \boldsymbol{\rho} \in \mathcal{C}} L_D(\mathbf{p}_D, \boldsymbol{\rho}, z, \zeta, \boldsymbol{\lambda}, \boldsymbol{\mu}), \tag{1.24}$$

where $\mathcal{X} = \{\mathbf{p}_D | p_{Dl}^k \leq P_{Dlk}^{\max}, \forall k \in \mathcal{N}, \forall l \in \mathcal{L}\}$, and $\mathcal{C} = \{\boldsymbol{\rho} | \sum_{l \in \mathcal{L}} \rho_l^k \leq 1, \forall k \in \mathcal{N}, \text{ and } \rho_l^k \in \{0, 1\}, \forall k \in \mathcal{K}, l \in \mathcal{L}\}$. Then, the dual problem can be stated as

$$\hat{L}_D(\zeta) \triangleq \min_{\boldsymbol{\lambda}, \boldsymbol{\mu} \geq 0} \bar{L}_D(\zeta, \boldsymbol{\lambda}, \boldsymbol{\mu}). \quad (1.25)$$

In order to solve the dual problem (1.25), we investigate problem (1.24) for the given $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$. In particular, we have

$$\begin{aligned} \bar{L}_D(\zeta, \boldsymbol{\lambda}, \boldsymbol{\mu}) &= \max_{z, \mathbf{p}_D \in \mathcal{X}, \boldsymbol{\rho} \in \mathcal{C}} L_D(\mathbf{p}_D, \boldsymbol{\rho}, z, \zeta, \boldsymbol{\lambda}, \boldsymbol{\mu}) \\ &= \max_{z, \mathbf{p}_D \in \mathcal{X}, \boldsymbol{\rho} \in \mathcal{C}} \sum_{k \in \mathcal{N}} \sum_{l \in \mathcal{L}} \rho_l^k f_l^k(p_{Dl}^k) + z(1 - \sum_{l \in \mathcal{L}} \mu_l) + \sum_{l \in \mathcal{L}} (\lambda_l P_{Dl}^{\max} - 2\zeta \mu_l P_0^l), \end{aligned} \quad (1.26)$$

where $f_l^k(p_{Dl}^k) \triangleq \mu_l w_l \hat{R}_{Dl}^k(\mathbf{p}_D, \boldsymbol{\rho}) - (\zeta \alpha_l \mu_l + \lambda_l) p_{Dl}^k$.

To obtain the nontrivial optimal solution of the dual problem (1.25), $\sum_{l \in \mathcal{L}} \mu_l = 1$ must hold. Moreover, problem (1.26) can be decomposed into N individual resource allocation problems for N subchannels where the resource allocation problem for subchannel $k \in \mathcal{N}$ can be stated as

$$\bar{L}_D^k(\zeta, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \max_{\mathbf{p}_D \in \mathcal{X}, \boldsymbol{\rho} \in \mathcal{C}} \sum_{l \in \mathcal{L}} \rho_l^k f_l^k(p_{Dl}^k). \quad (1.27)$$

For problem (1.27), suppose that D2D link l is allocated subchannel $k \in \mathcal{N}$ then we have

$$p_{Dl}^{k*} = \operatorname{argmax}_{p_{Dl}^k \in \mathcal{X}_l} f_l^k(p_{Dl}^k). \quad (1.28)$$

Note that we must have $\mu_l > 0$ because if $\mu_l = 0$, we have $p_{Dl}^{k*} = 0, \forall k \in \mathcal{N}$, which cannot be the optimal solution of problem (1.26). In addition, problem (1.28) can be addressed by solving $\frac{\partial f_l^k(p_{Dl}^k)}{\partial p_{Dl}^k} = 0$, where $\frac{\partial f_l^k(p_{Dl}^k)}{\partial p_{Dl}^k}$ is the first order derivative of $f_l^k(p_{Dl}^k)$. Then, it can be verified that solving $\frac{\partial f_l^k(p_{Dl}^k)}{\partial p_{Dl}^k} = 0$ is equivalent to solving $A_{kl}(p_{Dl}^k)^2 + 2B_{kl}p_{Dl}^k + C_{kl} = 0$. Consequently, the optimal solution of D2D link l that maximizes $f_l^k(p_{Dl}^k)$ is given by

$$p_{Dl}^{k*} = \left[\frac{-B_{kl}^d + \sqrt{\Delta_{kl}^d}}{A_{kl}^d} \right]_0^{P_{Dl}^{\max}}, \quad (1.29)$$

In summary, by solving problem (1.26) we can obtain the optimal power allocation for any D2D link on subchannel $k \in \mathcal{N}$. Recall that we have assumed that each subchannel can be allocated to at most one D2D link; therefore, for all subchannels $k \in \mathcal{N}$, we have

$$\rho_l^{k^*} = \begin{cases} 1 & \text{if } l = \operatorname{argmax}_{l \in \mathcal{L}} f_l^k(p_{Dl}^{k^*}) \\ 0 & \text{otherwise.} \end{cases} \quad (1.30)$$

So far we have presented the resource allocation solution for given λ, μ . Therefore, the remaining task is to solve problem (1.25), which can be completed by the sub-gradient method [21]. Finally, the performance achieved by Algorithm 1.1, which solves problem (1.20), is stated in the following proposition

Proposition 1.2. *Algorithm 1.1 returns a feasible solution of problem (1.20) with $\zeta^*, \mathbf{p}_D^*, \boldsymbol{\rho}^*, \boldsymbol{\lambda}^*$, and $\boldsymbol{\mu}^*$ at the end of its first phase. Moreover, if $\sum_{k \in \mathcal{N}} \rho_l^{k^*} p_{Dl}^{k^*} \leq P_{Dl}^{\max}, \lambda_l^*(P_D^{\max} - \sum_{k \in \mathcal{N}} \rho_l^{k^*} p_{Dl}^{k^*}) = 0$, and $R_{Dl}(\mathbf{p}_D^*, \boldsymbol{\rho}^*) - \zeta^* P_{Dl}^{\text{total}}(\mathbf{p}_D^*, \boldsymbol{\rho}^*) = 0, \forall l \in \mathcal{L}$, this feasible solution is the optimal solution of problem (1.20).*

1.2.2.4 Numerical Results

We consider the simulation setting with the base-station located at the center, $K = 20$ cellular users, and $L = 4$ D2D links randomly placed in 500m x 500m area, and $N = 20$ subchannels for uplink communications. We evaluate the performance of the proposed algorithm by compare it with the Algorithm in [22]. The RBR algorithm is obtained by solving the relaxation version of the original problem and performing intelligent rounding, which is described detail in Chapter 6. Fig. 1.5 indicates that the EE of D2D links achieved by our proposed algorithms are significantly higher than that of the conventional algorithm, e.g., at $d_{\max} = 150$ m, the proposed algorithms can achieve more than 90% of the upper-bound EE, which is about 300% that of the conventional algorithm and about 130% that of the SE-maximization solution.

Finally, Fig. 1.6 shows that the achieved EE of D2D links decreases as the number of D2D links increases. The performance gap between the proposed and the conventional algorithms also decreases as the number of D2D links increases. This is because as the system supports more D2D

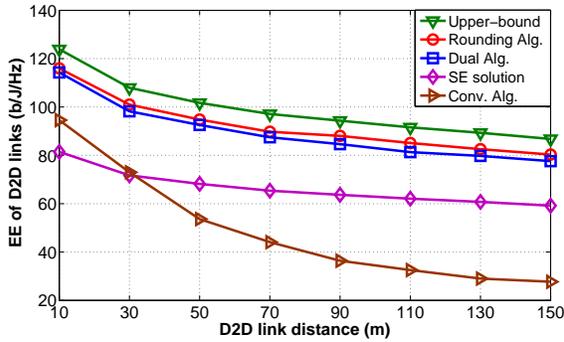


Figure 1.5 – Minimum EE of D2D links versus D2D link distance

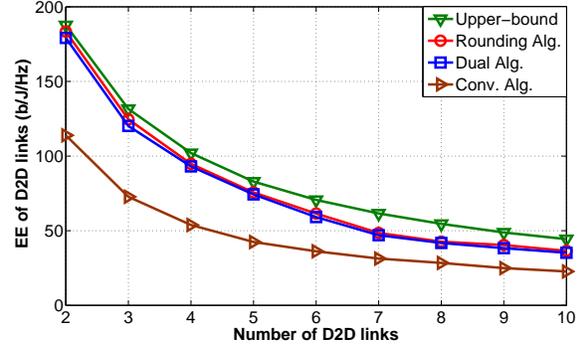


Figure 1.6 – Minimum EE of D2D links versus number of D2D links

links, the available resources for each D2D link becomes smaller, which results in the decrease in the achieved EE of D2D links.

1.2.3 Joint Mode Selection and Resource Allocation for Relay-based D2D Communications

In this contribution, we study the joint mode selection, resource group (RG) assignment, and power control problem for D2D underlaid cellular networks which aims at maximizing the system sum rate considering minimum rate constraints of cellular and D2D links. The resource allocation problem is formulated as an MINLP (Mixed-Integer Non-Linear Programming) problem. To solve this problem *optimally*, we first study the optimal power allocation for a given mode selection and RG assignment solution. Based on these results, the original resource allocation problem can be transformed into a RG allocation problem, which can be solved optimally by the Hungarian method. Extensive numerical studies demonstrate that the proposed design significantly outperforms existing D2D communication schemes using fixed direct or relay mode.

1.2.3.1 System Model

We consider the uplink of a single macro-cell system where K cellular links in the set $\mathcal{K} = \{1, \dots, K\}$ share the same spectrum of K resource group (RG) in the set $\mathcal{N} = \{1, \dots, K\}$ with L D2D links in the set $\mathcal{L} = \{1, \dots, L\}$. We assume that cellular link $k \in \mathcal{K}$ has been pre-allocated RG $k \in \mathcal{N}$,

which consists of m_k consecutive sub-channels.³ We also assume that each D2D link reuse the resource of one RG, and each RG is assigned to at most one D2D link.

Let $\boldsymbol{\rho}$ be a matrix capturing binary resource allocation decisions of the D2D links where $[\boldsymbol{\rho}]_{kl} = \rho_{kl} = 1$ if D2D link l is assigned RG k and $\rho_{kl} = 0$, otherwise. Moreover, each D2D transmitter can communicate with its corresponding receiver via either direct or relay mode (assisted by a relay). Let $\mathbf{x} = [x_1, \dots, x_L]$ be the binary mode selection decision vector for all D2D links where $x_l = 1$ if D2D link l operates in the direct mode and $x_l = 0$, otherwise. We also assume that the relay selection for each D2D link has been pre-determined where each D2D link $l \in \mathcal{L}$ can be assisted by its assigned relay. We assume that D2D link l is supported by relay r_l in the relay set $\mathcal{R} = \{r_1, \dots, r_L\}$. Denote h_{ab}^n as the channel gain from transmitter of link or relay b to the receiver of link or relay a on RG n .

We denote \mathbf{p}_C and \mathbf{p}_D are the power allocation vector of cellular and D2D links where $[\mathbf{p}_C]_k = p_{Ck}$ and $[\mathbf{p}_D]_l = p_{Dl}$ denote the transmit power of cellular link k and D2D link l , respectively. The data rate of cellular link k on its RG without any co-channel D2D links can be expressed as $R_{Ck}^{(o)} = m_k \log_2 \left(1 + p_{Ck} h_{kk}^k / \sigma^2 \right)$ where σ^2 denotes the thermal noise, and the data rate in b/s/Hz is normalized by the bandwidth of one sub-channel.

In this work, we allow D2D nodes in each D2D link to communicate to each other using either direct or relay mode. In the direct mode, the D2D transmitter communicates directly with its D2D receiver. However, in the relay mode, we assume that the Decode and Forward (DF) relaying strategy is employed where each communication period is divided into two equal intervals corresponding to the D2D transmitter to relay (D-R) communication phase and relay to D2D receiver (R-D) communication phase.

Direct Mode: If RG k is assigned to D2D link l , the data rates of cellular link k and D2D link l are described as

$$R_{Dlk}^{(d)} = m_k \log_2 \left(1 + \frac{p_{Dl} h_{ll}^k}{\sigma^2 + p_{Ck} h_{lk}^k} \right) \quad (1.31)$$

$$R_{Ckl}^{(d)} = m_k \log_2 \left(1 + \frac{p_{Ck} h_{kk}^k}{\sigma^2 + p_{Dl} h_{kl}^k} \right). \quad (1.32)$$

³This is the case in uplink LTE system using SC-FDMA, each subchannel is equivalent to one resource block (RB), and RG is a group of contiguous RBs assigned to one particular cellular link.

Relay Mode: Let $\mathbf{p}_R = [p_{R_1}, \dots, p_{R_L}]$ be the power allocation vector of the relays to support their D2D links. Then, the data rates of cellular link k in the first and second communications phases can be expressed as

$$R1_{Ckl}^{(r)} = \frac{m_k}{2} \log_2 \left(1 + \frac{p_{Ck} h_{kk}^k}{\sigma^2 + p_{Dl} h_{kl}^k} \right) \quad (1.33)$$

$$R2_{Ckl}^{(r)} = \frac{m_k}{2} \log_2 \left(1 + \frac{p_{Ck} h_{kk}^k}{\sigma^2 + p_{Rl} h_{kr_l}^k} \right). \quad (1.34)$$

Moreover, the data rates achieved on the D-R and R-D links in the first and second communications phases can be calculated, respectively, as

$$R1_{Dlk}^{(r)} = \frac{m_k}{2} \log_2 \left(1 + \frac{p_{Dl} h_{r_l l}^k}{\sigma^2 + p_{Ck} h_{r_l k}^k} \right) \quad (1.35)$$

$$R2_{Dlk}^{(r)} = \frac{m_k}{2} \log_2 \left(1 + \frac{p_{Rl} h_{l r_l}^k}{\sigma^2 + p_{Ck} h_{lk}^k} \right). \quad (1.36)$$

Finally, the data rates of cellular link k and D2D link l can be written, respectively, as

$$R_{Ckl}^{(r)} = R1_{Ckl}^{(r)} + R2_{Ckl}^{(r)} \quad (1.37)$$

$$R_{Dlk}^{(r)} = \min \left\{ R1_{Dlk}^{(r)}, R2_{Dlk}^{(r)} \right\}. \quad (1.38)$$

1.2.3.2 Problem Formulation

Using the above notations, the data rates of cellular link k and D2D link l can be expressed as

$$R_{Ck} = \left(1 - \sum_{l \in \mathcal{L}} \rho_{kl} \right) R_{Ck}^{(o)} + \sum_{l \in \mathcal{L}} \rho_{kl} \left[x_l R_{Ckl}^{(d)} + (1 - x_l) R_{Ckl}^{(r)} \right] \quad (1.39)$$

$$R_{Dl} = \sum_{n \in \mathcal{N}} \rho_{nl} \left[x_l R_{Dln}^{(d)} + (1 - x_l) R_{Dln}^{(r)} \right]. \quad (1.40)$$

We propose a joint mode selection, resource assignment, and power allocation problem which targets to maximize the sum rate of all communications links while the minimum data rate of

cellular and D2D links are satisfied in the following.

$$\max_{\mathbf{p}, \rho, \mathbf{x}} \sum_{k \in \mathcal{K}} R_{Ck} + \sum_{l \in \mathcal{L}} R_{Dl} \quad (1.41a)$$

$$\text{s.t. } R1_{Ckl}^{(r)} \geq \frac{1}{2} \rho_{kl} x_l R_{Ck}^{\min}, \quad R2_{Ckl}^{(r)} \geq \frac{1}{2} \rho_{kl} x_l R_{Ck}^{\min} \quad (1.41b)$$

$$R_{Ck} \geq R_{Ck}^{\min} \quad \forall k \in \mathcal{K}, \quad R_{Dl} \geq R_{Dl}^{\min}, \quad \forall l \in \mathcal{L} \quad (1.41c)$$

$$m_k p_{Ck} \leq P_{Ck}^{\max} \quad \forall k \in \mathcal{K} \quad (1.41d)$$

$$\sum_{n \in \mathcal{N}} \rho_{nl} m_n p_{Dln} \leq P_{Dl}^{\max} \quad \forall l \in \mathcal{L} \quad (1.41e)$$

$$\sum_{n \in \mathcal{N}} (1 - x_l) m_n p_{Rl} \leq P_{Rl}^{\max} \quad \forall r_l \in \mathcal{R} \quad (1.41f)$$

$$x_l \in \{0, 1\} \quad \forall l \in \mathcal{L}, \quad \rho_{kl} \in \{0, 1\} \quad \forall k \in \mathcal{K} \quad \forall l \in \mathcal{L}. \quad (1.41g)$$

In problem (1.41) constraints (1.41b) and (1.41c) are imposed to guarantee the minimum data rate of D2D and cellular links in all transmission intervals, and constraint (1.41g) requires each D2D link to work in one mode. In the following, we propose to solve problem (1.41) *optimally* through a solution approach with three phases, namely power allocation, mode selection, and resource group (RG) allocation. First, we solve the power allocation problem for each D2D link l in either relay or direct mode if it reuses the resource of cellular link k . Then, the mode selection is implemented to determine the optimal modes of D2D links. Finally, the original problem is transformed to RG assignment problem, which can be solved *optimally* by using the Hungarian method. We present this design in the following.

1.2.3.3 Power Allocation

Assume that D2D link l reuses the resource of cellular link k , we need to solve two power allocation problems corresponding to the direct and relay mode of D2D link l . The power allocation problem as D2D link l operates in the direct mode can be solved by using the algorithm in [23]. On the other hand, if D2D link l operates in the relay mode, we denote $\mathbf{p}_{kl} = [p_{Ck}, p_{Dl}, p_{Rl}]$, $P_c^m \triangleq P_{Ck}^{\max} / m_k$,

$P_d^m \triangleq P_{Dl}^{\max}/m_k$, $P_r^m \triangleq P_{Rl}^{\max}/m_k$. Then, we have the following power allocation problem.

$$\max_{\mathbf{p}_{kl}} w_{kl}^{(r)}(\mathbf{p}_{kl}) \triangleq R_{Ckl}^{(r)} + R_{Dlk}^{(r)} \quad (1.42a)$$

$$\text{s.t. } R1_{Ckl}^{(r)} \geq \frac{1}{2}R_{Ck}^{\min}, R2_{Ckl}^{(r)} \geq \frac{1}{2}R_{Ck}^{\min}, R_{Dlk}^{(r)} \geq R_{Dl}^{\min} \quad (1.42b)$$

$$p_{Ck} \in [0, P_c^m], p_{Dl} \in [0, P_d^m], p_{Rl} \in [0, P_r^m]. \quad (1.42c)$$

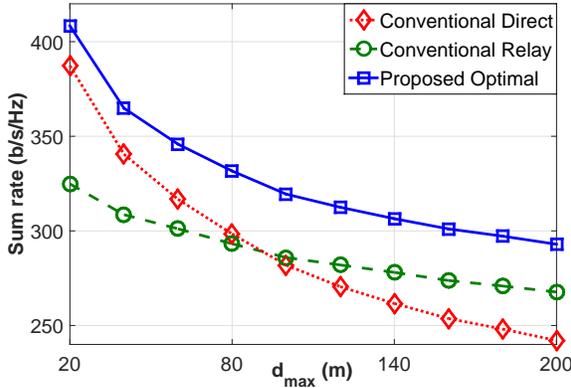
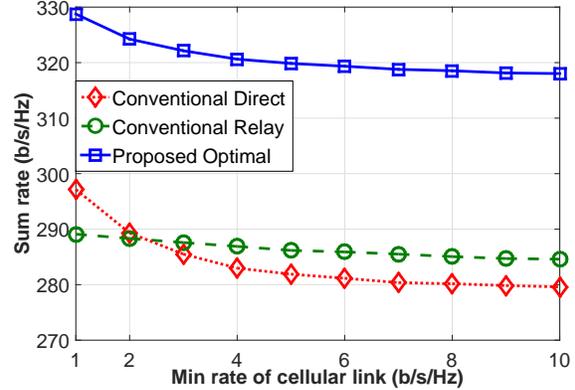
We characterize the optimal power allocation of the above problem in the following proposition.

Proposition 1.1. *If problem (1.42) is feasible then at optimality at least one node (D2D transmitter, relay, or cellular user) uses maximum transmit power and $R1_{Dlk}^{(r)} = R2_{Dlk}^{(r)}$.*

From Proposition 1.1, it can be verified that problem (1.42) can achieve its optimum if $p_{Ck} = P_c^m$, $p_{Dl} = P_d^m$, or $p_{Rl} = P_r^m$. Hence, we can determine the optimal solution of problem by evaluating all three cases. In each case, we need to determine the power allocation of two transmitters. For each case, i.e., $p_{Ck} = P_c^m$, $p_{Dl} = P_d^m$, or $p_{Rl} = P_r^m$, by applying the constraint $R1_{Dlk}^{(r)} = R2_{Dlk}^{(r)}$ from Proposition 1.1, we can transform problem (1.42) to an optimization problem of one variable p_{Ck} , p_{Dl} , or p_{Rl} . For the optimization of one variable, the local points can be determined by evaluating the first derivation of the objective function. Then the optimal solution can be obtained by comparing the objective values of all local and extreme points. Therefore, by using this procedure we can achieve the optimal solution of problem (1.42).

1.2.3.4 Joint Mode Selection and RG Allocation

Denote $w_{kl}^{(d)*}$ and $w_{kl}^{(r)*}$ as the optimal total rates when D2D link l reuses the resource of cellular link k in direct and relay modes, respectively. The optimal mode selection can be determined as follows. Assume that D2D link l is assigned RG k , if $w_{kl}^{(d)*} \geq w_{kl}^{(r)*}$ then D2D link l should operate in the direct mode, otherwise, it should operate in the relay mode. As D2D link l is assigned RG k , the rate increase due to D2D resource reuse is $w_{kl}^* = \max\{w_{kl}^{(d)*}, w_{kl}^{(r)*}\} - R_{Ck}^{(o)}$. Therefore, problem (1.41) can be transformed into the following problem, which can be solved optimally by

Figure 1.7 – Sum rate versus d_{\max} Figure 1.8 – Sum rate versus R_c^{\min}

the Hungarian method [24]

$$\max_{\rho} \sum_{l \in \mathcal{L}} \sum_{k \in \mathcal{N}} w_{kl}^* \rho_{kl} \quad (1.43a)$$

$$\text{s.t.} \quad \sum_{k \in \mathcal{N}} \rho_{kl} = 1, \forall l \in \mathcal{L}, \sum_{l \in \mathcal{L}} \rho_{kl} \leq 1, \forall k \in \mathcal{N} \quad (1.43b)$$

$$\rho_{kl} \in \{0, 1\}, \forall l \in \mathcal{L}, k \in \mathcal{N}. \quad (1.43c)$$

1.2.3.5 Numerical Results

We consider the system where there are $L = 15$ D2D links reusing the resource of 20 cellular links, and each cellular link is allocated one subchannel. Cellular users and relays are randomly distributed in the cell area of radius 500m. Moreover, each D2D transmitter and receiver are located randomly whose distance to its relay varies within d_{\max} , where $d_{\max} = 100m$. Finally, our proposed design, which is denoted as “Proposed Optimal”, is compared with two existing schemes developed in [23] and [25] denoted as “Conventional Direct” and “Conventional Relay”, respectively.

Fig. 1.7 presents the system sum rate versus d_{\max} . As d_{\max} is small, our proposed scheme performs similarly to the “Conventional Direct” scheme and significantly outperforms the “Conventional Relay” scheme. This is because as d_{\max} is small, the optimal D2D mode is usually the direct mode. However, as d_{\max} increases, D2D links tend to operate in the relay mode more frequently since the relay D2D mode can outperform the direct D2D mode. As a result, the proposed design performs much better than the other two existing schemes thanks to the benefits of adaptively switching between the direct D2D and relay D2D modes.

Fig. 1.8 shows the system sum rate versus R_c^{\min} . As R_c^{\min} increases, the system sum rate decreases moderately. This is because as R_c^{\min} becomes higher, cellular links must increase their transmit powers to meet the data rate requirements. Moreover, with optimal RG assignment, D2D links which suffer from low co-channel interference from certain cellular links tend to reuse the frequency resources of these cellular links. Thus, the higher transmit powers of cellular links would not degrade the data rates of D2D links significantly. As a result, the system sum rate decreases gradually as R_c^{\min} increases.

1.2.4 Joint Prioritized Scheduling and Resource Allocation for OFDMA-based Wireless Networks

In this contribution, we study the joint prioritized scheduling, subchannel assignment, and power allocation problem for multiple wireless links in the OFDMA-based wireless networks, which simultaneously (i) maximizes the number of scheduled non-prioritized links and (ii) maximizes their sum rate subject to the minimum rate requirements of prioritized and scheduled non-prioritized links. In particular, our work makes the following novel contributions.

- We formulate the scheduling and resource allocation design as a single-stage optimization problem considering QoS constraints of the prioritized and scheduled non-prioritized links.
- We develop a monotonic based optimal approaching (MBOA) algorithm to solve the above problem which asymptotically achieves the optimal set of scheduled non-prioritized links and their maximum sum rate.
- We propose another low-complexity iterative convex approximation (ICA) algorithm which sequentially performs power allocation and link removal in each iteration.
- The numerical studies demonstrate that the MBOA algorithm performs the best among all algorithms.

1.2.4.1 System Model

We consider uplink communications in a single-cell wireless system where K prioritized wireless links in the set $\mathcal{K} = \{1, \dots, K\}$ share the same spectrum comprising N orthogonal subchannels in

the set $\mathcal{N} = \{1, \dots, N\}$ with L non-prioritized wireless links in the set $\mathcal{L} = \{K + 1, \dots, K + L\}$. Let $\mathcal{M} = \mathcal{K} \cup \mathcal{L}$ denote the set of all the links. We denote p_m^n as the transmit power of link $m \in \mathcal{M}$ on subchannel n and we represent the transmit power vector of all links in the system as $\mathbf{p} = [\mathbf{p}_m]_{\forall m \in \mathcal{M}}$ where $\mathbf{p}_m = [p_m^n]_{\forall n \in \mathcal{N}}$ is the power allocation vector of link $m \in \mathcal{M}$ over the subchannels. We assume that the prioritized links utilize subchannels orthogonally; however, the non-prioritized links are allowed to reuse all subchannels to improve the spectrum-efficiency through exploiting the spatial diversity. We define the following subchannel assignment vector $\boldsymbol{\rho}^n = [\rho_k^n]_{\forall k \in \mathcal{K}, \forall n \in \mathcal{N}}$, where ρ_k^n denotes whether channel n is assigned to link k .

It was proved in [26] that if the mutual interference between two interfering links is strong enough, they should utilize the spectrum orthogonally to maximize the sum rate. Motivated by this result, we allow each prioritized link to exploit all subchannels mathematically; however, the *virtual* channel gains among the prioritized links are set very high. Such setting of the high interfering channel gains will indeed force prioritized links to use the subchannels orthogonally to avoid strong co-channel interference. Specifically, by setting the channel gains among prioritized links to a sufficiently large value η , the SINR of prioritized link k non-prioritized link l on subchannel n can be expressed as

$$\bar{\Gamma}_k^n(\mathbf{p}) = \frac{p_k^n h_{kk}^n}{\sigma_k^n + \sum_{k' \in \mathcal{K} \setminus k} p_{k'}^n \eta + \sum_{l \in \mathcal{L}} p_l^n h_{kl}^n} \quad (1.44)$$

$$\bar{\Gamma}_l^n(\mathbf{p}) = \frac{p_l^n h_{ll}^n}{\sigma_l^n + \sum_{k \in \mathcal{K}} p_k^n h_{lk}^n + \sum_{l' \in \mathcal{L} \setminus l} p_{l'}^n h_{ll'}^n}. \quad (1.45)$$

The data rates of prioritized link $k \in \mathcal{K}$ and non-prioritized link $l \in \mathcal{L}$ can be re-expressed, respectively, as

$$\bar{R}_k(\mathbf{p}) = \sum_{n \in \mathcal{N}} \log_2 \left(1 + \bar{\Gamma}_k^n(\mathbf{p}) \right) \quad (1.46)$$

$$\bar{R}_l(\boldsymbol{\rho}, \mathbf{p}) = \sum_{n \in \mathcal{N}} \log_2 \left(1 + \bar{\Gamma}_l^n(\mathbf{p}) \right). \quad (1.47)$$

We would like to design the joint link scheduling, subchannel assignment, and power control for all the links considering the following design objectives and constraints: (i) the minimum required data rates of prioritized links must be maintained, (ii) the number of scheduled non-prioritized links is maximized, and (iii) for the given set of scheduled non-prioritized links, the sum rate of the

scheduled non-prioritized links is maximized. We consider the centralized design where the CSI of all links is available for the optimization. To capture the scheduling decision, we introduce a binary link scheduling vector $\mathbf{s}_{\mathcal{L}} = [s_1, \dots, s_L]^T$, where $s_l = 1$ if the non-prioritized link $l \in \mathcal{L}$ is scheduled and $s_l = 0$, otherwise. Therefore, we have the following optimization problem

$$\max_{\mathbf{p}, \mathbf{s}_{\mathcal{L}}} \alpha \sum_{l \in \mathcal{L}} s_l + \sum_{l \in \mathcal{L}} \bar{R}_l(\mathbf{p}) \quad (1.48a)$$

$$\text{s.t.} \quad \bar{R}_k(\mathbf{p}) \geq R_k^{\min} \quad \forall k \in \mathcal{K} \quad (1.48b)$$

$$\bar{R}_l(\mathbf{p}) \geq s_l R_l^{\min} \quad \forall l \in \mathcal{L} \quad (1.48c)$$

$$\sum_{n \in \mathcal{N}} p_k^n \leq P_{\max} \quad \forall k \in \mathcal{K} \quad (1.48d)$$

$$\sum_{n \in \mathcal{N}} p_l^n \leq P_{\max} \quad \forall l \in \mathcal{L} \quad (1.48e)$$

$$s_l \in \{0, 1\} \quad \forall l \in \mathcal{L}. \quad (1.48f)$$

We develop an algorithm to solve problem (1.48) in the following sections.

1.2.4.2 Monotonic Based Optimal Approaching (MBOA) Algorithm

Note that the binary nature of scheduling vector $\mathbf{s}_{\mathcal{L}}$ makes problem (1.48) difficult to solve. To overcome this challenge, we approximate a discrete variable s_l by a continuous function $q : [0, 1] \rightarrow [0, 1]$ which is defined as $q(s_l) = (e^{Qs_l} - 1)/(e^Q - 1)$. Function $q(s_l)$ has the following properties: (i) $q(s_l)$ is a continuous and increasing function, (ii) $q(0) = 0$, and $q(1) = 1$. Moreover, as Q increases, the $q(\cdot)$ function approaches the step function more closely. Using the approximation function $q(\cdot)$, we arrive at the following problem

$$\begin{aligned} \max_{\mathbf{p}, \mathbf{s}_{\mathcal{L}}} \quad & \alpha \sum_{l \in \mathcal{L}} q(s_l) + \sum_{l \in \mathcal{L}} \bar{R}_l(\mathbf{p}) \\ \text{s.t.} \quad & (1.48b), (1.48c), (1.48d), (1.48e) \\ & s_l \in [0, 1] \quad \forall l \in \mathcal{L}. \end{aligned} \quad (1.49)$$

We now show that problem (1.49) can be transformed into a monotonic optimization problem in the followings [27]. It can be seen that the objective function of problem (1.49) is increasing in

$\mathbf{s}_{\mathcal{L}}$; however, it is non-increasing in \mathbf{p} . We define new variables $z_{mn} = \bar{\Gamma}_m^n(\mathbf{p}) \forall m \in \mathcal{M}, \forall n \in \mathcal{N}$, which denote the achievable SINR of link $m \in \mathcal{M}$ on subchannel $n \in \mathcal{N}$. We also define $\mathbf{z} = [z_{mn}]_{\forall m \in \mathcal{M}, \forall n \in \mathcal{N}}$. We introduce an auxiliary vector $\mathbf{t} = [t_l]_{\forall l \in \mathcal{L}}$ and $t_l \in [0, R_l^{\min}]$ then constraint (1.48c) for each non-prioritized link $l \in \mathcal{L}$ becomes equivalent to following constraints

$$\sum_{n \in \mathcal{N}} \log_2(1 + z_{ln}) + t_l \geq R_l^{\min} \quad (1.50)$$

$$t_l + s_l R_l^{\min} \leq R_l^{\min} \quad (1.51)$$

$$t_l \in [0, R_l^{\min}]. \quad (1.52)$$

Denote $\mathbf{x} = (\mathbf{t}, \mathbf{s}_{\mathcal{L}}, \mathbf{z})$ as the optimization vector which has $D = 2L + (K + L)N$ dimensions and $\mathcal{P} \triangleq \{\mathbf{p} | \sum_{n \in \mathcal{N}} p_m^n \leq P_{\max} \forall m \in \mathcal{M}\}$. Then, problem (1.49) can be transformed into the following one

$$\max_{\mathbf{x} \geq \mathbf{0}} f(\mathbf{x}) \triangleq \alpha \sum_{l \in \mathcal{L}} q(s_l) + \sum_{l \in \mathcal{L}} \sum_{n \in \mathcal{N}} \log_2(1 + z_{ln}) \quad (1.53a)$$

$$\text{s.t. } s_l \leq 1 \quad \forall l \in \mathcal{L} \quad (1.53b)$$

$$z_{mn} \leq \bar{\Gamma}_m^n(\mathbf{p}) \quad \forall m \in \mathcal{M} \quad \forall n \in \mathcal{N} \quad \forall \mathbf{p} \in \mathcal{P} \quad (1.53c)$$

$$t_l + s_l R_l^{\min} - R_l^{\min} \leq 0 \quad \forall l \in \mathcal{L} \quad (1.53d)$$

$$t_l \leq R_l^{\min} \quad \forall l \in \mathcal{L} \quad (1.53e)$$

$$\sum_{n \in \mathcal{N}} \log_2(1 + z_{kn}) - R_k^{\min} \geq 0 \quad \forall k \in \mathcal{K} \quad (1.53f)$$

$$\sum_{n \in \mathcal{N}} \log_2(1 + z_{ln}) + t_l - R_l^{\min} \geq 0 \quad \forall l \in \mathcal{L}. \quad (1.53g)$$

We characterize the property of problem (1.53) in the following proposition.

Proposition 1.3. *Problem (1.53) is a monotonic optimization problem.*

As problem (1.53) is a monotonic optimization problem, we can develop *polyblock outer approximation* algorithm [27] to solve it optimally. Assume that $\mathbf{x}^* = (\mathbf{t}^*, \mathbf{s}_{\mathcal{L}}^*, \mathbf{z}^*)$ and \mathbf{p}^* are the optimal solution of problem (1.53) and its corresponding power allocation, respectively. Nevertheless, $\mathbf{s}_{\mathcal{L}}^*$ can be fractional which is not a feasible solution of problem (1.48). We propose to schedule the non-prioritized link l if its scheduling solution $s_l^* \geq 1 - \epsilon$. Then, the performance of the MBOA algorithm is characterized in the following theorem.

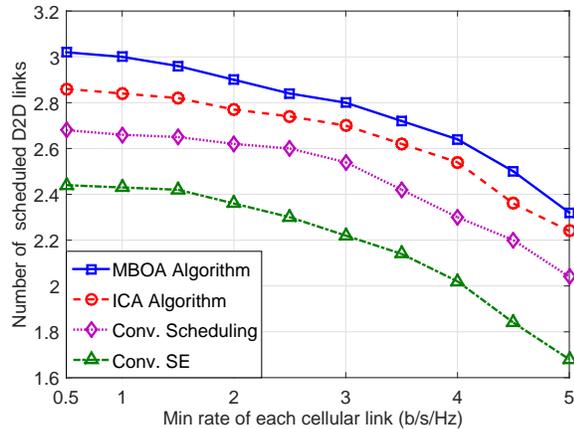


Figure 1.9 – Number of admitted D2D links versus R_c^{\min}

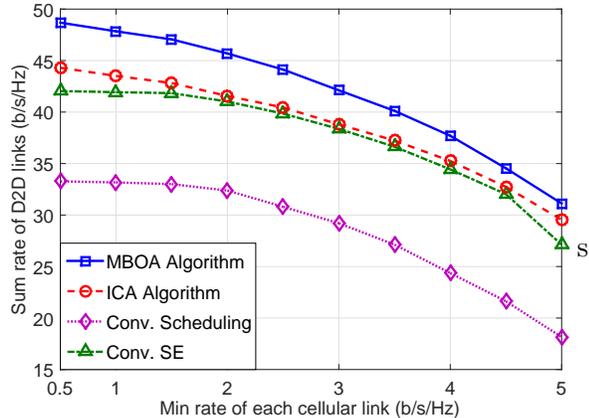


Figure 1.10 – Sum rate of D2D links versus minimum rate of cellular link R_c^{\min}

Theorem 1.4. *If the minimum required data rate of each non-prioritized link can be reduced by a small and controller number ϵR_l^{\min} , by choosing $Q \geq \ln L/\epsilon$, the proposed MBOA algorithm schedules the maximum number of non-prioritized links.⁴*

1.2.4.3 Numerical Results

We evaluate the performance of our proposed algorithms for wireless cellular networks supporting D2D communications which consist of $K = 4$ prioritized cellular links and $L = 5$ non-prioritized D2D links sharing the resource of $N = 10$ subchannels. We then evaluate the performance of the proposed algorithms and compare them with the “Conv. SE” and “Conv. Scheduling” algorithms, which are adopted from [28]. Figs. 1.9 and 1.10 show the number of scheduled D2D links and their sum rate versus the cellular links’ minimum required rate R_c^{\min} , respectively. As R_c^{\min} increases, the system has to assign more resources to the cellular links; therefore, a smaller number of D2D links and a smaller sum rate of the scheduled D2D links can be achieved, which can be observed for all algorithms. These results confirm that the MBOA algorithm performs the best in term of both number of scheduled D2D links and their sum rate. This is because the MBOA algorithm can schedule the maximum number of D2D links as stated in Theorem 1.4.

⁴If we consider the studied problem where the minimum rate of non-prioritized link l equals to $(1 + \epsilon)R_l^{\min}$ then the MBOA algorithm can guarantee the required QoS of all scheduled non-prioritized links.

1.2.5 Concluding Remarks

In this doctoral dissertation, we have developed various novel resource management techniques and algorithms for D2D communication in cellular networks. In particular, we have made four important research contributions. First, we have developed a spectrum-efficient resource allocation algorithms which perform significantly better than the other state-of-the-art designs in the literature. Second, we have proposed a general energy-efficient RA framework for D2D communication in cellular networks which targets to maximize the minimum weighted EE of the D2D links while the minimum data rates of individual cellular links are ensured.

Third, we have developed an optimal RA scheme for relay-based D2D communication in cellular network, which enable D2D communications to be operated in either the relay or direct mode. Moreover, the proposed algorithm significantly outperforms existing D2D communication schemes. Finally, we have formulated the scheduling and resource allocation design for D2D communication in cellular network where the cellular links is more prioritized than the D2D links. The proposed scheme allows the system to dynamically select the set of scheduled D2D links and to optimize the spectrum-efficiency of the system. We have proposed a monotonic-based algorithm which asymptotically achieves the optimal solution.

Chapter 2

Résumé Long

2.1 Contexte et motivation

Les opérateurs de téléphonie cellulaire sans fil ont vu la demande toujours croissante des applications à grande vitesse et du nombre de périphériques connectés qui s'accroît rapidement. A cette fin, les réseaux sans fil du futur devraient fournir une capacité beaucoup plus grande et supporter des taux de communication nettement plus élevés. Plus précisément, il est prédit que le système sans fil 5G devrait permettre une augmentation de 1000 fois de la capacité du système et une augmentation de 100 fois du débit de données des dispositifs connectés [3]. La communication Device-to-device (D2D) (périphérique à périphérique) a récemment été proposée comme une technologie importante pour atteindre ces objectifs [4] en soutenant le trafic local par des communications directes parmi les appareils mobiles. Dans la communication D2D, deux dispositifs proches peuvent établir une liaison de communication directe sous le contrôle de la station de base cellulaire (BS). Différents aspects de la conception de la communication D2D tels que l'interface matérielle, la découverte D2D et l'allocation des ressources ont été étudiés dans les communautés académiques et industrielles pour permettre les communications D2D et soutenir les futurs scénarios et applications du système [5].

En général, la communication D2D peut aider à améliorer considérablement la capacité du système en exploitant les avantages de la communication de proximité. Grâce à la faible distance de communication entre les périphériques proches, une communication robuste avec un débit de données élevé peut être établie pour répondre aux exigences rigoureuses des applications sans fil haut

débit émergentes telles que le partage de vidéos et les jeux en ligne. Dans la communication D2D, les dispositifs de proximité peuvent communiquer directement l'un à l'autre au lieu de contourner la BS; par conséquent, le délai de transmission peut être considérablement réduit. La communication D2D peut supporter non seulement les services traditionnels de voix et de données locaux, mais aussi de nombreuses applications émergentes basées sur D2D, telles que les réseaux sociaux, le partage de vidéos, les jeux en ligne et les applications de sécurité publique.

Third Generation Partnership Project (3GPP) considère la communication D2D comme une technologie essentielle pour l'avenir des systèmes cellulaires [6, 7]. La standardisation initiale de la communication D2D a été proposée dans la version 12 de 3GPP pour les applications de sécurité publique [8]. En outre, la communication D2D joue un rôle plus important dans de nombreux scénarios différents tels que l'Internet massif des choses, les communications extrêmes en temps réel, les communications de ligne de vie, les communications ultra-fiables et les services de diffusion [9]. En fait, les scénarios et les applications potentiels de la communication D2D sont toujours en discussion pour le système 5G, qui devrait être déployé en 2020.

Un défi majeur pour permettre la communication D2D dans les réseaux cellulaires est lié au développement de techniques efficaces de gestion des ressources radio. En particulier, il est souhaitable d'utiliser un système d'allocation de ressources pour la communication D2D qui peut exploiter efficacement les ressources limitées telles que la fréquence et la puissance d'émission. Différentes approches générales d'allocation de ressources pour la communication D2D peuvent être résumées comme dans la Fig. 2.1. Plus précisément, la communication D2D peut être utilisée en utilisant la fréquence attribuée au spectre cellulaire, c'est-à-dire en D2D en bande, ou en utilisant un spectre non-licence, c'est-à-dire une bande hors-ligne D2D [10]. Cependant, dans le D2D hors bande, la qualité du service (QoS) des liens D2D ne serait pas garantie en raison d'interférences incontrôlables provenant d'autres sources. Dans le D2D en bande, les liens D2D peuvent être utilisés dans des scénarios de superposition ou de sous-couche. Dans le scénario de superposition, la ressource de fréquence est réservée orthogonalement pour les liens D2D et les liens cellulaires; néanmoins, fonctionner dans le scénario de superposition pourrait entraîner une utilisation inefficace des ressources, car certaines ressources de fréquence pourraient être inutilisées en raison de l'indisponibilité du lien D2D ou du lien cellulaire. Enfin, dans le scénario de sous-couche, les ressources de fréquence sont partagées entre les liens D2D et cellulaires, ce qui peut potentielle-

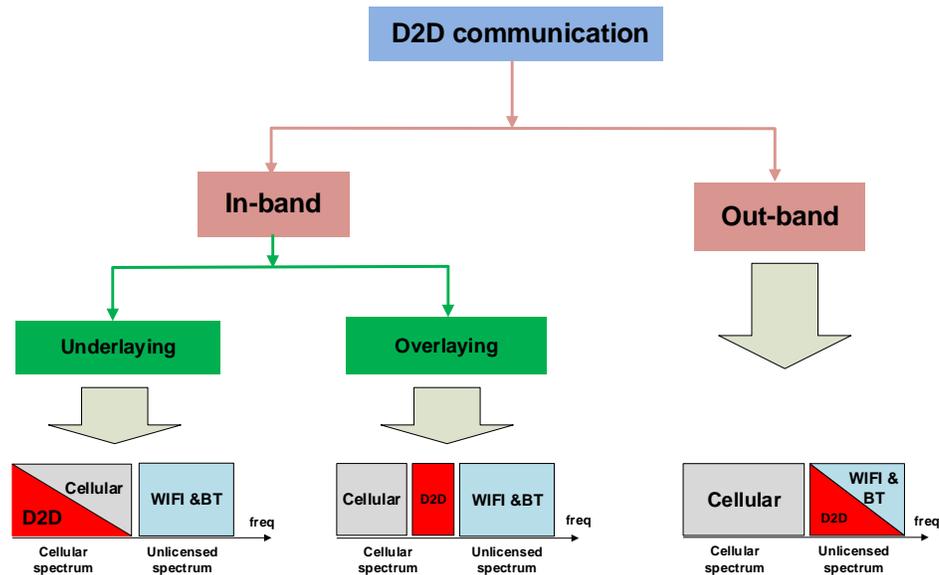


Figure 2.1 – Communication D2D dans les réseaux cellulaires

ment améliorer les performances du système car la ressource de fréquence et la diversité spatiale du système peuvent être exploitées efficacement.

Les bénéfices potentiels obtenus à partir du scénario de sous-couche viennent avec le prix des algorithmes d'allocation de ressources plus compliqués. En particulier, ces algorithmes sont nécessaires pour allouer efficacement les ressources de fréquence à chaque lien et gérer les interférences entre les liens. Cependant, pour permettre la communication D2D dans le réseau cellulaire, chaque opérateur peut employer différentes approches qui peuvent être choisies en considérant différents modèles du système, un ensemble différent des paramètres d'optimisation, différents fonctions d'utilité et schémas d'utilisation de ressources.

Du point de vue de l'opérateur, la maximisation de l'efficacité du spectre système (SE) peut être plus souhaitable puisque le revenu est proportionnel à la débit du système. Néanmoins, du point de vue des utilisateurs mobiles, l'optimisation de l'efficacité énergétique du système (EE) peut être plus appropriée car cette conception peut aider à prolonger la durée de vie de la batterie des périphériques. Dans ce but, la communication directe D2D peut être bénéfique si la paire sous-jacente

d'appareils est proche l'un de l'autre. Cependant, si deux périphériques sont éloignés l'un de l'autre, ils peuvent établir le chemin de communication entre eux à travers la BS ou un noeud de relais. Par conséquent, l'allocation de ressources pour la communication D2D par relais est un problème de conception important pour permettre la communication D2D dans les réseaux cellulaires [11]. Outre l'affectation des ressources, le contrôle de planification, qui détermine l'ensemble des liens à transmettre à un moment donné, est également essentiel à la réalisation d'une utilisation efficace des ressources. En outre, le cellulaire et le D2D peuvent être traités comme des liens primaires et secondaires, respectivement, où les liens primaires sont priorisés sur les liens secondaires. Par conséquent, la planification conjointe avec la conception de l'allocation des ressources pour ce modèle d'accès prioritaire sont un problème de conception important.

2.2 Contributions de la recherche

L'objectif général de ma recherche de doctorat est de développer des algorithmes efficaces d'allocation de ressources pour la communication D2D qui contribuent à permettre une intégration efficace de la communication D2D dans les réseaux cellulaires sans pour autant causer une grave dégradation des performances des liaisons cellulaires existantes. Plus précisément, nos principales contributions peuvent être décrites comme suit.

2.2.1 Affectation des ressources pour la communication D2D Réseaux cellulaires sous-appliqués à l'aide de l'approche graphique

Cette contribution se concentre sur l'allocation des ressources radioélectriques efficace pour les communications D2D dans les réseaux cellulaires qui détermine simultanément l'attribution de sous-bande et le contrôle de puissance pour les liaisons cellulaires et D2D. Plus précisément, notre travail fait les contributions suivantes.

- Nous formulons un problème conjoint de l'allocation de ressources pour la sélection de liens D2D, l'attribution de sous-bande et le problème de contrôle de puissance. Pour résoudre ce problème, nous dérivons d'abord l'allocation de puissance optimale pour une attribution de sous-bande donnée pour une paire de liaisons cellulaires et D2D. Ensuite, l'affectation de sous-

bande et l'allocation de puissance d'origine sont transformées en un problème d'affectation de sous-bande.

- Par la suite, nous développons un algorithme d'arrondi itératif pour résoudre le problème de l'attribution de sous-bande en utilisant la théorie des graphes. Les résultats numériques démontrent que l'algorithme d'arrondi itératif atteint presque le même taux d'addition comme celui obtenu par l'algorithme BnB optimal, qui donne nettement une performance supérieure à celle des algorithmes conventionnels.

2.2.1.1 Modèle du système

Nous considérons le problème du partage de spectre parmi les multiples liaisons D2D et cellulaires dans la direction de liaison montante. Laissez $\mathcal{N} = \{1, \dots, N\}$ être l'ensemble de sous-bandes dans le système.¹ Nous désignons $\mathcal{K}_c = \{1, \dots, K_c\}$ comme l'ensemble de liens cellulaires, $\mathcal{K}_d = \{K_c + 1, \dots, K_c + K_d\}$ comme l'ensemble de liens D2D et $\mathcal{K} = \mathcal{K}_c \cup \mathcal{K}_d$ comme l'ensemble de tous les liens de communication.

Laissez h_{kl}^n être le gain du canal de l'émetteur du lien l au destinataire du lien k sur la sous-bande n . Nous désignons les vecteurs d'attribution de puissance et d'attribution de sous-bande comme $\mathbf{p} = [p_k^n]_{\forall k \in \mathcal{K}, \forall n \in \mathcal{N}}$ et $\boldsymbol{\rho} = [\rho_k^n]_{\forall k \in \mathcal{K}, \forall n \in \mathcal{N}}$ où p_k^n et ρ_k^n sont, respectivement, l'allocation de puissance et l'affectation de sous-bande du lien k sur la sous-bande n . Pour plus de commodité, nous adoptons les notations suivantes: $\mathcal{K}_k \equiv \mathcal{K}_c$ si $k \in \mathcal{K}_c$ et $\mathcal{K}_k \equiv \mathcal{K}_d$ si $k \in \mathcal{K}_d$. Le rapport du signal sur interférence plus bruit (SINR) obtenu par la liaison $k \in \mathcal{K}$ sur la sous-bande n peut être exprimé comme $\Gamma_k^n(\mathbf{p}^n, \boldsymbol{\rho}^n) = \frac{\rho_k^n p_k^n h_{kk}^n}{\sigma_k^n + \sum_{l \in \mathcal{K} \setminus \mathcal{K}_k} \rho_l^n p_l^n h_{kl}^n}$. Par conséquent, les taux réalisables du lien $k \in \mathcal{K}$ sur la sous-bande n et toutes les sous-bandes peuvent être exprimés comme $R_k^n(\mathbf{p}^n, \boldsymbol{\rho}^n) = \log_2(1 + \Gamma_k^n(\mathbf{p}^n, \boldsymbol{\rho}^n))$, et $R_k(\mathbf{p}, \boldsymbol{\rho}) = \sum_{n \in \mathcal{N}} R_k^n(\mathbf{p}^n, \boldsymbol{\rho}^n)$.

Nous supposons que chaque lien cellulaire ou lien D2D actif est attribué à une sous-bande, ce qui convient aux communications de liaison montante [13]. Notre objectif de conception est de maximiser le taux de somme pondéré de tous les liens D2D sélectionnés et les liens cellulaires tout en satisfaisant les taux minimum requis de liens cellulaires et des liens D2D actifs. Le problème de l'allocation des ressources envisagé peut maintenant être formulé comme

¹Chaque sous-bande peut être une porteuse et un sous-canal dans les réseaux sans fil à multi-porteuses (Ex. , Réseaux sans fil basés sur LTE).

$$\max_{\mathbf{p}, \boldsymbol{\rho}} \mathcal{R} = \sum_{k \in \mathcal{K}_c} \alpha R_k(\mathbf{p}, \boldsymbol{\rho}) + \sum_{k \in \mathcal{K}_d} (1 - \alpha) R_k(\mathbf{p}, \boldsymbol{\rho}) \quad (2.1a)$$

$$\text{s.t. } R_k(\mathbf{p}, \boldsymbol{\rho}) \geq R_k^{\min} \quad \forall k \in \mathcal{K}_c \quad (2.1b)$$

$$R_k(\mathbf{p}, \boldsymbol{\rho}) \geq \mathbf{I}\left\{\sum_{n \in \mathcal{N}} \rho_k^n = 1\right\} R_k^{\min} \quad \forall k \in \mathcal{K}_d, \quad (2.1c)$$

$$\sum_{n \in \mathcal{N}} p_k^n \leq P_k^{\max} \quad \forall k \in \mathcal{K} \quad (2.1d)$$

$$\sum_{k \in \mathcal{K}_c} \rho_k^n \leq 1, \quad \sum_{k \in \mathcal{K}_d} \rho_k^n \leq 1 \quad \forall n \in \mathcal{N} \quad (2.1e)$$

$$\sum_{n \in \mathcal{N}} \rho_k^n = 1 \quad \forall k \in \mathcal{K}_c, \quad \sum_{n \in \mathcal{N}} \rho_k^n \leq 1 \quad \forall k \in \mathcal{K}_d \quad (2.1f)$$

$$\rho_k^n \in \{0, 1\} \quad \forall k \in \mathcal{K} \quad \forall n \in \mathcal{N}, \quad (2.1g)$$

où α est un paramètre de poids qui contrôle le partage de spectre des liens cellulaires et D2D. Dans le problème (2.1), les contraintes (2.1b) et (2.1c) sont pour assurer le débit de données minimum des liaisons cellulaires et D2D, et $\mathbf{I}\{A\}$ désigne la fonction indicatrice, qui équivaut à 1 si A est vrai et équivaut à 0, sinon. En outre, les contraintes (2.1e), (2.1f) et (2.1g) sont imposées pour garantir que chaque sous-bande est affectée à un cellulaire et un lien D2D, et chaque lien peut exploiter une sous-bande. Pour résoudre ce problème, nous étudions d'abord la solution optimale d'allocation de puissance pour une affectation de sous-bande donnée. Ensuite, le problème original est transformé en un problème d'affectation de sous-bande, qui peut être résolu par l'Algorithme d'arrondi itératif proposé.

2.2.1.2 Algorithme optimal d'allocation de puissance

Notez que nous autorisons chaque lien cellulaire et lien D2D actif à utiliser une seule sous-bande dans la formulation du problème (2.1). Par conséquent, si le lien $m \in \mathcal{K}$ est attribué la sous-bande n exclusivement, alors la puissance optimale pour ce lien est P_m^{\max} et la contribution correspondante

de ce lien vers la valeur objective est

$$w_m^n \triangleq \begin{cases} \alpha \log_2 \left(1 + \frac{P_m^{\max} h_{mm}^n}{\sigma_m^n} \right) & \text{if } m \in \mathcal{K}_c \\ (1 - \alpha) \log_2 \left(1 + \frac{P_m^{\max} h_{mm}^n}{\sigma_m^n} \right) & \text{if } m \in \mathcal{K}_d. \end{cases} \quad (2.2)$$

Cependant, si le lien cellulaire k et le lien D2D l partagent la sous-bande n , l'allocation optimale de puissance doit être déterminée à partir du problème d'optimisation suivant

$$\begin{aligned} \max_{p_{Ck}^n, p_{Dl}^n} \quad & w_{kl}^n \triangleq \alpha R_{Ck}^n + (1 - \alpha) R_{Dl}^n \\ \text{s.t.} \quad & R_{Ck}^n \geq R_k^{\min}, R_{Dl}^n \geq R_l^{\min} \\ & p_{Ck}^n \in [0, P_k^{\max}], p_{Dl}^n \in [0, P_l^{\max}], \end{aligned} \quad (2.3)$$

where $R_{Ck}^n = \log_2 \left(1 + \frac{p_{Ck}^n h_{kk}^n}{\sigma_k^n + p_{Dl}^n h_{kl}^n} \right)$ and $R_{Dl}^n = \log_2 \left(1 + \frac{p_{Dl}^n h_{ll}^n}{\sigma_l^n + p_{Ck}^n h_{lk}^n} \right)$.

Pour ce problème, il a été prouvé dans [14] que si le problème est faisable, alors les puissances d'émission optimales $\mathbf{P} = (p_{Ck}, p_{Dl})$ ont la forme $\mathbf{P} \in \{(P_k^{\max}, p_{Dl}), (p_{Ck}, P_l^{\max})\}$. Définissons maintenant les quantités suivantes:

$$P_{Dl}^{(1)} \triangleq \max \left\{ \frac{(2^{R_l^{\min}} - 1)(P_k^{\max} h_{lk}^n + \sigma_l^n)}{h_{ll}^n}, 0 \right\}, P_{Ck}^{(1)} \triangleq \max \left\{ \frac{(2^{R_k^{\min}} - 1)(P_l^{\max} h_{kl}^n + \sigma_k^n)}{h_{kk}^n}, 0 \right\} \quad (2.4)$$

$$P_{Dl}^{(2)} \triangleq \min \left\{ \frac{1}{h_{kl}^n} \left(\frac{P_k^{\max} h_{kk}^n}{2^{R_k^{\min}} - 1} - \sigma_k^n \right), P_l^{\max} \right\}, P_{Ck}^{(2)} \triangleq \min \left\{ \frac{1}{h_{lk}^n} \left(\frac{P_l^{\max} h_{ll}^n}{2^{R_l^{\min}} - 1} - \sigma_l^n \right), P_k^{\max} \right\} \quad (2.5)$$

$$P_{Dl}^{(3)} \triangleq (-B_{Dl} + \sqrt{\Delta_{Dl}})/A_{Dl}, P_{Ck}^{(3)} \triangleq (-B_{Ck} + \sqrt{\Delta_{Ck}})/A_{Ck}, \quad (2.6)$$

Ensuite, l'allocation de puissance optimale du problème (2.3) se caractérise dans le théorème suivant.

Theorem 2.1. *Si le problème (2.3) est faisable, sa solution optimale d'allocation de puissance appartient à l'ensemble $\mathcal{S}^* \triangleq \mathcal{S}_1 \cup \mathcal{S}_2$, où*

$$\mathcal{S}_1 \triangleq \begin{cases} \{(P_k^{\max}, P_{Dl}^{(1)}), (P_k^{\max}, P_{Dl}^{(2)}), (P_k^{\max}, P_{Dl}^{(3)})\}, & \text{si } P_{Dl}^{(3)} \in [0, P_l^{\max}] \\ \{(P_k^{\max}, P_{Dl}^{(1)}), (P_k^{\max}, P_{Dl}^{(2)})\}, & \text{autrement} \end{cases} \quad (2.7)$$

$$\mathcal{S}_2 \triangleq \begin{cases} \{(P_{Ck}^{(1)}, P_l^{\max}), (P_{Ck}^{(2)}, P_l^{\max}), (P_{Ck}^{(3)}, P_l^{\max})\}, & \text{si } P_{Ck}^{(3)} \in [0, P_k^{\max}] \\ \{(P_{Ck}^{(1)}, P_l^{\max}), (P_{Ck}^{(2)}, P_l^{\max})\}, & \text{autrement.} \end{cases} \quad (2.8)$$

Comme \mathcal{S}^* contient au maximum 6 solutions possibles d'allocation de puissance, nous pouvons déterminer la solution optimale en examinant facilement toutes les solutions potentielles dans \mathcal{S}^* .

2.2.1.3 Algorithme d'arrondi itératif basé sur les graphes

Étant donné que l'allocation de puissance optimale pour une attribution de sous-bande donnée peut être déterminée comme dans la section précédente, le problème (2.1) peut être transformé en un problème d'affectation de sous-bande. Nous proposons de résoudre le problème d'affectation de sous-bande en utilisant l'approche par graphes dans laquelle chaque lien ou sous-bande peut être modélisé en un sommet et une affectation de sous-bande correspond à un hyper-bord dans le graphe. Indiquez V^0 et E^0 comme l'ensemble des sommets et des hyper-bords correspondant au problème de l'assignation de sous-bande original. Pour un ensemble de bords E , laissez $V(E)$, $V_c(E)$, $V_d(E)$ et $V_n(E)$ les ensembles de sommets, les liens cellulaires réels, les liens D2D, et les sous-bandes associées à E , respectivement. Pour décrire la décision d'affectation de sous-bande, nous introduisons une variable binaire x_e où $x_e = 1$ signifie que le bord e est activé et $x_e = 0$, sinon. En outre, laissez \mathbf{x} désigner le vecteur dont les éléments sont des variables d'affectation de sous-bande x_e associées à

tous les arêtes possibles. Le problème d'affectation de sous-bande, c'est-à-dire $\mathbf{IP}(V, E)$, est décrit comme suit

$$\begin{aligned}
\max_{x_e} \quad & \mathcal{R} = \sum_{e \in E} w_e x_e \\
\text{s.t.} \quad & C1 : D(v, E) = 1 \quad \forall v \in V \cap K_C \\
& C2 : D(v, E) \leq 1 \quad \forall v \in V_d(E) \cup V_n(E) \\
& C3 : x_e \in \{0, 1\} \quad \forall e \in E,
\end{aligned} \tag{2.9}$$

où $D(v, E) \triangleq \sum_{e \in E(v)} x_e$ désigne le degré de vertex v dans l'ensemble des bords E associé à \mathbf{x} .

Notez que $\mathbf{LP}(V, E)$ correspond au problème d'affectation d'origine de la sous-bande. Désignez $\mathbf{LP}(V, E)$ comme version de relaxation linéaire du problème $\mathbf{IP}(V, E)$. On peut observer que le problème $\mathbf{LP}(V^0, E^0)$ peut être résolu facilement par des solutions d'optimisation standard [15]. Cependant, résoudre le problème $\mathbf{LP}(V^0, E^0)$ aboutit souvent à des valeurs fractionnaires pour certains bords e ($0 < x_e < 1$). Pour résoudre ce problème, nous proposons un algorithme d'arrondi itératif dans lequel nous résolvons un problème de relaxation linéaire et effectuons un arrondissement approprié pour les variables fractionnaires dans chaque itération.

L'Algorithme d'arrondi itératif effectue les opérations suivantes en trois phases de chaque itération t . Dans la phase 1, il résout le problème de relaxation linéaire pour les liens inactifs et les sous-bandes disponibles correspondant au graphe avec l'ensemble des sommets $V^{(t)}$ et l'ensemble des bords $E^{(t)}$, ce qui entraîne deux ensembles de variables égales à des valeurs fractionnaires ($0 < x_e < 1$) et une ($x_e = 1$), à savoir $E_a^{(t)}$ et $E_u^{(t)}$, respectivement. Nous disposons ensuite les bords de l'ensemble $E_u^{(t)}$ avec des variables fractionnaires d'affectation de sous-bande en phase 2 sur la base desquelles nous employons la méthode de ratio local en phase 3 pour déterminer l'ensemble des affectations de sous-bandes supplémentaires $E_g^{(t)}$. Les phases 2 et 3 ont été conçues de manière appropriée pour minimiser la perte de performance due à l'arrondissement des variables fractionnaires d'affectation de sous-bande. Les bords de $E_a^{(t)} \cup E_g^{(t)}$ seront utilisés pour effectuer les assignations de sous-bande correspondantes pour les liaisons cellulaires et / ou D2D à chaque itération. Enfin, nous mettons à jour l'ensemble des sous-bandes disponibles et des liens inactifs et revenons à la phase 1 de la prochaine itération jusqu'à la convergence. Les principales opérations de l'algorithme

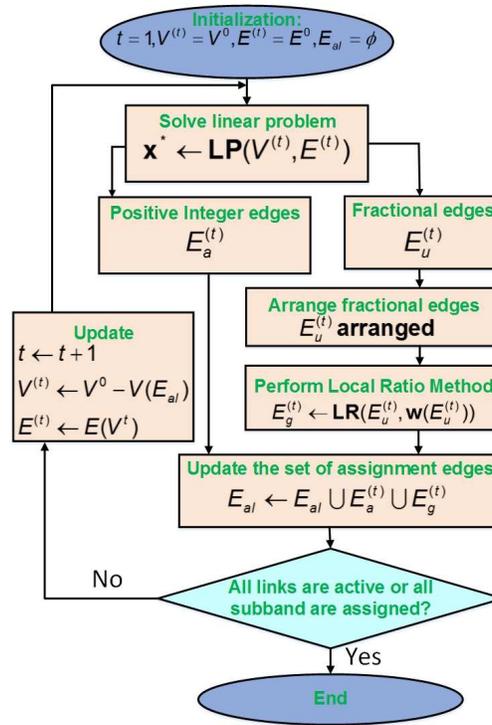


Figure 2.2 – L’organigramme de l’algorithme d’arrondi itératif

sont illustrées dans la Fig. 2.2. Nous présentons la performance de l’algorithme d’arrondi itératif dans le théorème suivant.

Theorem 2.2. *Une solution faisable du problème d’allocation de sous-bande original $\mathbf{IP}(V^0, E^0)$, offerte par Algorithme d’arrondi itératif, atteint au moins la moitié de la valeur d’objectif optimale du problème de relaxation linéaire $\mathbf{LP}(V^0, E^0)$.*

Ce théorème implique que nous pouvons toujours garantir que l’Algorithme d’arrondi itératif atteint au moins $1/2$ de la valeur objective optimale. Ceci est vrai même si la solution du problème $\mathbf{LP}(V^0, E^0)$ correspond à tous les arêtes fractionnaires.

2.2.1.4 Résultats numériques

Nous considérons la simulation où il existe une seule BS avec la zone de couverture de 500 m desservant $K_c = 20$ utilisateurs cellulaires distribués aléatoirement. En outre, il y a $N = 25$ sous-bandes, qui sont partagées par $K_c = 20$ liens cellulaires et $K_d = 30$ liens D2D, sauf indication du contraire. Nous comparons les performances de nos algorithmes proposés avec les algorithmes

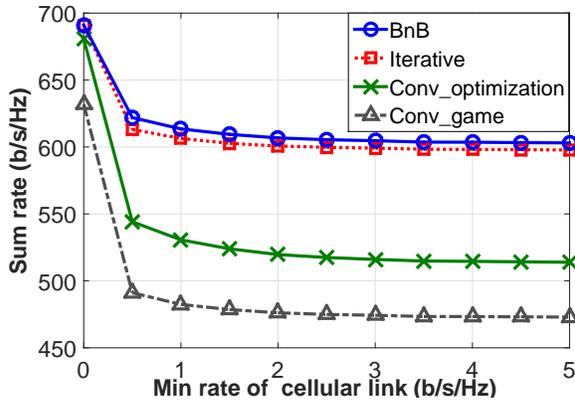


Figure 2.3 – Taux de somme par rapport au taux minimum de chaque lien cellulaire

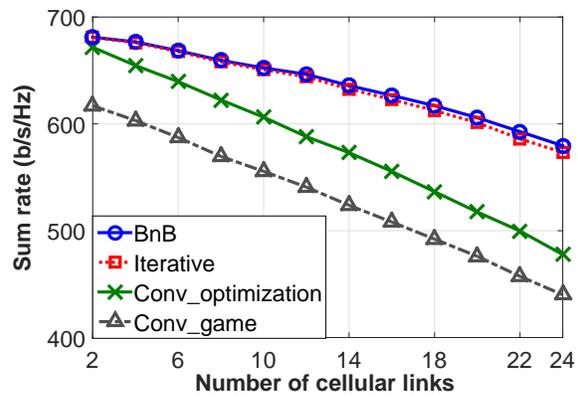


Figure 2.4 – Taux de somme du système par rapport à K_c comme $K_d = 30$

conventionnels développés pour le scénario I dans [16] et [17]. Le premier algorithme conventionnel, considéré comme l'algorithme conventionnel basé sur l'optimisation, contrairement au travail [17] qui a adopté une approche basée sur la théorie du jeu.

Dans la Fig. 2.3, nous démontrons le taux d'addition du système par rapport au taux minimal requis de liaisons cellulaires R_c^{\min} . On peut voir que le taux de somme du système atteint la valeur maximale comme $R_c^{\min} = 0$. C'est parce que lorsque en $R_c^{\min} = 0$, les liens D2D ont plus d'avantages que les liens cellulaires pour accéder à de bonnes sous-bandes grâce à la courte portée des liens D2D. Par conséquent, le taux des liaisons D2D devient plus élevé pour le plus petit taux minimum requis de chaque lien cellulaire. On peut également observer que le taux d'addition du système diminue de manière significative lorsque R_c^{\min} augmente de zéro avant d'être saturé à une valeur fixe.

Fig. 2.4 montre les variations du taux d'addition du système avec le nombre de liens cellulaires K_c lorsque nous choisissons $K_d = 30$. Cette figure démontre que le taux d'addition du système diminue avec le nombre de liens cellulaires. En fait, au fur et à mesure que K_c augmente, le nombre de liens D2D actifs est réduit, ce qui entraîne une diminution du taux de somme du système. Cependant, lorsque K_c est suffisamment grand, augmenter K_c conduit au scénario où les liens D2D actifs doivent partager des sous-bandes avec des liens cellulaires et le nombre de liens D2D actifs diminue.

2.2.2 Allocation de ressources éconergétiques pour les communications D2D dans les réseaux cellulaires

Dans cette contribution, nous étudions l'allocation conjointe de sous-canal et de puissance qui maximise l'EE minimale pondérée D2D et garantit les débits de données minimaux des liens cellulaires. Plus précisément, nous apportons les contributions suivantes.

- Nous formulons un problème général d'allocation de ressources éconergétiques en tenant compte de multiples liaisons cellulaires et D2D où chaque lien D2D peut réutiliser les ressources spectrales de plusieurs liaisons cellulaires. Nous caractérisons d'abord la solution optimale d'allocation de puissance pour un lien cellulaire en fonction de la puissance optimale de la liaison D2D co-canal. Sur la base de ce résultat, nous transformons le problème RA initial en problème RA uniquement pour les liens D2D.
- Nous proposons l'algorithme basé sur le dual qui résout le problème d'allocation de ressources dans le domaine dual. En particulier, nous adoptons la technique de programmation fractionnelle max-min pour transformer itérativement le problème de l'allocation de ressources en un problème de programmation non linéaire mixte (MINLP). Ensuite, nous résolvons le problème de MINLP sous-jacent en utilisant l'approche de décomposition duale.
- De nombreux résultats numériques sont présentés pour évaluer la performance des algorithmes développés. Plus précisément, il est démontré que les valeurs objectives obtenues par l'algorithme dual sont proches de celles de l'algorithme BnB optimal et significativement plus élevées que celles de l'algorithme conventionnel et la conception de l'allocation des ressources par spectralement efficace.

2.2.2.1 Modèle du système

Nous considérons le scénario d'allocation de ressources de liaison montante où les liaisons cellulaires partagent le même spectre avec de multiples liaisons D2D dans un seul système de macrocellules. Nous supposons que K liens cellulaires ascendants dans un ensemble $\mathcal{K} = \{1, \dots, K\}$ occupant des sous-canaux orthogonaux K dans l'ensemble $\mathcal{N} = \{1, \dots, K\}$ dans la cellule considérée. De plus,

nous supposons que l'ensemble $\mathcal{L} = \{1, \dots, L\}$ de liens D2D transmet des données en utilisant le même ensemble des sous-canaux.²

Nous présentons les vecteurs de puissance alloués comme $\mathbf{p} = [\mathbf{p}_C, \mathbf{p}_D]$ pour tous les liens, où $\mathbf{p}_C = [p_{Ck}^k]_{\forall k \in \mathcal{K}}$, $\mathbf{p}_D = [p_{Dl}^k]_{\forall l \in \mathcal{L}, \forall k \in \mathcal{K}}$. p_{Ck}^k et p_{Dl}^k désignent les puissances d'émission attribuées sur le sous-canal k du lien cellulaire $k \in \mathcal{K}$ et le lien D2D $l \in \mathcal{L}$, respectivement. Ensuite, le SINR réalisé par le lien cellulaire k et le lien D2D l sur le sous-canal k peuvent être exprimés, respectivement, comme $\Gamma_{Ck}^k(\mathbf{p}, \boldsymbol{\rho}) = \frac{p_{Ck}^k h_{kk}^k}{\sigma_k^k + \sum_{l \in \mathcal{L}} \rho_l^k p_{Dl}^k h_{kl}^k}$, et $\Gamma_{Dl}^k(\mathbf{p}, \boldsymbol{\rho}) = \frac{\rho_l^k p_{Dl}^k h_{ll}^k}{\sigma_l^k + p_{Ck}^k h_{lk}^k}$. Les taux de données du lien cellulaire $k \in \mathcal{K}$ sur son sous-canal k , du lien D2D $l \in \mathcal{L}$ sur le sous-canal k , et du lien D2D l sur tous les sous-canaux peuvent être calculés comme

$$R_{Ck}^k(\mathbf{p}, \boldsymbol{\rho}) = \log_2 \left(1 + \Gamma_{Ck}^k(\mathbf{p}, \boldsymbol{\rho}) \right) \quad (2.10)$$

$$R_{Dl}^k(\mathbf{p}, \boldsymbol{\rho}) = \log_2 \left(1 + \Gamma_{Dl}^k(\mathbf{p}, \boldsymbol{\rho}) \right) \quad (2.11)$$

$$R_{Dl}(\mathbf{p}, \boldsymbol{\rho}) = \sum_{k \in \mathcal{K}} \rho_l^k R_{Dl}^k(\mathbf{p}, \boldsymbol{\rho}). \quad (2.12)$$

Nous supposons que la puissance totale *consommée* du lien D2D l peut être exprimée comme [18, 19]

$$P_{Dl}^{\text{total}} = 2P_0^l + \alpha_l \sum_{k \in \mathcal{K}} \rho_l^k p_{Dl}^k, \quad (2.13)$$

où $2P_0^l$ représente la puissance du circuit fixe de l'émetteur et du récepteur de la liaison D2D l , et $\alpha_l > 1$ est un facteur qui tient compte de l'efficacité de l'amplificateur d'émission et des pertes de l'alimentation.

L'objectif de notre conception d'allocation de ressources est de maximiser l'EE pondéré minimale des liens D2D alors que le débit minimal de données des liaisons cellulaires est garanti. Par conséquent, cette conception peut être formulée comme le problème d'allocation de ressources éconergétique suivant pour atteindre l'équité max-min dans EE EE pesée pour les liens D2D

²Les sous-canaux orthogonaux considérés peuvent être des sous-porteuses ou des sous-canaux dans le système OFDMA ou simplement des canaux dans le système FDMA.

$$\max_{\mathbf{p}, \boldsymbol{\rho}} \min_{l \in \mathcal{L}} \frac{w_l R_{Dl}(\mathbf{p}, \boldsymbol{\rho})}{P_{Dl}^{\text{total}}} \quad (2.14a)$$

$$\text{s.t. } R_{Ck}^k(\mathbf{p}, \boldsymbol{\rho}) \geq R_{Ck}^{\min}, \forall k \in \mathcal{K} \quad (2.14b)$$

$$p_{Ck}^k \leq P_{Ck}^{\max}, \forall k \in \mathcal{K}, \quad (2.14c)$$

$$\sum_{k \in \mathcal{K}} \rho_l^k p_{Dl}^k \leq P_{Dl}^{\max}, \forall l \in \mathcal{L}, \quad (2.14d)$$

$$\sum_{l \in \mathcal{L}} \rho_l^k \leq 1, \forall k \in \mathcal{N} \quad (2.14e)$$

$$\rho_l^k \in \{0, 1\}, \forall k \in \mathcal{K}, l \in \mathcal{L}. \quad (2.14f)$$

2.2.2.2 Transformation du Problème

Pour résoudre le problème (2.14), décrivez d'abord l'allocation de puissance optimale du lien D2D $l \in \mathcal{L}$ sur le sous-canal $k \in \mathcal{N}$ dans la proposition suivante.

Proposition 2.1. *Si le lien D2D $l \in \mathcal{L}$ est autorisé à réutiliser le sous-canal $k \in \mathcal{N}$ de lien cellulaire k , alors sa puissance sur le sous-canal k , est $p_{Dl}^k = \frac{1}{h_{kl}^k} \left(\frac{p_{Ck}^k h_{kk}^k}{2^{R_{Ck}^{\min}} - 1} - \sigma_k^k \right) \in [0, P_{Dlk}^{\max}]$, où p_{Ck}^k est la puissance du lien cellulaire k , et $P_{Dlk}^{\max} = \min \left\{ P_{Dl}^{\max}, \frac{1}{h_{kl}^k} \left(\frac{P_{Ck}^{\max} h_{kk}^k}{2^{R_{Ck}^{\min}} - 1} - \sigma_k^k \right) \right\}$.*

De la Proposition 2.1, les taux de données du lien D2D l sur le sous-canal k et tous les sous-canaux peuvent être réécrits comme

$$\hat{R}_{Dl}^k(\mathbf{p}_D, \boldsymbol{\rho}) = \rho_l^k \log_2 \left(1 + \frac{p_{Dl}^k}{a_{kl} + b_{kl} p_{Dl}^k} \right) \quad (2.15)$$

$$\hat{R}_{Dl}(\mathbf{p}_D, \boldsymbol{\rho}) = \sum_{k \in \mathcal{N}} \hat{R}_{Dl}^k(\mathbf{p}_D, \boldsymbol{\rho}), \quad (2.16)$$

où

$$a_{kl} \triangleq \frac{\sigma_l^k}{h_{ll}^k} + \frac{(2^{R_k^{\min}} - 1) h_{lk}^k \sigma_k^k}{h_{kk}^k h_{ll}^k} \quad (2.17)$$

$$b_{kl} \triangleq \frac{(2^{R_k^{\min}} - 1) h_{lk}^k h_{kl}^k}{h_{kk}^k h_{ll}^k}, \quad (2.18)$$

et la puissance d'émission allouée doit satisfaire

$$p_{Dl}^k \leq P_{Dlk}^{\max}, \forall k \in \mathcal{N}, \forall l \in \mathcal{L}. \quad (2.19)$$

Par conséquent, le problème (2.14) équivaut à ce qui suit

$$\begin{aligned} \max_{(\mathbf{p}_D, \boldsymbol{\rho})} \min_{l \in \mathcal{L}} \frac{w_l \hat{R}_{Dl}(\mathbf{p}_{Dl}, \boldsymbol{\rho})}{P_{Dl}^{\text{total}}(\mathbf{p}_{Dl}, \boldsymbol{\rho})} \\ \text{s.t.} \quad (2.14d), (2.14f), (2.14e), (2.19). \end{aligned} \quad (2.20)$$

Pour résoudre le problème (2.20), nous considérons le problème d'optimisation suivant

$$\begin{aligned} \max_{\mathbf{p}_D, \boldsymbol{\rho}} \eta(\zeta, \mathbf{p}_D, \boldsymbol{\rho}) \triangleq \min_{l \in \mathcal{L}} \left[w_l \hat{R}_{Dl}(\mathbf{p}_D, \boldsymbol{\rho}) - \zeta P_{Dl}^{\text{total}}(\mathbf{p}_D, \boldsymbol{\rho}) \right] \\ \text{s.t.} \quad (2.14d), (2.14e), (2.14f), (2.19). \end{aligned} \quad (2.21)$$

Supposons que $\eta^*(\zeta) = \eta(\zeta, \mathbf{p}_D^*, \boldsymbol{\rho}^*)$ où $(\mathbf{p}_D^*, \boldsymbol{\rho}^*)$ est la solution optimale du problème (2.21) et \mathcal{D} désigne l'ensemble des solutions possibles du problème (2.20). Ensuite, nous pouvons caractériser la solution optimale du problème (2.21) dans le théorème suivant, qui est adopté à partir de [20].

Theorem 2.3. $\eta^*(\zeta)$ est une fonction décroissante de ζ . En outre, si nous avons

$$\begin{aligned} \max_{(\mathbf{p}_D, \boldsymbol{\rho}) \in \mathcal{D}} \min_{l \in \mathcal{L}} [w_l \hat{R}_{Dl}(\mathbf{p}_D, \boldsymbol{\rho}) - \zeta^* P_{Dl}^{\text{total}}(\mathbf{p}_D, \boldsymbol{\rho})] \\ = \min_{l \in \mathcal{L}} [w_l \hat{R}_{Dl}(\mathbf{p}_D^*, \boldsymbol{\rho}^*) - \zeta^* P_{Dl}^{\text{total}}(\mathbf{p}_D^*, \boldsymbol{\rho}^*)] = 0 \end{aligned} \quad (2.22)$$

alors $\zeta^* = \min_{l \in \mathcal{L}} \frac{w_l \hat{R}_{Dl}(\mathbf{p}_D^*, \boldsymbol{\rho}^*)}{P_{Dl}^{\text{total}}(\mathbf{p}_D^*, \boldsymbol{\rho}^*)}$ est la solution optimale de (2.20).

Theorem 2.3 nous permet de transformer un problème fractionnaire max-min général (2.20) en un problème d'optimisation non fractionnaire avec le paramètre ζ . En outre, la solution optimale du problème (2.20), ζ^* , peut être trouvée si $\eta^*(\zeta^*) = 0$. Puisque $\eta^*(\zeta)$ est une fonction décroissante de ζ , on peut voir que ζ^* peut être effectivement déterminé par la méthode de bisection.

2.2.2.3 Algorithme basé sur le dual

Algorithm 2.1. Algorithme basé sur le dual

- 1: Initialisation: $\zeta^{\max}, \zeta^{\min}$
 - 2: **repeat**
 - 3: Initialisation: Choisir $\zeta = \frac{1}{2}(\zeta^{\min} + \zeta^{\max})$, $\boldsymbol{\lambda}^{(0)}, \mu_l^{(0)} = \frac{1}{L}$, taille de pas $\theta^{(0)}$, et $\kappa^{(0)}$
 - 4: **repeat**
 - 5: Étape 1: pour tous $k \in \mathcal{K}, l \in \mathcal{L}$, calculer p_{Dl}^{k*} selon (2.29)
 - 6: Étape 2: pour tous $k \in \mathcal{K}$, effectuer l'allocation de sous-canal en suivant (2.30)
 - 7: Étape 3: mise à jour des variables duales $\boldsymbol{\lambda}$, and $\boldsymbol{\mu}$ par la méthode sous-gradient.
 - 8: **until** Convergence
 - 9: Sortie $z^* = \min_{l \in \mathcal{L}} \left[w_l \hat{R}_{Dl}(\mathbf{p}_D^*, \boldsymbol{\rho}^*) - \zeta P_{Dl}^{\text{total}}(\mathbf{p}_D^*, \boldsymbol{\rho}^*) \right]$
 - 10: Si $z^* > 0$, $\zeta^{\min} = \zeta$; autrement $\zeta^{\max} = \zeta$
 - 11: **until** Convergence de ζ
 - 12: Sortie $\mathbf{p}_{Dl}^*, \boldsymbol{\rho}^*$, et $\zeta^* = \min_{l \in \mathcal{L}} \frac{w_l \hat{R}_{Dl}(\mathbf{p}_{Dl}^*, \boldsymbol{\rho}^*)}{P_{Dl}^{\text{total}}(\mathbf{p}_{Dl}^*, \boldsymbol{\rho}^*)}$
-

Dans cette section, nous proposons un algorithme basé sur le dual pour résoudre le problème (2.20), qui est résumé dans Algorithm2.1. L'algorithme comprend deux boucles itératives. Dans la boucle externe, nous adoptons la technique de programmation fractionnaire max-min étudiée dans Theorem 2.3 pour atteindre la valeur optimale de ζ pour le problème (2.20). Dans la boucle interne (lignes 4-8), nous résolvons le problème (2.21) pour un ζ donné en utilisant la méthode de décomposition duale.

Dans ce qui suit, nous montrons comment résoudre le problème (2.21) pour une valeur donnée de ζ . Tout d'abord, on peut remarquer que le problème (2.21) équivaut au problème suivant

$$\begin{aligned}
 & \max_{z, \mathbf{p}_D, \boldsymbol{\rho}} \quad z \\
 & \text{s.t.} \quad w_l \hat{R}_{Dl}(\mathbf{p}_{Dl}, \boldsymbol{\rho}) - \zeta P_{Dl}^{\text{total}}(\mathbf{p}_D, \boldsymbol{\rho}) \geq z, \forall l \in \mathcal{L} \\
 & \quad (2.14d), (2.14e), (2.14f), (2.19).
 \end{aligned} \tag{2.23}$$

Pour aborder le problème (2.23), nous considérons son Lagrangien comme

$$\begin{aligned}
 & L_D(\mathbf{p}_D, \boldsymbol{\rho}, z, \zeta, \boldsymbol{\lambda}, \boldsymbol{\mu}) \\
 & = z(1 - \sum_{l \in \mathcal{L}} \mu_l) + \sum_{l \in \mathcal{L}} \mu_l \left[w_l \hat{R}_{Dl}(\mathbf{p}_D, \boldsymbol{\rho}) - \zeta P_{Dl}^{\text{total}}(\mathbf{p}_D, \boldsymbol{\rho}) \right] + \sum_{l \in \mathcal{L}} \lambda_l (P_{Dl}^{\max} - \sum_{k \in \mathcal{N}} \rho_l^k p_{Dl}^k), \text{ où } \boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_L]^T \\
 & \text{et } \boldsymbol{\mu} = [\mu_1, \dots, \mu_L]^T \text{ représentent les multiplicateurs de Lagrange.}
 \end{aligned}$$

Ensuite, la fonction duale peut être écrite comme

$$\bar{L}_D(\zeta, \boldsymbol{\lambda}, \boldsymbol{\mu}) \triangleq \max_{z, \mathbf{p}_D \in \mathcal{X}, \boldsymbol{\rho} \in \mathcal{C}} L_D(\mathbf{p}_D, \boldsymbol{\rho}, z, \zeta, \boldsymbol{\lambda}, \boldsymbol{\mu}), \tag{2.24}$$

où $\mathcal{X} = \{\mathbf{p}_D | p_{Dl}^k \leq P_{Dlk}^{\max}, \forall k \in \mathcal{N}, \forall l \in \mathcal{L}\}$, et $\mathcal{C} = \{\boldsymbol{\rho} | \sum_{l \in \mathcal{L}} \rho_l^k \leq 1, \forall k \in \mathcal{N}, \text{ et } \rho_l^k \in \{0, 1\}, \forall k \in \mathcal{K}, l \in \mathcal{L}\}$.

Ensuite, le problème dual peut être déclaré comme

$$\hat{L}_D(\zeta) \triangleq \min_{\boldsymbol{\lambda}, \boldsymbol{\mu} \geq 0} \bar{L}_D(\zeta, \boldsymbol{\lambda}, \boldsymbol{\mu}). \quad (2.25)$$

Pour résoudre le problème dual (2.25), nous étudions le problème (2.24) pour le $\boldsymbol{\lambda}$ et $\boldsymbol{\mu}$. En particulier, nous avons

$$\begin{aligned} \bar{L}_D(\zeta, \boldsymbol{\lambda}, \boldsymbol{\mu}) &= \max_{z, \mathbf{p}_D \in \mathcal{X}, \boldsymbol{\rho} \in \mathcal{C}} L_D(\mathbf{p}_D, \boldsymbol{\rho}, z, \zeta, \boldsymbol{\lambda}, \boldsymbol{\mu}) \\ &= \max_{z, \mathbf{p}_D \in \mathcal{X}, \boldsymbol{\rho} \in \mathcal{C}} \sum_{k \in \mathcal{N}} \sum_{l \in \mathcal{L}} \rho_l^k f_l^k(p_{Dl}^k) + z(1 - \sum_{l \in \mathcal{L}} \mu_l) + \sum_{l \in \mathcal{L}} (\lambda_l P_{Dl}^{\max} - 2\zeta \mu_l P_0^l), \end{aligned} \quad (2.26)$$

où $f_l^k(p_{Dl}^k) \triangleq \mu_l w_l \hat{R}_{Dl}^k(\mathbf{p}_D, \boldsymbol{\rho}) - (\zeta \alpha_l \mu_l + \lambda_l) p_{Dl}^k$.

Pour obtenir la solution optimale non triviale du problème dual (2.25), l'identité $\sum_{l \in \mathcal{L}} \mu_l = 1$ doit être toujours maintenue. En outre, le problème (2.26) peut être décomposé en N problèmes individuels d'allocation de ressources pour N sous-canaux où le problème d'allocation de ressources pour le sous-canal $k \in \mathcal{N}$ peut être déclaré comme

$$\bar{L}_D^k(\zeta, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \max_{\mathbf{p}_D \in \mathcal{X}, \boldsymbol{\rho} \in \mathcal{C}} \sum_{l \in \mathcal{L}} \rho_l^k f_l^k(p_{Dl}^k). \quad (2.27)$$

Pour le problème(2.27), supposons que le lien D2D l soit affecté sous-canal $k \in \mathcal{N}$ alors nous avons

$$p_{Dl}^{k*} = \operatorname{argmax}_{p_{Dl}^k \in \mathcal{X}_l} f_l^k(p_{Dl}^k). \quad (2.28)$$

Notez que nous devons avoir $\mu_l > 0$ parce que si $\mu_l = 0$, nous avons $p_{Dl}^{k*} = 0, \forall k \in \mathcal{N}$, ce qui ne peut être la solution optimale de problème (2.26). En outre, le problème (2.28) peut être résolu en résolvant $\frac{\partial f_l^k(p_{Dl}^k)}{\partial p_{Dl}^k} = 0$, où $\frac{\partial f_l^k(p_{Dl}^k)}{\partial p_{Dl}^k}$ est la dérivée de premier ordre de $f_l^k(p_{Dl}^k)$. Ensuite, on peut vérifier que la résolution de $\frac{\partial f_l^k(p_{Dl}^k)}{\partial p_{Dl}^k} = 0$ équivaut à résoudre $A_{kl}(p_{Dl}^k)^2 + 2B_{kl}p_{Dl}^k + C_{kl} = 0$. Par

conséquent, la solution optimale du lien D2D l qui maximise $f_l^k(p_{Dl}^k)$ est donnée par

$$p_{Dl}^{k*} = \left[\frac{-B_{kl}^d + \sqrt{\Delta_{kl}^d}}{A_{kl}^d} \right]_0^{P_{Dl}^{\max}}, \quad (2.29)$$

En résumé, en résolvant le problème (2.26), nous pouvons obtenir l'allocation de puissance optimale pour tout lien D2D sur le sous-canal $k \in \mathcal{N}$. Rappelons que nous avons supposé que chaque sous-canal peut être attribué au plus un lien D2D; donc, pour tous les sous-canaux $k \in \mathcal{N}$, nous avons

$$\rho_l^{k*} = \begin{cases} 1 & \text{if } l = \operatorname{argmax}_{l \in \mathcal{L}} f_l^k(p_{Dl}^{k*}) \\ 0 & \text{otherwise.} \end{cases} \quad (2.30)$$

Jusqu'à présent, nous avons présenté la solution d'allocation de ressources pour $\boldsymbol{\lambda}$, $\boldsymbol{\mu}$ données. Donc, la tâche restante est de résoudre le problème (2.25), qui peut être complété par la méthode du sous-gradient [21]. Enfin, la performance obtenue par Algorithm 2.1, qui résout le problème (2.20), est indiquée dans le proposition suivante

Proposition 2.2. *Algorithme 2.1 retourne une solution possible du problème*

(2.20) avec ζ^* , \mathbf{p}_D^* , $\boldsymbol{\rho}^*$, $\boldsymbol{\lambda}^*$, et $\boldsymbol{\mu}^*$ à la fin de sa première phase. De plus, si $\sum_{k \in \mathcal{N}} \rho_l^{k*} p_{Dl}^{k*} \leq P_{Dl}^{\max}$, $\lambda_l^*(P_D^{\max} - \sum_{k \in \mathcal{N}} \rho_l^{k*} p_{Dl}^{k*}) = 0$, et $R_{Dl}(\mathbf{p}_D^*, \boldsymbol{\rho}^*) - \zeta^* P_{Dl}^{\text{total}}(\mathbf{p}_D^*, \boldsymbol{\rho}^*) = 0, \forall l \in \mathcal{L}$, cette solution faisable est la solution optimale du problème (2.20).

2.2.2.4 Résultats numériques

Nous considérons le paramètre de simulation avec la station de base située au centre, $K = 20$ d'utilisateurs cellulaires et $L = 4$ liaisons D2D réparties aléatoirement dans une zone de 500m x 500m, et $N = 20$ sous-canaux pour les communications de liaison montante. Nous évaluons la performance de l'algorithme proposé en le comparant avec l'algorithme dans [22]. L'algorithme RBR est obtenu en résolvant la version de relaxation du problème d'origine et en effectuant un arrondi intelligent, qui est décrit en détail dans le chapitre 6. Fig. 2.5 indique que l'EE des liens D2D obtenus par nos algorithmes proposés est significativement plus élevée que celle de l'algorithme conventionnel, par exemple, à $d_{\max} = 150$ m, les algorithmes proposés peuvent atteindre plus de 90

% de l'EE supérieure, ce qui représente environ 300% celui de l'algorithme conventionnel et environ 130% celui de la solution de maximisation SE.

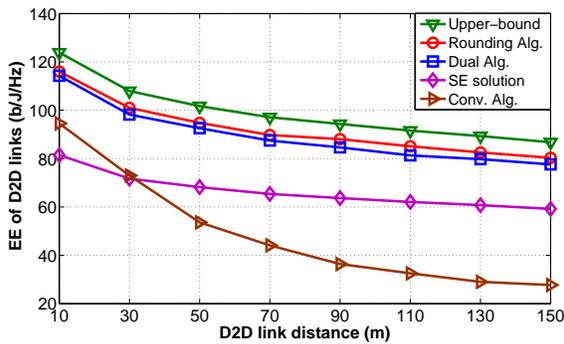


Figure 2.5 – EE minimum de liens D2D par rapport à la distance de liaison D2D

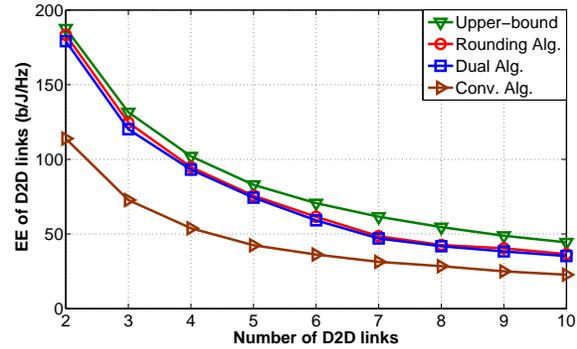


Figure 2.6 – EE minimum des liens D2D par rapport au nombre de liens D2D

Enfin, la Fig. 2.6 montre que l'EE atteinte des liens D2D diminue à mesure que le nombre de liens D2D augmente. L'écart de performance entre les algorithmes proposés et les algorithmes conventionnels diminue également à mesure que le nombre de liens D2D augmente. En effet, comme le système prend en charge plus de liens D2D, les ressources disponibles pour chaque lien D2D deviennent plus petites, ce qui entraîne une diminution de la EE obtenue des liens D2D.

2.2.3 Sélection du mode et allocation de la ressource conjointes pour les communications D2D relais

Dans cette contribution, nous étudions la sélection du mode, l'attribution du groupe de ressources (RG) et le problème de contrôle de puissance conjointement pour les réseaux cellulaires sous-étendus D2D qui vise à maximiser le taux de somme du système en considérant les contraintes de débit minimum des liaisons cellulaires et D2D. Le problème de l'allocation des ressources est formulé sous la forme d'un problème MINLP (programmation mixte en nombre entier non linéaire). Pour résoudre ce problème *de manière optimale*, nous étudions d'abord l'allocation de puissance optimale pour une sélection de mode et une solution d'affectation RG données. Sur la base de ces résultats, le problème initial de l'allocation des ressources peut être transformé en un problème d'allocation RG, qui peut être résolu de manière optimale par la méthode Hongroise. Des études numériques approfondies démontrent que la conception proposée dépasse considérablement les systèmes de communication D2D existants en mode fixe direct ou relais.

2.2.3.1 Modèle du système

Nous considérons le lien ascendant d'un seul système de macrocellules où K liens cellulaires dans l'ensemble $\mathcal{K} = \{1, \dots, K\}$ partagent le même spectre de K groupes de ressource (RG) dans l'ensemble $\mathcal{N} = \{1, \dots, K\}$ avec L liens D2D dans l'ensemble $\mathcal{L} = \{1, \dots, L\}$. Nous supposons que le lien cellulaire $k \in \mathcal{K}$ a été pré-attribué RG $k \in \mathcal{N}$, qui comprend m_k sous-canaux consécutifs.³ Nous supposons également que chaque lien D2D réutilise la ressource d'un RG, et chaque RG est affecté à au plus un lien D2D.

Laissez $\boldsymbol{\rho}$ être une matrice de capture des décisions d'allocation de ressources binaires des liens D2D où $[\boldsymbol{\rho}]_{kl} = \rho_{kl} = 1$ si lien D2D l est attribué RG k et $\rho_{kl} = 0$, sinon. En outre, chaque émetteur D2D peut communiquer avec son récepteur correspondant via le mode direct ou relais (assisté par un relais). Soit $\mathbf{x} = [x_1, \dots, x_L]$ le vecteur de décision de sélection de mode binaire pour tous les liens D2D où $x_l = 1$ si le lien D2D l fonctionne en mode direct et $x_l = 0$, sinon. Nous supposons également que la sélection de relais pour chaque liaison D2D a été prédéterminée où chaque liaison D2D $l \in \mathcal{L}$ peut être assisté par son relais assigné. Nous supposons que D2D link l est pris en charge par le relais r_l dans l'ensemble de relais $\mathcal{R} = \{r_1, \dots, r_L\}$. Indiquez h_{ab}^n en tant que gain de canal de l'émetteur du lien ou du relais b au destinataire du lien ou du relais a sur RG n .

Nous désignons \mathbf{p}_C et \mathbf{p}_D sont le vecteur d'allocation de puissance des liens cellulaires et D2D où $[\mathbf{p}_C]_k = p_{Ck}$ et $[\mathbf{p}_D]_l = p_{Dl}$ indiquent la puissance d'émission du lien cellulaire k et du lien D2D l , respectivement. Le taux de données du lien cellulaire k sur son RG sans aucun lien D2D co-canal peut être exprimé comme $R_{Ck}^{(o)} = m_k \log_2 \left(1 + p_{Ck} h_{kk}^k / \sigma^2 \right)$ où σ^2 indique le bruit thermique, et le taux de données en b/s/Hz est normalisé par la bande passante d'un sous-canal.

Dans ce travail, nous autorisons les nœuds D2D dans chaque lien D2D à communiquer les uns aux autres en mode direct ou en mode relais. En mode direct, l'émetteur D2D communique directement avec son récepteur D2D. Cependant, en mode relais, nous supposons que la stratégie de relais Decode and Forward (DF) est utilisée où chaque période de communication est divisée en deux intervalles égaux correspondant à l'émetteur D2D pour relayer (D-R) phase de communication et relais vers le récepteur D2D (R-D) phase de communication.

³Ceci est le cas dans le système LTE de liaison montante utilisant SC-FDMA, chaque sous-canal est équivalent à un bloc de ressource (RB) et RG est un groupe de RB contigus affecté à un lien cellulaire particulier.

Mode direct: Si RG k est affecté au lien D2D l , les taux de données du lien cellulaire k et D2D lien l sont décrits comme

$$R_{Dlk}^{(d)} = m_k \log_2 \left(1 + \frac{p_{Dl} h_{ll}^k}{\sigma^2 + p_{Ck} h_{lk}^k} \right) \quad (2.31)$$

$$R_{Ckl}^{(d)} = m_k \log_2 \left(1 + \frac{p_{Ck} h_{kk}^k}{\sigma^2 + p_{Dl} h_{kl}^k} \right). \quad (2.32)$$

Mode de relais: Laissez $\mathbf{p}_R = [p_{R_1}, \dots, p_{R_L}]$ être le vecteur d'allocation de puissance des relais pour supporter leurs liens D2D. Ensuite, les débits de données du lien cellulaire k dans les première et deuxième phases de communication peuvent être exprimés comme

$$R1_{Ckl}^{(r)} = \frac{m_k}{2} \log_2 \left(1 + \frac{p_{Ck} h_{kk}^k}{\sigma^2 + p_{Dl} h_{kl}^k} \right) \quad (2.33)$$

$$R2_{Ckl}^{(r)} = \frac{m_k}{2} \log_2 \left(1 + \frac{p_{Ck} h_{kk}^k}{\sigma^2 + p_{R_l} h_{kr_l}^k} \right). \quad (2.34)$$

En outre, les débits de données obtenus sur les liaisons D-R et R-D dans les première et deuxième phases de communication peuvent être calculés, respectivement, comme

$$R1_{Dlk}^{(r)} = \frac{m_k}{2} \log_2 \left(1 + \frac{p_{Dl} h_{r_l l}^k}{\sigma^2 + p_{Ck} h_{r_l k}^k} \right) \quad (2.35)$$

$$R2_{Dlk}^{(r)} = \frac{m_k}{2} \log_2 \left(1 + \frac{p_{R_l} h_{lr_l}^k}{\sigma^2 + p_{Ck} h_{lk}^k} \right). \quad (2.36)$$

Enfin, les taux de données du lien cellulaire k et D2D lien l peuvent être écrits, respectivement, comme

$$R_{Ckl}^{(r)} = R1_{Ckl}^{(r)} + R2_{Ckl}^{(r)} \quad (2.37)$$

$$R_{Dlk}^{(r)} = \min \left\{ R1_{Dlk}^{(r)}, R2_{Dlk}^{(r)} \right\}. \quad (2.38)$$

2.2.3.2 Formulation du problème

En utilisant les notes ci-dessus, les taux de données du lien cellulaire k et du lien D2D l peuvent être exprimés comme

$$R_{Ck} = (1 - \sum_{l \in \mathcal{L}} \rho_{kl}) R_{Ck}^{(o)} + \sum_{l \in \mathcal{L}} \rho_{kl} [x_l R_{Ckl}^{(d)} + (1 - x_l) R_{Ckl}^{(r)}] \quad (2.39)$$

$$R_{Dl} = \sum_{n \in \mathcal{N}} \rho_{nl} [x_l R_{Dln}^{(d)} + (1 - x_l) R_{Dln}^{(r)}]. \quad (2.40)$$

Nous proposons de faire conjointement le problème de sélection du mode, l'affectation de ressources et l'allocation de puissance qui vise à maximiser le taux de somme de toutes les communications, tandis que le débit de données minimal des liaisons cellulaires et D2D est satisfait dans les conditions suivantes.

$$\max_{\mathbf{p}, \rho, \mathbf{x}} \sum_{k \in \mathcal{K}} R_{Ck} + \sum_{l \in \mathcal{L}} R_{Dl} \quad (2.41a)$$

$$\text{s.t. } R1_{Ckl}^{(r)} \geq \frac{1}{2} \rho_{kl} x_l R_{Ck}^{\min}, \quad R2_{Ckl}^{(r)} \geq \frac{1}{2} \rho_{kl} x_l R_{Ck}^{\min} \quad (2.41b)$$

$$R_{Ck} \geq R_{Ck}^{\min} \quad \forall k \in \mathcal{K}, \quad R_{Dl} \geq R_{Dl}^{\min}, \quad \forall l \in \mathcal{L} \quad (2.41c)$$

$$m_k p_{Ck} \leq P_{Ck}^{\max} \quad \forall k \in \mathcal{K} \quad (2.41d)$$

$$\sum_{n \in \mathcal{N}} \rho_{nl} m_n p_{Dln} \leq P_{Dl}^{\max} \quad \forall l \in \mathcal{L} \quad (2.41e)$$

$$\sum_{n \in \mathcal{N}} (1 - x_l) m_n p_{Rl} \leq P_{Rl}^{\max} \quad \forall r_l \in \mathcal{R} \quad (2.41f)$$

$$x_l \in \{0, 1\} \quad \forall l \in \mathcal{L}, \quad \rho_{kl} \in \{0, 1\} \quad \forall k \in \mathcal{K} \quad \forall l \in \mathcal{L}. \quad (2.41g)$$

Dans le problème (2.41), les contraintes(2.41b) et (2.41c) sont imposées pour garantir le débit de données minimal des liens D2D cellulaires dans tous les intervalles de transmission et la contrainte (2.41g) nécessite que chaque lien D2D fonctionne en un seul mode.

Dans ce qui suit, nous proposons de résoudre le problème (2.41) de manière *optimale* par une approche de solution avec trois phases, à savoir l'allocation de puissance, la sélection de mode et l'allocation de groupe de ressources (RG). Tout d'abord, nous résolvons le problème d'allocation de puissance pour chaque lien D2D l dans le relais ou le mode direct s'il réutilise la ressource du lien cellulaire k . Ensuite, la sélection de mode est implémentée pour déterminer les modes optimaux des liens D2D. Enfin, le problème d'origine est transformé en problème d'affectation RG, qui peut être

résolu de manière *optimale* en utilisant la méthode Hongroise. Nous présentons cette conception dans ce qui suit.

2.2.3.3 Allocation de puissance

Supposons que le lien D2D l réutilise la ressource du lien cellulaire k , nous devons résoudre deux problèmes d'allocation de puissance correspondant au mode direct et relais du lien D2D l . Le problème d'allocation de puissance étant donné que le lien D2D l fonctionne en mode direct peut être résolu en utilisant l'algorithme dans [23]. D'autre part, si le lien D2D l fonctionne en mode relais, nous désignons $\mathbf{p}_{kl} = [p_{Ck}, p_{Dl}, p_{Rl}]$, $P_c^m \triangleq P_{Ck}^{\max}/m_k$, $P_d^m \triangleq P_{Dl}^{\max}/m_k$, $P_r^m \triangleq P_{Rl}^{\max}/m_k$. Ensuite, nous avons le problème d'allocation de puissance suivant.

$$\max_{\mathbf{p}_{kl}} w_{kl}^{(r)}(\mathbf{p}_{kl}) \triangleq R_{Ckl}^{(r)} + R_{Dlk}^{(r)} \quad (2.42a)$$

$$\text{s.t. } R1_{Ckl}^{(r)} \geq \frac{1}{2}R_{Ck}^{\min}, R2_{Ckl}^{(r)} \geq \frac{1}{2}R_{Ck}^{\min}, R_{Dlk}^{(r)} \geq R_{Dl}^{\min} \quad (2.42b)$$

$$p_{Ck} \in [0, P_c^m], p_{Dl} \in [0, P_d^m], p_{Rl} \in [0, P_r^m]. \quad (2.42c)$$

Nous caractérisons l'allocation de puissance optimale du problème ci-dessus dans la proposition suivante.

Proposition 2.1. *Si le problème (2.42) est réalisable, alors dans l'optimalité au moins un noeud (émetteur D2D, relais ou cellulaire) utilise la puissance d'émission maximale et $R1_{Dlk}^{(r)} = R2_{Dlk}^{(r)}$.*

De la Proposition 2.1, on peut vérifier que le problème (2.42) peut atteindre son optimum si $p_{Ck} = P_c^m$, $p_{Dl} = P_d^m$, ou $p_{Rl} = P_r^m$. Par conséquent, nous pouvons déterminer la solution optimale du problème en évaluant les trois cas. Dans chaque cas, nous devons déterminer l'attribution de puissance de deux émetteurs. Pour chaque cas, c'est-à-dire i.e., $p_{Ck} = P_c^m$, $p_{Dl} = P_d^m$, ou $p_{Rl} = P_r^m$, en appliquant la contrainte $R1_{Dlk}^{(r)} = R2_{Dlk}^{(r)}$ de Proposition 2.1, nous pouvons transformer le problème (2.42) à un problème d'optimisation d'une variable p_{Ck} , p_{Dl} ou p_{Rl} . Pour l'optimisation d'une variable, les points locaux peuvent être déterminés en évaluant la première dérivation de la fonction objectif. Ensuite, la solution optimale peut être obtenue en comparant les valeurs objectives de tous les points locaux et extrêmes. Par conséquent, en utilisant cette procédure, nous pouvons obtenir la solution optimale du problème (2.42).

2.2.3.4 Conjointe Sélection du mode et allocation RG

Indiquez $w_{kl}^{(d)*}$ et $w_{kl}^{(r)*}$ en tant que taux totaux optimaux lorsque le lien D2D l réutilise la ressource du lien cellulaire k en mode direct et relais, respectivement. La sélection optimale du mode peut être déterminée comme suit. Supposons que le lien D2D l soit affecté RG k , si $w_{kl}^{(d)*} \geq w_{kl}^{(r)*}$ alors le lien D2D l devrait opérer en mode direct, sinon, il devrait fonctionner en mode relais. Comme le lien D2D l est affecté à RG k , l'augmentation de taux due à la réutilisation des ressources D2D est $w_{kl}^* = \max\{w_{kl}^{(d)*}, w_{kl}^{(r)*}\} - R_{Ck}^{(o)}$. Par conséquent, le problème (2.41) peut être transformé en le problème suivant, qui peut être résolu de manière optimale par la méthode Hongroise [24]

$$\max_{\rho} \sum_{l \in \mathcal{L}} \sum_{k \in \mathcal{N}} w_{kl}^* \rho_{kl} \quad (2.43a)$$

$$\text{s.t.} \quad \sum_{k \in \mathcal{N}} \rho_{kl} = 1, \forall l \in \mathcal{L}, \sum_{l \in \mathcal{L}} \rho_{kl} \leq 1, \forall k \in \mathcal{N} \quad (2.43b)$$

$$\rho_{kl} \in \{0, 1\}, \forall l \in \mathcal{L}, k \in \mathcal{N}. \quad (2.43c)$$

2.2.3.5 Résultats numériques

Nous considérons le système où il existe $L = 15$ liens D2D réutilisant la ressource de 20 de liaisons cellulaires, et chaque lien cellulaire est attribué à un sous-canal. Les utilisateurs et les relais cellulaires sont distribués au hasard dans la zone cellulaire du rayon 500m. De plus, chaque émetteur et récepteur D2D sont situés de façon aléatoire dont la distance à son relais varie en d_{\max} , où $d_{\max} = 100m$. Enfin, notre conception proposée, appelée «Proposed Optimal», est comparée à deux schémas existants développés dans [23] et [25] dénommé «Conventional Direct» et «Conventional Relay», respectivement.

Fig. 2.7 présente le taux de somme du système par rapport à d_{\max} . Comme d_{\max} est petit, notre schéma proposé fonctionne de même que le système «Conventional Direct» et surpasse significativement le schéma «Conventional Relay». C'est parce que d_{\max} est petit, le mode D2D optimal est généralement le mode direct. Cependant, au fur et à mesure que d_{\max} augmente, les liaisons D2D ont tendance à fonctionner plus fréquemment dans le mode relais puisque le mode relais D2D peut surpasser le mode D2D direct. En conséquence, la conception proposée s'améliore bien mieux que

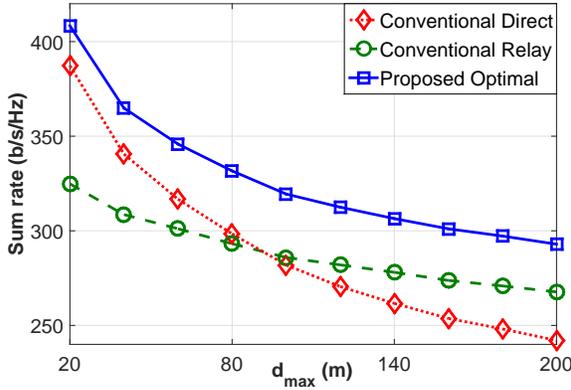


Figure 2.7 – Taux de somme par rapport contre d_{\max}

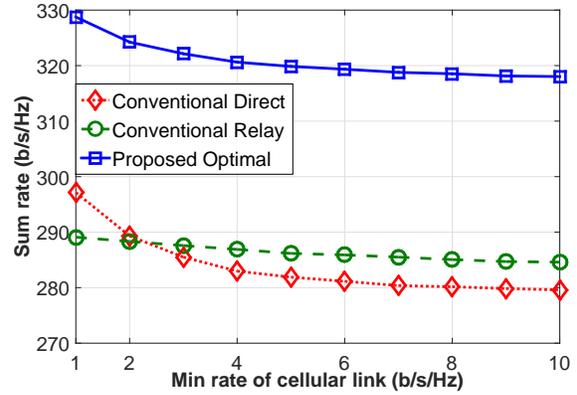


Figure 2.8 – Taux de somme par rapport contre R_c^{\min}

les deux autres systèmes existants grâce aux avantages d'un changement adaptatif entre les modes D2D direct et relais D2D.

Fig. 2.8 montre le taux de somme du système par rapport à R_c^{\min} . Comme R_c^{\min} augmente, le taux de somme du système diminue de manière modérée. C'est parce que quand R_c^{\min} devient plus élevé, les liens cellulaires doivent augmenter leurs puissances d'émission pour répondre aux exigences de débit de données. En outre, avec une répartition optimale des RG, les liaisons D2D qui souffrent de faibles interférences de co-canal à partir de certaines liaisons cellulaires ont tendance à réutiliser les ressources de fréquence de ces liens cellulaires. Ainsi, les puissances d'émission supérieures des liaisons cellulaires ne réduirait pas significativement les débits de données des liaisons D2D. Par conséquent, le taux de somme du système diminue progressivement à mesure que R_c^{\min} augmente.

2.2.4 La conjointe Planification partagée et allocation de ressources pour les réseaux sans fil basés sur OFDMA

Dans cette contribution, nous étudions conjointement l'agencement prioritaire, l'affectation du sous-canal et l'allocation de puissance pour les multiples réseaux sans fil dans les réseaux sans fil OFDMA, qui simultanément (i) maximise le nombre de liens non prioritaires planifiés et (ii) maximise leur taux de somme sujet aux exigences minimales en matière de taux des liens prioritaires et non prioritaires. En particulier, notre travail fait les nouvelles contributions.

- Nous formulons la conception de l'ordonnancement et de l'allocation des ressources comme un problème d'optimisation en une seule étape en tenant compte des contraintes QoS des liens prioritaires et des liens non priorisés programmés.
- Nous développons un algorithme monotone basé sur l'approche optimale (MBOA) pour résoudre le problème ci-dessus qui réalise de manière asymptotique l'ensemble optimal de liens non priorisés programmés et leur taux de sommation maximum.
- Nous proposons un autre algorithme itératif d'approximation convexe (ICA) à faible complexité qui effectue séquentiellement l'allocation de puissance et l'élimination des liens dans chaque itération.
- Les études numériques démontrent que l'algorithme MBOA est le meilleur parmi tous les algorithmes.

2.2.4.1 Modèle du système

Nous considérons les communications de liaison montante dans un système sans fil à une seule cellule où K liens sans fil prioritaires dans l'ensemble $\mathcal{K} = \{1, \dots, K\}$ partagent le même spectre comprenant N sous-canaux orthogonaux dans l'ensemble $\mathcal{N} = \{1, \dots, N\}$ avec L liens sans fil non prioritaires dans l'ensemble $\mathcal{L} = \{K+1, \dots, K+L\}$. Laissez $\mathcal{M} = \mathcal{K} \cup \mathcal{L}$ désigner l'ensemble de tous les liens. Nous désignons p_m^n comme la puissance d'émission du lien $m \in \mathcal{M}$ sur le sous-canal n et nous représentons le vecteur de puissance d'émission de tous les liens dans le système en tant que $\mathbf{p} = [\mathbf{p}_m]_{\forall m \in \mathcal{M}}$ où $\mathbf{p}_m = [p_m^n]_{\forall n \in \mathcal{N}}$ est le vecteur d'allocation de puissance du lien $m \in \mathcal{M}$ sur les sous-canaux. Nous supposons que les liens prioritaires utilisent les sous-canaux orthogonalement; cependant, les liens non prioritaires sont autorisés à réutiliser tous les sous-canaux pour améliorer l'efficacité du spectre en exploitant la diversité spatiale. Nous définissons le vecteur d'affectation de sous-canal suivante $\mathbf{rho}^n = [\rho_k^n]_{\forall k \in \mathcal{K}, \forall n \in \mathcal{N}}$, où ρ_k^n indique si le canal n est affecté au lien k .

Il a été prouvé dans [26] que si l'interférence mutuelle entre deux liens interférants est assez forte, ils devraient utiliser le spectre orthogonal pour maximiser le taux de somme. Motivés par ce résultat, nous autorisons chaque lien prioritaire à exploiter mathématiquement tous les sous-canaux; cependant, les gains des canaux *virtuels* parmi les liens prioritaires sont très élevés. Un tel réglage des gains élevés des canaux interférents imposera effectivement des liens prioritaires pour utiliser les

sous-canaux orthogonalement pour éviter les fortes interférences inter-canal. Plus précisément, en définissant les gains de canal parmi les liens priorisés à une valeur suffisamment grande η , le SINR de lien prioritaire k lien non priorisé l sur le sous-canal n peut être exprimé comme

$$\bar{\Gamma}_k^n(\mathbf{p}) = \frac{p_k^n h_{kk}^n}{\sigma_k^n + \sum_{k' \in \mathcal{K} \setminus k} p_{k'}^n \eta + \sum_{l \in \mathcal{L}} p_l^n h_{kl}^n} \quad (2.44)$$

$$\bar{\Gamma}_l^n(\mathbf{p}) = \frac{p_l^n h_{ll}^n}{\sigma_l^n + \sum_{k \in \mathcal{K}} p_k^n h_{lk}^n + \sum_{l' \in \mathcal{L} \setminus l} p_{l'}^n h_{ll'}^n}. \quad (2.45)$$

Les taux de données du lien priorisé $k \in \mathcal{K}$ et le lien non priorisé $l \in \mathcal{L}$ peuvent être ré-exprimés, respectivement, comme

$$\bar{R}_k(\mathbf{p}) = \sum_{n \in \mathcal{N}} \log_2 \left(1 + \bar{\Gamma}_k^n(\mathbf{p}) \right) \quad (2.46)$$

$$\bar{R}_l(\boldsymbol{\rho}, \mathbf{p}) = \sum_{n \in \mathcal{N}} \log_2 \left(1 + \bar{\Gamma}_l^n(\mathbf{p}) \right). \quad (2.47)$$

Nous souhaitons concevoir conjointement la planification des liens, l'affectation des sous-canaux et le contrôle de puissance pour tous les liens en tenant compte des objectifs et contraintes de conception suivants: (i) les taux de données minimum requis des liens prioritaires doivent être maintenus, (ii) le nombre de liens non priorisés planifiés est maximisé, et (iii) pour l'ensemble donné de liens non priorisés planifiés, le taux en somme de la planification des liens non priorisés est maximisé. Nous considérons la conception centralisée dans laquelle le CSI de tous les liens est disponible pour l'optimisation.

Pour capturer la décision de planification, nous introduisons un vecteur de planification de liaison binaire $\mathbf{s}_{\mathcal{L}} = [s_1, \dots, s_L]^T$, où $s_l = 1$ si le lien non priorisé $l \in \mathcal{L}$ est prévu et $s_l = 0$, sinon. Par conséquent, nous avons le problème d'optimisation suivant

$$\max_{\mathbf{p}, \mathbf{s}_{\mathcal{L}}} \alpha \sum_{l \in \mathcal{L}} s_l + \sum_{l \in \mathcal{L}} \bar{R}_l(\mathbf{p}) \quad (2.48a)$$

$$\text{s.t.} \quad \bar{R}_k(\mathbf{p}) \geq R_k^{\min} \quad \forall k \in \mathcal{K} \quad (2.48b)$$

$$\bar{R}_l(\mathbf{p}) \geq s_l R_l^{\min} \quad \forall l \in \mathcal{L} \quad (2.48c)$$

$$\sum_{n \in \mathcal{N}} p_k^n \leq P_{\max} \quad \forall k \in \mathcal{K} \quad (2.48d)$$

$$\sum_{n \in \mathcal{N}} p_l^n \leq P_{\max} \quad \forall l \in \mathcal{L} \quad (2.48e)$$

$$s_l \in \{0, 1\} \quad \forall l \in \mathcal{L}. \quad (2.48f)$$

Nous développons un algorithme pour résoudre le problème (2.48) dans les sections suivantes.

2.2.4.2 Algorithme d'approche optimale monotone (MBOA)

Notez que la nature binaire du vecteur d'ordonnement $\mathbf{s}_{\mathcal{L}}$ rend difficile la résolution du problème (2.48). Pour surmonter ce défi, on approxime une variable discrète s_l par une fonction continue $q : [0, 1] \rightarrow [0, 1]$ qui est définie comme $q(s_l) = (e^{Qs_l} - 1)/(e^Q - 1)$. La fonction $q(s_l)$ a les propriétés suivantes: (i) $q(s_l)$ est une fonction continue et croissante, (ii) $q(0) = 0$ et $q(1) = 1$. En outre, à mesure que Q augmente, la fonction $q(\cdot)$ approxime la fonction étape plus étroitement. En utilisant la fonction d'approximation $q(\cdot)$, nous arrivons au problème suivant

$$\begin{aligned} \max_{\mathbf{p}, \mathbf{s}_{\mathcal{L}}} \quad & \alpha \sum_{l \in \mathcal{L}} q(s_l) + \sum_{l \in \mathcal{L}} \bar{R}_l(\mathbf{p}) \\ \text{s.t.} \quad & (2.48b), (2.48c), (2.48d), 2.48e \\ & s_l \in [0, 1] \quad \forall l \in \mathcal{L}. \end{aligned} \quad (2.49)$$

Nous montrons maintenant que le problème (2.49) peut être transformé en un problème d'optimisation monotone dans les suivantes [27]. On peut voir que la fonction objective du problème (2.49) augmente dans $\mathbf{s}_{\mathcal{L}}$; cependant, il est sans augmentation dans \mathbf{p} . Nous définissons de nouvelles variables $z_{mn} = \bar{\Gamma}_m^n(\mathbf{p}) \quad \forall m \in \mathcal{M}, \forall n \in \mathcal{N}$, qui désignent le SINR réalisable du lien $m \in \mathcal{M}$ sur le sous-canal

$n \in \mathcal{N}$. Nous définissons également $\mathbf{z} = [z_{mn}]_{\forall m \in \mathcal{M}, \forall n \in \mathcal{N}}$. Nous introduisons un vecteur auxiliaire $\mathbf{t} = [t_l]_{\forall l \in \mathcal{L}}$ et $t_l \in [0, R_l^{\min}]$ alors la contrainte (2.48c) pour chaque lien non priorisé $l \in \mathcal{L}$ devient équivalente aux contraintes suivantes

$$\sum_{n \in \mathcal{N}} \log_2(1 + z_{ln}) + t_l \geq R_l^{\min} \quad (2.50)$$

$$t_l + s_l R_l^{\min} \leq R_l^{\min} \quad (2.51)$$

$$t_l \in [0, R_l^{\min}]. \quad (2.52)$$

Désignez $\mathbf{x} = (\mathbf{t}, \mathbf{s}_{\mathcal{L}}, \mathbf{z})$ comme vecteur d'optimisation qui a $D = 2L + (K + L)N$ dimensions et $\mathcal{P} \triangleq \{\mathbf{p} \mid \sum_{n \in \mathcal{N}} p_m^n \leq P_{\max} \forall m \in \mathcal{M}\}$. Ensuite, le problème (2.49) peut être transformé en le problème suivant

$$\max_{\mathbf{x} \succeq \mathbf{0}} f(\mathbf{x}) \triangleq \alpha \sum_{l \in \mathcal{L}} q(s_l) + \sum_{l \in \mathcal{L}} \sum_{n \in \mathcal{N}} \log_2(1 + z_{ln}) \quad (2.53a)$$

$$\text{s.t.} \quad s_l \leq 1 \quad \forall l \in \mathcal{L} \quad (2.53b)$$

$$z_{mn} \leq \bar{\Gamma}_m^n(\mathbf{p}) \quad \forall m \in \mathcal{M} \quad \forall n \in \mathcal{N} \quad \forall \mathbf{p} \in \mathcal{P} \quad (2.53c)$$

$$t_l + s_l R_l^{\min} - R_l^{\min} \leq 0 \quad \forall l \in \mathcal{L} \quad (2.53d)$$

$$t_l \leq R_l^{\min} \quad \forall l \in \mathcal{L} \quad (2.53e)$$

$$\sum_{n \in \mathcal{N}} \log_2(1 + z_{kn}) - R_k^{\min} \geq 0 \quad \forall k \in \mathcal{K} \quad (2.53f)$$

$$\sum_{n \in \mathcal{N}} \log_2(1 + z_{ln}) + t_l - R_l^{\min} \geq 0 \quad \forall l \in \mathcal{L}. \quad (2.53g)$$

Nous caractérisons la propriété du problème (2.53) dans la proposition suivante.

Proposition 2.3. *Problème (2.53) est un problème d'optimisation monotone*

Comme problème (2.53) est un problème d'optimisation monotone, nous pouvons développer l'algorithme *polyblock external approximation* [27] pour le résoudre de manière optimale. Supposons que $\mathbf{x}^* = (\mathbf{t}^*, \mathbf{s}_{\mathcal{L}}^*, \mathbf{z}^*)$ et \mathbf{p}^* sont la solution optimale du problème (2.53) et son allocation de puissance correspondante, respectivement. Néanmoins, $\mathbf{s}_{\mathcal{L}}^*$ peut être fractionnaire, ce qui n'est pas une solution possible de problème (2.48). Nous proposons de planifier le lien non priorisé l si sa solution de

planification $s_l^* \geq 1 - \epsilon$. Ensuite, la performance de l'algorithme MBOA se caractérise dans le théorème suivant.

Theorem 2.4. *Si le taux de données minimum requis de chaque lien non priorisé peut être réduit d'un petit nombre et d'un numéro de contrôleur ϵR_l^{\min} , en choisissant $Q \geq \ln L/\epsilon$, l'algorithme MBOA planifie le nombre maximal de liens non priorisés.*⁴

2.2.4.3 Résultats numériques

Nous évaluons la performance de nos algorithmes proposés pour les réseaux cellulaires sans fil prenant en charge les communications D2D qui consistent en des liaisons cellulaires prioritaires de $K = 4$ et des liaisons D2D non priorisées sur $L = 5$ partageant la ressource de $N = 10$ sous-canaux. Nous évaluons ensuite la performance des algorithmes proposés et les comparons avec les algorithmes “Conv. SE” et “Conv. Scheduling”, qui sont adoptés à partir de [28]. Figs. 2.9 et 2.10 montrent le nombre de liens D2D planifiés et leur taux de somme par rapport au taux minimum requis R_c^{\min} , respectivement. Quand R_c^{\min} augmente, le système doit affecter plus de ressources aux liens cellulaires; par conséquent, un plus petit nombre de liens D2D et un plus petit taux de somme des liens D2D programmés peut être atteint, ce qui peut être observé pour tous les algorithmes. Ces résultats confirment que l'algorithme MBOA effectue le meilleur en terme de nombre de liaisons D2D planifiées et de leur taux d'addition. C'est parce que l'algorithme MBOA peut planifier le nombre maximum de liens D2D comme indiqué dans Theorem 2.4.

2.2.5 Remarques finales

Dans cette thèse de doctorat, nous avons développé diverses techniques et algorithmes novateurs de gestion des ressources pour la communication D2D dans les réseaux cellulaires. En particulier, nous avons effectué quatre contributions de recherche importantes. Tout d'abord, nous avons développé des algorithmes d'allocation de ressources efficaces pour le spectre qui se comportent de manière significativement supérieure aux autres techniques dans la littérature. Deuxièmement, nous avons proposé un cadre RA général énergétiquement efficace pour la communication D2D dans les réseaux

⁴Si l'on considère le problème étudié où le taux minimum de lien non priorisé l équivaut à $(1 + \epsilon)R_l^{\min}$ alors l'algorithme MBOA peut garantir les QoS nécessaires de tous les liens non priorisés programmés.

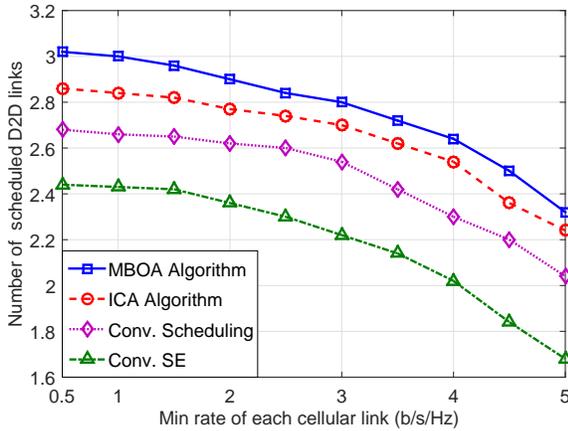


Figure 2.9 – Nombre de liens D2D admis par rapport à R_c^{\min}

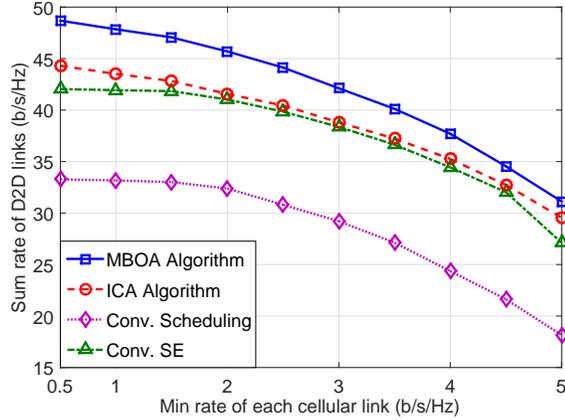


Figure 2.10 – Taux de somme des liens D2D par rapport au taux minimum de liaison cellulaire R_c^{\min}

cellulaires qui vise à maximiser l'EE pondérée minimale des liaisons D2D tandis que les débits minimaux de données des liaisons cellulaires individuelles sont assurés.

Troisièmement, nous avons développé un schéma RA optimal pour la communication D2D à relais dans le réseau cellulaire, ce qui permet de faire fonctionner les communications D2D en mode relais ou direct. De plus, l'algorithme proposé surpasse de manière significative les systèmes de communication D2D existants. Enfin, nous avons formulé la planification et la conception de l'allocation des ressources pour la communication D2D dans le réseau cellulaire où les liens cellulaires sont plus prioritaires que les liens D2D. Le schéma proposé permet au système de sélectionner dynamiquement l'ensemble des liaisons D2D planifiées et d'optimiser l'efficacité du spectre du système. Nous avons proposé un algorithme à base de monotonie qui permet d'atteindre asymptotiquement la solution optimale.

Chapter 3

Introduction

3.1 Background and Motivation

Wireless cellular operators have seen the ever increasing demand from high-speed applications and rapidly growing number of connected devices. Toward this end, future wireless networks are expected to deliver much larger capacity and support significantly higher communication rates. Specifically, it is predicted that the 5G wireless system should accommodate 1000-fold increase in the system capacity and 100-fold increase in the data rate of connected devices [3]. Device-to-device (D2D) communication, which is illustrated in Fig. 3.1, has been recently proposed as an important technology toward achieving these objectives [4] by supporting local traffic through direct communications among mobile devices. In D2D communication, two nearby devices can establish a direct communication link under the control of the cellular base station (BS). Different design aspects of D2D communication such as hardware interface, D2D discovery, and resource allocation, have been investigated in both academic and industry communities to enable D2D communications and support future system scenarios and applications [5].

In general, D2D communication can help significantly improve the system capacity by exploiting the advantages of proximity communication. Thanks to the short communication distance among the nearby devices, robust communication with a high data rate can be established to meet stringent requirements of emerging broadband wireless applications such as video sharing and online gaming. In D2D communication, the proximity devices can communicate directly to each other

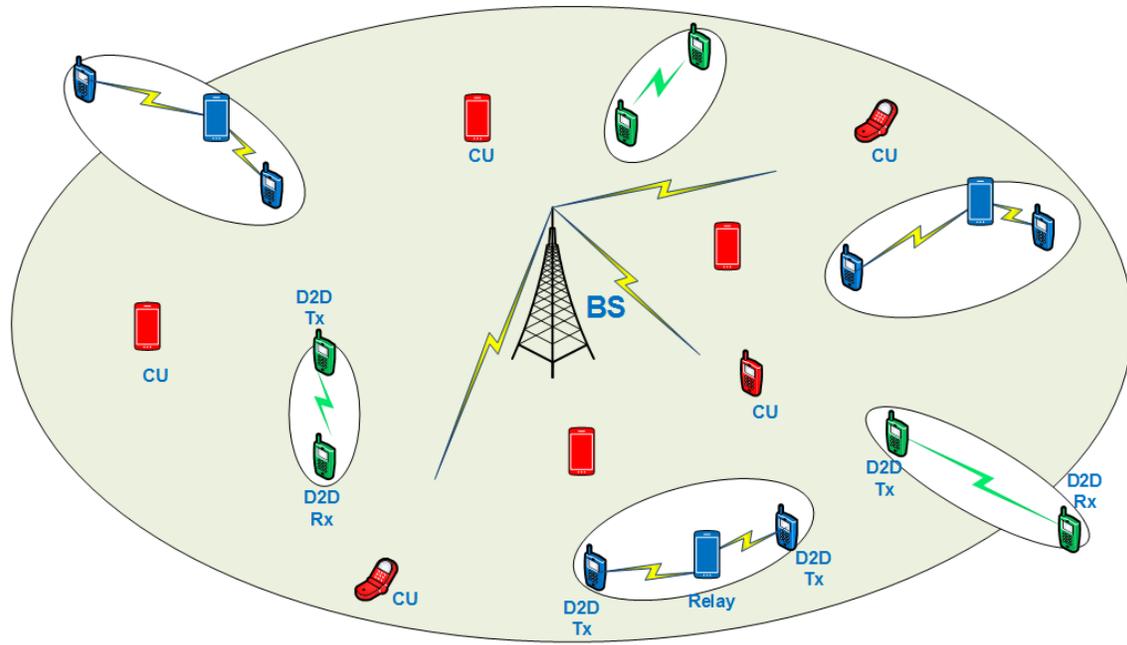


Figure 3.1 – D2D communication in cellular networks

instead of bypassing through the BS; hence, transmission delay can be reduced significantly. D2D communication can support not only the traditional local voice and data services but also many emerging D2D based applications such as social-aware networking, video sharing, online gaming, and public safety applications.

Third Generation Partnership Project (3GPP) considers D2D communication as an essential technology for the future of cellular system [6, 7]. The initial standardization of D2D communication was proposed in Release 12 of 3GPP for public safety applications [8]. Moreover, D2D communication plays more important roles in many different scenarios such as massive internet of thing (IoT), extreme real-time communication, lifeline communications, ultra-reliable communications, and broadcast-like service [9]. In fact, potential scenarios and applications of D2D communication are still being discussed for 5G system, which is expected to be deployed in 2020.

One major challenge for enabling D2D communication in cellular networks is related to the development of efficient radio resource management techniques. In particular, it is desirable to employ a resource allocation scheme for D2D communication that can exploit the limited resources such as frequency and transmit power efficiently. Different general resource allocation approaches for D2D communication can be summarized as in Fig. 3.2. Specifically, D2D communication can be employed by using the frequency allocated to the cellular spectrum, i.e., in-band D2D, or by

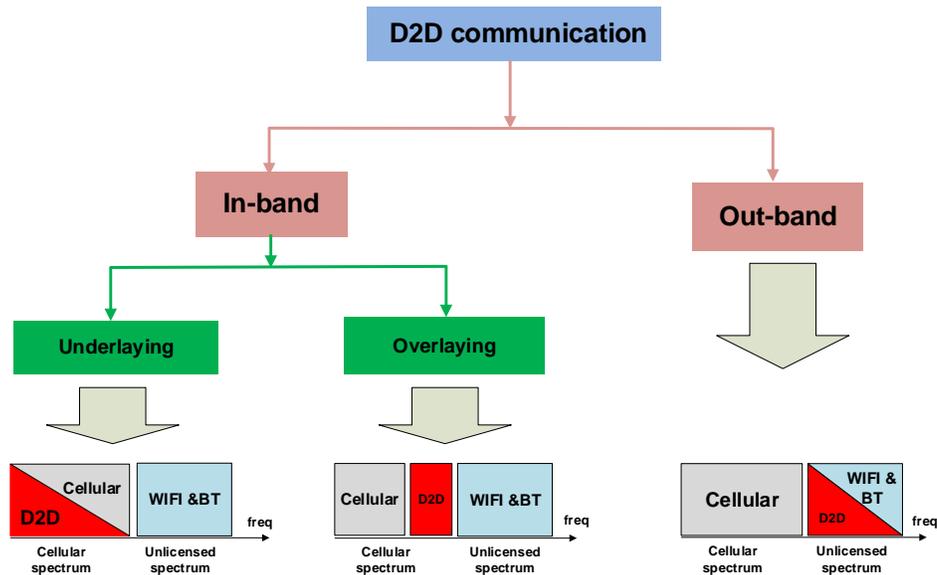


Figure 3.2 – D2D communication in cellular networks

using unlicensed spectrum, i.e., out-band D2D [10]. However, in the out-band D2D, the quality of service (QoS) of D2D links would not be guaranteed due to uncontrollable interferences from other sources. In the in-band D2D, D2D links can be operated in either overlaying or underlying scenarios. In the overlaying scenario, the frequency resource is orthogonally reserved for D2D and cellular links; nevertheless, operating in the overlaying scenario could result in inefficient resource utilization as some frequency resources might be unused due to the unavailability of either D2D or cellular links. Finally, in the underlying scenario, frequency resources are shared between D2D and cellular links, which can potentially improve the system performance since the frequency resource and spatial diversity of the system can be exploited efficiently.

The potential benefits obtained from the underlying scenario come with the price of more complicated resource allocation algorithms. In particular, these algorithms are required to efficiently allocate the frequency resources to each link and manage the interference between the links. However, to enable D2D communication in the cellular network, each operator may employ different approaches which can be chosen while considering different system models, set of optimization parameters, utility functions, and resource utilization schemes.

From the operator’s perspective, maximization of the system spectrum-efficiency (SE) can be more desirable since the revenue is proportional to the system throughput. Nevertheless, from the mobile users’ viewpoint, optimization of the system energy-efficiency (EE) can be more appropriate because this design can help prolong devices’ battery life. Toward this end, enabling direct D2D communication can be beneficial if the underlying pair of devices is close to each other. However, if two devices are far away from each other, they can establish the communication path with each other through the BS or a relay node. Therefore, resource allocation for relay-based D2D communication is an important design issue to enable D2D communication in cellular networks [11]. Besides resource allocation, scheduling control, which determines the set of links to transmit at a particular time, is also essential in achieving efficient resource utilization. Moreover, cellular and D2D can be treated as primary and secondary links, respectively, where the primary links are prioritized over the secondary links. Hence, joint scheduling and resource allocation design for this prioritized access model is an important design problem.

The overall objective of this dissertation is to develop novel resource allocation optimization models and propose innovative algorithms to enable D2D communication in wireless cellular networks. Toward this end, we consider three fundamental design aspects to allow harmonious coexistence of D2D and cellular communication links, namely spectrum and energy-efficient resource allocation for single hop D2D communication, joint mode selection and resource allocation for relay-based D2D communication, and joint scheduling and resource allocation for D2D communications. In the following, we discuss research challenges related to these design problems.

3.2 Research Challenges

A D2D-enabled cellular network is a highly complicated hybrid system, which requires sophisticated resource management design to enable robust and efficient coexistence between D2D and cellular links. However, various research challenges arise as one considers different system models, communications setting, design objectives, and constraints. We discuss some of these major challenges in the following subsections.

3.2.1 Challenges in Resource Allocation for Single Hop D2D Communications

Spectrum and energy-efficient resource allocation problems are formulated to enable efficient system operation and utilization of radio resources. However, there are many technical challenges to address for these resource allocation problems.

In general, there is a complexity-efficiency tradeoff for the design of resource allocation algorithms. Specifically, to implement a resource allocation algorithm, the BS must estimate/collect the necessary Channel State Information (CSI) of the the involved wireless links in the system, which is the input of resource allocation algorithms. The outputs of these algorithms are the values of optimization variables related to different allocation decisions such as the set of scheduled links, transmission power, and subchannel allocation decisions, which must be sent to the corresponding entities (BS and wireless links) for realization. All the involved procedures (i.e., CSI estimation/collection, running of a resource allocation algorithm, and realization of the obtained solution) should be implemented repeatedly within the channel coherence time, which can be generally smaller than $100ms$ [29]. Hence, it is essential to design a sufficiently fast algorithm to cope with the dynamics of the channel gains of all involved wireless and interfering links in the system.

The energy-efficient resource allocation has some specific challenge arising from its design objective. Specifically, this objective function has the fractional form, where the numerator and denominator correspond to the data rate and consumed power, respectively, of the involved wireless links. In general, the fractional optimization problem is much more challenging to address compared to those with non-fractional objective functions. In particular, suitable transformation techniques must be employed to solve the underlying energy-efficient resource allocation problem.

The final challenge concerns joint subchannel assignment and power allocation for cellular and D2D links, which involves a mixed integer optimization problem. Moreover, the channel gains of individual links are different and the interference levels among the links can be very different. Therefore, the designed resource allocation algorithms must intelligently allocate the appropriate subchannels for each link which can efficiently exploit favorable channel conditions and mitigate the co-channel interference. Finally, the channel assignment and power allocation for D2D and cellular links must be optimized jointly to achieve the best performance.

3.2.2 Challenges in Relay-based D2D Communication

It can be seen that one hop D2D is an efficient solution if the two devices are close to each other. Nevertheless, if they are far away from each other, the two devices can communicate with each other by using the relay-based D2D communication. Intelligent relay-based D2D communication design can lead to various benefits such as enhancing the connections quality and improving the system's coverage. However, besides the challenges encountered in single hop D2D communication, we have to deal with the further challenges in relay-based D2D communication.

The first challenge is related to the system's modeling. In particular, relay-based D2D communications involve in two transmission phases which are D2D transmitter to relay and relay to D2D receiver phases. Therefore, wireless links in the system can experience different interference patterns during the two transmission phases. Consequently, a proper system model is needed to capture the key characteristics of the system and this relay-based communication.

The second challenge concerns the mode selection for each pair of D2D devices in the system. Generally, each D2D link can operate in either the direct or relay mode, where in the direct mode, the two D2D devices executes the direct communication while in the relay mode, the pair of D2D devices performs two hop communication through a relay. Mode selection is required to determine the operation mode for each D2D pair. Since the mode selection decision of each D2D link is coupled with the mode selection decisions of other links, the mode decision optimization may have its complexity growing exponentially with the number of D2D links.

Last but not least, an efficient resource allocation algorithm must balance the resource utilization for the two hops of D2D communication in the relay mode. In fact, the channel quality of two hops in D2D communication, namely channel between D2D transmitter to relay and relay to D2D receiver, can be very different. Moreover, the end to end data rate between two devices would be maximized as the data rates of two hops are balanced. Therefore, the proposed algorithm must suitably determine the transmit power of each device and transmission time of each communication hop to in the relay mode.

3.2.3 Challenges in Scheduling and Resource Allocation for D2D Communication

Generally, to guarantee the required quality of service (QoS), one must determine the set of users to be scheduled since the wireless system may not be able to support all users concurrently. Hence, the joint scheduling and resource allocation, which decides the set of scheduled users and performs subchannel assignment and power allocation for the set of scheduled users, is a vital design. Toward this end, we have to deal with the various challenges, which are discussed in the following.

Firstly, one should develop a model that properly handles the access priorities for different users in the system. In general, the users in D2D communication can be considered as the secondary users, who have lower priority in accessing the radio resource compared to the cellular users. Nevertheless, due to short communication range in D2D communication, the data rate achieved by each D2D link can be much higher than the rates of cellular links. Therefore, an efficient design must properly capture the access priorities and exploits the diverse channel conditions of the D2D and cellular links.

Secondly, solving the joint scheduling and resource allocation problem can be very challenging. In general, even the scheduling problem for the non-prioritized network where all users have the same priority is already an NP-Hard problem [30]. The scheduling problem for the prioritized network is much more complicated to solve since we have to deal with different access priorities. Toward this end, a developed algorithm should determine the set of scheduled users and the allocated resources for them so that the best system performance can be achieved.

Finally, any resource allocation algorithm requires the knowledge about channel state information (CSI) which may be erroneous due to different factors such as estimation errors, quantization errors, measurement errors, and feedback delay. Therefore, it is necessary that the CSI errors must be effectively accounted for the design. In addition, it is desired that the required QoS of all users still be satisfied even with CSI errors. Moreover, the proposed design should balance between the system efficiency and the users' satisfaction.

3.3 Literature Review

In the following, we present the survey of the existing literature to our research studies. Firstly, we describe the related works on spectrum-efficient and energy-efficiency resource allocation for single hop D2D communication. Secondly, some recent works considering relay-based D2D communication are investigated. Finally, we review the recent papers on admission control and resource allocation for D2D communication in cellular networks.

3.3.1 Resource Allocation for Single Hop D2D Communication

Spectrum-efficient resource allocation design for D2D communication in cellular networks has received lots of research interests over the past few years [31–36]. The work [33] develops a fair resource allocation strategy for D2D links while assuring the QoS of cellular links. Nonetheless, the employed decomposed power and subchannel allocation for cellular and D2D links may not achieve the optimal performance in general. In [34], the sum-rate optimization for D2D and cellular links is considered for the setting with multiple D2D and cellular links; however, the authors assume that each cellular link has been pre-allocated one subchannel and the authors simply optimize the matching design for each D2D link with one cellular link. In addition, the work [35] employs the dynamic programming approach to solve the resource allocation for the D2D underlaid cellular system but the proposed greedy interference-based scheduling strategy might not provide any performance guarantee. Moreover, the work [36] aims to maximize the resource allocated to the cellular links given the QoS guarantees of D2D links but the proposed column generation based algorithm has exponential computational complexity.

The existing resource allocation designs mostly consider very simple settings such as the scenarios with only one cellular link and one D2D link [31] or one D2D link sharing resources with multiple cellular links [32], or the setting where the resource allocations for all cellular links are assumed to be predetermined [34]. Moreover, sub-optimal algorithms developed in several existing works may not provide strong performance guarantees with respect to the optimal solution [33, 35, 37] or require very high computational complexity [36].

There have been also some research works on energy-efficient resource allocation for D2D underlaying cellular networks. In [38], the authors propose a resource allocation solution based on

the non-cooperative game theory in which each D2D link selfishly performs power and subchannel allocation to maximize its own energy-efficiency considering the fixed resource allocation of other links. Therefore, in this approach, the co-channel interference among the links may not be managed adequately. A coalition game formulation is proposed to tackle the energy-efficient design in [39], in which the authors propose a joint mode selection and resource allocation for D2D and cellular links; however, in this work, the authors assume that each D2D link only achieves its minimum required rate, which may not fully exploit the advantage of short range D2D communication. Moreover, a simple energy-efficient design minimizing the total power consumption is also considered in [40, 41]. However, the resource allocation solutions in these works may not fully exploit the advantages of D2D communication in balancing both spectrum and energy-efficiency. To the best of our knowledge, all existing solutions devised for the energy-efficient D2D communication in cellular networks only permit the D2D links to reuse limited resource of cellular links, e.g., each D2D link can reuse resource of one cellular link [39, 40, 42]. Nevertheless, the efficiency of the existing solutions are not verified by either mathematical or experimental results since the optimal solutions cannot be determined in most cases.

3.3.2 Resource Allocation for Relay-based D2D Communication in Cellular Networks

Relay-based D2D communication in cellular networks has been investigated recently by both academia and standard bodies [43–47]. In [43], the authors propose an innovative spectrum sharing strategy for the D2D and cellular links, in which a D2D device can act as a relay to assist cellular communication via the cooperative communication and it also can involve in its D2D communication. However, this approach considers an oversimplified setting with one cellular link and one D2D link. The works in [44, 45] consider a more general setting with multiple cellular and D2D links. The joint subchannel and power allocation for relay-based D2D are studied in both works. While [44] solves the problem by employing the distributed message passing game, the authors in [45] tackle the problem by performing the Hungarian method iteratively. Nonetheless, the models studied in [44, 45] always force D2D links to communicate through relays, which may adversely degrade the performance of strong direct D2D links. As far as we can tell, none of the existing literature tackles the mode selection, each D2D link can either communicate directly or through a relay node, for relay-based D2D communication in cellular networks. Therefore, joint mode selection and resource

allocation optimization for relay-based D2D communication would enable to improve the system performance.

3.3.3 Scheduling and Resource Allocation for D2D Communication in Cellular Networks

Joint scheduling and resource allocation has been initially investigated for the interference-limited single-channel system. Specifically, in this model, one has to determine the set of scheduled links and their transmit powers to optimize certain desired objectives such as maximizing the number of scheduled links and weighted sum rate of the scheduled links [28, 48, 49]. In [28], J. Zander *et al.* propose an iterative algorithm which sequentially performs power allocation and greedy link removal until the minimum required rates of active links are satisfied. The authors in [48, 49] consider the problem which simultaneously maximizes the number of scheduled users and minimizes the power consumption. Both [49] and [48] propose linear programming (LP) based deflation algorithms which solve an LP approximation of the original problem whose solution is used to remove the “worst” link in each iteration. Although the above works study the joint admission control and resource allocation, differentiated priorities among the links are not considered and the underlying single-channel problem is obviously easier to deal with compared to the multiple-channel one.

Joint scheduling and resource allocation has also been investigated for cognitive radio networks (CRN), where primary users (PU) are granted higher priority in spectrum access compared to secondary users (SU) [50–52]. In particular, the downlink scheduling, channel assignment, and power control problem for multi-channel CRNs is addressed in [50] where the authors propose a dynamic interference graph algorithm (DIGA) which greedily assigns a channel to the best SU in each assignment step. As an extension of [50], joint channel assignment and power control for both uplink and downlink scheduling in multi-channel CRNs is addressed in [51] where the authors propose to transform the formulated problem into a maximal weighted bipartite matching problem and develop a greedy algorithm to solve it. Nguyen *et al.* in [52] tackle the link scheduling and resource allocation problem in CRNs by applying the coloring approach to the interference graph, which is shown to be more efficient than the algorithms in [51]. However, all above works assume that each SU can reuse only one channel from certain PU, which may limit their applicability and achievable performance.

Recently, joint scheduling and resource allocation for D2D-enabled communication systems are investigated in [17, 23, 53, 54]. Although these works consider both scheduling and resource allocation, the scheduling design to maximize the number of admitted D2D links is not their primary design objective; instead, the scheduling result is simply the bi-product of the studied spectrum-efficiency or energy-efficiency optimization problems. Nevertheless, the subchannel assignments for the cellular links are accounted for in these works, which permit each D2D link to exploit the resource of only one cellular link. It can be observed that none of the existing works discussed above addresses all key design aspects for the general multi-channel wireless system, i.e., the joint admission control and resource allocation with different access priorities where each wireless link can exploit multiple channels. In fact, consideration of all these aspects requires tackling a very challenging interference management problem.

3.4 Research Objectives and Contributions

The general objective of my Ph.D research is to develop efficient resource allocation algorithms for D2D communication which contribute to enable efficient integration of the D2D communication in cellular network while not causing severe performance degradation of the existing cellular links. Specifically, our main contributions, which are highlighted in Fig. 3.3, can be described as follows.

1. *Spectrum-efficient resource allocation for D2D communications:*

We formulate the resource allocation problem for joint subchannel assignment and power control that aims at maximizing the weighted sum-rate while guaranteeing the minimum required rates of individual cellular and D2D links. To solve this problem, we first derive the optimal power allocation for a given subchannel assignment for one pair of cellular and D2D links. Based on this result, we transform the original resource allocation problem into a subchannel assignment problem using the graph-based approach. We then develop a novel iterative rounding algorithm to solve the subchannel assignment problem. We show that the proposed Iterative Rounding algorithm significantly outperforms conventional spectrum sharing algorithms.

2. *Energy-efficient resource allocation for D2D communications:*

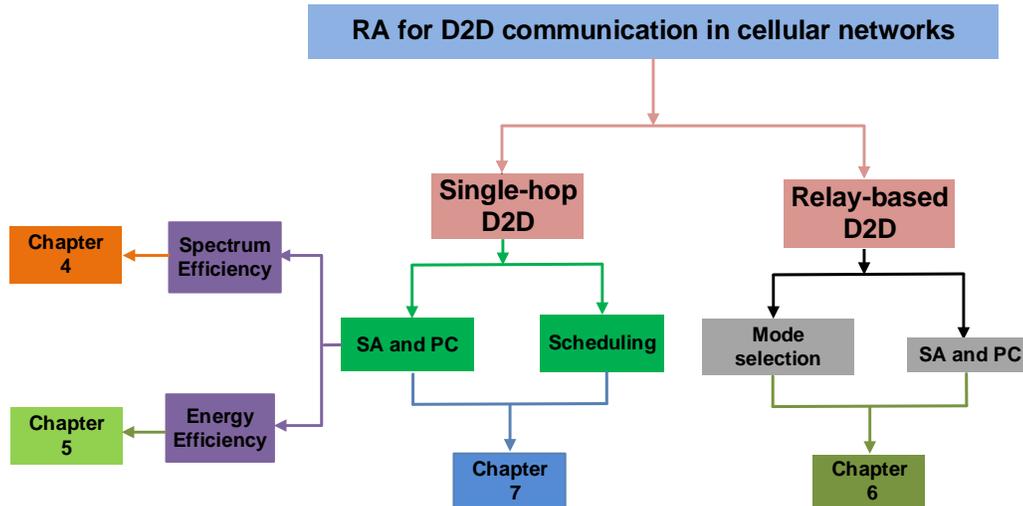


Figure 3.3 – D2D communication in cellular networks

We propose a general energy-efficient resource allocation design for D2D communication in cellular networks problem considering multiple cellular and D2D links where each D2D link can reuse the spectrum resources of multiple cellular links. We first characterize the optimal power allocation solution for a cellular link to transform the original resource allocation problem into an equivalent resource allocation problem for D2D links. We then propose a dual-based algorithm that solves the resource allocation problem in the dual domain. Particularly, we adopt the max-min fractional programming technique to iteratively transform the resource allocation problem into a Mixed Integer Nonlinear Programming (MINLP) problem which is solved by using the dual decomposition approach. Finally, the centralized and distributed implementations with limited message passing are proposed.

3. *Resource allocation for relay-based D2D communication:*

We study the joint mode selection, subchannel assignment, and power control problem for relay-based D2D communication in cellular networks which aims at maximizing the system sum rate considering minimum rate constraints of cellular and D2D links. The mode selection and resource allocation problem is formulated as an MINLP (Mixed-Integer Non-Linear

Programming) problem. We then propose an algorithm which obtains the optimal solution. We first derive the optimal power allocation for a given mode selection and subchannel allocation. Based on these results, the original resource allocation problem can be transformed into a resource allocation problem, which can be solved optimally by the Hungarian method [55]. Finally, extensive numerical studies demonstrate that the proposed design significantly outperforms existing D2D communication schemes.

4. *Joint prioritized scheduling and resource allocation for D2D communication in cellular network:*

We formulate the scheduling and resource allocation design for D2D communication in cellular networks as a single-stage optimization problem considering QoS constraints of the cellular and D2D links. We develop a monotonic-based algorithm to solve the proposed problem which asymptotically achieves the optimal solution. We also propose another low-complexity iterative convex approximation algorithm which sequentially performs power allocation and link removal in each iteration. We then describe how the proposed algorithms can be implemented in the future cellular network system according to the 3rd Generation Partnership Project (3GPP) standard for D2D communication in the cellular network. We also describe the responsibilities and operations of both user equipment (UE) and the BS during the channel estimation and feedback processes to obtain the necessary CSI at the BS to execute the proposed algorithms. Finally, we then describe a conservative design so that the BS can execute the proposed algorithm considering the CSI errors.

3.5 Dissertation Outline

The remaining of this dissertation is organized as follows. Chapter 4 reviews some fundamental mathematical optimization background and resource allocation for OFDMA-based system. In Chapter 5, we present the spectrum-efficient resource allocation for single hop D2D communication in the cellular network. Then, we discuss the energy-efficient resource allocation for single hop D2D communication in a cellular network in Chapter 6. The relay-based D2D communication in a cellular network is investigated in Chapter 7. In Chapter 8, we describe the admission control and resource allocation for D2D communication in the cellular network. Chapter 9 summarizes the main contributions of the dissertation and makes some recommendations for future research.

Chapter 4

Background

This chapter presents some fundamentals of optimization and resource allocation for OFDMA-based system. The resource allocation for OFDMA-based is of critical importance as OFDM is a medium access scheme adopted by most current wireless standards. In addition, some optimization tools which are utilized to model and solve various resource allocation problems in this dissertation are also presented.

4.1 Mathematical Optimization

4.1.1 Basic Concepts

The standard form of an optimization problem can be described as follows [21]:

$$\begin{aligned} \min_x \quad & f_0(x) \\ \text{s.t.} \quad & f_i(\mathbf{x}) \leq 0, \quad i = 1, \dots, m \\ & h_i(x) = 0, \quad i = 1, \dots, p \end{aligned} \tag{4.1}$$

where $\mathbf{x} \in R^n$ is a vector of optimization variables and $f_0(\mathbf{x}) \in R$ is an objective function. The set of \mathbf{x} that satisfies all m inequality and p equality constraints is the feasible set. If the feasible set is empty, the problem is infeasible. The optimal value of the problem is $f_0^*(\mathbf{x}) = \inf\{f_0(\mathbf{x}) \mid g_i(\mathbf{x}) \leq$

$0, i = 1, \dots, m, h_i(\mathbf{x}) = 0, i = 1, \dots, p\}$. If $f_0^*(\mathbf{x}) = -\infty$, the problem is unbounded. If $f_0^*(\mathbf{x}) = f_0(\mathbf{x}^*) \in \mathbb{R}$ then \mathbf{x}^* is the optimal solution.

4.1.2 Convex Optimization Problem

Convex optimization, which studies the problem of minimizing convex functions over convex sets, is well-known in the field of optimization. The convexity makes optimization easier to address compared to the general case since local minimum must be a global minimum, and first-order conditions are sufficient conditions for optimality [21]. In the following, the fundamental of convex optimization is briefly introduced.

4.1.2.1 Definition

A set S is convex if it satisfies the following condition [21]

$$\alpha\mathbf{x} + \beta\mathbf{y} \in S, \forall \mathbf{x}, \mathbf{y} \in S \quad (4.2)$$

such that $\alpha + \beta = 1, \alpha > 0, \beta > 0$. Then a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex if the domain \mathcal{D} of f is a convex set and f satisfies

$$f(\alpha\mathbf{x} + \beta\mathbf{y}) \leq \alpha f(\mathbf{x}) + \beta f(\mathbf{y}) \quad (4.3)$$

for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and all $\alpha, \beta \in \mathbb{R}$, with $\alpha + \beta = 1, \alpha > 0, \beta > 0$. A convex optimization problem has the following form

$$\begin{aligned} \min_{\mathbf{x}} \quad & f_0(\mathbf{x}) \\ \text{s.t.} \quad & f_i(\mathbf{x}) \leq 0, \quad i = 1, \dots, m \\ & h_i(\mathbf{x}) = 0, \quad i = 1, \dots, p \end{aligned} \quad (4.4)$$

where the functions $f_0, f_1, \dots, f_m : \mathbb{R}^n \rightarrow \mathbb{R}$ are convex, the functions $h_1, \dots, h_p : \mathbb{R}^n \rightarrow \mathbb{R}$ are linear.

4.1.2.2 Lagrange Dual Function and Lagrange Dual Problem

The Lagrangian $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}$ of problem 4.4 has the following form

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f_0(\mathbf{x}) + \sum_{i=1}^m \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^p \nu_i h_i(x) \quad (4.5)$$

where $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_m]$ and $\boldsymbol{\nu} = [\nu_1, \dots, \nu_p]$ are the Lagrangian multiplier vectors. The Lagrange dual function $g : \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}$ is defined as follows:

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\mathbf{x} \in \mathcal{D}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\mathbf{x} \in \mathcal{D}} \left(f_0(\mathbf{x}) + \sum_{i=1}^m \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^p \nu_i h_i(x) \right). \quad (4.6)$$

The Lagrange dual problem is then defined as

$$\begin{aligned} \max_{\boldsymbol{\lambda}, \boldsymbol{\nu}} \quad & g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \\ \text{s.t.} \quad & \boldsymbol{\lambda} \geq 0. \end{aligned} \quad (4.7)$$

Problem (4.4) is the primal problem. We denote p^* and d^* as the optimal values of problem (4.4) and problem (4.7), respectively. The difference $p^* - d^*$ is referred as the optimal duality gap. When $p^* = d^*$, the optimal duality gap is zero, and the problem has strong duality.

4.1.2.3 Karush-Kuhn-Tucker Conditions

Let \mathbf{x}^* and $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ be the primal and dual optimal solutions under the strong duality condition. Then, they will satisfy the KKT conditions which are expressed as follows:

$$\begin{aligned} f_i(\mathbf{x}^*) &\leq 0, \quad \forall i = 1, \dots, m \\ h_i(\mathbf{x}^*) &= 0, \quad \forall i = 1, \dots, p \\ \lambda_i^* &\geq 0, \quad \forall i = 1, \dots, m \\ \lambda_i^* f_i(\mathbf{x}^*) &= 0, \quad \forall i = 1, \dots, m \\ \Delta f_0(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i^* \Delta f_i(\mathbf{x}^*) + \sum_{i=1}^p \nu_i^* \Delta h_i(\mathbf{x}^*) &= 0 \end{aligned} \quad (4.8)$$

where Δf_i is the gradient of function f_i .

Theorem 4.1. *The condition 4.8 are sufficient conditions to obtain an optimal solution for a convex optimization problem.*

It means that any points \mathbf{x}^* and $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ satisfying KKT conditions are the primal and dual optimal solutions.

4.1.3 Monotonic Optimization

The convex optimization problem can be solved easily by many well-known optimization algorithms such as the interior point method [21]; however, many optimization problems are non-convex by their nature. An example is a power control problem for throughput maximization in interference-limited wireless networks, which cannot be transformed to a convex optimization one by the common change-of-variable techniques. These nonconvex optimization problems might have hidden monotonic structures such as the data rate of any link monotonically increases with SINR. These structures might help transform the original problem into a monotonic optimization problem, which can be solved optimally by a *polyblock approximation* algorithm[56–58]. In the following, the basic theory about monotonic optimization problem is presented.

Definition 4.1 (Vector). *For any two vectors $\mathbf{x}, \mathbf{y} \in R^M$, we write $\mathbf{x} \succeq \mathbf{y}$ and say that \mathbf{x} dominates \mathbf{y} if $x_i \geq y_i, \forall i = 1, \dots, M$, where x_i and y_i are the i^{th} dimension of \mathbf{x} and \mathbf{y} , respectively.*

Definition 4.2 (Box). *For any vertex $\mathbf{b} \in R_+^M$, the hyper rectangle $[\mathbf{0}, \mathbf{b}] = \{\mathbf{x} \in R_+^M | \mathbf{0} \preceq \mathbf{x} \preceq \mathbf{b}\}$ is referred as a box with vertex \mathbf{b} where R_+^M denotes the M -dimensional non-negative real domain.*

Definition 4.3 (Normal set). *A set $\mathcal{G} \subset R_+^M$ is called normal if, for any two points $\mathbf{x}, \mathbf{x}' \in R_+^M$ such that $\mathbf{x}' \preceq \mathbf{x}$, if $\mathbf{x} \in \mathcal{G}$, then $\mathbf{x}' \in \mathcal{G}$ too.*

Definition 4.4 (Reverse normal set). *A set $\mathcal{H} \subset R_+^M$ is reverse normal in box $[\mathbf{0}, \mathbf{b}]$ if $\mathbf{b} \succeq \mathbf{x}' \succeq \mathbf{x} \succeq \mathbf{0}$, then $\mathbf{x} \in \mathcal{H}$ implies $\mathbf{x}' \in \mathcal{H}$.*

Definition 4.5 (Polyblock). *Given any finite set of vertices $\mathcal{V} \in R_+^M$, the union of all boxes $[\mathbf{0}, \mathbf{x}]$, $\mathbf{x} \in \mathcal{V}$, is a polyblock with vertex set \mathcal{V} .*

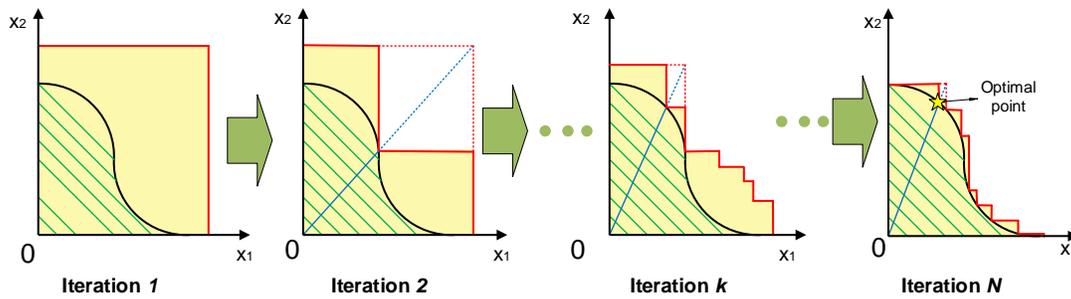


Figure 4.1 – The procedure of polyblock approximation algorithm

Definition 4.6 (Monotonic Optimization). *A canonical monotonic optimization problem has the following form:*

$$\begin{aligned} \max_{\mathbf{x} \succeq \mathbf{0}} \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{x} \in \mathcal{G} \cap \mathcal{H} \end{aligned} \tag{4.9}$$

where \mathcal{G} and \mathcal{H} are nonempty normal and closed reverse normal sets, respectively, and $f(\mathbf{x})$ is an increasing function.

After applying some change-of-variable techniques, a nonconvex optimization problem can be transformed into a monotonic optimization problem. Then the *polyblock approximation algorithm* can be executed to obtain the optimal solution of the original problem. The *polyblock approximation algorithm* is presented in the following.

4.1.3.1 Polyblock Approximation Algorithm

We denote \mathcal{D} and $\partial\mathcal{D}$ as the feasible region and its boundary, respectively, of problem (4.9). The overall framework of polyblock approximation algorithm to solve problem (4.9) optimally is described as follows.

The procedure of the polyblock approximation is depicted in Fig. 4.1. In general, it determines the optimal solution by an iterative-based approach, where in each iteration, the algorithm determines a lower-bound \mathbf{LB} and an upper-bound \mathbf{UB} of the original problem. The lower-bound and upper-bound of the algorithm satisfy the following constraint $\mathbf{LB} \leq f(\mathbf{x}^*) \leq \mathbf{UB}$, where $f(\mathbf{x}^*)$ is the optimal objective value of problem (4.9). The algorithm is terminated as $|\mathbf{UB} - \mathbf{LB}| \leq \epsilon$, where ϵ is a small tolerant parameter.

The values of \mathbf{LB} and \mathbf{UB} in iteration k are determined as follows. In each iteration k , let \mathcal{P}_k be a polyblock containing \mathcal{D} , which is an outer-approximated region of \mathcal{D} and is depicted as the yellow region within the red boundary lines in Fig. 4.1. Moreover, \mathcal{P}_k becomes smaller as the algorithm proceeds, which means that $\mathcal{P}_0 \supset \mathcal{P}_1 \supset \cdots \supset \mathcal{P}_{k+1} \supset \mathcal{P}_k$. In this algorithm, the upper-bound \mathbf{UB} is the optimal value of $f(\mathbf{x})$ in \mathcal{P}_k , which is obtained at one of the vertexes in \mathcal{P}_k . We denote \mathbf{v}_k as the vertex corresponding to the optimal value of $f(\mathbf{x})$ in the polyblock \mathcal{P}_k , and $\boldsymbol{\pi}(\mathbf{v}_k)$ as the projection of \mathbf{v}_k on $\partial\mathcal{D}$. Since $f(\mathbf{x})$ is an increasing function of \mathbf{x} , the optimal solution of problem (4.6) is located in $\partial\mathcal{D}$. Therefore, the lower-bound in iteration k , \mathbf{LB} , can be determined as $\max\{\mathbf{LB}_{k-1}, f(\boldsymbol{\pi}(\mathbf{v}_k))\}$, where \mathbf{LB}_{k-1} is the lower-bound obtained from the previous iteration $k - 1$.

4.1.4 Branch and Bound Algorithm

Branch and bound (BnB) algorithms are methods for global optimization in nonconvex optimization problems. The BnB algorithms maintain an upper and lower bounds on the optimal objective value of the investigated problem. They terminate when the suboptimal point found is ϵ -suboptimal. Note that if we can choose any positive value of ϵ , we can state that the designed algorithm is the optimal one. The BnB algorithm we describe here determines the global minimum of a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ over an n -dimensional rectangle Q_0 , to within some prescribed accuracy ϵ . We let f^*

denote the optimal value, i.e., $f^* = \inf_{\mathbf{x} \in \mathcal{Q}_0} f(\mathbf{x})$. For a rectangle $\mathcal{Q} \in \mathcal{Q}_0$, we define

$$\Phi_{\min}(\mathcal{Q}) = \inf_{\mathbf{x} \in \mathcal{Q}} f(\mathbf{x}) \quad (4.10)$$

so $f^* = \Phi_{\min}(\mathcal{Q}_0)$. In the algorithm, we use two functions $\Phi_{lb}(\mathcal{Q})$ and $\Phi_{ub}(\mathcal{Q})$ defined for any rectangle $\mathcal{Q} \subset \mathcal{Q}_0$. These functions must satisfy the following conditions. First, they are lower and upper bounds of $\Phi_{\min}(\mathcal{Q})$, respectively, for any $\mathcal{Q} \subset \mathcal{Q}_0$:

$$\Phi_{lb}(\mathcal{Q}) \leq \Phi_{\min}(\mathcal{Q}) \leq \Phi_{ub}(\mathcal{Q}). \quad (4.11)$$

The second condition is that the bounds become tight as the rectangle shrinks to a point. Finally, the functions $\Phi_{ub}(\mathcal{Q})$ and $\Phi_{lb}(\mathcal{Q})$ should be cheap to compute since the algorithm involves computation of these functions a huge number of times. We now describe the procedure of BnB algorithm. It starts by computing

$$L_1 = \Phi_{lb}(\mathcal{Q}_0) \quad (4.12)$$

$$U_1 = \Phi_{ub}(\mathcal{Q}_0), \quad (4.13)$$

which are lower and upper bounds on f^* , respectively. If $U_1 - L_1 \leq \epsilon$, the algorithm terminates. Otherwise, we partition \mathcal{Q}_0 into two rectangles, $\mathcal{Q}_0 = \mathcal{Q}_1 \cup \mathcal{Q}_2$, and compute $\Phi_{lb}(\mathcal{Q}_i)$ and $\Phi_{ub}(\mathcal{Q}_i)$, $i = 1, 2$. Then, we have new lower and upper bounds on f^* :

$$L_2 = \min\{\Phi_{lb}(\mathcal{Q}_1), \Phi_{lb}(\mathcal{Q}_2)\} \leq \Phi_{\min}(\mathcal{Q}_0) \quad (4.14)$$

$$U_2 = \min\{\Phi_{ub}(\mathcal{Q}_1), \Phi_{ub}(\mathcal{Q}_2)\}. \quad (4.15)$$

If $U_2 - L_2 \leq \epsilon$, the algorithm terminates. Otherwise, we partition one of \mathcal{Q}_1 and \mathcal{Q}_2 into two rectangles, to obtain a new partition of \mathcal{Q}_0 into three rectangles, and compute Φ_{lb} and Φ_{ub} for these new rectangles. We update the lower bound L_3 as the minimum over the lower bounds over the partition of \mathcal{Q}_0 , and similarly for the upper bound U_3 .

At each iteration we split one rectangle into two, so after k iterations, we have a partition of \mathcal{Q}_0 into k rectangles, $\mathcal{Q}_0 = \cup_{i=0}^k \mathcal{Q}_i$. The associated lower and upper bounds on f^* are expressed as

$$L_k = \min_{i=1, \dots, k} \Phi_{lb}(\mathcal{Q}_i) \quad (4.16)$$

$$U_k = \min_{i=1, \dots, k} \Phi_{ub}(\mathcal{Q}_i). \quad (4.17)$$

We now give the rule for choosing which rectangle to split at each step, and we have to specify which edge along which the rectangle is to be split. One standard method for choosing the rectangle in the current partition to be split is to choose one with the smallest lower bound, i.e., a rectangle that satisfies $\Phi_{lb}(\mathcal{Q}) = L_k$. Once we choose the rectangle to split, we split it along any of its longest edges. We summary the algorithm as below.

Algorithm 4.1. BRANCH AND BOUND ALGORITHM

- 1: Initialization: $k = 0$; $\mathcal{L}_0 = \{\mathcal{Q}_0\}$; $L_0 = \Phi_{lb}(\mathcal{Q}_0)$; $U_0 = \Phi_{ub}(\mathcal{Q}_0)$
 - 2: **while** $U_k - L_k \geq \epsilon$ **do**
 - 3: Pick a rectangle $\mathcal{Q} \in \mathcal{L}_k$ in which $\Phi_{lb}(\mathcal{Q}) = L_k$
 - 4: Split \mathcal{Q} into \mathcal{Q}_I and \mathcal{Q}_{II}
 - 5: Form \mathcal{L}_{k+1} from \mathcal{L}_k by removing \mathcal{Q}_k and adding \mathcal{Q}_I and \mathcal{Q}_{II} .
 - 6: $L_{k+1} = \min_{\mathcal{Q} \in \mathcal{L}_{k+1}} \Phi_{lb}(\mathcal{Q})$
 - 7: $U_{k+1} = \min_{\mathcal{Q} \in \mathcal{L}_{k+1}} \Phi_{ub}(\mathcal{Q})$
 - 8: $k = k+1$
 - 9: **end while**
-

As the algorithm proceeds, we can eliminate some rectangles from consideration; they can be pruned, i.e., pruning procedure, as $\Phi_{min}(\mathcal{Q}_0)$ cannot be achieved in them. This is done as follows. At each iteration, we eliminate from the list \mathcal{L}_k any rectangles that satisfy $\Phi_{lb}(\mathcal{Q}) > U_k$, since every point in such a rectangle is worse than the current upper bound on f^* . In general, the choice of the upper-bound and lower-bound is the most important task for any BnB algorithm. The tight bound can speed up the algorithm dramatically as we can eliminate many rectangles in each iteration.

4.2 Resource Allocation for OFDMA-based System

In this section, we provide the background on the resource allocation for OFDMA-based communication systems.

4.2.1 System Model

The OFDMA-based systems provide three dimensions of diversity which are time, frequency, and users, for a more efficient resource allocation. As the system involves many dimensions of resource allocation, a typical resource allocation algorithm must determine all of these parameters. In this section, we study a general type of resource allocation for an OFDMA-based system which includes subchannel assignment and power control. In particular, we consider a system with $M = \{1, \dots, m, \dots, M\}$ links sharing the resource of $N = \{1, \dots, N\}$ subchannels. We denote $\mathcal{S} = \{1, \dots, s, \dots, S\}$ as the set of all sources in the system, i.e., the set of the transmitters. Moreover, we also have $\mathcal{M} = \{\mathcal{S}_1, \dots, \mathcal{S}_s, \dots, \mathcal{S}_S\}$, where \mathcal{S}_s is the set of links originated from source $s \in \mathcal{S}$. Let h_{kl}^n be the channel gain from the transmitter of link l to the receiver of link k on subchannel n . We denote p_m^n as the transmit power of link $m \in \mathcal{M}$ on subchannel n and we represent the transmit power vector of all links in the system as $\mathbf{p} = [\mathbf{p}_m]_{\forall m \in \mathcal{M}}$ where $\mathbf{p}_m = [p_m^n]_{\forall n \in \mathcal{N}}$ is the power allocation vector of link $m \in \mathcal{M}$ over the subchannels. Moreover, we denote $\boldsymbol{\rho} = [\rho_1^1, \dots, \rho_1^N, \dots, \rho_M^1, \dots, \rho_M^N]$ as the subchannel assignment vector, where $\rho_m^n = 1$ if subchannel n is allocated to link m and $\rho_m^n = 0$, otherwise.

4.2.2 System Utilities

The system utility is the overall objective which the system desire to obtain. In general, the choices of the system utilities can be categorized into one of the following potential forms:

- (i) Weighted sum rates of users' individual rates: $U(\boldsymbol{\rho}, \mathbf{p}) = \sum_{m=1}^M w_m R_m(\boldsymbol{\rho}, \mathbf{p})$;
- (ii) Geometric mean of the rates of users' individual rates: $U(\boldsymbol{\rho}, \mathbf{p}) = \left(\prod_{m=1}^M R_m(\boldsymbol{\rho}, \mathbf{p}) \right)^{\frac{1}{M}}$;
- (iii) Harmonic mean of the rates of users' individual rates: $U(\boldsymbol{\rho}, \mathbf{p}) = \left(\sum_{m=1}^M (R_m(\boldsymbol{\rho}, \mathbf{p}))^{-1} \right)^{-1}$;
- (iv) Minimum rate of all users' individual rates: $U(\boldsymbol{\rho}, \mathbf{p}) = \min_{m \in \mathcal{M}} R_m(\boldsymbol{\rho}, \mathbf{p})$.

Depending on the design objective, we can have different choices of the objective functions. For example, if the main objective is to maximize the spectrum-efficiency of the system we should choose the first utility function while if the objective is to maintain the fairness among the links in the system, we should choose the fourth utility function.

4.2.3 Resource Allocation in Interference-free OFDMA System

In the interference-free OFDMA-based system, i.e., noise-limited system, each subchannel is assigned to at most one link. Therefore, there is no interference among the links in the system. The signal to noise ratio (SNR) achieved by link m on subchannel n can be expressed as

$$\gamma_m^n(\boldsymbol{\rho}, \mathbf{p}) = \frac{\rho_m^n p_m^n h_{mm}^n}{\sigma_m^n}, \quad (4.18)$$

where σ_m^n denotes the noise power on subchannel n of user k . Therefore, the data rate in bit/s/Hz (i.e., normalized by the subchannel bandwidth) of link $m \in \mathcal{M}$ on subchannel n and all subchannels can be expressed, respectively, as

$$R_m^n(\boldsymbol{\rho}, \mathbf{p}) = \log_2(1 + \gamma_m^n(\boldsymbol{\rho}, \mathbf{p})) \quad (4.19)$$

$$R_m(\boldsymbol{\rho}, \mathbf{p}) = \sum_{n \in \mathcal{N}} R_m^n(\boldsymbol{\rho}, \mathbf{p}). \quad (4.20)$$

The resource allocation for the interference-free OFDMA system can be described as follows:

$$\max_{(\boldsymbol{\rho}, \mathbf{p})} U(\boldsymbol{\rho}, \mathbf{p}) \quad (4.21a)$$

$$\text{s.t. } R_m(\boldsymbol{\rho}, \mathbf{p}) \geq R_m^{\min}, \forall m \in \mathcal{M} \quad (4.21b)$$

$$\sum_{m \in \mathcal{S}_s} \sum_{n \in \mathcal{N}} \rho_m^n p_m^n \leq P_{\mathcal{S}_s}^{\max}, \forall s \in \mathcal{S} \quad (4.21c)$$

$$\sum_{m \in \mathcal{M}} \rho_m^n \leq 1, \forall n \in \mathcal{N} \quad (4.21d)$$

$$\rho_m^n \in \{0, 1\} \forall m \in \mathcal{M}, \forall n \in \mathcal{N}. \quad (4.21e)$$

In problem (4.21), constraint (4.21c) restricts the total transmit power of each source, and constraint (4.21d) ensures that each subchannel is allocated to at most one link, which guarantees no interference among the links in the system. Note that in the above problem, if $|\mathcal{S}| = 1$, i.e., the problem has one transmitter and multiple receivers, the modeled system corresponds to the downlink of a single-cell cellular communication system. Moreover, if $|\mathcal{S}_s| = 1$, the modeled system is equivalent to the uplink of a single-cell cellular communication system.

It is proved in [30] that for $M > 2$, the resource problem involving subchannel assignment and power allocation is NP-Hard. These resource allocation problems have been studied in several works [59–64] where mostly heuristic and suboptimal algorithms have been proposed to solve them.

4.2.4 Resource Allocation for Interference-Limited OFDMA-Based System

In the interference-limited system, each subchannel can be reused by multiple links. Each link is suffered from interference from other links using the same subchannel. Therefore, the signal to interference plus noise ratio (SINR) achieved by link m on subchannel n can be expressed as

$$\Gamma_m^n(\boldsymbol{\rho}, \mathbf{p}) = \frac{\rho_m^n p_m^n h_{mm}^n}{\sigma_m^n + \sum_{l \neq m} \rho_l^n p_l^n h_{ml}^n} \quad (4.22)$$

where $\sum_{l \neq m} \rho_l^n p_l^n h_{ml}^n$ represents the interference from the other links using subchannel n . Hence, the data rate of link $m \in \mathcal{M}$ on subchannel n and all subchannels can be expressed, respectively, as

$$R_m^n(\boldsymbol{\rho}, \mathbf{p}) = \log_2(1 + \Gamma_m^n(\boldsymbol{\rho}, \mathbf{p})) \quad (4.23)$$

$$R_m(\boldsymbol{\rho}, \mathbf{p}) = \sum_{n \in \mathcal{N}} R_m^n(\boldsymbol{\rho}, \mathbf{p}). \quad (4.24)$$

The resource allocation for the interference-limited OFDMA system can be described as follows:

$$\max_{(\boldsymbol{\rho}, \mathbf{p})} U(\boldsymbol{\rho}, \mathbf{p}) \quad (4.25a)$$

$$\text{s.t. } R_m(\boldsymbol{\rho}, \mathbf{p}) \geq R_m^{\min}, \forall m \in \mathcal{M} \quad (4.25b)$$

$$\sum_{m \in \mathcal{S}_s} \sum_{n \in \mathcal{N}} \rho_m^n p_m^n \leq P_{\mathcal{S}_s}^{\max}, \forall s \in \mathcal{S} \quad (4.25c)$$

$$\rho_m^n \in \{0, 1\}, \forall m \in \mathcal{M}, \forall n \in \mathcal{N}. \quad (4.25d)$$

Problem (4.25) is much more challenging to tackle compared to problem (4.21) since besides the challenges encountered in problem (4.21), in problem (4.25), we have to deal with the interference among the links using each subchannel. Depending on specific systems and applications, we might have more constraints on subchannel assignment variables. For example, for the downlink multi-cell

OFDMA-based system, in which $s \in \mathcal{S}$ and \mathcal{S}_s correspond to base station (BS) s and the set of links from BS s , respectively, each subchannel is allocated to at most one link in each cell. Therefore, we might have the following additional constraint:

$$\sum_{m \in \mathcal{S}_s} \rho_m^n \leq 1, \forall s \in \mathcal{S}. \quad (4.26)$$

These resource allocation problems have been studied in several works [65–68]. Most of these works tackle the problem by using an iterative approach, where the original problem is solved by alternating between optimizing the subchannel assignment for a given power allocation solution and optimizing the power allocation for a given sub-channel assignment solution. For a given power allocation solution, generally, problem (4.21) is equivalent to an integer problem (IP); therefore, some works solve this problem by applying suitable graph based methods [69–72]. On the other hand, the power allocation problem for a given sub-channel assignment solution is a non-convex optimization problem in general. Most existing works develop some successive convex approximation algorithms (SCA), which solve the original power allocation problem by sequentially tackling an approximated convex optimization problem iteratively [73–75].

4.3 Summary

This chapter discussed some fundamental concepts of optimization and resource allocation for OFDMA-based systems. Firstly, the basic concepts of optimization and convex optimization were introduced. In particular, we briefly described the optimization problem, convex optimization, Lagrange dual problem, and KKT conditions for a particular optimization problem. Then, monotonic optimization and polybock approximation algorithm were presented, which can be used to solve some nonconvex optimization problem with hidden monotonic properties. Afterward, we introduced a general branch and bound algorithm, which returns an optimal solution for a general optimization problem. Finally, we described some general resource allocation formulations for an OFDMA-based system. These optimization techniques and OFDMA-based system models have been used in modeling and solving different resource allocation problems in this dissertation.

Chapter 5

Resource Allocation for D2D Communication Underlaid Cellular Networks Using Graph-based Approach

The content of this chapter was published in IEEE Transactions on Wireless Communication in the following paper:

T. D. Hoang, L. B. Le, and T. Le-Ngoc, “Resource allocation for D2D underlaid cellular networks using graph-based approach,” *IEEE Transaction on Wireless Communication*, vol. 15, no. 10, pp. 7099–7113, October 2016.

5.1 Abstract

In this chapter, we study the non-orthogonal dynamic spectrum sharing for device-to-device (D2D) communications in the D2D underlaid cellular network. Our design aims to maximize the weighted system sum-rate under the constraints that (i) each cellular or active D2D link is assigned one subband and (ii) the required minimum rates for cellular and active D2D links are guaranteed. To

solve this problem, we first characterize the optimal power allocation solution for a given subband assignment. Based on this result, we formulate the subband assignment problem by using the graph-based approach, in which each link corresponds to a vertex and each subband assignment is represented by a hyper-edge. We then propose an Iterative Rounding algorithm and an optimal Branch-and-Bound (BnB) algorithm to solve the resulting graph-based problem. We prove that the Iterative Rounding algorithm achieves at least $1/2$ of the optimal weighted sum-rate. Extensive numerical studies illustrate that the proposed Iterative Rounding algorithm significantly outperforms conventional spectrum sharing algorithms and attains almost the same system sum-rate as the optimal BnB algorithm.

5.2 Introduction

The fast growth of mobile traffic has motivated the development of enabling technologies for significant network capacity enhancement in future wireless networks [3]. Device-to-device communications has been proposed as a mean to improve the system spectral and energy-efficiency and reduce traffic load in the core network [27, 76, 77]. Specifically, in the dense networks, D2D communications can improve the system spectral efficiency significantly since spatial spectrum reuse is exploited effectively through enabling short-range D2D communication links. Efficient radio resource management for D2D communications is essential to realize these benefits.

In general, spectrum assignment for cellular and D2D links can be performed in the orthogonal or non-orthogonal manner [78]. Moreover, non-orthogonal resource allocation for D2D and cellular links can be divided into three scenarios as described in the following.

Scenario I: Each active (admitted) D2D link is assigned one subband and each subband is exploited by at most one D2D link. This scenario allows us to design efficient and low-complexity algorithms, which can be used to address the design in more general settings. This scenario is especially beneficial for the dense deployment of D2D communications.

Scenario II: Each active (admitted) D2D link can be assigned multiple subbands and each subband is exploited by at most one D2D link. This scenario requires more complicated resource allocation design compared to scenario I. This is because beside the design issues of scenario I, scenario II requires to determine the number of subbands allocated to each D2D link.

Scenario III: Each active (admitted) D2D link can be assigned multiple subbands and each subband can be exploited by multiple D2D links. Resource allocation for this scenario is certainly very challenging. In fact, even if the subband assignment solution can be determined, the power allocation problem is still strongly NP-Hard [79]; therefore, only heuristic algorithms can be developed to obtain a feasible solution with practically affordable computation complexity. Furthermore, to solve the resource allocation in this scenario, the channel state information (CSI) of the interfering channels among D2D links over all subbands must be available. Estimation of such CSI may not be feasible in many practical D2D applications, especially in the dense D2D communications setting, due to the large CSI estimation and signaling overhead.

Due to the potential benefits of studying scenario I, many existing works focus on designing efficient resource allocation algorithms for this scenario [16, 17, 23, 53, 80–83]. Early works consider simple network settings such as the system with only one cellular link and one D2D link [80]. Moreover, most existing resource allocation designs assume that channel allocations for cellular links have been predetermined [16, 17, 23, 53, 54, 80–90]. From the admission control perspective, the current literature either ignores the link selection issue or proposes only greedy link selection algorithms [16, 54, 83, 86]. Generally, D2D communications can be assisted and controlled by the cellular base-station (BS) [78] through which optimization of the subband assignment, power allocation, and link selection for both cellular and D2D links would lead to the best system performance.

Solving this joint design problem for any aforementioned scenarios requires us to deal with the nonlinear power allocation and optimization of integer variables related to the subband assignments and link selection. Even for a given power allocation and link selection solution, we still need to tackle an integer subband assignment problem, which is NP-Hard in general. Therefore, it is very challenging to tackle this joint design problem even for scenario I. Moreover, development of an efficient and low-complexity resource allocation algorithm is of great interest for practical implementation. The current work focuses on resource allocation design for scenarios I and II and we reserve the study of scenario III for our future works.

5.2.1 Related Works

Resource allocation design for the setting in which D2D and cellular links share a single channel is investigated in [80–83, 85]. In particular, the authors of [80] consider the power allocation and mode

Tableau 5.1 – Summary of related works and our current work

Ref.	Scenario	Approach	Objective	QoS	Model	SA for cellular links	Link selection	Multi-D2D and cellular links	Theoretical performance analysis	Optimal solution
[80]	I	Optimization	Sum-rate	No	PA	No	No	No	Yes	Yes
[81]	I	Optimization	Sum-rate	Yes	PA	No	No	No	Yes	No
[82]	I	Optimization	Energy efficiency	Yes	PA	No	No	No	No	No
[83]	I	Optimization	Sum-rate	Yes	PA	No	Yes	No	Yes	Yes
[53]	I	Optimization	Sum-rate	Yes	SA, PA	No	No	Yes	Yes	Yes
[23]	I	Optimization	Sum-rate	Yes	SA, PA	No	No	Yes	No	No
[16]	I	Optimization	Weighted sum-rate	Yes	SA, PA	Yes	Yes	No	No	No
[17]	I	Game theory	Sum-rate	Yes	PA, SA	No	No	Yes	No	No
[84]	I	Optimization	Sum-rate	Yes	PA, SA	No	No	No	No	No
[85]	III	Optimization	Sum-rate	No	PA	No	No	No	No	No
[86]	III	Dynamic programming	Number of required subchannels	Yes	No	No	Yes	No	No	No
[87]	III	Dynamic programming	Sum-rate	No	SA, PA	No	No	Yes	No	No
[88]	III	Graph based	Sum utility	Yes	SA, PA	No	No	Yes	No	No
[89]	III	Optimization and game theory	Sum rate	Yes	PA	No	No	Yes	No	No
[54]	III	Optimization	Sum-rate	No	SA, PA	No	Yes	No	No	No
[90]	III	Graph based	Sum-rate	No	SA, PA	No	Yes	No	No	No
[91]	III	Graph based	Sum-rate	No	NA	Yes	No	Yes	No	No
[92]	I	Game theory	Sum-rate	Yes	PA, SA	No	No	Yes	No	No
[93, 94]	III	Game theory	Sum-rate	Yes	PA, SA	No	No	Yes	No	No
[1]	II	Game theory	Energy-efficient	Yes	PA, SA	No	No	Yes	No	No
[2]	II	Game theory	Sum-rate	Yes	PA, SA	No	No	Yes	No	No
[95]	III	Game theory	Sum utility	Yes	PA, SA	No	No	Yes	No	No
Our work	I, II	Graph based and optimization	Weighted sum-rate	Yes	SA, PA	Yes	Yes	Yes	Yes	Yes

selection problem to maximize the sum-rate where they study the power allocation problem for each mode where there are only one cellular link and one D2D link. Joint power and rate control of D2D

and cellular links is studied in [81], and the mode switching problem for D2D communication is investigated in [82]. Both works [81] and [82], however, consider the system with one D2D link and one cellular link. In [83], the power control design is pursued to optimize the spectrum efficiency of D2D links in vehicular systems while the joint admission control, mode selection, and power control problem is studied for a general D2D communication system in [85]. Both works [83] and [85] consider the systems with multiple cellular and D2D links and a single channel.

In general, joint optimization of subband and power allocation is required for efficient resource utilization in multi-channel wireless systems. The setting with multiple D2D and cellular links sharing multiple channels is considered in several recent works [16, 17, 23, 53, 84]. In [53], the system sum-rate optimization for D2D and cellular links is considered. Nonetheless, the authors assume that each cellular link has been pre-allocated one subband and the work aims at optimizing the matching of each D2D link with one cellular link so that the sum rate is maximized. The joint D2D mode selection and resource allocation framework is proposed in [23] where each D2D link can either reuse the resource of cellular links or exploit the dedicated resource assuming that the resource allocation for cellular links is pre-determined. Resource allocation for D2D communication is investigated in [16] where the heuristic matching design between D2D and cellular links based on their relative distance is adopted. Moreover, the stable marriage matching algorithms are adopted in [17, 84] to determine the efficient matching between cellular and D2D links. Although all these existing works focus on scenario I mentioned above, link selection is not studied, these works assume that the resource allocations of the cellular links are pre-determined.

Resource allocation designs, which allow multiple D2D links to reuse the same resource, are studied in several recent papers [54, 86–91, 96, 97]. In particular, the works in [86] and [87] employ the dynamic programming approach to solve the resource allocation for the D2D underlying cellular network. The joint subchannel and power allocation design using an interference-graph method is conducted in [91], a semi-distributed resource allocation algorithm is proposed in [88], and the iterative algorithms are developed in [96, 97]. However, the authors propose to share the downlink cellular resources with D2D links, which is not recommended in the LTE-A standard. In [89], the authors propose a joint spectrum and power allocation algorithm for the D2D underlying cellular system assuming that the power of each D2D link is fixed and the interference from cellular links to D2D links is negligible. Moreover, in [54], a two-step resource allocation algorithm is proposed where greedy subband assignment is performed in the second step after the power allocation in the

first step. Finally, in [90] the graph coloring algorithm is proposed to match cellular resources with one or two D2D links; nevertheless, the power allocation is not studied. For most aforementioned works, the channel assignments for cellular links are assumed to be pre-determined, only heuristic algorithms are proposed, and link selection for D2D links is not investigated.

Game theory has also been employed for D2D resource allocation design in several existing works [1, 2, 92–95]. Auction based resource allocation for D2D communications is studied in [92]. The two-stage Stackelberg game is employed to engineer the resource allocation in [93] where the cellular BS is the leader and D2D links are the followers. Non-cooperative game formulation is adopted to design the resource allocation for D2D links in [94], and the coalition game approach is employed to solve the joint mode selection and resource allocation in [2]. For the game theory approach adopted in [1, 2, 92–95], D2D and cellular links usually act as the players and the obtained stable solution could be satisfactory for all users but it may not be necessarily the most efficient solution. Moreover, these works do not consider link selection for D2D links and subband assignment optimization for cellular links.

We summarize these related works and their characteristics in Table I where PA and SA stand for power allocation and subband allocation, respectively. It can be observed that none of these existing works addresses all following design aspects: consideration of a general setting with multiple cellular and D2D links, joint subband allocation optimization for cellular and D2D links, D2D link selection, QoS guarantees for both cellular and D2D links, and theoretical performance analysis of developed sub-optimal algorithms.

5.2.2 Contributions and Novelty of the Current Work

This chapter focuses on the radio resource allocation for D2D communications in cellular networks for the first scenario and the developed algorithm for scenario I is employed to tackle the resource allocation for scenario II. Specifically, our work makes the following contributions.

- We formulate the resource allocation problem for joint D2D link selection, subband assignment, and power control that aims at maximizing the weighted sum-rate while guaranteeing the minimum rate requirements of individual cellular and active (selected) D2D links. The D2D link selection is indeed embedded into the considered joint optimization problem in our

design. Moreover, to solve this problem, we first derive the optimal power allocation for a given subband assignment for one pair of cellular and D2D links, which enables us to determine the contribution of each subband assignment to the optimization objective. Based on this result, we transform the original resource allocation problem into the subband assignment problem.

- We formulate the subband assignment problem by employing the graph-based approach. Since each link can exploit a subband orthogonally or non-orthogonally, we introduce the concept of virtual cellular and D2D links to capture all possible types of subband assignments. We then formulate a graph-based problem where each link/subband and subband assignment correspond to one vertex and one hyper-edge in the underlying graph, respectively. This problem belongs to the family of three-dimensional matching problems, which are generally NP-Hard.
- We develop a novel Iterative Rounding algorithm to solve the subband assignment problem based on the combination of linear programming and efficient rounding techniques. Specifically, in each iteration we solve a relaxed version of the subband assignment problem for unallocated subbands and network links. Then, we develop a sophisticated mechanism to arrange fractional-weight edges of the underlying graph in an appropriate order and employ the *Local Ratio Method* [12] to determine some subband assignments. Moreover, the unassigned subbands and network links are used to form the network graph based on which we can decide further subband assignments in the next iteration by using the same procedure. We prove that the weighted sum-rate achieved by this Iterative Rounding algorithm is at least half of the optimal weighted sum-rate. In addition, we present an optimal Branch and Bound (BnB) algorithm, which has significantly lower computational complexity compared to the optimal exhaustive search algorithm.
- Numerical results demonstrate that the Iterative Rounding algorithm achieves almost the same sum-rate as that attained by the optimal BnB algorithm. In addition, these two algorithms result in up to 40% sum-rate gain compared to the conventional algorithms in [16] and [17]. Moreover, numerical results also confirm that in the dense D2D communications scenario, resource allocation design under scenario I can achieve reasonably good performance. Finally, we show that by using the Iterative Rounding algorithm in the two-step design approach

to address scenario II, we can achieve dramatically higher sum-rate than those due to the conventional algorithms in [1, 2].

The remaining of this chapter is organized as follows. In Section 5.3, we describe the system model and problem formulation. The optimal power allocation is described in Section 5.4, followed by the description of subband assignment algorithms in Section 5.5. Discussions of algorithm complexity, signaling, and further extensions are given in Section 5.6. In Section 5.7, we present the numerical results, and Section 5.8 concludes the chapter.

5.3 System Model and Problem Formulation

5.3.1 System Model

We consider the spectrum sharing problem among multiple D2D and cellular links in the uplink direction. Let $\mathcal{N} = \{1, \dots, N\}$ with size $|\mathcal{N}| = N$ be the set of subbands in the system.¹ We denote $\mathcal{K}_c = \{1, \dots, K_c\}$ as the set of cellular links, $\mathcal{K}_d = \{K_c + 1, \dots, K_c + K_d\}$ as the set of D2D links, and $\mathcal{K} = \mathcal{K}_c \cup \mathcal{K}_d$ as the set of all communications links with size $|\mathcal{K}| = K_c + K_d = K$. We assume that each subband can be allocated to at most one cellular and one D2D link, which means that cellular and D2D links utilize available subbands orthogonally within its tier.

Let h_{kl}^n be the channel gain from the transmitter of link l to the receiver of link k on subband n . We denote P_k^{\max} as the maximum transmit power of link $k \in \mathcal{K}$. In addition, we denote the power vectors on subband n and all the subbands as $\mathbf{p}^n = [p_1^n, \dots, p_K^n]^T$, and $\mathbf{p} = \text{vec}[\mathbf{p}^1, \dots, \mathbf{p}^N]$, respectively. For clarity, we will also use p_{Ck}^n and p_{Dl}^n to explicitly denote the powers of cellular link k and D2D link l on subband n , respectively. To represent assignment decision of subband n to link $k \in \mathcal{K}$, we define a binary variable ρ_k^n where $\rho_k^n = 1$ if the subband n is assigned to link $k \in \mathcal{K}$, and $\rho_k^n = 0$, otherwise. We also define the subband assignment vectors $\boldsymbol{\rho}^n = [\rho_1^n, \dots, \rho_K^n]^T$ and $\boldsymbol{\rho} = \text{vec}[\boldsymbol{\rho}^1, \dots, \boldsymbol{\rho}^N]$. For convenience, we adopt the following notations: $\mathcal{K}_k \equiv \mathcal{K}_c$ if $k \in \mathcal{K}_c$ and $\mathcal{K}_k \equiv \mathcal{K}_d$ if $k \in \mathcal{K}_d$.

¹Each subband can be a carrier and sub-channel in the multi-carrier wireless networks (e.g., LTE-based wireless networks).

The signal to interference plus noise ratio (SINR) achieved by link $k \in \mathcal{K}$ on subband n can be expressed as

$$\Gamma_k^n(\mathbf{p}^n, \boldsymbol{\rho}^n) = \frac{\rho_k^n p_k^n h_{kk}^n}{\sigma_k^n + \sum_{l \in \mathcal{K} \setminus \mathcal{K}_k} \rho_l^n p_l^n h_{kl}^n}, \quad (5.1)$$

where σ_k^n denotes the noise power for link k on subband n , and $\sum_{l \in \mathcal{K} \setminus \mathcal{K}_k} \rho_l^n p_l^n h_{kl}^n$ is the interference due to other links in the $\mathcal{K} \setminus \mathcal{K}_k$ set. The achievable rates of link $k \in \mathcal{K}$ on subband n and all the subbands can be expressed as

$$R_k^n(\mathbf{p}^n, \boldsymbol{\rho}^n) = \log_2(1 + \Gamma_k^n(\mathbf{p}^n, \boldsymbol{\rho}^n)), \quad (5.2)$$

$$R_k(\mathbf{p}, \boldsymbol{\rho}) = \sum_{n \in \mathcal{N}} R_k^n(\mathbf{p}^n, \boldsymbol{\rho}^n), \quad (5.3)$$

where the rate is calculated in b/s/Hz, which is normalized by the bandwidth of one subband.

5.3.2 Problem Formulation

We assume that each cellular link or active D2D link is allocated one subband, which is suitable for uplink communications [13].² Our design objective is to maximize the weighted sum-rate of all selected D2D links and cellular links while satisfying the minimum required rates of cellular links and active D2D links. Specifically, the QoS requirements of cellular links are expressed as³

$$R_k(\mathbf{p}, \boldsymbol{\rho}) \geq R_k^{\min} \quad \forall k \in \mathcal{K}_c. \quad (5.4)$$

In addition, the minimum rate requirement of an D2D link when it is selected is described as

$$R_k(\mathbf{p}, \boldsymbol{\rho}) \geq \mathbf{I}\left\{\sum_{n \in \mathcal{N}} \rho_k^n = 1\right\} R_k^{\min} \quad \forall k \in \mathcal{K}_d, \quad (5.5)$$

where $\mathbf{I}\{A\}$ denotes the indicator function, which equals to 1 if A is true and equals 0, otherwise.

Here, we have $\sum_{n \in \mathcal{N}} \rho_k^n = 1$ for each selected D2D link k , which is assigned exactly one subband. The

²Extension to the case where each link can be allocated multiple subbands is discussed in Section 5.6.2.

³In practice, it can be infeasible to support these minimum rate constraints. In this work, we assume, however, that these constraints can always be supported.

power constraints of all the links are given as

$$\sum_{n \in \mathcal{N}} p_k^n \leq P_k^{\max} \quad \forall k \in \mathcal{K}. \quad (5.6)$$

Moreover, we assume that each subband can be allocated to only one cellular link and one D2D link, which is captured by the following constraints

$$\sum_{k \in \mathcal{K}_c} \rho_k^n \leq 1 \quad \forall n \in \mathcal{N} \quad (5.7)$$

$$\sum_{k \in \mathcal{K}_d} \rho_k^n \leq 1 \quad \forall n \in \mathcal{N}. \quad (5.8)$$

In addition, each cellular link is assigned one subband and each D2D link can use one subband, which are captured by the following subband assignment constraints

$$\sum_{n \in \mathcal{N}} \rho_k^n = 1 \quad \forall k \in \mathcal{K}_c \quad (5.9)$$

$$\sum_{n \in \mathcal{N}} \rho_k^n \leq 1 \quad \forall k \in \mathcal{K}_d \quad (5.10)$$

$$\rho_k^n \in \{0, 1\} \quad \forall k \in \mathcal{K} \quad \forall n \in \mathcal{N}. \quad (5.11)$$

Note that we only achieve equality in the constraint (5.10) if the corresponding D2D link is selected (active) and assigned one subband accordingly. The considered resource allocation problem can now be formulated as

$$\begin{aligned} \max_{\mathbf{p}, \boldsymbol{\rho}} \quad & \mathcal{R} = \sum_{k \in \mathcal{K}_c} \alpha R_k(\mathbf{p}, \boldsymbol{\rho}) + \sum_{k \in \mathcal{K}_d} (1 - \alpha) R_k(\mathbf{p}, \boldsymbol{\rho}) \\ \text{subject to} \quad & (5.4), (5.5), (5.6), (5.7), (5.8), (5.9), (5.10), (5.11), \end{aligned} \quad (5.12)$$

where α is a weight parameter that controls spectrum sharing of cellular and D2D links.

To solve this problem, we first investigate the optimal power allocation solution for a given subband assignment based on which we can develop subband assignment algorithms.

5.4 Optimal Power Allocation Algorithm

Note that we allow each cellular and active D2D link to use only one subband in the problem formulation (5.12). Therefore, if link $m \in \mathcal{K}$ is allocated subband n exclusively then the optimal power for this link is P_m^{\max} and the corresponding contribution of this link to the objective value is

$$w_m^n \triangleq \begin{cases} \alpha \log_2 \left(1 + \frac{P_m^{\max} h_{mm}^n}{\sigma_m^n} \right) & \text{if } m \in \mathcal{K}_c \\ (1 - \alpha) \log_2 \left(1 + \frac{P_m^{\max} h_{mm}^n}{\sigma_m^n} \right) & \text{if } m \in \mathcal{K}_d. \end{cases} \quad (5.13)$$

However, if cellular link k and D2D link l share subband n then the optimal power allocation must be determined from the following optimization problem⁴

$$\begin{aligned} \max_{p_{Ck}^n, p_{Dl}^n} \quad & w_{kl}^n \triangleq \alpha R_{Ck}^n + (1 - \alpha) R_{Dl}^n \\ \text{s.t.} \quad & R_{Ck}^n \geq R_k^{\min}, R_{Dl}^n \geq R_l^{\min} \\ & p_{Ck}^n \in [0, P_k^{\max}], p_{Dl}^n \in [0, P_l^{\max}], \end{aligned} \quad (5.14)$$

where $R_{Ck}^n = \log_2 \left(1 + \frac{p_{Ck}^n h_{kk}^n}{\sigma_k^n + p_{Dl}^n h_{kl}^n} \right)$ and $R_{Dl}^n = \log_2 \left(1 + \frac{p_{Dl}^n h_{ll}^n}{\sigma_l^n + p_{Ck}^n h_{lk}^n} \right)$.

In fact, (5.14) presents the power allocation problem for weighted sum-rate maximization of two communication links. For this problem, it has been proved in [14] that if the problem is feasible then the optimal transmit powers $\mathbf{P} = (p_{Ck}, p_{Dl})$ have the form

$$\mathbf{P} \in \{(P_k^{\max}, p_{Dl}), (p_{Ck}, P_l^{\max})\}. \quad (5.15)$$

In [14], the authors consider maximizing the sum-rate of two interfering links without minimum rate constraints. For each possible optimal power allocation solution given in (5.15), the sum-rate objective function becomes a concave function of one variable and the optimal solution to maximize such a concave function belongs to the set of extreme points in the power set, i.e., $(0, P_l^{\max})$, $(P_k^{\max}, 0)$, or (P_k^{\max}, P_l^{\max}) .

In contrast to [14], we deal with the maximization of the weighted sum rate of one D2D link and one cellular link with minimum rate constraints. To address this more complicated problem, we

⁴This power allocation problem aims at maximizing the contribution of the underlying subband assignment to the system weighted sum-rate.

have to transform the weighted sum-rate objective function and the transformed function is neither concave nor convex as described in (5.36) of Appendix A. In addition, the feasible region of this constrained problem is more complicated than that of the problem in [14]. Hence, characterization of the optimal power allocation solution in this setting is more challenging.

Let us now define the following quantities:

$$P_{Dl}^{(1)} \triangleq \max \left\{ \frac{(2^{R_l^{\min}} - 1)(P_k^{\max} h_{lk}^n + \sigma_l^n)}{h_{ll}^n}, 0 \right\} \quad (5.16)$$

$$P_{Dl}^{(2)} \triangleq \min \left\{ \frac{1}{h_{kl}^n} \left(\frac{P_k^{\max} h_{kk}^n}{2^{R_k^{\min}} - 1} - \sigma_k^n \right), P_l^{\max} \right\} \quad (5.17)$$

$$P_{Dl}^{(3)} \triangleq (-B_{Dl} + \sqrt{\Delta_{Dl}})/A_{Dl}, \quad (5.18)$$

where A_{Dl} , B_{Dl} , and Δ_{Dl} are specified in Appendix 5.9.1. Then, the optimal power allocation solution of problem (5.14) is characterized in the following proposition whose proof is given in Appendix 5.9.1.

Proposition 5.1. *If the optimal power allocation solution of problem (5.14) is in the form (P_k^{\max}, p_{Dl}) then it belongs the following set*

$$\mathcal{S}_1 \triangleq \begin{cases} \{(P_k^{\max}, P_{Dl}^{(1)}), (P_k^{\max}, P_{Dl}^{(2)}), (P_k^{\max}, P_{Dl}^{(3)})\}, \\ \quad \text{if } P_{Dl}^{(3)} \in [0, P_l^{\max}] \\ \{(P_k^{\max}, P_{Dl}^{(1)}), (P_k^{\max}, P_{Dl}^{(2)})\}, \text{ otherwise.} \end{cases} \quad (5.19)$$

Similarly, let us define

$$P_{Ck}^{(1)} \triangleq \max \left\{ \frac{(2^{R_k^{\min}} - 1)(P_l^{\max} h_{kl}^n + \sigma_k^n)}{h_{kk}^n}, 0 \right\} \quad (5.20)$$

$$P_{Ck}^{(2)} \triangleq \min \left\{ \frac{1}{h_{lk}^n} \left(\frac{P_l^{\max} h_{ll}^n}{2^{R_l^{\min}} - 1} - \sigma_l^n \right), P_k^{\max} \right\} \quad (5.21)$$

$$P_{Ck}^{(3)} \triangleq (-B_{Ck} + \sqrt{\Delta_{Ck}})/A_{Ck}, \quad (5.22)$$

where A_{Ck} , B_{Ck} , and Δ_{Ck} are calculated as

$$\begin{aligned} A_{Ck} &= \frac{1}{\beta} h_{kk} h_{lk}^2, \quad B_{Ck} = \frac{1-\beta}{2\beta} P_l^{\max} h_{ll} h_{lk} h_{kk} + \frac{1}{\beta} \sigma_l h_{lk} h_{kk}, \\ C_{Ck} &= \frac{1}{\beta} \sigma_l h_{kk} (\sigma_l + P_l^{\max} h_{ll}) - P_l^{\max} h_{ll} h_{lk} (\sigma_l + P_l^{\max} h_{kl}), \\ \Delta_{Ck} &= B_{Ck}^2 - A_{Ck} C_{Ck}, \quad \beta = \frac{1-\alpha}{\alpha}. \end{aligned}$$

Then, we have following results.

Proposition 5.2. *If the optimal power allocation solution of problem (5.14) is in the form (p_{Ck}, P_l^{\max}) , then it belongs to the following set*

$$\mathcal{S}_2 \triangleq \begin{cases} \{(P_{Ck}^{(1)}, P_l^{\max}), (P_{Ck}^{(2)}, P_l^{\max}), (P_{Ck}^{(3)}, P_l^{\max})\}, \\ \quad \text{if } P_{Ck}^{(3)} \in [0, P_k^{\max}] \\ \{(P_{Ck}^{(1)}, P_l^{\max}), (P_{Ck}^{(2)}, P_l^{\max})\}, \text{ otherwise.} \end{cases} \quad (5.23)$$

Proof. The proof is omitted due to the space constraint. \square

Combining the results in Propositions 5.1 and 5.2, we can characterize the optimal solution structure of problem (5.14) in the following theorem.

Theorem 5.1. *If the problem (5.14) is feasible then its optimal power allocation solution belongs to the set $\mathcal{S}^* \triangleq \mathcal{S}_1 \cup \mathcal{S}_2$.*

Since \mathcal{S}^* contains at most 6 possible power allocation solutions, we can determine the optimal solution by examining all potential solutions in \mathcal{S}^* easily. Therefore, the optimal contribution to the system weighted sum-rate for each subband n due to exclusive and co-sharing solutions in (5.13) and (5.14), denoted as w_m^n and w_{kl}^n , respectively, can be determined accordingly. If problem (5.13) or (5.14) is not feasible, we will set $w_m^n = -Q$ or $w_{kl}^n = -Q$ where Q is a sufficiently large number so that link m or the pair of D2D link k and cellular link l is not assigned subband n .

5.5 Subband Assignment Algorithms

5.5.1 Graph-based Resource Allocation Formulation

Since optimal power allocation for a given subband assignment can be determined as in the previous section, problem (5.12) can be transformed to the subband assignment problem. We propose to solve the subband assignment problem by using the graph-based approach where each link or subband can be modeled as a vertex, and one subband assignment corresponds to one hyper-edge in the graph. This design is presented in more details in the following.

5.5.1.1 Graph-based Model

We now describe how all design requirements and constraints of the resource allocation problem can be modeled. The constraint (5.9) means that each cellular link must be allocated one subband. To fulfill this requirement, we introduce $\mathcal{K}_{cv} = \{0_{c_1}, 0_{c_2}, \dots, 0_{c_{N_d}}\}$ as the set of $N_d = (N - K_c)$ virtual cellular links. Then, the number of cellular links (actual and virtual cellular links) is equal to the number of subbands. The introduction of virtual cellular links, therefore, enables us to model the subband assignment problem as the one-one matching between cellular links and subbands, which guarantees that each cellular link is assigned one subband. Furthermore, we introduce a single virtual D2D link 0_d and also define $\mathcal{K}_{dv} = \{0_d\}$. For convenience, we also define the sets $\mathcal{K}_C \triangleq \mathcal{K}_c \cup \mathcal{K}_{cv}$ and $\mathcal{K}_D \triangleq \mathcal{K}_d \cup \mathcal{K}_{dv}$.

Now, we define the sets of vertexes and hyper-edges of the graph as follows:

$$V^0 = \{k, l, n \mid k \in \mathcal{K}_C, l \in \mathcal{K}_D, n \in \mathcal{N}\} \quad (5.24)$$

$$E^0 = \{e = (k, l, n) \mid k \in \mathcal{K}_C, l \in \mathcal{K}_D, n \in \mathcal{N}\}, \quad (5.25)$$

where V^0 is the set of vertexes whose elements are the cellular links, D2D links, and subbands; and E^0 is the set of hyper-edges where each hyper-edge $e = (k, l, n) \in E^0$ corresponds to the assignment of subband n to cellular link k and D2D link l . For simplicity, we call *edges* instead of *hyper-edges* in the sequel. Determination of the subband assignment solution is then equivalent to determining a subset of edges in this graph, which satisfies all constraints of the resource allocation problem and maximizes the weighted sum-rate. It is clear that if the final solution chooses an edge

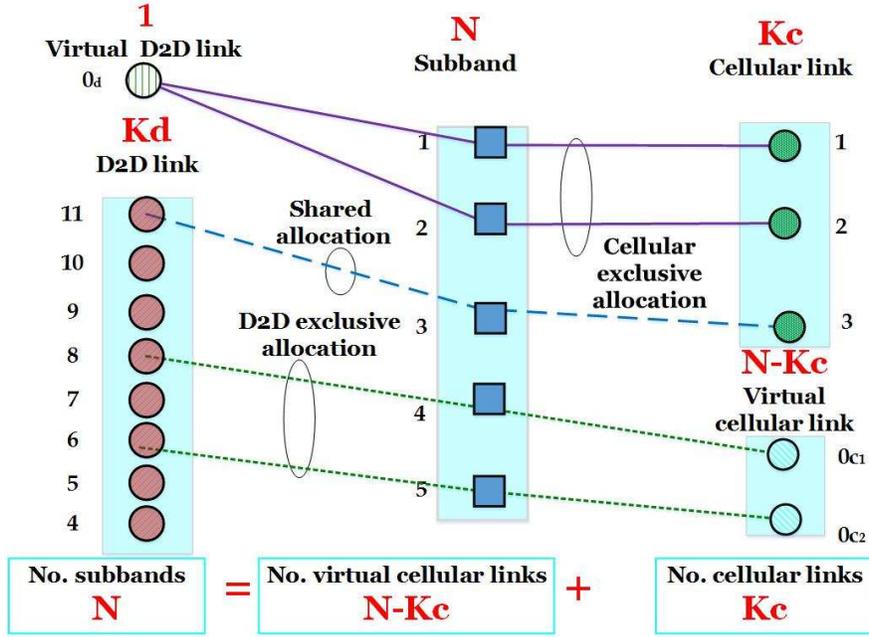


Figure 5.1 – Hyper-edge and vertex presentation of the subband assignment problem

corresponding to cellular link k , virtual D2D link 0_d , and subband n then this cellular link k uses subband n exclusively. It can be observed that a single virtual D2D link 0_d is sufficient for our design purpose if this virtual D2D link 0_d can be matched with multiple cellular links on the corresponding different subbands in the final solution. Similarly, if a particular D2D link l is matched with one virtual cellular link on subband n then this D2D link l uses subband n exclusively.

Fig. 5.1 illustrates this graph representation where the edges represent the corresponding subband assignments. Recall that edge $e = (k, l, n)$ corresponds to the assignment of subband n to cellular link k and D2D link l whose contribution w_e to the weighted sum-rate (the design objective) can be determined through the optimal power allocation in (5.13) and (5.14) for exclusive and sharing subband assignments, respectively.⁵ By using the results in the previous section, we can determine the weights w_e for all possible subband assignments.

5.5.1.2 Subband Assignment Problem

Let $V(E)$ be the set of vertexes associated with the set of edges E . We denote $V_c(E)$, $V_d(E)$, and $V_n(E)$ are the sets of actual cellular links, D2D links, and subbands in the set of vertexes $V(E)$,

⁵Specifically, we have $w_e = w_m^n$ for exclusive subband assignment in (5.13) and $w_e = w_{kl}^n$ for sharing subband assignment in (5.14).

respectively. To describe the subband assignment decision, we introduce a binary variable x_e where $x_e=1$ means edge e is activated (i.e., the corresponding subband allocation is made) and $x_e=0$, otherwise. Moreover, let \mathbf{x} denote the vector whose elements are subband assignment variables x_e associated with all possible edges. In addition, the degree of vertex v in the set of edges E associated with \mathbf{x} can be defined as

$$D(v, E) \triangleq \sum_{e \in E(v)} x_e, \quad (5.26)$$

where $E(v)$ is the set of edges containing vertex v .

Suppose that we have determined the subband assignments for all subbands and let E denote the set of only active edges with $x_e = 1$. Then, $D(v, E), v \in V_d(E)$ and $D(v, E), v \in V_c(E)$ describe the number of subbands allocated to D2D link v and cellular link v , respectively. Similarly, $D(v, E), v \in \mathcal{N}$ is the number of link pairs (D2D and cellular links) using subband $v \in \mathcal{N}$. Therefore, the subband assignment problem can be reformulated into the following integer programming problem $\mathbf{IP}(V, E)$

$$\begin{aligned} \max_{x_e} \quad & \mathcal{R} = \sum_{e \in E} w_e x_e \\ \text{s.t.} \quad & C1 : D(v, E) = 1 \quad \forall v \in V \cap K_C \\ & C2 : D(v, E) \leq 1 \quad \forall v \in V_d(E) \cup V_n(E) \\ & C3 : x_e \in \{0, 1\} \quad \forall e \in E. \end{aligned} \quad (5.27)$$

In problem (5.27), constraints C1 implies that each cellular link v (both actual and virtual links) must be allocated exactly one subband, which guarantees its required minimum rate. In addition, constraints C2 ensure that each D2D link v is assigned at most one subband, and each subband is shared by at most one pair of cellular-D2D links. Constraints C2 allow us to capture two possible outcomes for each D2D link, i.e., it is allocated one subband if it is selected (active) or it is not allocated any subband if it is not selected (inactive). This explains the need to use inequality instead of equality constraints in C2. Note also that the constraint C2 is not applied to the virtual D2D link 0_d since its degree can be greater than 1, i.e., there are multiple cellular links using their allocated subbands exclusively.

For problem $\mathbf{IP}(V, E)$, if we relax the integer allocation constraint $x_e \in \{0, 1\} \quad \forall e \in E$ to $x_e \in [0, 1] \quad \forall e \in E$ then we obtain the corresponding linear relaxation problem, which is referred

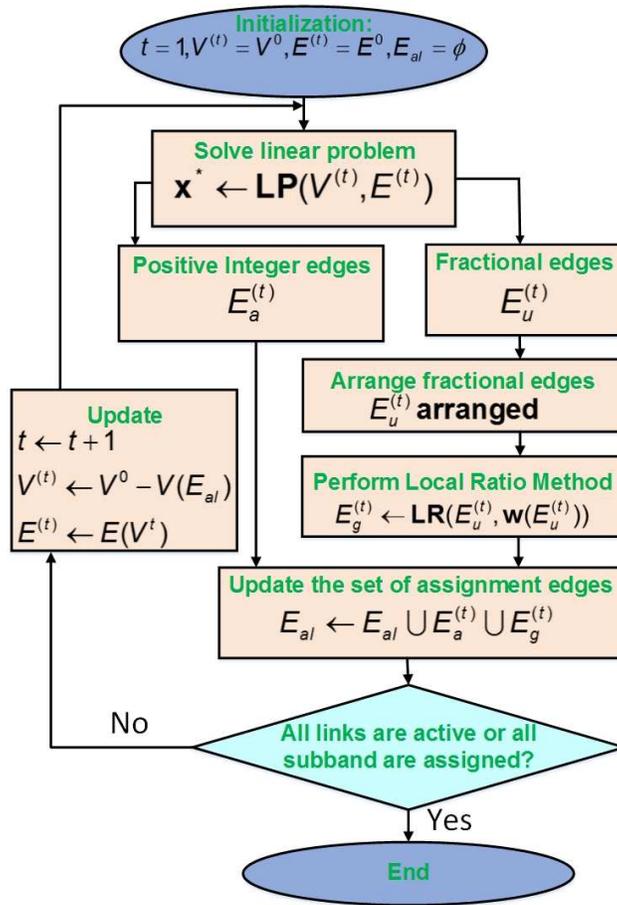


Figure 5.2 – The flowchart of Iterative Rounding Algorithm

to as $\mathbf{LP}(V, E)$ in the sequel. Note that $\mathbf{IP}(V^0, E^0)$ corresponds to the original subband allocation problem, which is challenging to address since there may not exist a polynomial-time algorithm to solve it. To overcome this challenge, we propose two algorithms to solve problem $\mathbf{IP}(V^0, E^0)$ in the following, namely Iterative Rounding algorithm, and optimal BnB algorithm.

5.5.2 Iterative Rounding Algorithm

5.5.2.1 Outline of Design Principals

It can be observed that the linear relaxation problem $\mathbf{LP}(V^0, E^0)$ can be solved easily by standard optimization solutions, i.e., simplex or interior point method [15]. However, solving problem $\mathbf{LP}(V^0, E^0)$ often results in fractional values for some edges e ($0 < x_e < 1$). To address this issue, we propose an Iterative Rounding Algorithm in which we solve a linear relaxation problem and

perform suitable rounding for fractional variables in each iteration. This is repeated until all links are active or all subband are allocated.

Specifically, the Iterative Rounding Algorithm performs the following operations in three phases of each iteration t . In phase 1, it solves the linear relaxation problem for inactive links and available subbands corresponding to the graph with the set of vertexes $V^{(t)}$ and the set of edges $E^{(t)}$, which results in two sets of variables equal to fractional values ($0 < x_e < 1$) and one ($x_e = 1$), namely $E_a^{(t)}$ and $E_u^{(t)}$, respectively. We then arrange the edges in the set $E_u^{(t)}$ with fractional subband assignment variables in phase 2 based on which we employ the Local Ratio Method in phase 3 to determine the set of additional subband assignments $E_g^{(t)}$. Phases 2 and 3 have been indeed appropriately designed to minimize the performance loss due to rounding of the fractional subband assignment variables. The edges in $E_a^{(t)} \cup E_g^{(t)}$ will be used to perform the corresponding subband assignments for cellular and/or D2D links in each iteration. Finally, we update the set of available subbands and inactive links and go back to phase 1 of the next iteration until convergence. The main operations of the algorithm are illustrated in Fig. 5.2.

5.5.2.2 Resource Allocation Algorithm

Detailed operations of the Iterative Rounding Algorithm are presented in Algorithm 5.1. In each iteration t , we have to solve a subband assignment problem formulated in (5.27) whose underlying graph is formed by the sets of vertexes and edges $(V^{(t)}, E^{(t)})$, i.e., $V = V^{(t)}, E = E^{(t)}$, which is a sub-graph of original graph with the corresponding sets (V^0, E^0) . We initialize the algorithm with $V^{(t)} = V^0, E^{(t)} = E^0$ in the first iteration $t = 1$. Moreover, we use E_{al} to denote the set of edges corresponding to all subband assignments accumulated over iterations. The operations of three phases conducted in each iteration t are described in the following.

Linear Relaxation Phase In this phase, we solve the linear programming relaxation problem $\mathbf{LP}(V^{(t)}, E^{(t)})$ (line 4 in Algorithm 5.1), and obtain two sets of subband assignment variables whose values are fractional (smaller than 1) and equal to one, respectively. Specifically, let \mathbf{x}^* denote the

Algorithm 5.1. Iterative Rounding Algorithm

```

1: Initialization  $t = 1$ ,  $E_{al} = \emptyset$ ,  $V^{(t)} = V^0$ ,  $E^{(t)} = E^0$ 
2: while  $V_n(E_{al}) \neq \mathcal{N}$  and  $V_c(E_{al}) \cup V_d(E_{al}) \neq \mathcal{K}_c \cup \mathcal{K}_d$  do
3:   Phase 1:
4:   Solve  $\mathbf{x}^* = \text{argmax LP}(V^{(t)}, E^{(t)})$ .
      $E_a^{(t)} = \{e \in E^{(t)} | x_e^* = 1\}$ ,  $E_u^{(t)} = \{e \in E^{(t)} | 0 < x_e^* < 1\}$ .
5:   Phase 2:
     Set  $E_{ua}^{(t)} \leftarrow \emptyset$ ,  $E_{uu}^{(t)} \leftarrow E_u^{(t)}$ .
6:   for  $i = 1$  to  $|E_u^{(t)}|$  do
7:      $e^* = \text{argmin}_{e \in E_{uu}^{(t)}} c(E_{uu}^{(t)}, e)$ .
      $E_{ua}^{(t)} \leftarrow E_{ua}^{(t)} \cup e^*$ .
      $E_{uu}^{(t)} \leftarrow E_{uu}^{(t)} - e^*$ .
8:   end for
      $E_u^{(t)} \leftarrow E_{ua}^{(t)}$ 
9:   Phase 3:
10:   $E_g^{(t)} \leftarrow \text{LR}(E_u^{(t)}, \mathbf{w}(E_u^{(t)}))$ .
11:   $E_{al} = E_{al} \cup E_a^{(t)} \cup E_g^{(t)}$ .
      $t \leftarrow t + 1$ .
     Update  $V^{(t)} \leftarrow V^0 - V(E_{al})$  and  $E^{(t)} = E(V^{(t)})$ .
12: end while
13: Output  $E_{al}$  and perform subband assignments according to the edges in set  $E_{al}$ .

```

Algorithm 5.2. LR($E, \mathbf{w}(E)$)

```

1:  $E_{\text{temp}} \leftarrow \emptyset$ ,  $j = 1$ .
2: repeat
3:   Choose from  $E$  the smallest index edge  $e^*$ .
4:    $e_j \leftarrow e^*$ 
5:   Update  $E_{\text{temp}} \leftarrow E_{\text{temp}} \cup e_j$  and  $j \leftarrow j + 1$ 
6:   For each edge  $e' \in E \cap E(V(e))$ ,
     1. Update the weight value  $w_{e'} \leftarrow w_{e'} - w_{e^*}$ .
     2. If  $w_{e'} \leq 0$ ,  $E \leftarrow E - e'$ 
7: until  $E = \emptyset$ 
8:  $E_s^{(|E_{\text{temp}}|+1)} = \emptyset$ 
9: for  $j = |E_{\text{temp}}|$  to 1 do
10:  Choose edge  $e_j \in E_{\text{temp}}$  to do the following
     If  $V(e_j) \cap V(E_s^{(j+1)}) = \emptyset$ ,  $E_s^{(j)} \leftarrow E_s^{(j+1)} \cup e_j$ , else  $E_s^{(j)} \leftarrow E_s^{(j+1)}$ .
11: end for
12: Return  $E_s^{(1)}$ .

```

optimal solution of problem $\text{LP}(V^{(t)}, E^{(t)})$ then we define these two sets as

$$E_a^{(t)} = \{e \in E^{(t)} | x_e^* = 1\} \quad (5.28)$$

$$E_u^{(t)} = \{e \in E^{(t)} | 0 < x_e^* < 1\}, \quad (5.29)$$

which are associated with $\text{LP}(V^{(t)}, E^{(t)})$. We define $V_u^{(t)} = V(E_u^{(t)})$ and $V_a^{(t)} = V(E_a^{(t)})$ as the sets of vertexes associated with $E_a^{(t)}$ and $E_u^{(t)}$, respectively. We refer to edges in the set $E_u^{(t)}$ as fractional

edges in the sequel. To proceed further, let $E(V)$ represent the set of edges each of which has at least one of its vertexes in the set V . Based on \mathbf{x}^* , we obtain the assignment vectors \mathbf{x}_a^* and \mathbf{x}_u^* corresponding to edges in the sets $E(V_a^{(t)})$ and $E(V_u^{(t)})$, respectively. Now, $E_u^{(t)}$ and \mathbf{x}_u^* are used in the following two phases to determine additional subband assignments.

Fractional Edge Arrangement Phase The objective of this phase is to arrange the fractional edges, which helps select additional edges for further subband assignments (lines 6-8 Algorithm 5.1). Toward this end, we define a coupling parameter $c(e, E)$ for each fractional edge e as

$$c(e, E) \triangleq \sum_{e' \in E \cap E(V(e))} x_{e'}^*, \quad (5.30)$$

where $V(e)$ is the set of vertexes of edge e . In fact, $c(e, E)$ represents the sum of fractional values of edges in set E that have at least one common vertex with edge e . In this phase, we arrange the edges in $E_u^{(t)}$ in the ascending order of their coupling parameters. Specifically, the edge arrangement procedure is executed as follows. Initially, $c(e, E_u^{(t)})$ is calculated for all edges $e \in E_u^{(t)}$. Then, the edge having smallest coupling parameter is selected and removed from $E_u^{(t)}$. These procedure is repeated until all the edges in $E_u^{(t)}$ is arranged. Moreover, we assume that edges are indexed in the order they are added to set $E_{ua}^{(t,i)}$ in line 7 of Algorithm 5.1. The set of arranged edge $E_u^{(t)}$ is used for the next phase of the algorithm.

Local Ratio Rounding Phase In the third phase, we employ the *Local Ratio Method* [12] to assign the remaining bandwidth resources to fractional edges, which is described in Algorithm 5.2, $\mathbf{LR}(E, \mathbf{w}(E))$. The way these assignments are performed can be explained as follows. The rearranged edges obtained at the end of phase two of Algorithm 5.1 are considered sequentially to perform further subband assignments in the order of increasing indexes of $E_u^{(t)}$. The idea is that if we sequentially round up edges starting from those with smallest coupling parameters and perform corresponding subband assignments then this rounding operation would minimize the impacts to other unallocated edges and therefore the performance loss.

At the end of each iteration t , we obtain the set $E_g^{(t)}$, which contains edges corresponding further subband assignments. Finally, $E_a^{(t)}$ and $E_g^{(t)}$ are added to the “assignment” set E_{al} at the end of each iteration (line 11 in Algorithm 5.1). In Algorithm 5.2 $\mathbf{LR}(E, \mathbf{w}(E))$ where $\mathbf{w}(E)$ is the original

weighted vector corresponding to the edges in E , in which w_e is the weighted value of edge e obtained from the optimal power allocation described in Section 5.4. As we consider a particular edge e , we decrease the weight of each coupled edge e' of e by w_e (line 6). Here, edges e' and e are defined to be coupled if they share at least one vertex and we assume that e is coupled with itself. After these weight updates, any edges with non-positive weights will be excluded from being added to set $E_g^{(t)}$. In the end, we exclude some further edges so that each edge in the returned set $E_s^{(1)}$ is not coupled with any other edges in this set (lines 9-11 of Algorithm 5.2).

5.5.2.3 Properties of Iterative Rounding Algorithm

Firstly, we state an important property related to the linear relaxation phase.

Proposition 5.3. \mathbf{x}_u^* is also an optimal solution of problem $\mathbf{LP}(V_u^{(t)}, E(V_u^{(t)}))$.

Recall that $V_u^{(t)}$ is the set of vertexes corresponding to the fractional edges, and $E(V_u^{(t)})$ is the set of edges each of which has at least one of its vertexes in $V_u^{(t)}$.

Proof. The proof is given in Appendix 5.9.2. □

The results in Proposition 5.3 are very useful because it implies that one can split the linear programming problem $\mathbf{LP}(V^{(t)}, E^{(t)})$ into two linear programming problems $\mathbf{LP}(V_u^{(t)}, E(V_u^{(t)}))$ and $\mathbf{LP}(V_a^{(t)}, E(V_a^{(t)}))$ of smaller size. In addition, problem $\mathbf{LP}(V_a^{(t)}, E(V_a^{(t)}))$ has the integer optimal solution and problem $\mathbf{LP}(V_u^{(t)}, E(V_u^{(t)}))$ has the fractional optimal solution. Consequently, we can perform subband assignments corresponding to the edges in $E_a^{(t)}$ which provides one part of the final subband assignment solution. Moreover, we only need to solve problem $\mathbf{LP}(V_u^{(t)}, E(V_u^{(t)}))$ to determine further subband assignments.

In addition, we need to investigate the solution obtained from each iteration of the algorithm. To facilitate the description, let us denote $z(E) = \sum_{e \in E} w_e$ as total weight of all edges in E , and $l^*(V, E)$ be the optimal objective value of problem $\mathbf{LP}(V, E)$. At each iteration t , we perform the rounding operations to obtain set $E_g^{(t)}$ based on the solution of problem $\mathbf{LP}(V_u^{(t)}, E(V_u^{(t)}))$.

Proposition 5.4. The total weight of selected edges in set $E_g^{(t)}$ (line 10 of Algorithm 5.1) is at least half of the optimal objective of the linear relaxation problem, i.e., $z(E_g^{(t)}) \geq \frac{1}{2}l^*(V_u^{(t)}, E(V_u^{(t)}))$.

Proof. The proof is given in Appendix 5.9.3 □

The above proposition characterizes the achieved objective in each iteration t . We now state the main results that characterize the performance guarantee of Algorithm 5.1. The proof is given in Appendix 5.9.4.

Proposition 5.5. *A feasible solution of the original subband allocation problem $\mathbf{IP}(V^0, E^0)$, offered by Algorithm 5.1, achieves at least half of the optimal objective value of the linear relaxation problem $\mathbf{LP}(V^0, E^0)$.*

This proposition implies that we can always guarantee that Algorithm 5.1 achieves at least 1/2 optimal objective value. This is true even if the solution of problem $\mathbf{LP}(V^0, E^0)$ corresponds all fractional edges.

5.5.3 Branch and Bound Algorithm

We now describe an optimal Branch and Bound (BnB) algorithm with novel branching and bounding procedures to limit the search complexity.

5.5.3.1 Branching

We propose to branch the solution space over a tree comprising K_c levels where each level corresponds to one cellular link as demonstrated in Fig. 5.3. Each node n at level k in the tree represents one potential assignment of subband n to cellular link k . Since each subband can be allocated to at most one cellular link, we remove the underlying subbands from the next-level allocations associated with their child nodes. Therefore, the number of branches originated from each node decreases over the tree levels. The objective of the BnB algorithm is to determine an optimal path through the tree which corresponds to optimal subband assignments.

In general, at each level k , we can branch up to $K_d N$ nodes where each node corresponds to the sharing of cellular link k and D2D link l on subband n . Note that in our proposed tree structure, each node at level k presents the subband assignment for cellular link k , which is independent of the subband assignment for D2D links. Therefore, our branching procedure effectively reduces the search complexity.

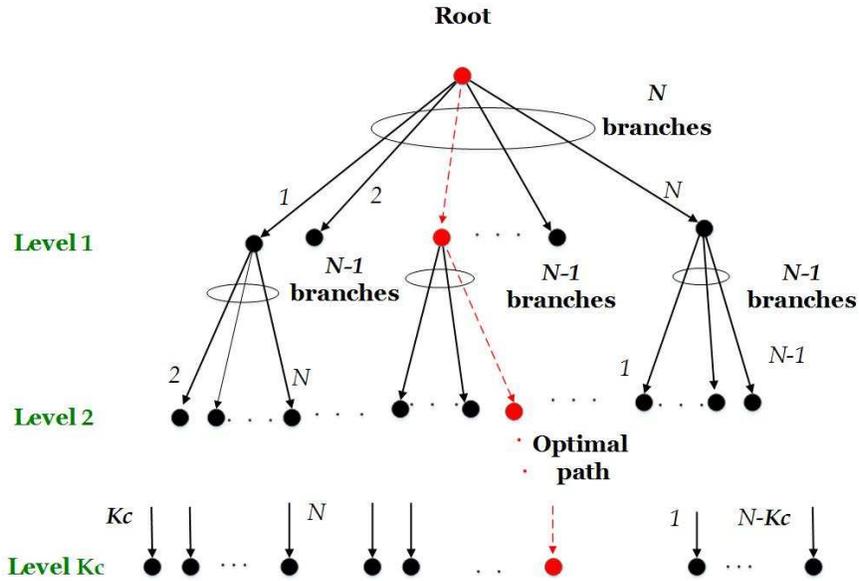


Figure 5.3 – Tree structure of BnB algorithm

5.5.3.2 Bounding

To reduce the search complexity, we remove sub-trees originating from some particular nodes if subband allocations corresponding to these sub-trees cannot belong to the potential optimal path. To facilitate these sub-tree removals, we maintain a global lower-bound Z_{int} which corresponds to the current best feasible solution. Moreover in each visited node, we calculate the local upper-bound B_{ul} which presents an upper-bound of the objective value for the optimal solution containing the visited node. If we have $Z_{\text{int}} > B_{ul}$ at any node then the sub-tree formed by the child nodes of the underlying node will be removed from future considerations.

For a particular node n at level k_0 , there is a unique path from this node to the root, which provides the sets of active cellular links and their allocated subbands, denoted as $V_{ca} \triangleq \{1, \dots, k_0\}$ and $V_{na} \triangleq \{n_1, \dots, n_{k_0}\}$, respectively. The local upper-bound B_{ul} for node n_k at level k can be calculated by solving problem $\mathbf{LP}(V^0, E^0)$ with the following constraints

$$D(k, E(n_k)) = 1 \quad \forall k \in V_{ca}, \quad (5.31)$$

which force subband n_k to be allocated to cellular link k for all $k \in V_{ca}$.

We develop an algorithm to find a feasible solution at level K_c , which can be used to update Z_{int} as follows. Note that at level K_c all cellular links are active. Therefore, problem $\mathbf{IP}(V^0, E^0)$ degenerates into the resource allocation problem of D2D links. We assume that subband $n \in \mathcal{N}$ is allocated for cellular link $k_n \in \mathcal{K}_C$ where k_n can be a virtual cellular link. Hence, if subband n is allocated to D2D link l , the weighted sum rate gain can be calculated as

$$q_l^n = \begin{cases} w_l^n & \text{if } k_n \in \mathcal{K}_{cv} \\ w_{k_n l}^n - w_{k_n}^n & \text{if } k_n \in \mathcal{K}_c, \end{cases} \quad (5.32)$$

where w_l^n , $w_{k_n}^n$ and $w_{k_n l}^n$ are the weighted value of D2D link l if it uses subband n exclusively; weighted value of cellular link k_n if it uses subband n exclusively; and weighted value for cellular link k_n and D2D link l on subband n , respectively. Hence, problem $\mathbf{IP}(V^0, E^0)$ can be transformed to the following problem

$$\begin{aligned} \max_{\mathbf{y}} \quad & \sum_{l \in \mathcal{K}_d} \sum_{n \in \mathcal{N}} q_l^n y_{ln} \\ \text{s.t.} \quad & C1 : \sum_{n \in \mathcal{K}_d} y_{ln} \leq 1 \quad \forall l \in \mathcal{K}_d \\ & C2 : \sum_{l \in \mathcal{N}} y_{ln} \leq 1 \quad \forall n \in \mathcal{N} \\ & C3 : y_{ln} \in \{0, 1\} \quad \forall n \in \mathcal{N} \quad \forall l \in \mathcal{K}_d, \end{aligned} \quad (5.33)$$

where \mathbf{y} represents the subband assignment vector of all D2D links. This problem belongs to the class of job assignment problem, which can be solved efficiently by the Hungarian method [55]. Denote \mathbf{y}^0 and \mathcal{R}^0 as the assignment vector of D2D links and objective value obtained by the Hungarian method. Hence, the objective value of problem $\mathbf{IP}(V^0, E^0)$ can be expressed as $Z_{\text{current}} = \mathcal{R}^0 + \sum_{n \in V_{na}} (\sum_{l \in \mathcal{K}_d} y_{ln}^0) w_{k_n}^n$. Therefore, if $Z_{\text{current}} > Z_{\text{int}}$, we can update the global lower-bound as $Z_{\text{int}} \leftarrow Z_{\text{current}}$. Finally, we search over the proposed tree until all nodes have been solved or excluded to determine the optimal path (i.e., optimal subband assignments).

5.6 Complexity Analysis and Extensions

5.6.1 Complexity Analysis

The computational complexity is analyzed by counting the number of operations required in the power allocation and subband assignment phases, but the complexity of the power allocation phase is indeed negligible. In the Iterative Rounding algorithm, the number of iterations is $O(1)$. In the first iteration, the computational complexity is dominated by the complexity required to solve the linear program. The primal dual interior-point method employed to solve this problem has complexity of $O((K_C K_D N)^{3.5})$, where $K_C K_D N$ is the number of variables [15, p. 324]. The complexity of fractional edge arrangement phase and local ratio rounding phase is negligible; Therefore, the overall computational complexity of the Iterative Rounding algorithm is $O((K_C K_D N)^{3.5})$.

The computational complexity of the BnB algorithm depends on the operations conducted by each node and the number of visited nodes. The complexity of the Hungarian method, which is employed to solve the job assignment and linear programming problems is $O(M^{3.5})$ where M is the number of variables. The worst-case complexity of the BnB algorithm can be calculated based on the maximum number of visited node, which is $\frac{N!}{(N-K_c)!}$. In addition, the complexity of running the Hungarian method are $O((\max\{K_c, K_d, N\})^3)$ [55], which is much smaller than $O(K_C K_D N)^{3.5}$. Hence, the complexity of the BnB algorithm is $O(\frac{N!(K_C K_D N)^{3.5}}{(N-K_c)!})$. Note that the complexity of the exhaustive search algorithm is $O(\frac{N!N!}{(N-K_c)!(N-K_d)!})$ if $N > K_d$ and $O(\frac{N!K_d!}{(N-K_c)!(K_d-N)!})$ if $N < K_d$. Therefore, for sufficiently large values of K_d and N (e.g., $N, K_d > 10$), the BnB algorithm is significantly more efficient than the exhaustive search algorithm.

5.6.2 Further Extensions

We now address the more general case where each cellular or D2D link can be allocated multiple subbands (scenario II). We adopt the two-step approach to perform resource allocation for this general case. In the first step, we determine the number of subbands that must be allocated to each link m , denoted as N_m , by using certain bandwidth allocation approach. Then, we can employ the proposed resource allocation framework presented previously to address the design in the second step as follows. We map each link m to N_m new equivalent links, each of which has the maximum

Tableau 5.2 – Simulation parameters

Description	Parameter	Value
Number of cellular links	K_c	10, 20
Number of D2D links	K_d	10, 25, 30, 50
Number of subbands	N	25
Path loss exponent	γ	3
Maximum distance between Tx and Rx of D2D links	d_{\max}	80m
Maximum transmitted power	P_c^{\max}, P_d^{\max}	0.5W
Minimum required rate	R_c^{\min}, R_d^{\min}	3b/s/Hz
Weight parameter	α	0.5
Noise power	σ_k	10^{-13}

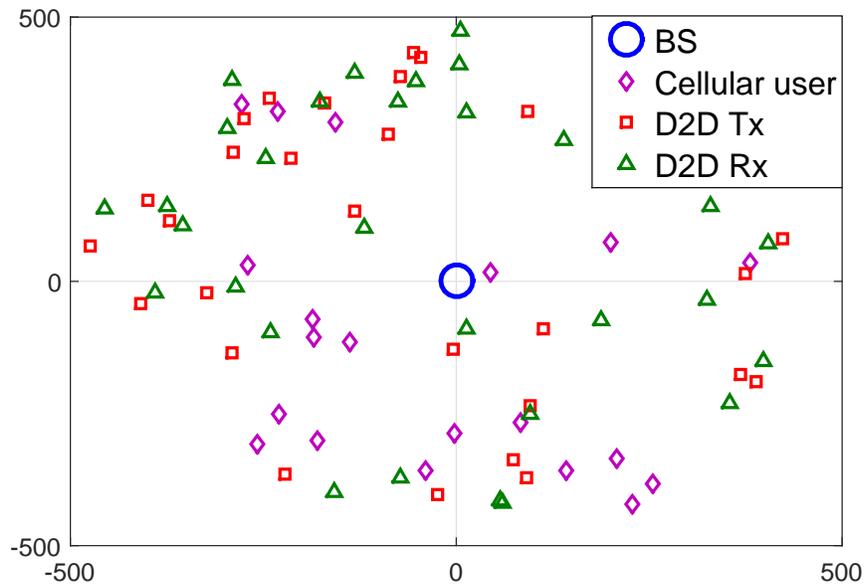
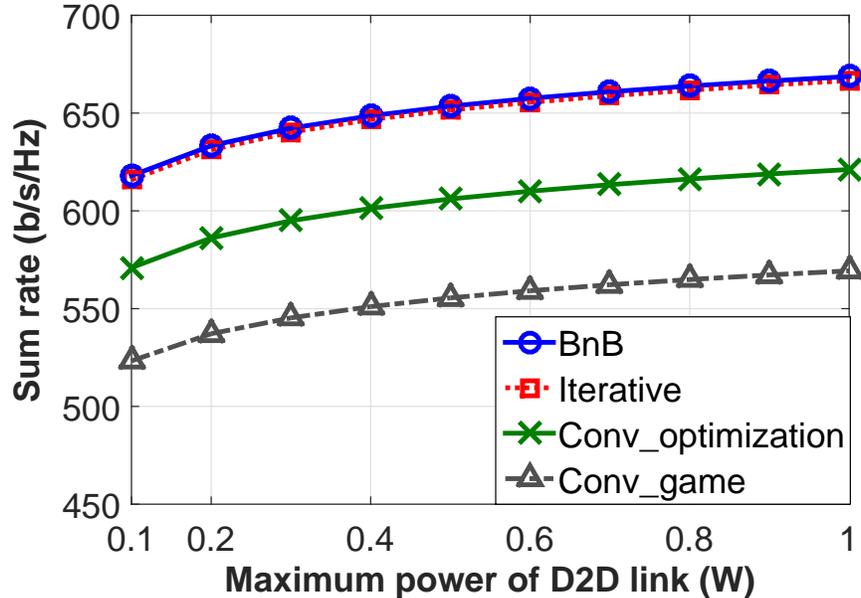


Figure 5.4 – Simulation setting

transmit power P_m^{\max}/N_m , and the minimum required data rate R_m^{\min}/N_m . We then apply the proposed resource allocation algorithms for the equivalent system with more links. Finally, the assigned subbands and the corresponding powers for each link m can be determined from the resource allocation solutions for its N_m links.

Figure 5.5 – System sum rate versus P_d^{\max} as $K_c = 10$

5.7 Numerical Results

5.7.1 Simulation Setting

We consider the simulation setting illustrated in Fig. 5.4 where there is a single BS with the coverage area of 500m serving $K_c = 20$ randomly distributed cellular users. Moreover, there are $N = 25$ subbands, which are shared by $K_c = 20$ cellular links and $K_d = 30$ D2D links unless stated otherwise. The D2D transmitters are randomly distributed in the cell area, and each D2D receiver is randomly located in the area close to its D2D transmitter within the maximum distance of d_{\max} . The channel power gain is modeled as $h_{kl}^n = d_{kl}^{-\gamma} \delta$ where d_{kl} is the distance between the receiver of link k and the transmitter of link l ; δ represents the Rayleigh fading, which follows exponential distribution with the mean value of 1. The values of parameters for our simulation are summarized in Table 5.2 unless stated otherwise.

5.7.2 Numerical Results for Scenario I

We compare the performance of our proposed algorithms with conventional algorithms developed for scenario I in [16] and [17]. The first conventional algorithm, referred to as the optimization-based conventional algorithm, performs the resource allocation in three phases. In the first phase,

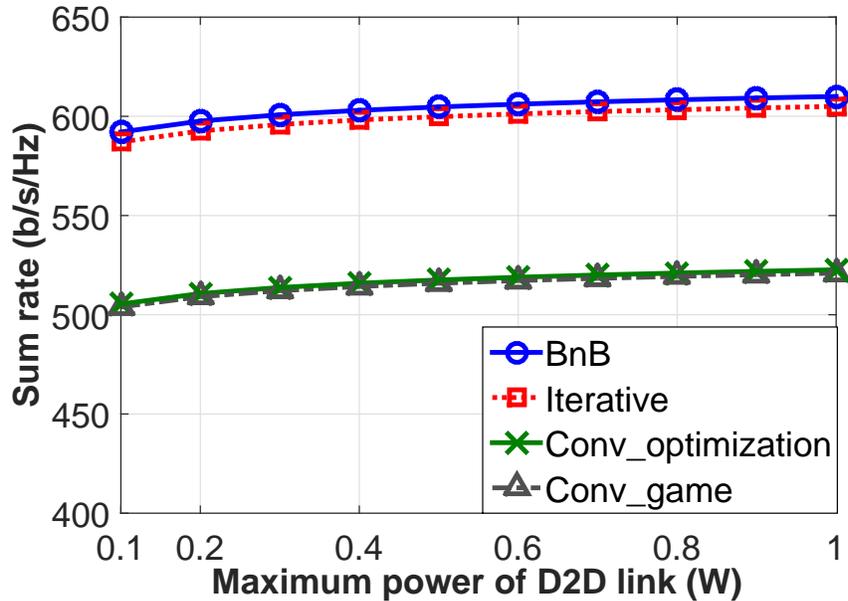


Figure 5.6 – System sum rate versus P_d^{\max} as $K_c = 20$

the matching between cellular and D2D links is executed based on the relative distance between cellular transmitter and D2D receiver. Then, in the second phase, the power allocation for a pair of cellular and D2D links is performed to achieve the optimal weighted sum-rate of both links. Finally, the matching between the subbands and different pairs of links is determined by using the Hungarian method. In contrast to the algorithm in [16], the work [17] adopted a game-based approach where the resource allocation for cellular links is determined first. Then, the optimal power allocation for each pair of cellular and D2D links is calculated. Finally, the matching between D2D links and the resources of cellular links is decided by using the stable matching and cheating technique [98].

In the following, all numerical results are obtained by averaging over 1000 random realizations of D2D, cellular locations, and channel gains. The results corresponding to the BnB, Iterative Rounding, optimization-based conventional, and game-based conventional algorithms are indicated by “BnB”, “Iterative”, “Conv_optimization”, and “Conv_game” in all figures, respectively.

Figs. 5.5 and 5.6 illustrate the system sum-rate versus the maximum power of each D2D link as $K_c = 10$ and $K_c = 20$, respectively. It shows that the system sum-rate increases moderately as the D2D maximum power becomes larger. This is because short-range D2D links do not require very large transmit power to achieve high data rates compared to the cellular links. Moreover, the D2D power budget is mainly used to combat the interference from cellular links, which interprets the moderate gain in system sum-rate. It is remarkable that the Iterative Rounding algorithm

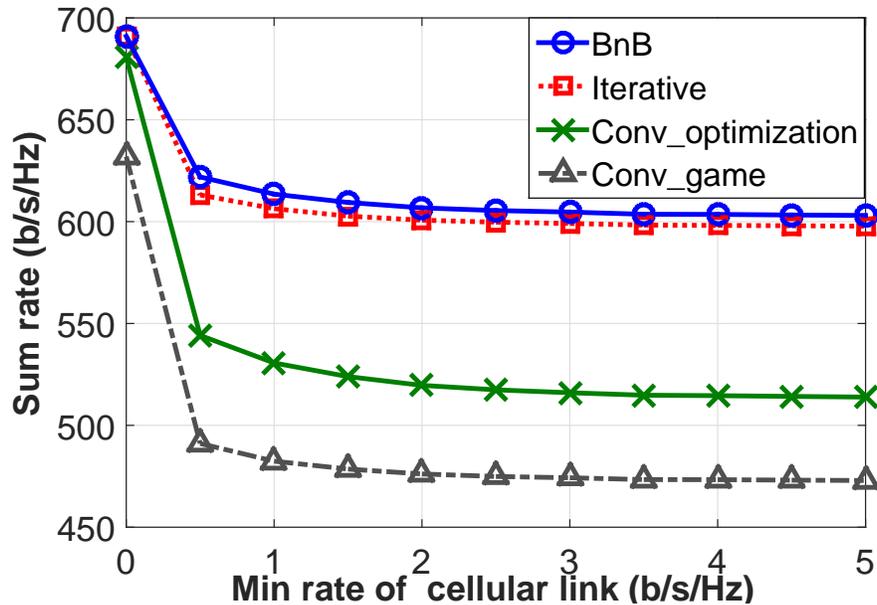


Figure 5.7 – Sum rate versus minimum rate of each cellular link

performs extremely well and its achieved sum-rate is almost the same as that due to the optimal BnB algorithm. The Iterative Rounding algorithm indeed performs quite better than the worst-case performance bound stated in Proposition 5.5. This is because the worst-case bound assumes that the linear relaxation phase results in all fractional edges and the worst performance loss in the rounding phase.

Moreover, the Optimization-based conventional algorithm performs better than the Game-based conventional algorithm as $K_c = 10$. This is due to the fact that the Optimization-based conventional algorithm jointly optimizes the subband assignments for both cellular and D2D links while the Game-based conventional algorithm performs subband assignments for cellular links and D2D links in two separate steps which results in the significant performance loss.

In Fig. 5.7, we demonstrate the system sum-rate versus the minimum required rate of cellular links R_c^{\min} . It can be seen that the system sum-rate reaches the maximum value as $R_c^{\min} = 0$. This is because when as $R_c^{\min} = 0$, D2D links have more advantages than cellular links in accessing good subbands thanks to the short-range of D2D links. Hence, the rate of D2D links become higher for the smaller minimum required rate of each cellular link. It can also be observed that the system sum-rate decreases significantly as R_c^{\min} increases from zero before getting saturated at fixed value.

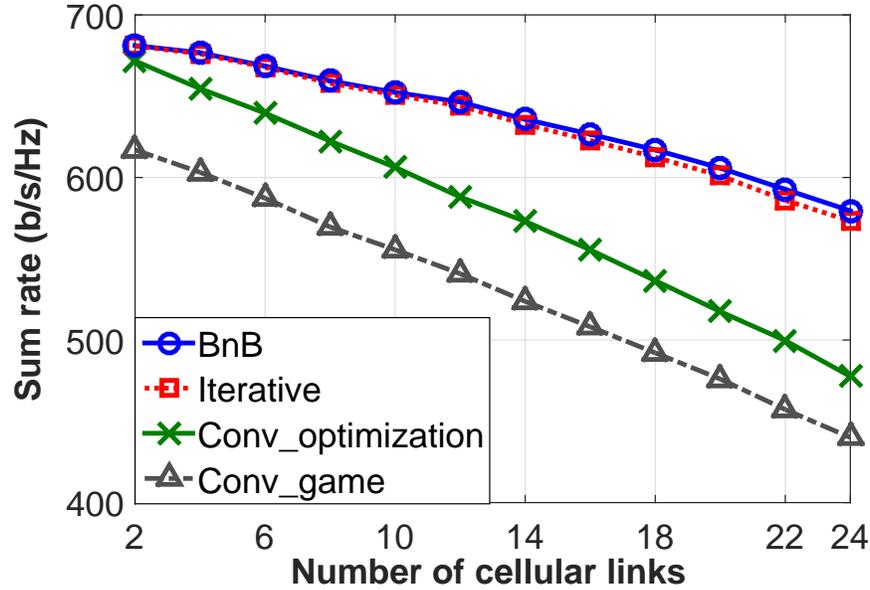
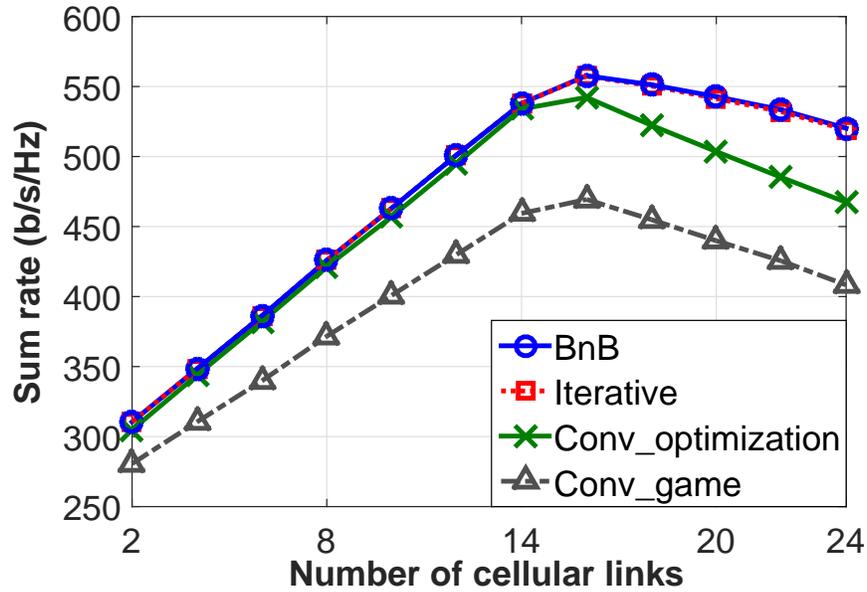
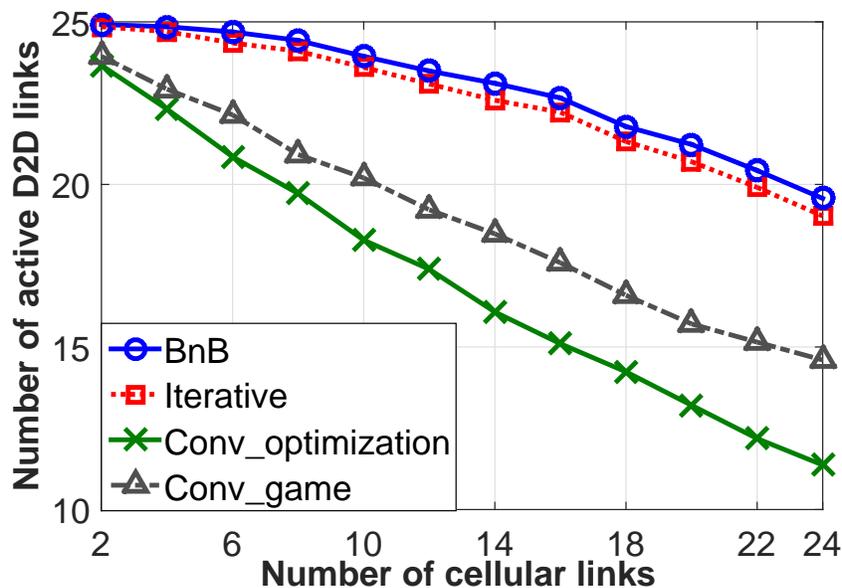


Figure 5.8 – System sum rate versus K_c as $K_d = 30$

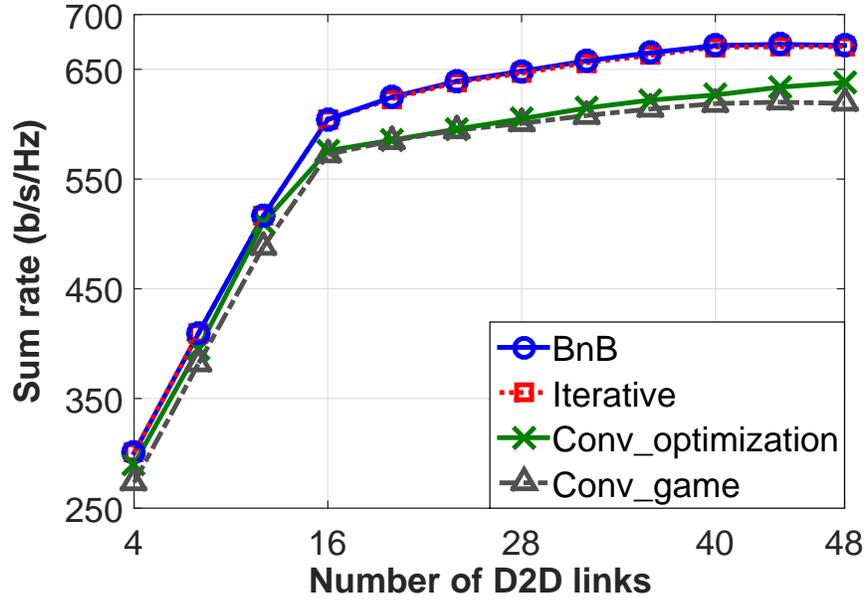
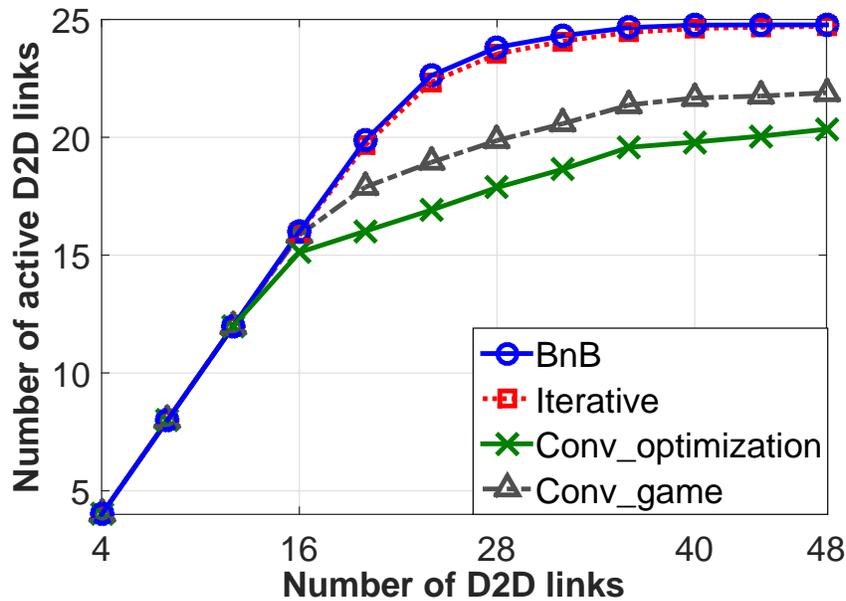
Fig. 5.8 shows the variations of the system sum-rate with the number of cellular links K_c as we fix $K_d = 30$. This figure demonstrates that the system sum-rate decreases with the number of cellular links. In fact, as K_c increases, the number of active D2D links is reduced, which results in the decrease in the system sum-rate. In addition, Fig. 5.9 demonstrates the system sum-rate versus K_c as we fix $K_d = 10$. As shown in this figure, the system sum-rate first increases then decreases with K_c , which can be explained as follows. As K_c is small, increasing K_c does not significantly impact the data rates of D2D links since all D2D links can still exploit the subbands exclusively. Moreover, larger number of cellular links can result in better spectrum utilization, which improves the system sum-rate. However, when K_c is sufficiently large, increasing K_c leads to the scenario where active D2D links must share subbands with cellular links and the number of active D2D links decreases, which is confirmed by Fig. 5.10. Therefore, it results in the decrease of the system sum-rate.

Fig. 5.11 presents the system sum-rate for varying number of D2D links K_d as we fix $K_c = 10$. It is shown that as K_d increases, the system sum-rate increases quite significantly for all algorithms. In fact, as K_d increases, number of active D2D links increases, which is confirmed by Fig. 5.12; therefore, the higher number of D2D links K_d leads to the greater system sum-rate. Moreover, since K_c is small, as K_d increases, D2D links can access the large bandwidth, which results in significant improvement of the system sum-rate.

Figure 5.9 – System sum rate versus K_c as $K_d = 10$ Figure 5.10 – Number of active D2D links versus K_c as $K_d = 30$

5.7.3 Numerical Results for Scenario II

We now study the achievable network performance when the Iterative Rounding Algorithm is utilized to address scenario II where each active D2D link can be assigned multiple subbands. We compare our proposed algorithms with the algorithms in [1] and [2] denoted as “Conventional 1” and “Conventional 2” algorithms. The former algorithm performs resource allocation by using the

Figure 5.11 – System sum rate versus K_d as $K_c = 10$ Figure 5.12 – Number of active D2D links versus K_d as $K_c = 10$

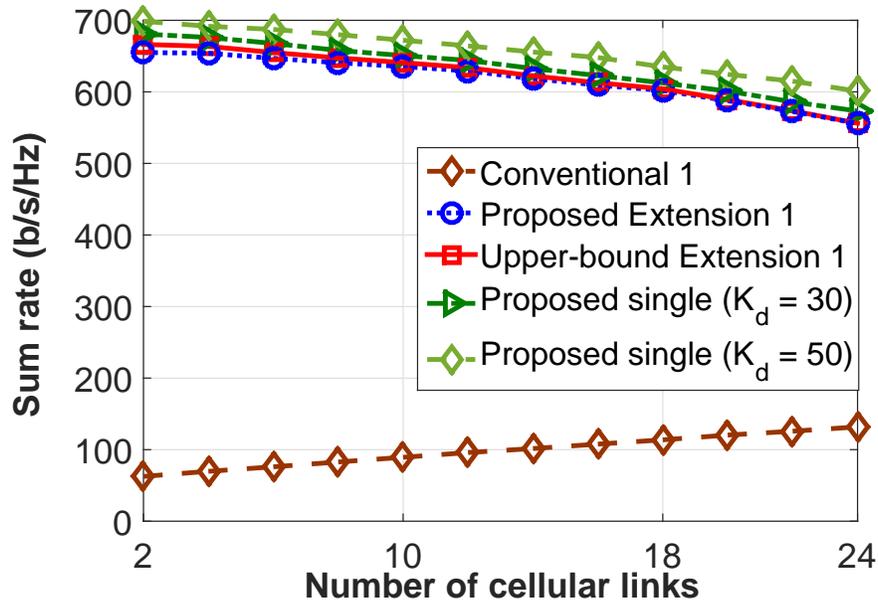
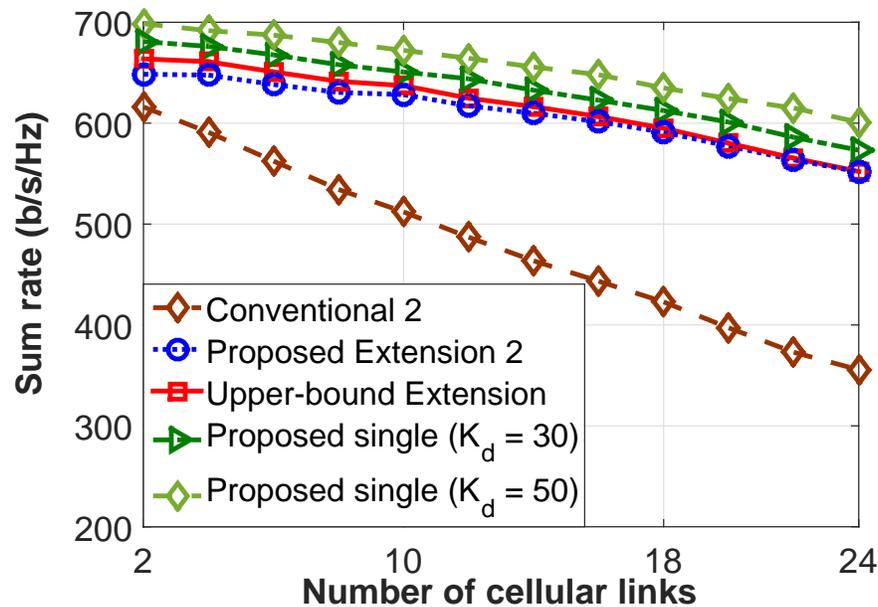
coalitional game approach where the coalitions are determined by a greedy algorithm. The latter algorithm implements the resource allocation by using merge-and-split coalitional game method.

Results for all proposed algorithms presented in the following employ the Iterative Rounding algorithm, and they are different in the methods employed to determine the number of subbands assigned to each D2D link in step one of the proposed two-step approach. While in the “Proposed single” algorithm, each D2D link is assigned one subband, the “Proposed Extension 1” and

“Proposed Extension 2” determine the number of subbands assigned to each D2D link by using the outcomes of “Conventional 1” and “Conventional 2” algorithms, respectively. Moreover, the “Proposed multi-fair” algorithm assigns the same number of subbands to all admitted D2D links. We also consider the “Full admission” scheme where all D2D links are admitted without performing adaptive D2D link selection. Finally, the “Proposed multi-sequential” algorithm determines the number of subbands for D2D links as follows. First, each D2D link is assigned one subband. Then, if there are still available subbands, we sequentially assign one more subband to one D2D link, which results in the highest improvement of the system sum-rate in each iteration.

Figs. 5.13 and 5.14 show the sum-rate versus the number of cellular links K_c as $K_d = 25$ (K_d is shown, otherwise). It can be seen that the proposed algorithms (“Proposed single”, “Proposed Extension 1”, and “Proposed Extension 2”) perform significantly better than the conventional algorithms, which confirms the efficacy of the proposed joint subband and power allocation design. The “Proposed single” algorithm even performs better than the “Proposed Extension 1” and “Proposed Extension 2” algorithms. This implies that in the dense D2D communication scenario, assignment of a single subband for each D2D link results in sufficiently good performance. In addition, the “Proposed single” algorithm achieves higher sum rate for $K_d = 50$ than for $K_d = 25$ thanks to the additional multiuser diversity gain. Finally, the “Proposed single” algorithm results in better sum-rate than that of the “Full admission”, which demonstrates the benefits of adaptive D2D link selection.

Figs. 5.15 and 5.16 show the sum-rate versus the number of D2D links K_d as $K_c = 20$. It is shown that as K_d is small, assignment of multiple subbands for each D2D link (scenario II) can result in better performance. However, as the number of D2D links increases, the performance gap among different designs for scenarios I and II decreases. Moreover, except for cases where K_d is in the interval $[8, 16]$, the performances of the “Proposed multi-fair” and “Proposed multi-sequential” algorithms are similar. This illustrates the regions where fair subband allocation can be adopted to enjoy both low computation complexity and good performance.

Figure 5.13 – Sum rate versus K_c (Comparing with [1])Figure 5.14 – Sum rate versus K_c (Comparing with [2])

5.8 Conclusion

In this chapter, we have developed efficient resource allocation algorithms for D2D underlaid cellular network systems. The proposed algorithms are differentiated mainly by their subband assignment designs, which are developed based on the optimal power allocation solution for individual pairs of cellular and D2D links on any subband. We have established the theoretical performance guarantee

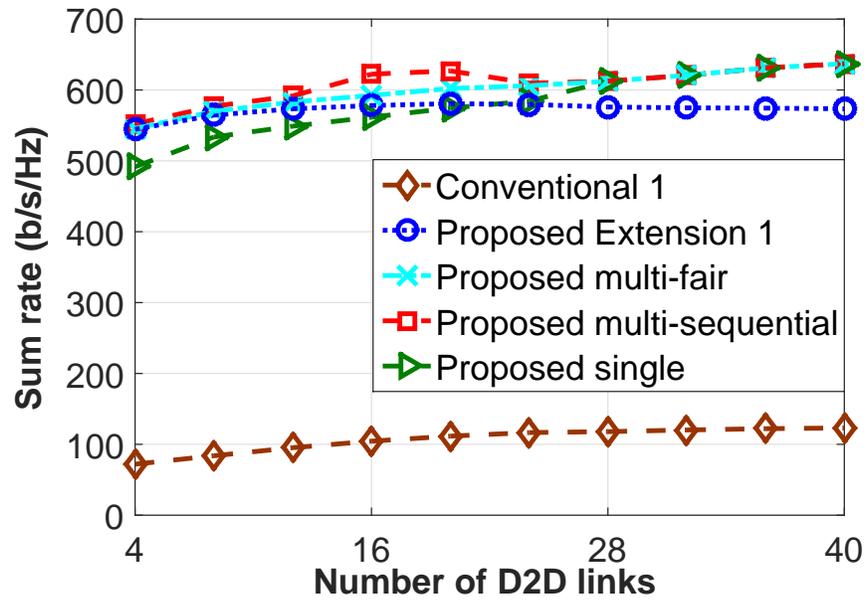


Figure 5.15 – Sum rate versus K_d (Comparing with [1])

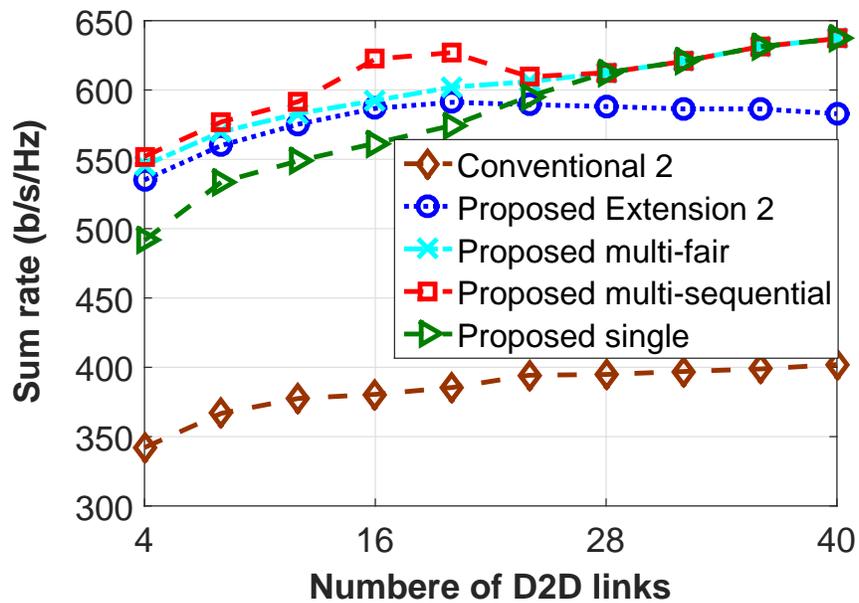


Figure 5.16 – Sum rate versus K_d (Comparing with [2])

for the Iterative Rounding algorithm and analyzed the computational complexity of the proposed algorithms. Numerical results have confirmed that the proposed designs result in better performance than state-of-the-art algorithms.

5.9 Appendices

5.9.1 Proof of Proposition 5.1

In problem (5.14), we only consider subband n . Therefore, we omit the index n in this Appendix for brevity. If the optimal power of problem (5.14) belongs to the set (P_k^{\max}, p_{Dl}) , then we only have to find the optimal power allocation of D2D link i.e., p_{Dl} . The rates achieved by cellular link k and D2D link l can be expressed, respectively, as $R_{Ck} = \log_2(1 + P_k^{\max}h_{kk}/(\sigma_k + p_{Dl}h_{kl}))$, and $R_{Dl} = \log_2(1 + p_{Dl}h_{ll}/(\sigma_l + P_k^{\max}h_{lk}))$.

It can be verified that R_{Ck} is monotonically decreasing with p_{Dl} ; hence, to satisfy minimum rate requirement of cellular link k we must have $p_{Dl} \leq (P_k^{\max}h_{kk}/(2^{R_k^{\min}} - 1) - \sigma_k)/h_{kl}$. Similarly, R_{Dl} is monotonically increasing with p_{Dl} . Hence, $p_{Dl} \geq (2^{R_l^{\min}} - 1)(P_k^{\max}h_{lk} + \sigma_l)/h_{ll}$. As a result, to obtain a feasible solution, we must have $p_{Dl} \in [P_{Dl}^{(1)}, P_{Dl}^{(2)}]$, where

$$P_{Dl}^{(1)} = \max \left\{ \frac{(2^{R_l^{\min}} - 1)(P_k^{\max}h_{lk} + \sigma_l)}{h_{ll}}, 0 \right\} \quad (5.34)$$

$$P_{Dl}^{(2)} = \min \left\{ \frac{1}{h_{kl}} \left(\frac{P_k^{\max}h_{kk}}{2^{R_k^{\min}} - 1} - \sigma_k \right), P_l^{\max} \right\}. \quad (5.35)$$

In addition, we modify w_{kl} as $w_{kl} = \alpha(R_{Ck} + \beta R_{Dl})$ where $\beta = (1 - \alpha)/\alpha$. Now we define

$$f(p_{Dl}) = \left(1 + \frac{P_k^{\max}h_{kk}}{\sigma_k + p_{Dl}h_{kl}} \right) \left(1 + \frac{p_{Dl}h_{ll}}{\sigma_l + P_k^{\max}h_{lk}} \right)^\beta. \quad (5.36)$$

Then, maximizing w_{kl} is *equivalent* to maximizing $f(p_{Dl})$. The optimal value of $f(p_{Dl})$ can be found by solving the equation $f'(p_{Dl}) = 0$. The derivative of $f(p_{Dl})$ can be expressed as $f'(p_{Dl}) = (A_{Dl}p_{Dl}^2 + 2B_{Dl}p_{Dl} + C_{Dl})/D_{Dl}$ where $A_{Dl} = \beta h_{ll}h_{kl}^2$, $B_{Dl} = 0.5\beta - 1P_k^{\max}h_{kk}h_{kl}h_{ll} + \beta\sigma_k h_{kl}h_{ll}$, $C_{Dl} = \beta\sigma_k h_{ll}(\sigma_k + P_k^{\max}h_{kk}) - P_k^{\max}h_{kk}h_{kl}(\sigma_l + P_k^{\max}h_{lk})$, and $D_{Dl} = (\sigma_l + P_k^{\max}h_{lk})(\sigma_k + p_{Dl}h_{kl})^2$.

If $\Delta_{Dl} = B_{Dl}^2 - A_{Dl}C_{Dl} > 0$, $f'(p_{Dl}) = 0$ has two solutions $(-B_{Dl} \pm \sqrt{\Delta_{Dl}})/A_{Dl}$. In addition, $(-B_{Dl} - \sqrt{\Delta_{Dl}})/A_{Dl} < 0$ and $(-B_{Dl} + \sqrt{\Delta_{Dl}})/A_{Dl}$ are respectively the minimum and maximum points of function $f(p_{Dl})$. Let us denote $P_{Dl}^{(3)} = (-B_{Dl} + \sqrt{\Delta_{Dl}})/A_{Dl}$. If $P_{Dl}^{(3)} \in [P_{Dl}^{(1)}, P_{Dl}^{(2)}]$, $P_{Dl}^{(3)}$ is the optimal solution of problem (5.14). On the other hand, if $P_{Dl}^{(3)} \notin [P_{Dl}^{(1)}, P_{Dl}^{(2)}]$, $f(p_{Dl})$ is a monotonic function in $[P_{Dl}^{(1)}, P_{Dl}^{(2)}]$; therefore, $P_{Dl}^{(1)}$ or $P_{Dl}^{(2)}$ is the optimal solution of problem (5.14).

If $\Delta_{Dl} < 0$, which means that $f'(p_{Dl}) > 0 \forall p_{Dl}$ or $f(p_{Dl})$ is a monotonic function of p_{Dl} then $f(p_{Dl})$ achieves its maximum at $P_{Dl}^{(1)}$ or $P_{Dl}^{(2)}$. Therefore, in any cases, the optimal solution of problem (5.14) belongs to \mathcal{S}_1 which completes the proof of the proposition.

5.9.2 Proof of Proposition 5.3

We prove the proposition by contradiction. Specifically, we show that if \mathbf{x}_u^* is not an optimal solution of problem $\mathbf{LP}(V_u^{(t)}, E(V_u^{(t)}))$, then \mathbf{x}^* is not the optimal solution of problem $\mathbf{LP}(V^{(t)}, E^{(t)})$. First, consider problem $\mathbf{LP}(V, E)$ and let us denote $l(V, E, \mathbf{x})$ as the objective value associated with \mathbf{x} and the optimal objective value is $l^*(V, E)$. Suppose that $\mathbf{x}^o \neq \mathbf{x}_u^*$ is the optimal solution of problem $\mathbf{LP}(V_u^{(t)}, E(V_u^{(t)}))$. Now we define a new vector \mathbf{x}' where the value of its element is set as $x'_e = 1$ if $e \in E_a^{(t)}$, $x'_e = x_e^o$ if $e \in E(V_u^{(t)})$, and $x'_e = 0$, otherwise. On the other hand, we have

$$l(V^{(t)}, E^{(t)}, \mathbf{x}^*) = l(V_a^{(t)}, E(V_a^{(t)}), \mathbf{x}_a^*) + l(V_u^{(t)}, E(V_u^{(t)}), \mathbf{x}_u^*) \quad (5.37)$$

$$l(V^{(t)}, E^{(t)}, \mathbf{x}') = l(V_a^{(t)}, E(V_a^{(t)}), \mathbf{x}_a^*) + l(V_u^{(t)}, E(V_u^{(t)}), \mathbf{x}^o). \quad (5.38)$$

Since $l(V_u^{(t)}, E(V_u^{(t)}), \mathbf{x}_u^*) < l(V_u^{(t)}, E(V_u^{(t)}), \mathbf{x}^o)$, we have $l(V^{(t)}, E^{(t)}, \mathbf{x}^*) < l(V^{(t)}, E^{(t)}, \mathbf{x}')$. Hence, \mathbf{x}^* is not the optimal solution of $\mathbf{LP}(V^{(t)}, E^{(t)})$, which is a contradiction.

5.9.3 Proof of proposition 5.4

The proof of Proposition 5.4 is equivalent to the proof of the following inequality

$$\sum_{e \in E_s^{(1)}} w_e \geq \frac{1}{2} \sum_{e \in E_u^{(t)}} x_e^* w_e, \quad (5.39)$$

where $E_s^{(1)}$ is the set of edges returned in Algorithm $\mathbf{LR}(E_u^{(t)}, \mathbf{w}(E_u^{(t)}))$, and x_e^* corresponds to edge e in \mathbf{x}_u^* which is the optimal solution of problem $\mathbf{LP}(V_u^{(t)}, E(V_u^{(t)}))$. To complete the proof of this proposition, we use the results in the following lemma whose proof can be adapted from that in [99] with some minor modifications.

Lemma 5.1. *In iteration t of Algorithm 5.1, the chosen edge e^* in each sub-iteration i always has the coupling parameter with the unarranged edges ($E_{uu}^{(t)}$) being smaller than 2, i.e., $c(e^*, E_{uu}^{(t)}) \leq 2$.*

Recall that we employ the *Local Ratio Method* in Algorithm 5.2 to allocate the resources to some fractional edges in $E_u^{(t)}$. Therefore, all the lines mentioned in the following correspond to Algorithm 5.2. We notice that each time we update E_{temp} (line 5), we also modify weighted vector of $E_u^{(t)}$ (line 6). Therefore, we have $|E_{\text{temp}}| + 1$ iterations updating the weight vector of $E_u^{(t)}$. We define $\mathbf{w}^{(j)}(E_u^{(t)})$ and $\bar{\mathbf{w}}^{(j)}(E_u^{(t)})$ respectively as the weight vectors of $E_u^{(t)}$ before and after weight update in iteration j ($1 \leq j \leq |E_{\text{temp}}| + 1$). We now define $\hat{\mathbf{w}}^{(j)}(E_u^{(t)}) = \mathbf{w}^{(j)}(E_u^{(t)}) - \bar{\mathbf{w}}^{(j)}(E_u^{(t)})$. In addition, we denote $w_e^{(j)}$, $\bar{w}_e^{(j)}$, and $\hat{w}_e^{(j)}$ are respectively the elements corresponding to edge e in $\mathbf{w}_e^{(j)}$, $\bar{\mathbf{w}}_e^{(j)}$, and $\hat{\mathbf{w}}_e^{(j)}$. Note that in lines 3 and 10 in Algorithm 5.2 we consider the same set of edges E_{temp} ; therefore, each index j ($1 \leq j \leq |E_{\text{temp}}|$) in lines 3 and 10 corresponds to a specific edge $e_j \in E_{\text{temp}}$.

We prove the proposition by induction on number of edges in $E_s^{(j)}$. At first, the statement in the proposition holds in the base case at iteration $j = |E_{\text{temp}}| + 1$ since at that iteration $E_s^{(|E_{\text{temp}}|+1)} = \emptyset$ and $\mathbf{w}^{(|E_{\text{temp}}|+1)}(E_u^{(t)}) \leq \mathbf{0}$, where $\mathbf{0}$ is a zero vector. We assume the induction hypothesis that at any iteration j ($1 < j \leq |E_{\text{temp}}| + 1$) we have

$$\sum_{e \in E_s^{(j)}} w_e^{(j)} \geq \frac{1}{2} \sum_{e \in E_u^{(t)}} x_e^* w_e^{(j)}. \quad (5.40)$$

We need to prove the following

$$\sum_{e \in E_s^{(j-1)}} w_e^{(j-1)} \geq \frac{1}{2} \sum_{e \in E_u^{(t)}} x_e^* w_e^{(j-1)}. \quad (5.41)$$

We assume that at iteration $j-1$ edge e^* is chosen (line 4 or 10). From the edge selection procedure in lines 9-11, we have

$$E_s^{(j-1)} = \begin{cases} E_s^{(j)} & \text{if } e^* \notin E_s^{(j-1)} \\ E_s^{(j)} \cup \{e^*\} & \text{if } e^* \in E_s^{(j-1)}. \end{cases} \quad (5.42)$$

Due to the employed update procedure for the weight vector, we have $\mathbf{w}^{(j)}(E_u^{(t)}) = \bar{\mathbf{w}}^{(j-1)}(E_u^{(t)})$ and $\bar{w}_{e^*}^{(j-1)} = 0$. Combining these arguments with (5.42) we have

$$\begin{aligned} \sum_{e \in E_s^{(j-1)}} \bar{w}_e^{(j-1)} &= \sum_{e \in E_s^{(j)}} w_e^{(j)} \\ \sum_{e \in E_u^{(t)}} x_e^* w_e^{(j)} &= \sum_{e \in E_u^{(t)}} x_e^* \bar{w}_e^{(j-1)}. \end{aligned} \quad (5.43)$$

Using the results in (5.40) and (5.43), we arrive at

$$\sum_{e \in E_s^{(j-1)}} \bar{w}_e^{(j-1)} \geq \frac{1}{2} \sum_{e \in E_u^{(t)}} x_e^* \bar{w}_e^{(j-1)}. \quad (5.44)$$

Assume that E_c is the set of edges coupled with the considered edge e^* in line 6. In addition, each element of $\hat{\mathbf{w}}^{(j-1)}(E_u^{(t)})$ satisfies

$$\hat{w}_e^{(j-1)} = \begin{cases} w_{e^*}^{(j-1)} & \forall e \in E_c \\ 0 & \text{otherwise.} \end{cases} \quad (5.45)$$

According to Lemma 1, we have $c(e^*, E_c) = \sum_{e \in E_c} x_e^* \leq 2$. Using this result and (5.45), we obtain

$$\sum_{e \in E_u^{(t)}} x_e^* \hat{w}_e^{(j-1)} \leq 2w_{e^*}^{(j-1)}. \quad (5.46)$$

In addition, according to the proposed edge selection procedure (line 10), at least one edge coupled with e^* must belong to $E_s^{(j-1)}$. Therefore,

$$\sum_{e \in E_s^{(j-1)}} \hat{w}_e^{(j-1)} \geq w_{e^*}^{(j-1)}. \quad (5.47)$$

From (5.46) and (5.47), the following inequality holds

$$\sum_{e \in E_s^{(j-1)}} \hat{w}_e^{(j-1)} \geq \frac{1}{2} \sum_{e \in E_u^{(t)}} x_e^* \hat{w}_e^{(j-1)}. \quad (5.48)$$

Combining the fact that $\mathbf{w}^{(j-1)}(E_u^{(t)}) = \bar{\mathbf{w}}^{(j-1)}(E_u^{(t)}) + \hat{\mathbf{w}}^{(j-1)}(E_u^{(t)})$ and the results from (5.44) and (5.48), we have

$$\sum_{e \in E_s^{(j-1)}} w_e^{(j-1)} \geq \frac{1}{2} \sum_{e \in E_u^{(t)}} x_e^* w_e^{(j-1)}, \quad (5.49)$$

which complies with (5.41). On the other hand, as $j = 1$, we have $\mathbf{w}(E_u^{(t)}) = \mathbf{w}^{(1)}(E_u^{(t)})$; therefore, the inequality (5.39) holds.

5.9.4 Proof of proposition 5.5

To address the considered subband assignment problem, we have reformulated an equivalent problem $\mathbf{IP}(V^0, E^0)$ where we have created $(N - K_c)$ cellular virtual links and the degree of each cellular virtual links is smaller than 1. Under this construction, there are at most $(N - K_c)$ D2D links, each of which can use a subband exclusively. Therefore, we can always guarantee that each of K_c cellular links is assigned one subband, which enables them to maintain the minimum rate constraints. In addition, after the first iteration of Algorithm 5.1, we admit $E_a^{(1)} \cup E_g^{(1)}$. In addition, we have (i): $z(E_a^{(1)} \cup E_g^{(1)}) = z(E_a^{(1)}) + z(E_g^{(1)})$. According to Proposition 5.3, we arrive at (ii): $l^*(V^{(1)}, E^{(1)}) = l^*(V_u^{(1)}, E(V_u^{(1)})) + z(E_a^{(1)})$. Also Proposition 5.4 suggests that (iii): $z(E_g^{(1)}) \geq \frac{1}{2}l^*(V_u^{(1)}, E(V_u^{(1)}))$. Combining (i), (ii), and (iii), we have

$$z(E_a^{(1)} \cup E_g^{(1)}) \geq \frac{1}{2}l^*(V^{(1)}, E^{(1)}) = \frac{1}{2}l^*(V^0, E^0). \quad (5.50)$$

Note that $E_a^{(1)} \cup E_g^{(1)} \subset E_{al}^f$ where E_{al}^f is the set E_{al} obtained at the end of Algorithm 5.1. Therefore, we have $z(E_{al}^f) \geq \frac{1}{2}l^*(V^0, E^0)$, which finishes the proof of the proposition.

Chapter 6

Energy-Efficient Resource Allocation for D2D Communications in Cellular Networks

The content of this chapter was published in IEEE Transactions on Vehicular Technology in the following paper:

T. D. Hoang, L. B. Le, and T. Le-Ngoc, “Energy-efficient resource allocation for D2D communications in cellular networks,” *IEEE Transaction on Vehicular Technology*, vol. 65, no. 9, pp. 6972–6986, September 2016.

6.1 Abstract

In this chapter, we study the energy-efficient resource allocation problem for device-to-device (D2D) communication underlying cellular networks which aims to maximize the minimum weighted energy-efficiency (EE) of D2D links while guaranteeing the minimum data rates for cellular links. We first characterize the optimal power allocation of the cellular links to transform the original resource allocation problem into the joint subchannel and power allocation problem for D2D links. We then propose three resource allocation algorithms with different complexity, namely Dual-Based, Branch-

and-Bound (BnB), and Relaxation-Based Rounding (RBR) algorithms. While the Dual-Based algorithm solves the problem by using dual decomposition method, the BnB and RBR algorithms tackle the problem by employing the relaxation approach. We establish the strong performance guarantees for the proposed algorithms through theoretical analysis. Extensive numerical studies demonstrate that the proposed algorithms achieve superior performance and significantly outperform a conventional algorithm.

6.2 Introduction

Deployment of device-to-device communications in the wireless cellular network has been expected to significantly increase the network throughput and reduce the traffic load in the core network [76, 97, 100–103]. Efficient resource allocation algorithms play a critical role in acquiring these benefits while limiting negative impacts on the performance of existing communications between users and base-stations (BSs). In general, we can perform orthogonal or non-orthogonal spectrum sharing for D2D and cellular communication links. Orthogonal spectrum sharing assumes cellular and D2D links using distinct parts of the spectrum, and consequently the system must reserve dedicated spectral resources for D2D links. On the other hand, the non-orthogonal spectrum sharing allows D2D links to reuse the resource of cellular links in order to improve the spectrum utilization and efficiency at the costs of co-channel interference between cellular and D2D links. Since, in a cellular system, the BS has more powerful processing capacity than mobile terminals to deal with interference experienced at the receiver’s side, it is more beneficial if the D2D links reuse the uplink resources of cellular links.

Green communication has attracted a lot of attention in recent years where maximization of EE has become an important design objective in engineering modern wireless systems [104–106]. In fact, development of energy-efficient resource allocation algorithms has been considered for 3GPP LTE systems [107, 108]. In general, downlink EE would be less critical than the uplink EE since BSs can have access to various energy sources while user devices are supported by energy-limited batteries [109, 110].

There are some existing works considering spectrum-efficient resource allocation for the D2D underlying cellular network [53, 80, 111–113] with various objectives and system constraints. In

[111], the authors develop a simple power control algorithm based on the signal to interference plus noise ratio (SINR) of the cellular link to guarantee its required performance and to maximize the sum-rate of D2D links. The authors of [80] propose a mode selection algorithm to maximize the sum-rate where they develop power control algorithms to attain the optimal solution for each mode. Power allocation for cellular and D2D links to maximize the rate of a single D2D link while guaranteeing the required rate of each cellular link is studied in [112]. In [53], the sum-rate optimization for D2D and cellular links is considered where the system with multiple D2D and multiple cellular links is studied. Finally, the work [113] develops a fair resource allocation for D2D links while assuring the required quality of service (QoS) of cellular links. In fact, to maximize the throughput or spectrum-efficiency, mobile devices would utilize large transmission powers, which may result in serious degradation of their EE. This motivates us to consider energy-efficient resource allocation for D2D communications in this current work.

There have been some initial efforts in developing energy-efficient resource allocation solutions for D2D underlying cellular networks [1, 22, 77, 114, 115]. In [22], a resource allocation solution based on non-cooperative game theory is proposed where each D2D link selfishly performs power and subchannel allocation to maximize its own EE considering the fixed resource allocation of other links. This design approach, however, may not lead to efficient utilization of the spectral resources. The authors in [114] solve the energy-efficient resource allocation problem by using the combinatorial auction game where the cellular BS acts as an auctioneer which sequentially decides the price of each resource and sells it to the set of D2D links achieving the highest utility. In this design, each D2D link is allowed to reuse the resource of only one cellular link, which may limit the achievable rates of D2D links. Moreover, a coalition game is employed to tackle the energy-efficient resource allocation problem in [1] where the authors address the joint mode selection and resource allocation for D2D and cellular links. Nevertheless, this work assumes that each D2D link only achieves its minimum required rate, which might not fully exploit the advantage of short-range D2D communications.

Other energy-efficient designs that aim to minimize the total power consumption are pursued in [116], [117]. However, resource allocation solutions in these papers may not fully exploit the advantages of D2D communications to achieve the optimum EE. In this chapter, we study the joint subchannel and power allocation that maximizes the minimum weighted EE of D2D links and guarantees the minimum data rates of cellular links. Specifically, we make the following contributions.

- We formulate a general energy-efficient resource allocation problem considering multiple cellular and D2D links where each D2D link can reuse the spectral resources of multiple cellular links. This model is, therefore, more general than most models studied in the literature [53, 80, 111, 112]. We first characterize the optimal power allocation solution for a cellular link as a function of the optimal power of the co-channel D2D link. Based on this important result, we transform the original resource allocation problem into the resource allocation problem for only D2D links.
- We propose the dual-based algorithm that solves the resource allocation problem in the dual domain. Particularly, we adopt the max-min fractional programming technique to iteratively transform the resource allocation problem into a Mixed Integer Nonlinear Programming (MINLP) problem. Then, we solve the underlying MINLP problem by using the dual decomposition approach. Theoretical analysis demonstrates that the algorithm converges to a feasible solution of the original problem. Moreover, the achieved solution is optimal if, at convergence, the duality gap of the underlying MINLP problem is zero. In addition, a distributed implementation with limited message exchange for the proposed algorithm is described, which can potentially reduce the computational burden of the BS and the system signaling overhead.
- We study the relaxation-based solution approach, which tackles the resource allocation problem by relaxing the subchannel allocation variables. In particular, we apply the branch-and-bound (BnB) approach to branch the subchannel allocation vector space to smaller sub-spaces in which some subchannel allocation variables are determined and others are undetermined. An upper-bound is calculated by solving a max-min fractional program of the relaxed problem where all undetermined subchannel allocation variables are relaxed. In particular, we sequentially transform the relaxed problem into a convex problem and solve it by using the interior-point method until convergence. Moreover, we obtain a lower-bound of the objective value by rounding the fractional subchannel allocation solution acquired in the upper-bound calculation. Motivated by the procedure to calculate the upper-bound in the BnB algorithm, we also propose a low-complexity Relaxation-based Rounding (RBR) algorithm. In this RBR algorithm, we first solve the relaxed problem of the original resource allocation problem for D2D links. Then, based on the obtained solution for the relaxed problem, we develop an efficient rounding procedure, which aims at minimizing the performance loss and maximize the design objective, to attain a feasible solution for the considered resource allocation problem.

- The computational complexity of the proposed algorithms is analyzed. Moreover, extensive numerical results are presented to evaluate the performance of the developed algorithms. Specifically, it is shown that the objective values achieved by the dual-based and RBR algorithms are very close to that of the optimal BnB algorithm. In addition, the proposed algorithms significantly outperform the conventional algorithm and the spectrum-efficient resource allocation design.

The remainder of the chapter is organized as follows. Section 6.3 presents the system model and problem formulation. The problem transformation is described in Section 6.4. Sections 6.5 and 6.6 develop the dual-based and relaxation-based algorithms, respectively. The computational complexity is analyzed in Section 6.7. Section 6.8 presents illustrative results, followed by conclusions in Section 6.9.

6.3 System Model and Problem Formulation

6.3.1 System Model

We consider uplink resource allocation scenario where cellular links share the same spectrum with multiple D2D links in a single macro-cell system. We assume that K uplink cellular links in a set $\mathcal{K} = \{1, \dots, K\}$ occupying K orthogonal subchannels in the set $\mathcal{N} = \{1, \dots, K\}$ in the considered cell. Moreover, we assume that the set $\mathcal{L} = \{1, \dots, L\}$ of D2D links transmits data using the same set of subchannels.¹ In these notations, $K = |\mathcal{K}|$, $L = |\mathcal{L}|$, and $N = |\mathcal{N}|$ denote the numbers of cellular links, D2D links, and subchannels, respectively, where $|A|$ denotes the cardinality of set A .

Let h_{kl}^n denote the channel gain from the transmitter of link l to the receiver of link k on subchannel n . We assume that the subchannel allocation for cellular links has been pre-determined and we are interested in allocating these subchannels to D2D links efficiently. Without loss of generality, we assume that cellular link k has been allocated subchannel k . We introduce vector $\boldsymbol{\rho}_l = [\rho_l^1, \dots, \rho_l^K]$ to describe subchannel allocation decisions for D2D link l where $\rho_l^k = 1$ if subchannel k is allocated for D2D link l and $\rho_l^k = 0$, otherwise. Let $\boldsymbol{\rho} = [\boldsymbol{\rho}_1, \dots, \boldsymbol{\rho}_L]$ denote the subchannel allocation variables for all D2D links.

¹The considered orthogonal subchannels can be sub-carriers or sub-channels in the OFDMA system or simply channels in the FDMA system.

We present the allocated power vectors as $\mathbf{p} = [\mathbf{p}_C, \mathbf{p}_D]$ for all the links, where $\mathbf{p}_C = [p_{C1}^1, \dots, p_{CK}^K]$ for K cellular links, $\mathbf{p}_D = [\mathbf{p}_{D1}, \dots, \mathbf{p}_{DL}]$, $\mathbf{p}_{Dl} = [p_{Dl}^1, \dots, p_{Dl}^K]$, for D2D links, and p_{Ck}^k and p_{Dl}^k denote the allocated transmit powers on subchannel k of cellular link $k \in \mathcal{K}$ and D2D link $l \in \mathcal{L}$, respectively. Then, the SINR achieved by cellular link k on its allocated subchannel k can be expressed as

$$\Gamma_{Ck}^k(\mathbf{p}, \boldsymbol{\rho}) = \frac{p_{Ck}^k h_{kk}^k}{\sigma_k^k + \sum_{l \in \mathcal{L}} \rho_l^k p_{Dl}^k h_{kl}^k}, \quad (6.1)$$

where $\sum_{l \in \mathcal{L}} \rho_l^k p_{Dl}^k h_{kl}^k$ represents the interference due to the D2D link using subchannel k and σ_k^k denotes the noise power on subchannel k . Similarly, the SINR of D2D link l on subchannel k can be written as

$$\Gamma_{Dl}^k(\mathbf{p}, \boldsymbol{\rho}) = \frac{\rho_l^k p_{Dl}^k h_{ll}^k}{\sigma_l^k + p_{Ck}^k h_{lk}^k}. \quad (6.2)$$

The data rates in b/s/Hz (i.e., normalized by the subchannel bandwidth) of cellular link $k \in \mathcal{K}$ on its subchannel k , D2D link $l \in \mathcal{L}$ on subchannel k , and D2D link l on all the subchannels can be calculated as

$$R_{Ck}^k(\mathbf{p}, \boldsymbol{\rho}) = \log_2 \left(1 + \Gamma_{Ck}^k(\mathbf{p}, \boldsymbol{\rho}) \right), \quad (6.3)$$

$$R_{Dl}^k(\mathbf{p}, \boldsymbol{\rho}) = \log_2 \left(1 + \Gamma_{Dl}^k(\mathbf{p}, \boldsymbol{\rho}) \right), \quad (6.4)$$

and

$$R_{Dl}(\mathbf{p}, \boldsymbol{\rho}) = \sum_{k \in \mathcal{K}} \rho_l^k R_{Dl}^k(\mathbf{p}, \boldsymbol{\rho}), \quad (6.5)$$

respectively. We assume that the total *consumed* power of D2D link l can be expressed as [18, 19]

$$P_{Dl}^{\text{total}} = 2P_0^l + \alpha_l \sum_{k \in \mathcal{K}} \rho_l^k p_{Dl}^k, \quad (6.6)$$

where $2P_0^l$ represents the fixed circuit power of both transmitter and receiver of D2D link l , and $\alpha_l > 1$ is a factor accounting for the transmit amplifier efficiency and feeder losses.

6.3.2 Problem Formulation

In this work, we consider the resource allocation design with the following constraints. First, it is required to maintain the minimum rate of each cellular link k (on its allocated subchannel k), i.e.,

$$R_{Ck}^k(\mathbf{p}, \boldsymbol{\rho}) \geq R_{Ck}^{\min}, \forall k \in \mathcal{K}. \quad (6.7)$$

Second, the power constraints of individual *links* are given as

$$p_{Ck}^k \leq P_{Ck}^{\max}, \forall k \in \mathcal{K}, \quad (6.8)$$

$$\sum_{k \in \mathcal{K}} \rho_l^k p_{Dl}^k \leq P_{Dl}^{\max}, \forall l \in \mathcal{L}, \quad (6.9)$$

where P_{Ck}^{\max} and P_{Dl}^{\max} are the maximum transmit powers of cellular link k and D2D link l , respectively.

Third, the subchannel allocation variables are binary, i.e.,

$$\rho_l^k \in \{0, 1\}, \forall k \in \mathcal{K}, l \in \mathcal{L}. \quad (6.10)$$

Finally, similar to [1, 23, 53, 80, 112, 114, 116], we require that each subchannel can be reused by at most one D2D link to limit the interference from D2D links to cellular links, and hence to guarantee the performance of the cellular links, i.e.,

$$\sum_{l \in \mathcal{L}} \rho_l^k \leq 1, \forall k \in \mathcal{K}. \quad (6.11)$$

The objective of our resource allocation design is to maximize the minimum weighted EE of the D2D links. Therefore, this design can be formulated as the following energy-efficient resource allocation problem to attain the max-min fairness in weighed EE for D2D links:

$$\begin{aligned} \max_{\mathbf{p}, \boldsymbol{\rho}} \quad & \min_{l \in \mathcal{L}} \frac{w_l R_{Dl}(\mathbf{p}, \boldsymbol{\rho})}{P_{Dl}^{\text{total}}} \\ \text{s.t.} \quad & (6.7), (6.8), (6.9), (6.11), (6.10), \end{aligned} \quad (6.12)$$

where $\frac{w_l R_{Dl}(\mathbf{p}, \boldsymbol{\rho})}{P_{Dl}^{\text{total}}}$ represents the weighted EE of the D2D links. The weight parameters w_l can be employed to control the relative priorities among different D2D links and $\sum_{l \in \mathcal{L}} w_l = L$.

The resource allocation design in this chapter allows each D2D link to share spectral resources with multiple cellular links but the spectral resource of each cellular link can be reused by at most one D2D link. This model allows us to (i) achieve good balance between excellent performance for D2D links and affordable interference management complexity, (ii) protect the QoS of cellular links efficiently, (iii) avoid large signaling overhead due to the Channel State Information (CSI) estimation and feedback of interfering channels among D2D links.

6.4 Problem Transformation

To solve problem (6.12), we first describe the optimal power allocation of D2D link $l \in \mathcal{L}$ on subchannel $k \in \mathcal{N}$ in the following proposition.

Proposition 6.1. *If D2D link $l \in \mathcal{L}$ is allowed to reuse subchannel $k \in \mathcal{N}$ of cellular link k , then its power on subchannel k , $p_{Dl}^k = \frac{1}{h_{kl}^k} \left(\frac{p_{Ck}^k h_{kk}^k}{2^{R_{Ck}^{\text{min}}} - 1} - \sigma_k^k \right) \in [0, P_{Dlk}^{\text{max}}]$, where p_{Ck}^k is the power of cellular link k , and $P_{Dlk}^{\text{max}} = \min \left\{ P_{Dl}^{\text{max}}, \frac{1}{h_{kl}^k} \left(\frac{P_{Ck}^{\text{max}} h_{kk}^k}{2^{R_{Ck}^{\text{min}}} - 1} - \sigma_k^k \right) \right\}$.*

Proof. The proof is given in Appendix 6.10.1. □

From Proposition 6.1, the data rate of D2D link l on subchannel k given in (6.4) can be re-written as

$$\begin{aligned} & \hat{R}_{Dl}^k(\mathbf{p}_D, \boldsymbol{\rho}) \\ &= \rho_l^k \log_2 \left(1 + \frac{p_{Dl}^k h_{ll}^k}{\sigma_l^k + \frac{(2^{R_{Ck}^{\text{min}}} - 1) h_{lk}^k}{h_{kk}^k} (\sigma_k^k + p_{Dl}^k h_{kl}^k)} \right). \end{aligned} \quad (6.13)$$

For convenience, let us define

$$a_{kl} \triangleq \frac{\sigma_l^k}{h_{ll}^k} + \frac{(2^{R_{Ck}^{\text{min}}} - 1) h_{lk}^k \sigma_k^k}{h_{kk}^k h_{ll}^k} \quad (6.14)$$

$$b_{kl} \triangleq \frac{(2^{R_{Ck}^{\text{min}}} - 1) h_{lk}^k h_{kl}^k}{h_{kk}^k h_{ll}^k}. \quad (6.15)$$

Then, the data rate of D2D link l on subchannel $k \in \mathcal{N}$, $\hat{R}_{Dl}^k(\mathbf{p}_D, \boldsymbol{\rho})$ in (6.13), and the total rate over all subchannels, $\hat{R}_{Dl}(\mathbf{p}_D, \boldsymbol{\rho})$ in (6.5), can be rewritten, respectively, as

$$\hat{R}_{Dl}^k(\mathbf{p}_D, \boldsymbol{\rho}) = \rho_l^k \log_2 \left(1 + \frac{p_{Dl}^k}{a_{kl} + b_{kl} p_{Dl}^k} \right), \quad (6.16)$$

and

$$\hat{R}_{Dl}(\mathbf{p}_D, \boldsymbol{\rho}) = \sum_{k \in \mathcal{N}} \hat{R}_{Dl}^k(\mathbf{p}_D, \boldsymbol{\rho}), \quad (6.17)$$

where the allocated transmit power must satisfy

$$p_{Dl}^k \leq P_{Dlk}^{\max}, \forall k \in \mathcal{N}, \forall l \in \mathcal{L}. \quad (6.18)$$

Therefore, problem (6.12) is equivalent to the following

$$\begin{aligned} & \max_{(\mathbf{p}_D, \boldsymbol{\rho})} \min_{l \in \mathcal{L}} \frac{w_l \hat{R}_{Dl}(\mathbf{p}_D, \boldsymbol{\rho})}{P_{Dl}^{\text{total}}(\mathbf{p}_D, \boldsymbol{\rho})} \\ & \text{s.t.} \quad (6.9), (6.10), (6.11), (6.18). \end{aligned} \quad (6.19)$$

In order to solve problem (6.19), we consider the following optimization problem

$$\begin{aligned} & \max_{\mathbf{p}_D, \boldsymbol{\rho}} \eta(\zeta, \mathbf{p}_D, \boldsymbol{\rho}) \triangleq \min_{l \in \mathcal{L}} \left[w_l \hat{R}_{Dl}(\mathbf{p}_D, \boldsymbol{\rho}) - \zeta P_{Dl}^{\text{total}}(\mathbf{p}_D, \boldsymbol{\rho}) \right] \\ & \text{s.t.} \quad (6.9), (6.10), (6.11), (6.18). \end{aligned} \quad (6.20)$$

Suppose that $\eta^*(\zeta) = \eta(\zeta, \mathbf{p}_D^*, \boldsymbol{\rho}^*)$ where $(\mathbf{p}_D^*, \boldsymbol{\rho}^*)$ is the optimal solution of problem (6.20), and \mathcal{D} denotes the set of feasible solutions of problem (6.19). Then, we can characterize the optimal solution of problem (6.20) in the following theorem, which is adopted from [20].

Theorem 6.1. $\eta^*(\zeta)$ is a decreasing function of ζ . In addition, if we have

$$\begin{aligned} & \max_{(\mathbf{p}_D, \boldsymbol{\rho}) \in \mathcal{D}} \min_{l \in \mathcal{L}} [w_l \hat{R}_{Dl}(\mathbf{p}_D, \boldsymbol{\rho}) - \zeta^* P_{Dl}^{\text{total}}(\mathbf{p}_D, \boldsymbol{\rho})] \\ & = \min_{l \in \mathcal{L}} \left[w_l \hat{R}_{Dl}(\mathbf{p}_D^*, \boldsymbol{\rho}^*) - \zeta^* P_{Dl}^{\text{total}}(\mathbf{p}_D^*, \boldsymbol{\rho}^*) \right] = 0 \end{aligned} \quad (6.21)$$

then $\zeta^* = \min_{l \in \mathcal{L}} \frac{w_l \hat{R}_{Dl}(\mathbf{p}_D^*, \boldsymbol{\rho}^*)}{P_{Dl}^{\text{total}}(\mathbf{p}_D^*, \boldsymbol{\rho}^*)}$ is the optimal solution of (6.19).

It is worth noting that the main theorem in [20] states the necessary and sufficient condition for ζ^* to be the optimal solution of the fractional programming problem. In Theorem 6.1, we only present the sufficient condition for ζ^* to be the optimal solution of problem (6.19). However, the theorem of [20] requires the set of feasible solutions of the fractional programming problem to be continuous, which is not required in our theorem. Importantly, Theorem 6.1 allows us to transform a general max-min fractional problem (6.19) to a non-fractional optimization problem with the parameter ζ . In addition, the optimal solution of problem (6.19), ζ^* , can be found if $\eta^*(\zeta^*) = 0$. Since $\eta^*(\zeta)$ is a decreasing function of ζ , it can be seen that ζ^* can be indeed determined by the gradient or bisection method.

A general algorithm solving problem (6.19) based on the solution of problem (6.20) is described in Algorithm 6.1, which iteratively solves problem (6.20) for given ζ and updates ζ until convergence. Therefore, the remaining challenge is how to solve problem (6.20) for a given ζ . In general, problem (6.20) is NP-hard, which implies that solving this problem optimally requires exponential complexity.

Algorithm 6.1. General Algorithm

- 1: Initialization: Set $\epsilon = 10^{-6}$, $\zeta = 0$, $\zeta_t = \epsilon$
 - 2: **while** $(\zeta - \zeta_t) \geq \epsilon$ **do**
 - 3: Solve problem (6.20) for given ζ to obtain $(\mathbf{p}_D^*, \boldsymbol{\rho}^*) = \underset{(\mathbf{p}_D, \boldsymbol{\rho})}{\operatorname{argmax}} \eta(\zeta, \mathbf{p}_D, \boldsymbol{\rho})$
 - 4: $\zeta_t = \zeta$, $\zeta = \min_{l \in \mathcal{L}} \frac{w_l \hat{R}_{Dl}(\mathbf{p}_D^*, \boldsymbol{\rho}^*)}{P_{Dl}^{\text{total}}(\mathbf{p}_D^*, \boldsymbol{\rho}^*)}$
 - 5: **end while**
 - 6: Output $(\mathbf{p}_D^*, \boldsymbol{\rho}^*)$, and ζ
-

6.5 Dual-Based Algorithm

In this section, we propose a dual-based algorithm to solve problem (6.19). Then, we will present the distributed implementation for this algorithm.

6.5.1 Algorithm Development

The dual-based resource allocation algorithm is summarized in Algorithm 6.2. The algorithm comprises two iterative loops. In the outer loop, we adopt the max-min fractional programming

Algorithm 6.2. Dual-Based Algorithm

-
- 1: Initialization: $\zeta^{\max}, \zeta^{\min}$
 - 2: **repeat**
 - 3: Initialization: Choose $\zeta = \frac{1}{2}(\zeta^{\min} + \zeta^{\max})$, $\boldsymbol{\lambda}^{(0)}, \mu_l^{(0)} = \frac{1}{L}$, step size $\theta^{(0)}$, and $\kappa^{(0)}$
 - 4: **repeat**
 - 5: Step 1: For all $k \in \mathcal{K}, l \in \mathcal{L}$, calculate p_{Dl}^{k*} according to (6.29)
 - 6: Step 2: For all $k \in \mathcal{K}$, perform subchannel allocation following (6.37)
 - 7: Step 3: Update dual variables $\boldsymbol{\lambda}$, and $\boldsymbol{\mu}$ as in (6.38) and (6.43)
 - 8: **until** Convergence
 - 9: Output $z^* = \min_{l \in \mathcal{L}} [w_l \hat{R}_{Dl}(\mathbf{p}_D^*, \boldsymbol{\rho}^*) - \zeta P_{Dl}^{\text{total}}(\mathbf{p}_D^*, \boldsymbol{\rho}^*)]$
 - 10: If $z^* > 0$, $\zeta^{\min} = \zeta$; otherwise $\zeta^{\max} = \zeta$
 - 11: **until** Convergence of ζ
 - 12: Output $\mathbf{p}_{Dl}^*, \boldsymbol{\rho}^*$, and $\zeta^* = \min_{l \in \mathcal{L}} \frac{w_l \hat{R}_{Dl}(\mathbf{p}_{Dl}^*, \boldsymbol{\rho}^*)}{P_{Dl}^{\text{total}}(\mathbf{p}_{Dl}^*, \boldsymbol{\rho}^*)}$
-

technique described in Algorithm 6.1 to attain the optimal value of ζ for problem (6.19). In the inner loop, we solve problem (6.20) for a given ζ by employing the dual decomposition method.

Algorithm 6.2 performs two main tasks, which are executed sequentially. Specifically, we solve problem (6.20) for a given ζ (lines 4-8) in the first task while we update the value of ζ based on the results of the first task by using the bisection method in the second task (line 10). In the following, we show how to solve problem (6.20) for a given value of ζ . First, it can be observed that problem (6.20) is equivalent to the following problem

$$\begin{aligned}
 & \max_{z, \mathbf{p}_D, \boldsymbol{\rho}} \quad z \\
 & \text{s.t.} \quad w_l \hat{R}_{Dl}(\mathbf{p}_{Dl}, \boldsymbol{\rho}) - \zeta P_{Dl}^{\text{total}}(\mathbf{p}_D, \boldsymbol{\rho}) \geq z, \forall l \in \mathcal{L} \\
 & \quad \quad (6.9), (6.10), (6.11), (6.18).
 \end{aligned} \tag{6.22}$$

To tackle problem (6.22), we consider its Lagrangian as follows:

$$\begin{aligned}
& L_D(\mathbf{p}_D, \boldsymbol{\rho}, z, \zeta, \boldsymbol{\lambda}, \boldsymbol{\mu}) \\
&= z + \sum_{l \in \mathcal{L}} \mu_l \left[w_l \hat{R}_{Dl}(\mathbf{p}_D, \boldsymbol{\rho}) - \zeta P_{Dl}^{\text{total}}(\mathbf{p}_D, \boldsymbol{\rho}) - z \right] \\
&\quad + \sum_{l \in \mathcal{L}} \lambda_l (P_{Dl}^{\text{max}} - \sum_{k \in \mathcal{N}} \rho_l^k p_{Dl}^k) \\
&= z(1 - \sum_{l \in \mathcal{L}} \mu_l) + \sum_{l \in \mathcal{L}} \mu_l \left[w_l \hat{R}_{Dl}(\mathbf{p}_D, \boldsymbol{\rho}) - \zeta P_{Dl}^{\text{total}}(\mathbf{p}_D, \boldsymbol{\rho}) \right] \\
&\quad + \sum_{l \in \mathcal{L}} \lambda_l (P_{Dl}^{\text{max}} - \sum_{k \in \mathcal{N}} \rho_l^k p_{Dl}^k), \tag{6.23}
\end{aligned}$$

where $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_L]^T$ and $\boldsymbol{\mu} = [\mu_1, \dots, \mu_L]^T$ represent the Lagrange multipliers. Then, the dual function can be written as

$$\bar{L}_D(\zeta, \boldsymbol{\lambda}, \boldsymbol{\mu}) \triangleq \max_{z, \mathbf{p}_D \in \mathcal{X}, \boldsymbol{\rho} \in \mathcal{C}} L_D(\mathbf{p}_D, \boldsymbol{\rho}, z, \zeta, \boldsymbol{\lambda}, \boldsymbol{\mu}), \tag{6.24}$$

where $\mathcal{X} = \{\mathbf{p}_D | p_{Dl}^k \leq P_{Dlk}^{\text{max}}, \forall k \in \mathcal{N}, \forall l \in \mathcal{L}\}$, and $\mathcal{C} = \{\boldsymbol{\rho} | \sum_{l \in \mathcal{L}} \rho_l^k \leq 1, \forall k \in \mathcal{N}, \text{ and } \rho_l^k \in \{0, 1\}, \forall k \in \mathcal{K}, l \in \mathcal{L}\}$. Then, the dual problem can be stated as

$$\hat{L}_D(\zeta) \triangleq \min_{\boldsymbol{\lambda}, \boldsymbol{\mu} \geq 0} \bar{L}_D(\zeta, \boldsymbol{\lambda}, \boldsymbol{\mu}). \tag{6.25}$$

In order to solve the dual problem (6.25), we investigate problem (6.24) for the given $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$. In particular, we have

$$\begin{aligned}
\bar{L}_D(\zeta, \boldsymbol{\lambda}, \boldsymbol{\mu}) &= \max_{z, \mathbf{p}_D \in \mathcal{X}, \boldsymbol{\rho} \in \mathcal{C}} L_D(\mathbf{p}_D, \boldsymbol{\rho}, z, \zeta, \boldsymbol{\lambda}, \boldsymbol{\mu}) \\
&= \max_{z, \mathbf{p}_D \in \mathcal{X}, \boldsymbol{\rho} \in \mathcal{C}} \\
&\quad \sum_{k \in \mathcal{N}} \sum_{l \in \mathcal{L}} \rho_l^k \left[\mu_l w_l \hat{R}_{Dl}^k(\mathbf{p}_D, \boldsymbol{\rho}) - (\zeta \alpha_l \mu_l + \lambda_l) p_{Dl}^k \right] \\
&\quad + z(1 - \sum_{l \in \mathcal{L}} \mu_l) + \sum_{l \in \mathcal{L}} \left(\lambda_l P_{Dl}^{\text{max}} - 2\zeta \mu_l P_0^l \right). \tag{6.26}
\end{aligned}$$

Note that z is an uncontrolled variable in problem (6.26). Thus, to obtain the nontrivial optimal solution of the dual problem (6.25), $\sum_{l \in \mathcal{L}} \mu_l = 1$ must hold. Moreover, problem (6.26) can be decomposed into N individual resource allocation problems for N subchannels where the resource

allocation problem for subchannel $k \in \mathcal{N}$ can be stated as

$$\begin{aligned} & \bar{L}_D^k(\zeta, \boldsymbol{\lambda}, \boldsymbol{\mu}) \\ &= \max_{\mathbf{p}_D \in \mathcal{X}, \boldsymbol{\rho} \in \mathcal{C}} \sum_{l \in \mathcal{L}} \rho_l^k \left[\mu_l w_l \hat{R}_{Dl}^k(\mathbf{p}_D, \boldsymbol{\rho}) - (\zeta \alpha_l \mu_l + \lambda_l) p_{Dl}^k \right]. \end{aligned} \quad (6.27)$$

Then,

$$\begin{aligned} & \bar{L}_D(\zeta, \boldsymbol{\lambda}, \boldsymbol{\mu}) \\ &= \sum_{k \in \mathcal{N}} \bar{L}_D^k(\zeta, \boldsymbol{\lambda}, \boldsymbol{\mu}) + \sum_{l \in \mathcal{L}} \left(\lambda_l P_{Dl}^{\max} - 2\zeta \mu_l P_0^l \right). \end{aligned} \quad (6.28)$$

We now define

$$f_l^k(p_{Dl}^k) \triangleq \mu_l w_l \hat{R}_{Dl}^k(\mathbf{p}_D, \boldsymbol{\rho}) - (\zeta \alpha_l \mu_l + \lambda_l) p_{Dl}^k. \quad (6.29)$$

For problem (6.27), suppose that D2D link l is allocated subchannel $k \in \mathcal{N}$ then we have

$$p_{Dl}^{k*} = \operatorname{argmax}_{p_{Dl}^k \in \mathcal{X}_l} f_l^k(p_{Dl}^k). \quad (6.30)$$

Note that we must have $\mu_l > 0$. This is because if $\mu_l = 0$, we have $p_{Dl}^{k*} = 0, \forall k \in \mathcal{N}$, which cannot be the optimal solution of problem (6.26). In addition, problem (6.30) can be addressed by solving $\frac{\partial f_l^k(p_{Dl}^k)}{\partial p_{Dl}^k} = 0$, where $\frac{\partial f_l^k(p_{Dl}^k)}{\partial p_{Dl}^k}$ is the first order derivative of $f_l^k(p_{Dl}^k)$, which can be written as

$$\frac{\partial f_l^k}{\partial p_{Dl}^k} = \frac{\mu_l w_l a_{kl} \log_2 e}{(a_{kl} + b_{kl} p_{Dl}^k)[a_{kl} + (b_{kl} + 1)p_{Dl}^k]} - (\zeta \alpha_l \mu_l + \lambda_l). \quad (6.31)$$

Then, it can be verified that solving $\frac{\partial f_l^k(p_{Dl}^k)}{\partial p_{Dl}^k} = 0$ is equivalent to solving $A_{kl}(p_{Dl}^k)^2 + 2B_{kl}p_{Dl}^k + C_{kl} = 0$ where

$$A_{kl}^d \triangleq \left(\zeta \alpha_l + \frac{\lambda_l}{\mu_l} \right) b_{kl} (b_{kl} + 1) \quad (6.32)$$

$$B_{kl}^d \triangleq \left(\zeta \alpha_l + \frac{\lambda_l}{\mu_l} \right) (a_{kl} b_{kl} + 0.5 a_{kl}) \quad (6.33)$$

$$C_{kl}^d \triangleq \left(\zeta \alpha_l + \frac{\lambda_l}{\mu_l} \right) a_{kl}^2 - w_l a_{kl} \log_2 e \quad (6.34)$$

$$\Delta_{kl}^d \triangleq (B_{kl}^d)^2 - A_{kl}^d C_{kl}^d. \quad (6.35)$$

Consequently, the optimal solution of D2D link l that maximizes $f_l^k(p_{Dl}^k)$ is given by

$$p_{Dl}^{k*} = \left[\frac{-B_{kl}^d + \sqrt{\Delta_{kl}^d}}{A_{kl}^d} \right]_0^{P_{Dl}^{\max}}, \quad (6.36)$$

where $[x]_a^b = b$ if $x > b$, $[x]_a^b = a$ if $x < a$, otherwise $[x]_a^b = x$.

In summary, by solving problem (6.26) we can obtain the optimal power allocation for any D2D link on subchannel $k \in \mathcal{N}$. Recall that we have assumed that each subchannel can be allocated to at most one D2D link; therefore, for all subchannels $k \in \mathcal{N}$, we have

$$\rho_l^{k*} = \begin{cases} 1 & \text{if } l = \operatorname{argmax}_{l \in \mathcal{L}} f_l^k(p_{Dl}^{k*}) \\ 0 & \text{otherwise.} \end{cases} \quad (6.37)$$

So far we have presented the resource allocation solution for given $\boldsymbol{\lambda}$, $\boldsymbol{\mu}$. Therefore, the remaining task is to solve problem (6.25), which can be completed by the sub-gradient method as described in the following. In the initial iteration $s = 0$, we solve problem (6.24) with the initial value of $\boldsymbol{\lambda}^{(0)}$ and $\boldsymbol{\mu}^{(0)}$. Then, in iteration $s + 1$, we update the dual variables $\boldsymbol{\lambda}^{(s+1)}$ and $\boldsymbol{\mu}^{(s+1)}$ based on the solution in iteration s , then we solve problem (6.24) with the updated value of $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$. The procedure to update $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ by using the sub-gradient method can be expressed as follows

$$\lambda_l^{(s+1)} = \left[\lambda_l^{(s)} + \theta_l^{(s)} \left(\sum_{k \in \mathcal{N}} \rho_l^{k*} p_{Dl}^{k*(s)} - P_{Dl}^{\max} \right) \right]^+, \forall l \in \mathcal{L} \quad (6.38)$$

$$\mu_l^{(s+1)} = \left[\mu_l^{(s)} - \kappa_l^{(s)} (z_l^{(s)} - z_{\min}^{(s)}) \right]^+, \forall l \in \mathcal{L}, \quad (6.39)$$

where $[x]^+ = \max\{x, 0\}$, and

$$z_l^{(s)} = w_l R_{Dl}(\mathbf{p}_D^{(s)}, \boldsymbol{\rho}^{(s)}) - \zeta(2P_0^l + \alpha_l \sum_{k \in \mathcal{N}} \rho_l^{k*} p_{Dl}^{k*(s)}) \quad (6.40)$$

$$z_{\min}^{(s)} = \min_{l \in \mathcal{L}} \left[w_l R_{Dl}(\mathbf{p}_D^{(s)}, \boldsymbol{\rho}^{(s)}) - \zeta(2P_0^l + \alpha_l \sum_{k \in \mathcal{N}} \rho_l^{k*} p_{Dl}^{k*(s)}) \right] \quad (6.41)$$

and $\theta_l^{(s)}$, $\kappa_l^{(s)}$ are step sizes, which can be chosen appropriately to ensure the convergence of the underlying iterative updates. Note that $\mathbf{p}_D^{(s)}$, $\boldsymbol{\rho}^{(s)}$ are, respectively, the transmit power and sub-

channel allocation solutions for given $\boldsymbol{\lambda}^{(s)}$, $\boldsymbol{\mu}^{(s)}$. Recall that $\boldsymbol{\mu}$ must satisfy the constraint $\sum_{l \in \mathcal{L}} \mu_l = 1$. Therefore, we normalize $\boldsymbol{\mu}^{(s+1)}$ as follows:

$$\mu_l^{(s+1)} = \frac{\mu_l^{(s+1)'}}{\sum_{l \in \mathcal{L}} \mu_l^{(s+1)'}}, \forall l \in \mathcal{L}. \quad (6.42)$$

It is shown in [118] that the dual decomposition procedure converges to the optimal solution of problem (6.25) for appropriately chosen $\theta_l^{(s)}$ and $\kappa_l^{(s)}$. Therefore, the iterative loop in the first task of Algorithm 6.2 always converges to the dual solution of problem (6.20) for any value of ζ . On the other hand, the performance achieved by Algorithm 6.2, which solves problem (6.19), is stated in the following proposition

Proposition 6.2. *Algorithm 6.2 returns a feasible solution of problem (6.19) with ζ^* , \mathbf{p}_D^* , $\boldsymbol{\rho}^*$, $\boldsymbol{\lambda}^*$, and $\boldsymbol{\mu}^*$ at the end of its first phase. Moreover, if $\sum_{k \in \mathcal{N}} \rho_l^{k*} p_{Dl}^{k*} \leq P_{Dl}^{\max}$, $\lambda_l^*(P_D^{\max} - \sum_{k \in \mathcal{N}} \rho_l^{k*} p_{Dl}^{k*}) = 0$, and $R_{Dl}(\mathbf{p}_D^*, \boldsymbol{\rho}^*) - \zeta^* P_{Dl}^{\text{total}}(\mathbf{p}_D^*, \boldsymbol{\rho}^*) = 0, \forall l \in \mathcal{L}$, this feasible solution is the optimal solution of problem (6.19).*

Proof. The proof is provided in Appendix 6.10.3. □

6.5.2 Distributed Implementation with Limited Message Passing

The distributed implementation with limited message exchange to execute Algorithm 6.2 is now described. In this implementation, instead of performing all the necessary tasks, the BS assigns some to the D2D links to reduce the computational burden on the BS. Due to the QoS requirements of cellular links and the strong interference coupling among wireless links, certain coordination among the BS and mobile devices via message passing deems necessary to achieve efficient spectrum sharing for D2D and cellular links.²

We can modify the procedure to update the dual-variable μ , which is introduced to adjust $z_l^{(s)}$ in each iteration, with $\sum_{l \in \mathcal{L}} \mu_l = 1$ as

$$\mu_l^{(s+1)} = \mu_l^{(s)} + \kappa_l^{(s)}, \quad (6.43)$$

²Distributed resource allocation algorithms can also be developed by using advanced game-theory and learning techniques [119], [120]. We will explore these solution approaches in our future works.

where $\kappa_l^{(s)} > 0$ if $z_l^{(s)} = z_{\min}^{(s)}$ and $\kappa_l^{(s)} < 0$, otherwise, and $\kappa_l^{(s)}$ is chosen to satisfy $\sum_{l \in \mathcal{L}} \kappa_l^{(s)} = 0$. A typical update of $\boldsymbol{\mu}$ with a fixed step-size can be implemented as

$$\kappa_l^{(s)} = \begin{cases} \kappa, & \text{if } z_l^{(s)} = z_{\min}^{(s)} \\ \frac{-\kappa}{L-1}, & \text{otherwise,} \end{cases} \quad (6.44)$$

where κ is a small value to guarantee the convergence of the updates [118]. The distributed procedure for Algorithm 6.2 can be described as follows.

Initialization:

Each D2D link l initializes the following system parameters: ζ^{\max} , ζ^{\min} , $\zeta = \frac{1}{2}(\zeta^{\min} + \zeta^{\max})$, $\lambda_l^{(0)} = 0$, $\mu_l^{(0)} = \frac{1}{L}$.

Step 1 (D2D):

For given ζ , λ_l , μ_l , each D2D link l calculates $p_{Dl}^{k*} = \left[\frac{-B_{kl}^d + \sqrt{\Delta_{kl}^d}}{A_{kl}^d} \right]_0^{P_{Dl}^{\max}}$, $\forall k \in \mathcal{K}$ and broadcasts the value of $f_l^k(p_{Dl}^{k*})$, which is defined in (6.29).

Step 2 (BS):

The BS after collecting all values $f_l^k(p_{Dl}^{k*}), \forall k \in \mathcal{K}, \forall l \in \mathcal{L}$ broadcasts $f_k^{\max} = \max_{l \in \mathcal{L}} f_l^k(p_{Dl}^{k*})$ to all users.

Step 3 (D2D):

Each D2D link l performs subchannel allocation by using the following rule

$$\rho_l^{k*} = \begin{cases} 1 & \text{if } f_l^k(p_{Dl}^{k*}) \geq f_k^{\max} \\ 0 & \text{otherwise.} \end{cases} \quad (6.45)$$

Moreover, it calculates $z_l = w_l \hat{R}_{Dl}(\mathbf{p}_{Dl}, \boldsymbol{\rho}_l^*) - \zeta(2P_0^l + \alpha_l \sum_{k \in \mathcal{N}} \rho_l^{k*} p_{Dl}^{k*})$, then broadcasts z_l . By using the received information, each D2D link l updates λ_l and μ_l according to (6.38) and (6.43),

and return to Step 1 until convergence.

Step 4 (BS):

The BS collects information of z_l , calculates $z_{\min} = \min_{l \in \mathcal{L}} z_l$, then broadcasts z_{\min} . If $z_{\min} > \zeta$, it updates $\zeta \leftarrow z_{\min}$ and broadcasts ζ . This procedure continues (go back to Step 1) and only terminates if there is no further increase in ζ . The BS then calculates the power allocation for each cellular link as

$$p_{Ck}^{k*} = \frac{\sigma_k^k + I_k}{h_{kk}^k} (2^{R_{Ck}^{\min}} - 1), \quad (6.46)$$

where I_k is the estimated interference caused by the co-channel D2D link on channel k . Then it broadcasts p_{Ck}^{k*} to other cellular links.

The main tasks performed by different network entities can be explained as follows. The BS collects the information regarding $f_l^k(p_{Dl}^{k*}), \forall k \in \mathcal{K}, \forall l \in \mathcal{L}$ then it broadcasts the maximum value $f_k^{\max} = \max_{l \in \mathcal{L}} f_l^k(p_{Dl}^{k*})$ for each subchannel (step 2). The BS is also responsible for calculating and broadcasting the updated value of $z_{\min} = \min_{l \in \mathcal{L}} z_l$ and ζ (step 4). Moreover, as the algorithm converges, the BS calculates the transmit powers of all cellular links and broadcasts the results to the cellular links. Each D2D link l is responsible for calculating the possible power allocation in each subchannel and performing subchannel allocation based on the obtained information. Moreover, it broadcast the values of $f_l^k(p_{Dl}^{k*})$ and z_l (step 1 and step 3).

6.6 Relaxation-Based Algorithms

The dual-based Algorithm 6.2 has polynomial time complexity; however, it may not achieve the optimal solution. In this section, we propose the optimal BnB algorithm and the low-complexity Relaxation-Based Rounding (RBR) algorithm.

6.6.1 Optimal Branch-and-Bound Algorithm

In this section, we apply the Branch-and-Bound (BnB) approach [121] to develop an algorithm that attains the optimal solution of the original problem. Although the BnB algorithm may not achieve the polynomial time complexity, it can significantly reduce the complexity compared to the exhaustive search algorithm. Since any feasible subchannel allocation variable is binary, we propose the BnB algorithm by branching the feasible set of the subchannel allocation vectors where each branching iteration is executed by setting an undetermined subchannel allocation variable to a binary value 0 or 1. Specifically, the algorithm determines the optimal path in the search tree that corresponds to the optimal subchannel allocations for all D2D links. In addition, this optimal path is decided by iteratively visiting potential nodes in the search tree where each node m is associated with some already determined subchannel allocation variables (corresponding to part of the underlying path connecting node m with the root node) and other undetermined subchannel allocation variables.

Let $\bar{\mathcal{Q}}_m$ be the set of all feasible subchannel allocation vectors $\boldsymbol{\rho}$ related to node m where each vector $\boldsymbol{\rho} \in \bar{\mathcal{Q}}_m$ contains corresponding determined and undetermined subchannel allocation variables associated with node m . For convenience, we use m to indicate the iteration index of the searching procedure, and hence $m = 1$ indicates the root node (i.e., we start our search from the root node). Note that each element of $\bar{\mathcal{Q}}_1$ contains all the undetermined subchannel allocation variables.

In each iteration with the corresponding parent node, we consider one of its two child nodes by choosing one undetermined element ρ_i^k of subchannel assignment vector $\boldsymbol{\rho}$ and set it to a binary value 0 or 1 (called node m). In node m , the local upper-bound, BU_m , and lower-bound, BL_m , must be calculated. We also maintain the global upper-bound, BU^* , and lower-bound, BL^* , which are, respectively, the highest local upper-bound and local lower-bound of all active nodes in the searching procedure. In a particular node m , if the calculated local upper-bound satisfies $\text{BU}_m < \text{BL}^*$ then we can remove this node from future consideration because it cannot lead to the optimal solution. On the other hand, if the calculated local lower-bound satisfies $\text{BL}_m > \text{BL}^*$, we can update $\text{BL}^* = \text{BL}_m$. Furthermore, if the global lower-bound BL^* is sufficiently close to the global upper-bound BU^* , we can terminate the algorithm, and output BL^* . In the following, we present the procedures to find the local upper-bound and lower-bound in each node m of the algorithm.

6.6.1.1 Upper-bound Calculation

To obtain the upper-bound of node m , we take the following procedure. First, we define the set \mathcal{Q}_m corresponding to set $\bar{\mathcal{Q}}_m$ but any undetermined subchannel allocation variable ρ_l^k for each element of \mathcal{Q}_m is relaxed as $\rho_l^k \in [0, 1]$. Then, we consider the following problem

$$\begin{aligned} \max_{\mathbf{p}_D, \boldsymbol{\rho} \in \mathcal{Q}_m} \quad & \min_{l \in \mathcal{L}} \frac{w_l \hat{R}_{Dl}(\mathbf{p}_D, \boldsymbol{\rho})}{P_{Dl}^{\text{total}}} \\ \text{s.t.} \quad & (6.9), (6.11), (6.18), \end{aligned} \quad (6.47)$$

whose optimal objective value provides the local upper-bound BU_m of the resource allocation solution in node m . The difference-form of problem (6.47) can be expressed as

$$\begin{aligned} \max_{z, \mathbf{p}_D, \boldsymbol{\rho} \in \mathcal{Q}_m} \quad & z \\ \text{s.t.} \quad & w_l \hat{R}_{Dl}(\mathbf{p}_D, \boldsymbol{\rho}) - \zeta P_{Dl}^{\text{total}}(\mathbf{p}_D, \boldsymbol{\rho}) \geq z, \forall l \in \mathcal{L}, \\ & \text{and (6.9), (6.11), (6.18)}. \end{aligned} \quad (6.48)$$

We now introduce a new vector \mathbf{s}_D that corresponds to the power vector of D2D links \mathbf{p}_D and consider the following optimization problem

$$\max_{z, \mathbf{s}_D, \boldsymbol{\rho} \in \mathcal{Q}_m} \quad z \quad (6.49a)$$

$$\text{s.t.} \quad w_l \bar{R}_{Dl}(\mathbf{s}_{Dl}, \boldsymbol{\rho}) - \zeta \bar{P}_{Dl}^{\text{total}}(\mathbf{s}_{Dl}, \boldsymbol{\rho}) \geq z, \forall l \in \mathcal{L} \quad (6.49b)$$

$$\sum_{k \in \mathcal{K}} s_{Dl}^k \leq P_D^{\text{max}}, \forall l \in \mathcal{L} \quad (6.49c)$$

$$\sum_{l \in \mathcal{L}} \rho_l^k \leq 1, \forall k \in \mathcal{N} \quad (6.49d)$$

$$s_{Dl}^k \leq \rho_l^k P_{Dlk}^{\text{max}}, \forall k \in \mathcal{N}, \forall l \in \mathcal{L} \quad (6.49e)$$

where

$$\bar{R}_{Dl}(\mathbf{s}_{Dl}, \boldsymbol{\rho}) = \sum_{k \in \mathcal{N}} \rho_l^k \log_2 \left(1 + \frac{s_{Dl}^k}{a_{kl} \rho_l^k + b_{kl} s_{Dl}^k} \right) \quad (6.50)$$

$$\bar{P}_{Dl}^{\text{total}} = 2P_0^l + \alpha_l \sum_{k \in \mathcal{K}} s_{Dl}^k. \quad (6.51)$$

We state one important result in the following proposition.

Proposition 6.3. *Problem (6.48) and (6.49) are equivalent, and problem (6.49) is convex.*

Proof. The proof can be found in Appendix 6.10.4. □

Proposition 6.3 implies that the optimum solution of (6.48) can be obtained by solving the convex problem (6.49) using the standard interior point method [21]. We propose Algorithm 3 to solve problem (6.47), in which we iteratively solve problem (6.49) for a given ζ (line 3) and update ζ (line 4) until convergence. If $(\zeta^*, \mathbf{s}_D^*, \boldsymbol{\rho}^*)$ is the solution obtained from Algorithm 6.3 then the components of \mathbf{p}_D^* are given in (6.73) detailed in Appendix 6.3. We now state another important result in the following proposition.

Proposition 6.4. *The obtained solution $(\mathbf{p}_D^*, \boldsymbol{\rho}^*)$ in Algorithm 6.3 is the optimal solution of (6.47).*

Proof. See proof in Appendix 6.10.5. □

Proposition 6.4 implies that ζ obtained from Algorithm 6.3 is an upper-bound of the resource allocation problem associated with node m .

Algorithm 6.3. Upper-bound Calculation

- 1: Initialization: Set $\epsilon = 10^{-6}$, $\zeta_t = 0$, $t = 0$, $\zeta = \epsilon$
- 2: **while** $|\zeta_t - \zeta| \geq \epsilon$ **do**
- 3: Solve problem (6.49) as $\zeta = \zeta_t$ by interior point method to get $(z^{(t)}, \mathbf{s}_D^{(t)}, \boldsymbol{\rho}^{(t)})$
- 4: $\zeta = \zeta_t$, $\zeta_t = \min_{l \in \mathcal{L}} \frac{w_l \bar{R}_{Dl}(\mathbf{s}_D^{(t)}, \boldsymbol{\rho}^{(t)})}{\bar{P}_{Dl}^{\text{total}}(\mathbf{s}_D^{(t)}, \boldsymbol{\rho}^{(t)})}$.
- 5: $t \leftarrow t + 1$
- 6: **end while**
- 7: Output $(\zeta, \mathbf{s}_D^*, \boldsymbol{\rho}^*) = (\zeta, \mathbf{s}_D^{(t)}, \boldsymbol{\rho}^{(t)})$
- 8: Perform power allocation for all D2D links

$$p_{Dl}^{k*} = \begin{cases} 0, & \text{if } \rho_l^{k*} = 0 \\ \frac{s_{Dl}^k}{\rho_l^{k*}}, & \text{otherwise} \end{cases} \quad (6.52)$$

- 9: Output $(\zeta, \mathbf{p}_D^*, \boldsymbol{\rho}^*)$
-

6.6.1.2 Lower-bound Calculation

Note that the local lower-bound in a particular node m can be the objective value achieved by a feasible solution. In node m , while determining the local upper-bound, we obtain $(\mathbf{s}_D^*, \boldsymbol{\rho}^*)$ and $(\mathbf{p}_D^*, \boldsymbol{\rho}^*)$ in lines 7 and 9 of Algorithm 6.3, respectively. Since $\boldsymbol{\rho}^*$ can contain fractional components, it might not be a feasible solution of problem (6.47). The local lower-bound in node m , BL_m , can be obtained by rounding off the values of the fractional subchannel allocation variables. The new feasible resource allocation vector $(\hat{\mathbf{p}}_D, \hat{\boldsymbol{\rho}})$ can be obtained by the following rules

$$\hat{\rho}_l^k = \begin{cases} 1, & \text{if } \rho_l^{k*} = \max_{l \in \mathcal{L}} \rho_l^{k*} \\ 0, & \text{otherwise,} \end{cases} \quad (6.53)$$

$$\hat{p}_{Dl}^k = \begin{cases} 0, & \text{if } \hat{\rho}_l^k = 0 \\ s_{Dl}^{k*}, & \text{if } \hat{\rho}_l^k = 1. \end{cases} \quad (6.54)$$

Specifically, subchannel k is assigned to D2D link l with highest value of ρ_l^{k*} . Moreover, the power allocated to subchannel k is equal to s_{Dl}^{k*} to ensure the feasibility of the resulting solution. This feasible solution $(\hat{\mathbf{p}}_D, \hat{\boldsymbol{\rho}})$ is then used to calculate the local lower-bound.

6.6.2 Relaxation-Based Rounding Algorithm

The BnB algorithm may require, in some cases, to visit a large number of nodes. In the following, we propose the Relaxation-Based-Rounding (RBR) algorithm (Algorithm 6.4), which requires to solve only one relaxed problem and execute the rounding procedure only once. Specifically, we run Algorithm 6.3 for the root node, which is employed by the BnB algorithm, to obtain the initial solution in line 1. Based on the obtained result, we perform subchannel allocations for all subchannel k and D2D link l with $\rho_l^{k*} = 1$ then we execute the rounding procedure (lines 5-10), which is designed to minimize the performance loss as follows. Let $\mathcal{S}_l = \{k \in \mathcal{N} | \rho_l^{k*} = 1\}$, $\mathcal{S}_{fl} = \{k \in \mathcal{N} | \rho_l^{k*} \in (0, 1)\}$ be the sets of exclusive and shared subchannels allocated to D2D link l , respectively, and $\mathcal{U}_k = \{l \in \mathcal{L} | \rho_l^{k*} > 0\}$ be the set of D2D links with positive subchannel allocation variables on subchannel k . In line 2, we calculate the EE of D2D link $l \in \mathcal{L}$ contributed by its

exclusive subchannels in set \mathcal{S}_l as

$$\zeta_l = \frac{\sum_{k \in \mathcal{S}_l} \log_2 \left(1 + \frac{s_{Dl}^{k*}}{a_{kl} + s_{Dl}^{k*}} \right)}{2P_0 + \alpha_l \sum_{k \in \mathcal{S}_l} s_{Dl}^{k*}}. \quad (6.55)$$

In lines 8-9, we allocate each shared subchannel to a unique D2D link. First, we calculate the possible EE improvement of each D2D link $l \in \mathcal{L}$ over its shared subchannel $k \in \mathcal{C}_f$ as

$$\begin{aligned} \Delta_l^k(\mathcal{S}_l) = & \frac{\sum_{n \in \mathcal{S}_l \cup \{k\}} w_l \log_2 \left(1 + \frac{s_l^{k*}}{a_{kl} + b_{kl} s_l^{k*}} \right)}{2P_0 + \alpha_l \sum_{k \in \mathcal{S}_l \cup \{k\}} s_{Dl}^{k*}} \\ & - \frac{\sum_{n \in \mathcal{S}_l} w_l \log_2 \left(1 + \frac{s_l^{k*}}{a_{kl} + b_{kl} s_l^{k*}} \right)}{2P_0 + \alpha_l \sum_{k \in \mathcal{S}_l} s_{Dl}^{k*}}, \end{aligned} \quad (6.56)$$

where \mathcal{S}_l is the set of subchannels assigned to D2D link l before we consider subchannel k and \mathcal{C}_f is the set of unallocated subchannels defined in Algorithm 6.4.

Now define $\mathcal{S} = \{\mathcal{S}_1, \dots, \mathcal{S}_L\}$ where \mathcal{S}_l denotes the set of subchannels allocated to D2D link l , and $\mathcal{A}_k^{\mathcal{S}} = \{l \in \mathcal{N} \mid \Delta_l^k(\mathcal{S}_l) > 0\}$ as the set of D2D links which can improve its EE if these links are assigned subchannels $k \in \mathcal{C}_f$ for a given set \mathcal{S} . In the proposed rounding procedure, we sequentially allocate one subchannel $k \in \mathcal{C}_f$ to the D2D link in $\mathcal{A}_k^{\mathcal{S}}$ that has the minimum EE (line 8). After each assignment, we update the set of assigned subchannels and EE for each D2D link (line 9). The rounding procedure is terminated when all subchannels are allocated.

6.7 Complexity Analysis

In this section, we analyze the complexity of the proposed algorithms in term of the number of required arithmetic operations. Since in BnB algorithm, the number of visited nodes is not fixed, its complexity cannot be exactly determined. On the other hand, the dual-based algorithm requires to solve problem (6.20) iteratively by the dual decomposition method for given ζ with complexity of $O(NL)$. Number of iterations required to update ζ has complexity of $O(1)$. Therefore, the complexity of the dual-based algorithm is $O(NL)$.

The Relaxation-Based Rounding algorithm comprises two phases. In the relaxation phase, we iteratively solve problem (6.49) for given ζ by the interior-point method with complexity of

Algorithm 6.4. Relaxation-Based Rounding Algorithm

-
- 1: Run Algorithm 6.3 for the root node of BnB algorithm to obtain $(\mathbf{s}_D^*, \boldsymbol{\rho}^*)$.
 - 2: Perform subchannel allocations for all subchannel k and D2D link l with $\rho_l^{k*} = 1$
 - 3: Define following sets
 $\mathcal{C}_f = \{k \in \mathcal{N} | \rho_l^{k*} \in [0, 1), \forall l \in \mathcal{L}\}$
 $\mathcal{S}_l = \{k \in \mathcal{N} | \rho_l^{k*} = 1\}$, $\mathcal{S} = \{\mathcal{S}_1, \dots, \mathcal{S}_L\}$
 - 4: Calculate EE for D2D link l :

$$\zeta_l = \frac{\sum_{n \in \mathcal{S}_l} w_l \log_2 \left(1 + \frac{s_l^{k*}}{a_{kl} + b_{kl} s_l^{k*}} \right)}{2P_0 + \alpha_l \sum_{k \in \mathcal{S}_l} s_{Dl}^{k*}}, \forall l \in \mathcal{L}$$
 - 5: **while** $\mathcal{C}_f \neq \emptyset$ **do**
 - 6: Select subchannel $k \in \mathcal{C}_f$
 $\mathcal{U}_k = \{l \in \mathcal{L} | \rho_l^{k*} > 0\}$
 - 7: Calculate $\Delta_l^k(\mathcal{S}_l), \forall l \in \mathcal{U}_k$, according to (6.56)
 $\mathcal{A}_k^{\mathcal{S}} = \{l \in \mathcal{L} | \Delta_l^k(\mathcal{S}_l) > 0\}$
 - 8: $l^* = \underset{l \in \mathcal{A}_k^{\mathcal{S}}}{\operatorname{argmin}} \zeta_l$.
 Assign subchannel k to D2D link l^* , update EE of D2D link l^* , $\zeta_l \leftarrow \zeta_l + \Delta_l^k(\mathcal{S}_l)$
 - 9: Update $\mathcal{S}_{l^*} \leftarrow \mathcal{S}_{l^*} \cup \{k\}$, $s_{Dl^*}^{k*} \leftarrow 0, \forall l \in \mathcal{L} \setminus \{l^*\}$, and $\mathcal{C}_f \leftarrow \mathcal{C}_f - \{k\}$
 - 10: **end while**
 - 11: Output $\mathbf{p}_{Dl}^* = \mathbf{s}_D^*$, and \mathcal{S}
-

$O\left(m^{\frac{1}{2}}(m+n)n^2\right)$, where m is the number of inequality constraints and n is number of variables [122]. Therefore, the complexity of solving problem (6.49) and also of the relaxation phase is $O(N^{3.5})$. In addition, the rounding phase has complexity of $O(L^2)$. Finally, the complexity of the RBR algorithm is $O(N^{3.5})$.

6.8 Numerical Results

We consider the simulation setting shown in Fig. 6.1 with the base-station located at the center, $K = 20$ cellular users, and $L = 2$ or 4 D2D links randomly placed in 500m x 500m area, and $N = 20$ subchannels for uplink communications. The summary of parameter settings used in the simulations is presented in Table I.

The subchannel power gain is modeled as $h_{kl}^n = \left(\frac{d_{kl}}{d_0}\right)^{-3} \delta$ where $d_0 = 1$ m is the reference distance, and $d_{kl} > d_0$ is the distance between the receiver of link k and the transmitter of link l , and δ represents the Rayleigh fading coefficient, which follows the exponential distribution with the mean value of 1. We set the noise power equal to 10^{-12} W for every link. The circuit power of each cellular link P_0 is 0.5 W, the factor α_l is 1.5 for each D2D link, and the maximum transmit powers

Tableau 6.1 – Simulation parameters

Description	Parameter	Value
Number of D2D links	L	2 or 4
Number of cellular links	K	20
Number of subchannels	N	20
Maximum distance between Tx and Rx of D2D links	d_{\max}	50m
Maximum transmit power of cellular link	P_{Ck}^{\max}	0.5W
Maximum transmit power of D2D link	P_{Dl}^{\max}	0.5W
Circuit power	P_0	0.5W
Scaling factor	α_l	1.5
Minimum required rate of cellular links	R_{Ck}^{\min}	2 b/s/Hz
Weighting parameter	w_l	1
Noise power	σ_k	10^{-12}

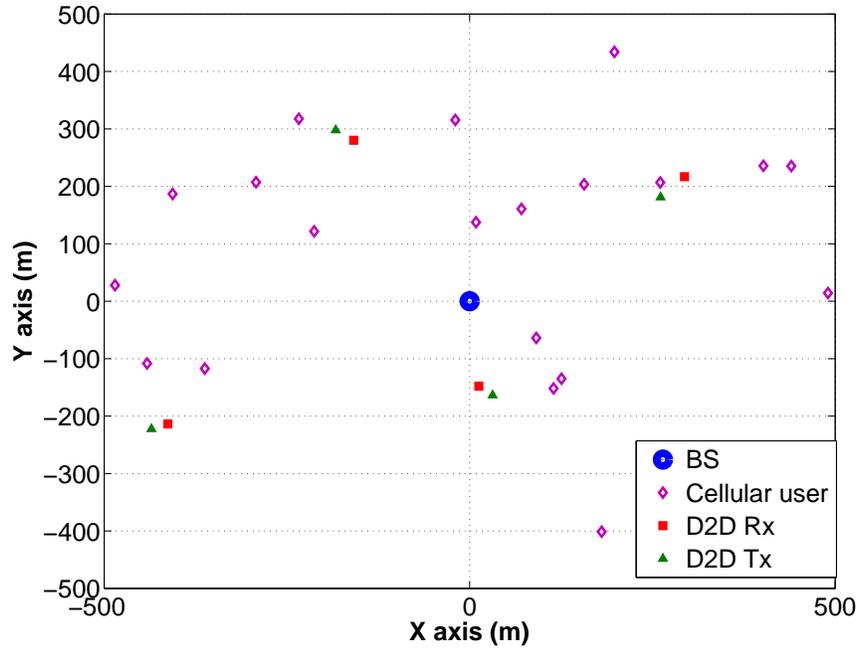


Figure 6.1 – Simulation setting

of each cellular link k and cellular link l are $P_{Ck}^{\max} = P_{Dl}^{\max} = 0.5$ W, $\forall k \in \mathcal{K}, \forall l \in \mathcal{L}$. In addition, the weighting parameters of D2D links are set as $w_l = 1, \forall l \in \mathcal{L}$, the maximum distance of D2D links d_{\max} is 50 m, and the minimum required rate of each cellular link k is $R_{Ck}^{\min} = R_C^{\min} = 2$ b/s/Hz, $\forall k \in \mathcal{K}$.

We evaluate the performance of the proposed algorithms and that [22] with minor modification (called “conventional” algorithm) since our work and [22] consider the similar network settings, non-orthogonal spectrum sharing between cellular and D2D links, and link EE maximization objective.

However, there are some differences between two works. In [22], each D2D or cellular link performs power allocation for all subchannels to maximize the EE while in this work, we consider the joint subchannel assignment and power allocation to maximize the minimum weighted EE of D2D links. Since we focus on maximizing the minimum weighted EE of D2D links while satisfying the minimum cellular link data rates, we modify the algorithm developed in [22] so as to maximize the EE of D2D links, and minimizing the total transmit power of cellular links while maintaining the minimum cellular-link rate requirement.

We also consider the spectrum-efficient solution as a reference, which is obtained by solving problem (6.48) for the case where all subchannel allocation variables are undetermined and $\zeta = 0$. In addition, to verify the efficiency of our algorithms, we compare the objective values achieved by our algorithms with their corresponding upper-bounds achieved by solving the relaxed version of problem (6.19). All numerical results are acquired by averaging over 1000 random realizations of D2D and cellular locations, and channel gains. The EE of D2D links corresponding to the BnB algorithm, dual-based algorithm, RBR algorithm, upper bound as well as the spectrum-efficiency (SE) maximization solution are indicated by "BnB Alg.", "Dual Alg.", "Rounding Alg.", "Upper-bound", and "SE solution", respectively. In all figures in the following, we show the minimum achieved EE of all D2D links (i.e., the design objective) versus different parameters. For brevity, the minimum EE of all D2D links is simply referred to as EE of D2D links in the figures and following discussions.

Figs. 6.2 and 6.3 show the achieved EE of D2D links versus d_{\max} and R_C^{\min} , respectively, for $L = 2$, which allows us to obtain the optimal solution of problem (6.12) through the BnB algorithm described in Section 6.6.1 within reasonable time. It can be seen that the conventional algorithm and SE-maximization solution achieve much lower EE than our proposed energy-efficient algorithms. Moreover, the EE gap between the proposed and conventional algorithms significantly increases as d_{\max} increases. Therefore, it confirms that the proposed algorithms can effectively manage the co-channel interference among the links and make efficient subchannel assignments for D2D links. On the other hand, in the conventional algorithm, since each D2D link selfishly optimizes its EE, the interference among the links may not be well managed. As a result, some D2D links could achieve low EE values, which explains the inferior performance of the conventional design.

It is remarkable that the RBR algorithm performs extremely well with its achieved EE very close to that of the optimal BnB algorithm. In fact, when the number of D2D links is small, the

number of shared subchannels is small as compared to the number of available subchannels, which leads to small performance loss during the rounding phase. Figs. 6.2 and 6.3 also indicate that the Dual-Based Algorithm 6.2 can offer performance close to that of the BnB algorithm since with a small number of D2D links, the duality gap of problem (6.22) for given ζ is also small.

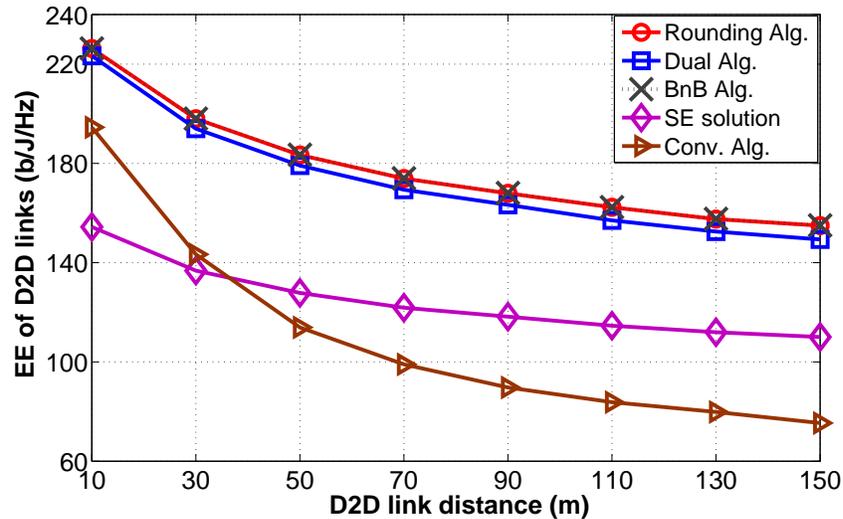


Figure 6.2 – Minimum EE of D2D links versus d_{\max} for $L = 2$

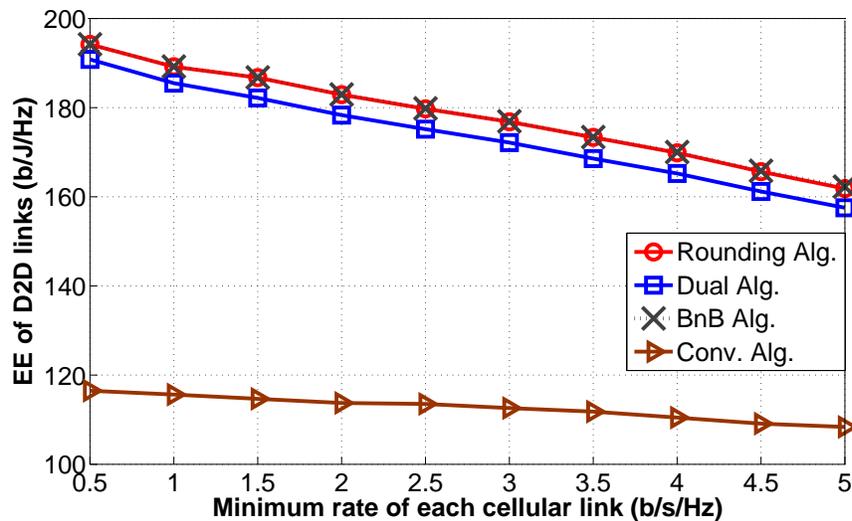


Figure 6.3 – Minimum EE of D2D links versus minimum rate of cellular links for $L = 2$

The characteristics and performance of different algorithms are further investigated with larger number of D2D links, $L = 4$. Figs. 6.4 and 6.5 show the convergence of the dual-based Algorithm 6.2 and the initial relaxation phase of the RBR Algorithm 6.4 for $d_{\max} = 10$ m and 100 m, respectively. They confirm that the gradient-based method used in Algorithm 6.4 to determine the optimal solution converges faster than the bisection method employed by Algorithm 6.2.

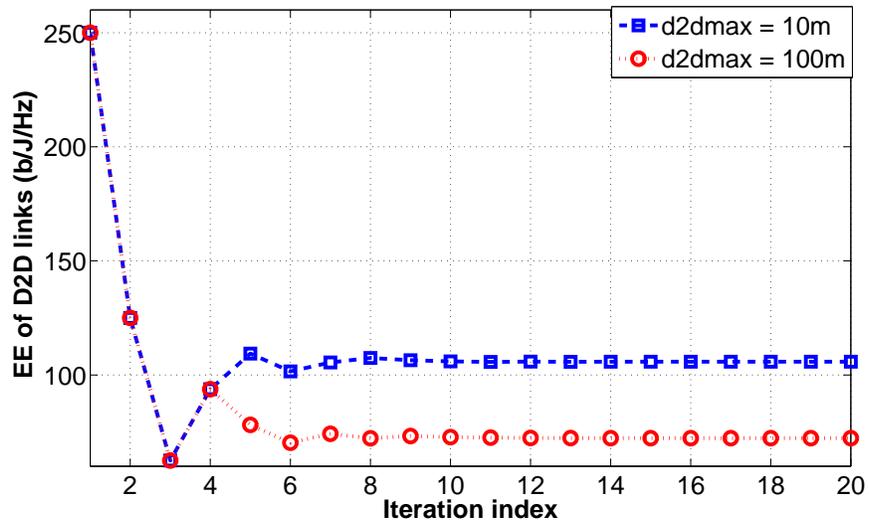


Figure 6.4 – Convergence behavior of dual-based Algorithm 6.2

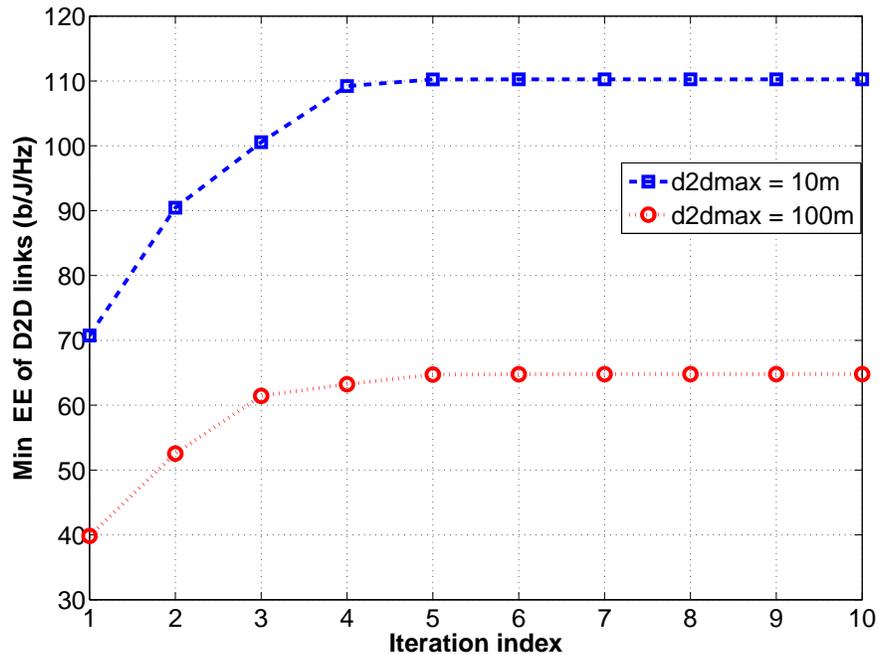


Figure 6.5 – Convergence behavior of the relaxation phase of RBR Algorithm 6.4

Fig. 6.6 indicates that the EE of D2D links achieved by Algorithms 2 and 4 are significantly higher than that of the conventional algorithm, e.g., at $d_{max} = 150$ m, Algorithms 2 and 4 can achieve more than 90% of the upper-bound EE, which is about 300% that of the conventional algorithm and about 130% that of the SE-maximization solution.

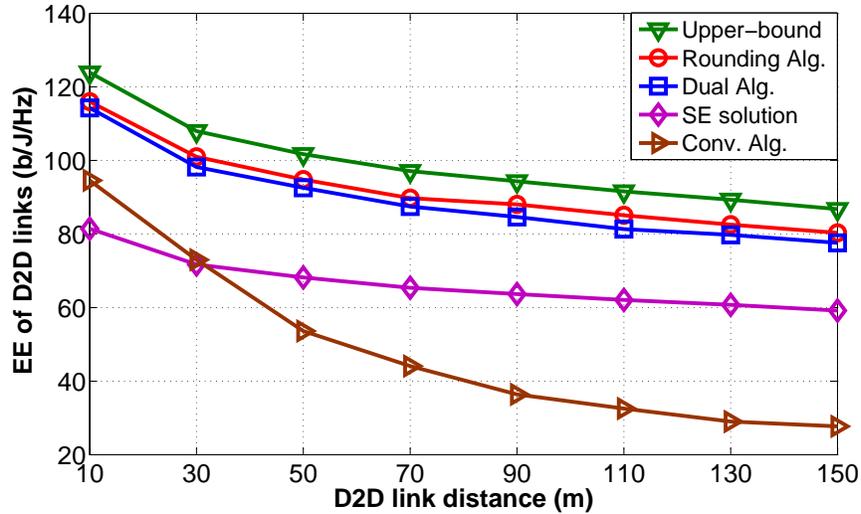


Figure 6.6 – Minimum EE of D2D links versus D2D link distance when $L = 4$

The achieved EE of D2D links versus the minimum required rate of cellular links, plotted in Fig. 6.7, indicates that as the required cellular-link rate increases, the achieved EE of D2D links is reduced. This observation can be explained as follows. As the minimum required rate of each cellular link increases, each cellular user has to increase its transmit power to maintain the required rate, which results in stronger interference for the co-channel D2D links. Moreover, since D2D links are relatively robust against interference due to their short communication distances, the minimum EE is moderately impacted as the minimum rate of cellular links increases.

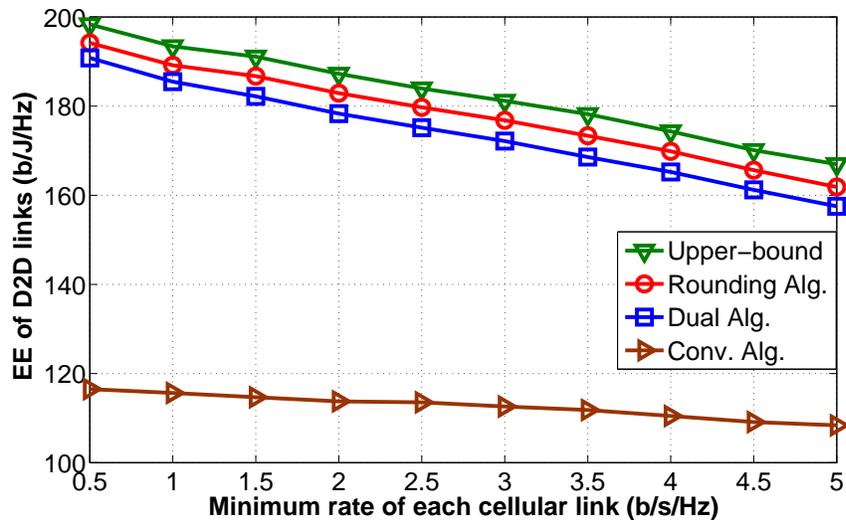


Figure 6.7 – Minimum EE of D2D links versus minimum required cellular-link rate with $L = 4$

Fig. 6.8 demonstrates the achieved EE of D2D links as a function of the circuit power. Both Algorithm 6.2 and Algorithm 6.4 offer excellent performance, which is very close to the upper bound. For small circuit power, their achieved EE is about 150% of the EE due to the conventional algorithm. However, when the circuit power increases, the performance gap between the proposed and the conventional algorithms is reduced since the total consumed power is dominated by the circuit power.

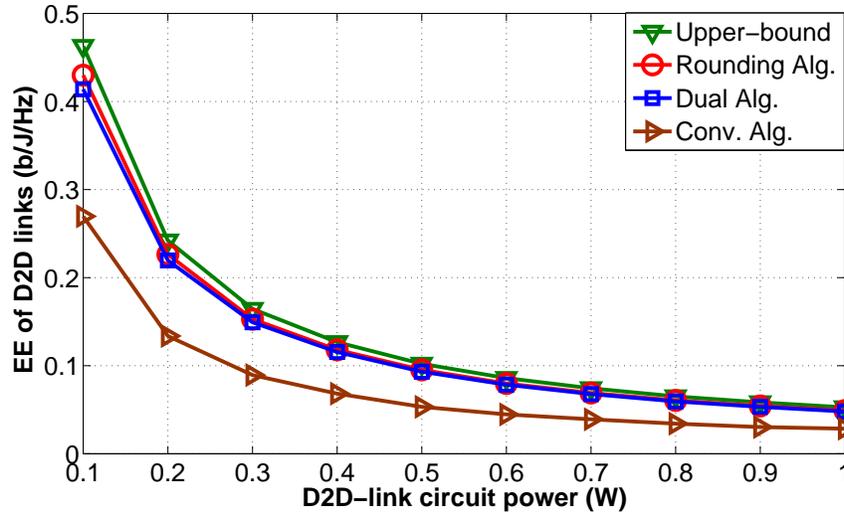


Figure 6.8 – Minimum EE of D2D links versus circuit power

The achieved EE of D2D links versus the noise power, plotted in Fig. 6.9 indicates that, for lower noise power, both Algorithms 2 and 4 achieve higher EE, much better than the conventional algorithm. This is because for the proposed algorithms, when σ reduces, each D2D or cellular link decreases its transmit power. As a result, the co-channel interference between D2D and cellular links also decreases, which leads to the improvement in the EE. In contrast, in the conventional algorithm, all links would operate in the high-interference regime; therefore, the noise power is more negligible compared to the interference and its variation does not impact the EE achieved by D2D links.

Finally, Fig. 6.10 shows that the achieved EE of D2D links decreases as the number of D2D links increases. The performance gap between the proposed and the conventional algorithms also decreases as the number of D2D links increases. This is because as the system supports more D2D links, the available resources for each D2D link becomes smaller, which results in the decrease in the achieved EE of D2D links. Fig. 6.10 also illustrates that as the number of D2D links increases, the gap between the proposed dual-based algorithm and the upper-bound becomes larger since the

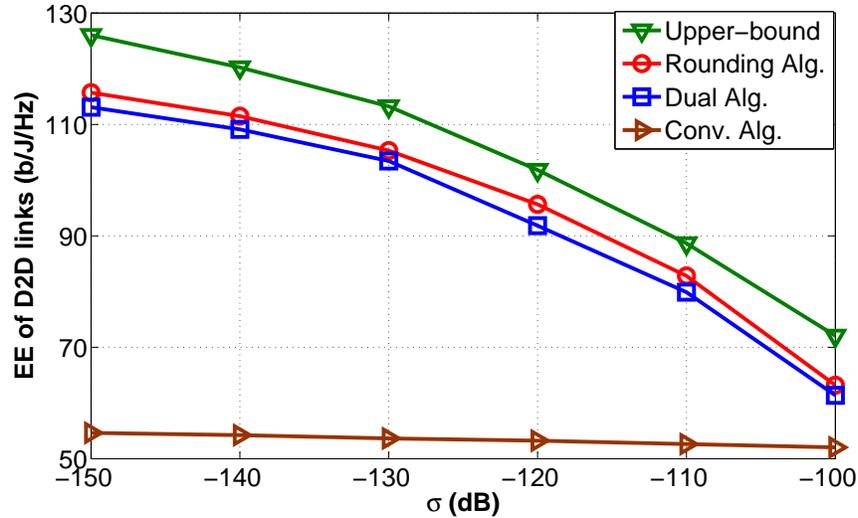


Figure 6.9 – Minimum EE of D2D links versus the noise power

duality gap of the underlying MINLP at convergence under Algorithm 6.2 becomes higher, which results in larger performance loss for this algorithm.

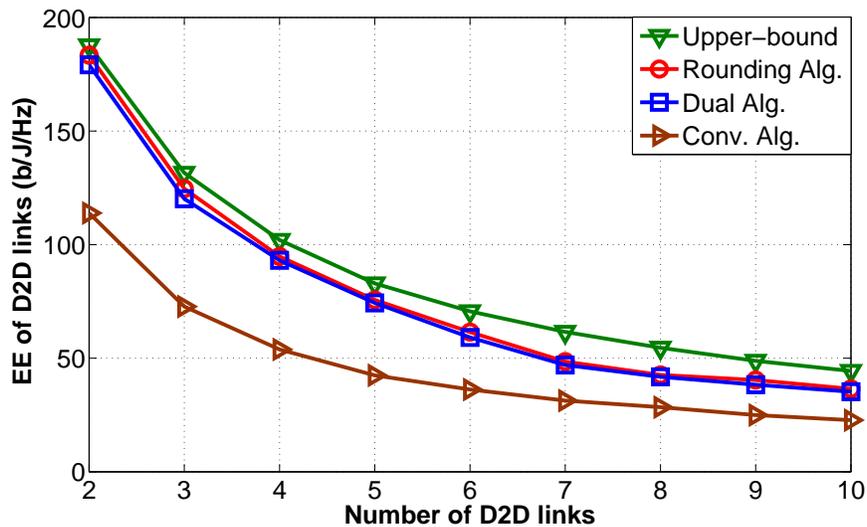


Figure 6.10 – Minimum EE of D2D links versus number of D2D links

6.9 Conclusions

In this chapter, we have developed efficient resource allocation algorithms for D2D underlying cellular systems, which maximizes the minimum weighted EE of D2D links while guaranteeing the QoS of cellular links. In particular, we have proposed the optimal BnB algorithm based on the

novel branching and bounding procedures, and proposed two low-complexity algorithms: (i) the dual-based Algorithm 6.2 solves the resource allocation problem in the dual domain, and (ii) the Relaxed-Based Rounding (RBR) Algorithm 6.4 solves the relaxed version first and then applies a rounding procedure to obtain a feasible solution for the considered resource allocation problem. We have studied the theoretical performance of the proposed low-complexity algorithms and analyzed their computational complexity. Numerical results have confirmed that both proposed Algorithms 2 and 4 can achieve excellent performance, which is close to that due to the optimal BnB algorithm and the upper bound.

6.10 Appendices

6.10.1 Proof of Proposition 6.1

We first show that the min-rate constraints of cellular link k must be met at equality as follows:

$$\log_2 \left(1 + \frac{p_{Ck}^k h_{kk}^k}{\sigma_k^k + p_{Dl}^k h_{kl}^k} \right) = R_{Ck}^{\min}. \quad (6.57)$$

Note that if $\frac{\sigma_k^k}{h_{kk}^k} (2^{R_{Ck}^{\min}} - 1) > P_{Ck}^{\max}$, the required minimum rate of cellular link k cannot be supported; hence, problem (6.12) is infeasible. On the other hand, if $\frac{\sigma_k^k}{h_{kk}^k} (2^{R_{Ck}^{\min}} - 1) \leq P_{Ck}^{\max}$, the cellular link k can allow D2D link l to reuse its resource. Therefore, the power of cellular link k can be expressed as the power of D2D link l on subchannel k as follows:

$$p_{Ck}^k = \frac{\sigma_k^k + p_{Dl}^k h_{kl}^k}{h_{kk}^k} (2^{R_{Ck}^{\min}} - 1) \quad (6.58)$$

$$p_{Dl}^k = \frac{1}{h_{kl}^k} \left(\frac{p_{Ck}^k h_{kk}^k}{2^{R_{Ck}^{\min}} - 1} - \sigma_k^k \right) \quad (6.59)$$

Moreover, the power of cellular link k must satisfy the maximum power constraint, which can be expressed as

$$p_{Dl}^k \leq \frac{1}{h_{kl}^k} \left(\frac{P_{Ck}^{\max} h_{kk}^k}{2^{R_{Ck}^{\min}} - 1} - \sigma_k^k \right). \quad (6.60)$$

Let us now define

$$P_{Dlk}^{\max} \triangleq \min \left\{ P_{Dl}^{\max}, \frac{1}{h_{kl}^k} \left(\frac{P_{Ck}^{\max} h_{kk}^k}{2^{R_{Ck}^{\min}} - 1} - \sigma_k^k \right) \right\} \quad (6.61)$$

Then, it can be verified that if D2D link $l \in \mathcal{L}$ reuses the resource of cellular link k , we have $p_{Dl}^k \in [0, P_{Dlk}^{\max}]$, and $p_{Dl}^k = \frac{1}{h_{kl}^k} \left(\frac{p_{Ck}^k h_{kk}^k}{2^{R_{Ck}^{\min}} - 1} - \sigma_k^k \right)$.

6.10.2 Proof of Theorem 6.1

First, we prove that $\eta^*(\zeta)$ is a decreasing function of ζ . Suppose we have $\zeta_1 > \zeta_2$, and $\eta^*(\zeta_1)$ and $\eta^*(\zeta_2)$ are the optimal objective value of problem (6.20) corresponding to the parameters ζ_1 and ζ_2 , respectively. Now, we define the following

$$(\mathbf{p}_D^o, \boldsymbol{\rho}^o) \triangleq \underset{(\mathbf{p}_D, \boldsymbol{\rho}) \in \mathcal{D}}{\operatorname{argmax}} \min_{l \in \mathcal{L}} [w_l \hat{R}_{Dl}(\mathbf{p}_D, \boldsymbol{\rho}) - \zeta_1 P_{Dl}^{\text{total}}(\mathbf{p}_D, \boldsymbol{\rho})]. \quad (6.62)$$

Then, we have

$$w_l \hat{R}_{Dl}(\mathbf{p}_D^o, \boldsymbol{\rho}^o) - \zeta_1 P_{Dl}^{\text{total}}(\mathbf{p}_D^o, \boldsymbol{\rho}^o) \geq \eta^*(\zeta_1), \forall l \in \mathcal{L}. \quad (6.63)$$

Because $P_{Dl}^{\text{total}}(\mathbf{p}_D^o, \boldsymbol{\rho}^o) \geq 0$ and $\zeta_1 > \zeta_2$, we have $\zeta_1 P_{Dl}^{\text{total}}(\mathbf{p}_D^o, \boldsymbol{\rho}^o) \geq \zeta_2 P_{Dl}^{\text{total}}(\mathbf{p}_D^o, \boldsymbol{\rho}^o), \forall l \in \mathcal{L}$. Consequently, we have $w_l \hat{R}_{Dl}(\mathbf{p}_D^o, \boldsymbol{\rho}^o) - \zeta_1 P_{Dl}^{\text{total}}(\mathbf{p}_D^o, \boldsymbol{\rho}^o) \leq w_l \hat{R}_{Dl}(\mathbf{p}_D^o, \boldsymbol{\rho}^o) - \zeta_2 P_{Dl}^{\text{total}}(\mathbf{p}_D^o, \boldsymbol{\rho}^o), \forall l \in \mathcal{L}$. Finally, we arrive at the following

$$\begin{aligned} \eta^*(\zeta_2) &= \max_{(\mathbf{p}_D, \boldsymbol{\rho}) \in \mathcal{D}} \min_{l \in \mathcal{L}} [w_l \hat{R}_{Dl}(\mathbf{p}_D, \boldsymbol{\rho}) - \zeta_2 P_{Dl}^{\text{total}}(\mathbf{p}_D, \boldsymbol{\rho})] \\ &\geq \min_{l \in \mathcal{L}} [w_l \hat{R}_{Dl}(\mathbf{p}_D^o, \boldsymbol{\rho}^o) - \zeta_2 P_{Dl}^{\text{total}}(\mathbf{p}_D^o, \boldsymbol{\rho}^o)] \\ &\geq \min_{l \in \mathcal{L}} [w_l \hat{R}_{Dl}(\mathbf{p}_D^o, \boldsymbol{\rho}^o) - \zeta_1 P_{Dl}^{\text{total}}(\mathbf{p}_D^o, \boldsymbol{\rho}^o)] \\ &= \eta^*(\zeta_1). \end{aligned} \quad (6.64)$$

From these results, we have $\eta^*(\zeta_2) \geq \eta^*(\zeta_1), \forall \zeta_1 > \zeta_2$, which implies that $\eta^*(\zeta)$ is a decreasing function of ζ .

Assume that (6.21) holds, we need to prove that ζ^* is the optimal solution of problem (6.19). We prove this by contradiction as follows. If ζ^* is not the optimal solution of problem (6.19), then we have $\exists \zeta^o = \min_{l \in \mathcal{L}} \frac{w_l \hat{R}_{Dl}(\mathbf{p}_D^o, \boldsymbol{\rho}^o)}{P_{Dl}^{\text{total}}(\mathbf{p}_D^o, \boldsymbol{\rho}^o)}$, where $\zeta^o > \zeta^*$ be the optimal solution of problem (6.19). This means

that $\frac{w_l \hat{R}_{Dl}(\mathbf{p}_D^o, \boldsymbol{\rho}^o)}{P_{Dl}^{\text{total}}(\mathbf{p}_D^o, \boldsymbol{\rho}^o)} > \zeta^*, \forall l \in \mathcal{L}$, which implies that $w_l \hat{R}_{Dl}(\mathbf{p}_D^o, \boldsymbol{\rho}^o) - \zeta^* P_{Dl}^{\text{total}}(\mathbf{p}_D^o, \boldsymbol{\rho}^o) > 0, \forall l \in \mathcal{L}$.

Therefore, we have

$$\begin{aligned} & \max_{\mathbf{p}_D, \boldsymbol{\rho} \in \mathcal{D}} \min_{l \in \mathcal{L}} [w_l \hat{R}_{Dl}(\mathbf{p}_D, \boldsymbol{\rho}) - \zeta^* P_{Dl}^{\text{total}}(\mathbf{p}_D, \boldsymbol{\rho})] \\ & \geq \min_{l \in \mathcal{L}} [w_l \hat{R}_{Dl}(\mathbf{p}_D, \boldsymbol{\rho}) - \zeta^* P_{Dl}^{\text{total}}(\mathbf{p}_D, \boldsymbol{\rho})] \\ & > 0, \end{aligned} \tag{6.65}$$

which contradicts with the assumption in (6.21). Therefore, $\zeta^* = \min_{l \in \mathcal{L}} \frac{w_l \hat{R}_{Dl}(\mathbf{p}_D^*, \boldsymbol{\rho}^*)}{P_{Dl}^{\text{total}}(\mathbf{p}_D^*, \boldsymbol{\rho}^*)}$ is the optimal solution of problem (6.12).

6.10.3 Proof of Proposition 6.2

Since the dual decomposition algorithm proposed to solve problem (6.20) for given ζ always converges if we choose the step sizes $\boldsymbol{\theta}$ and $\boldsymbol{\kappa}$ appropriately [118], and the bisection method to update ζ always converges, the iterative loops of Algorithm 6.2 always converge. Hence, Algorithm 6.2 returns a feasible solution of problem (6.19).

From the dual decomposition procedure, we have

$$\begin{aligned} (\mathbf{p}_D^*, \boldsymbol{\rho}^*) = \operatorname{argmax}_{\mathbf{p}_D \in \mathcal{X}, \boldsymbol{\rho} \in \mathcal{C}} & \sum_{l \in \mathcal{L}} \mu_l^* [w_l \hat{R}_{Dl}(\mathbf{p}_D, \boldsymbol{\rho}) - \zeta^* P_{Dl}^{\text{total}}] \\ & + \sum_{l \in \mathcal{L}} \lambda_l^* (P_D^{\max} - \sum_{k \in \mathcal{N}} \rho_l^k p_{Dl}^k). \end{aligned} \tag{6.66}$$

We denote $\mathcal{X}_f = \{\mathbf{p}_D \in \mathcal{X} \mid \sum_{k \in \mathcal{N}} \rho_l^k p_{Dl}^k \leq P_{Dl}^{\max}, \boldsymbol{\rho} \in \mathcal{C}\}$ as the set of feasible power allocation solutions of problem (6.19); hence, $\mathcal{X}_f \subset \mathcal{X}$. Then, we have

$$\max_{\mathbf{p}_D \in \mathcal{X}_f, \boldsymbol{\rho} \in \mathcal{C}} \sum_{l \in \mathcal{L}} \mu_l^* \left[w_l \hat{R}_{Dl}(\mathbf{p}_D, \boldsymbol{\rho}) - \zeta^* P_{Dl}^{\text{total}}(\mathbf{p}_D, \boldsymbol{\rho}) \right] \quad (6.67)$$

$$\stackrel{(a)}{\leq} \max_{\mathbf{p}_D \in \mathcal{X}_f, \boldsymbol{\rho} \in \mathcal{C}} \sum_{l \in \mathcal{L}} \mu_l^* \left[w_l \hat{R}_{Dl}(\mathbf{p}_D, \boldsymbol{\rho}) - \zeta^* P_{Dl}^{\text{total}}(\mathbf{p}_D, \boldsymbol{\rho}) \right] + \sum_{l \in \mathcal{L}} \lambda_l^* (P_D^{\max} - \sum_{k \in \mathcal{N}} \rho_l^k p_{Dl}^k) \quad (6.68)$$

$$\stackrel{(b)}{\leq} \max_{\mathbf{p}_D \in \mathcal{X}, \boldsymbol{\rho} \in \mathcal{C}} \sum_{l \in \mathcal{L}} \mu_l^* \left[w_l \hat{R}_{Dl}(\mathbf{p}_D, \boldsymbol{\rho}) - \zeta^* P_{Dl}^{\text{total}}(\mathbf{p}_D, \boldsymbol{\rho}) \right] + \sum_{l \in \mathcal{L}} \lambda_l^* (P_D^{\max} - \sum_{k \in \mathcal{N}} \rho_l^k p_{Dl}^k) \quad (6.69)$$

$$\stackrel{(c)}{=} 0. \quad (6.70)$$

Inequality (a) holds because of $\sum_{l \in \mathcal{L}} \lambda_l^* (P_D^{\max} - \sum_{k \in \mathcal{N}} \rho_l^k p_{Dl}^k) \geq 0$; inequality (b) is the result of $\mathcal{X}_f \subset \mathcal{X}$; and equality (c) is due to the assumption of Proposition 6.2. Therefore, we have

$$\max_{\mathbf{p}_D \in \mathcal{X}_f, \boldsymbol{\rho} \in \mathcal{C}} \sum_{l \in \mathcal{L}} \mu_l^* \left[w_l \hat{R}_{Dl}(\mathbf{p}_D, \boldsymbol{\rho}) - \zeta^* P_{Dl}^{\text{total}}(\mathbf{p}_D, \boldsymbol{\rho}) \right] = 0. \quad (6.71)$$

Since we have $\mu_l^* > 0, \forall l \in \mathcal{L}$, the following holds

$$\max_{\mathbf{p}_D \in \mathcal{X}_f, \boldsymbol{\rho} \in \mathcal{C}} \min_{l \in \mathcal{L}} \left[w_l R_{Dl}(\mathbf{p}_D, \boldsymbol{\rho}) - \zeta^* P_{Dl}^{\text{total}}(\mathbf{p}_D, \boldsymbol{\rho}) \right] = 0. \quad (6.72)$$

Therefore, by using the results of Theorem 6.1, ζ^* is the optimal solution of problem (6.19).

6.10.4 Proof of Proposition 6.3

If $(z^*, \mathbf{p}_D^*, \boldsymbol{\rho}^*)$ is the optimal solution of problem (6.48), we can express \mathbf{s}_D^* as $s_{Dl}^{k*} = \rho_l^{k*} p_{Dl}^{k*}, \forall l \in \mathcal{L}, \forall k \in \mathcal{N}$. Therefore, $(z^*, \mathbf{s}_D^*, \boldsymbol{\rho}^*)$ is a feasible solution of problem (6.49). On the other hand, if $(z^*, \mathbf{s}_D^*, \boldsymbol{\rho}^*)$ is the optimal solution of problem (6.49), \mathbf{p}_D^* is given as

$$p_{Dl}^{k*} = \begin{cases} 0, & \text{if } \rho_l^{k*} = 0 \\ \frac{s_{Dl}^{k*}}{\rho_l^{k*}}, & \text{otherwise.} \end{cases} \quad (6.73)$$

Consequently, $(z^*, \mathbf{p}_D^*, \boldsymbol{\rho}^*)$ is a feasible solution of problem (6.48). Hence, z^* is the optimal objective value of problem (6.48) iff it is the optimal objective value of problem (6.49), which means that problems (6.48) and (6.49) are equivalent.

In the following, we prove that (6.48) is convex. It can be seen that in problem (6.48), the objective is a linear function of variable z , and (6.49c)-(6.49e) are the linear constraints. Therefore, we only need to prove that $g_l(\mathbf{s}_{Dl}, \boldsymbol{\rho})$ are concave functions of $(\mathbf{s}_{Dl}, \boldsymbol{\rho}), \forall l \in \mathcal{L}$ where

$$g_l(\mathbf{s}_{Dl}, \boldsymbol{\rho}) \triangleq w_l \bar{R}_{Dl}(\mathbf{s}_{Dl}, \boldsymbol{\rho}) - \zeta \bar{P}_{Dl}^{\text{total}}(\mathbf{s}_{Dl}, \boldsymbol{\rho}). \quad (6.74)$$

Moreover, in $g_l(\mathbf{s}_{Dl}, \boldsymbol{\rho})$, $\bar{P}_{Dl}^{\text{total}}(\mathbf{s}_{Dl}, \boldsymbol{\rho}) = 2P_0^l + \alpha_l \sum_{k \in \mathcal{K}} s_{Dl}^k$ is a linear combination of \mathbf{s}_{Dl} . Therefore, the remaining task is to prove that $\bar{R}_{Dl}(\mathbf{s}_{Dl}, \boldsymbol{\rho})$ are concave functions of $(\mathbf{s}_{Dl}, \boldsymbol{\rho}), \forall l \in \mathcal{L}$.

We also have $\bar{R}_{Dl}(\mathbf{s}_{Dl}, \boldsymbol{\rho}) = \sum_{k \in \mathcal{K}} \bar{R}_{Dl}^k(s_{Dl}^k, \rho_l^k)$ where

$$\bar{R}_{Dl}^k(s_{Dl}^k, \rho_l^k) = \begin{cases} \rho_l^k \log_2 \left(1 + \frac{s_{Dl}^k}{a_{kl} \rho_l^k + b_{kl} s_{Dl}^k} \right), \\ \quad \text{if } \rho_l^k \text{ undetermined} \\ \log_2 \left(1 + \frac{s_{Dl}^k}{a_{kl} + b_{kl} s_{Dl}^k} \right), \text{ if } \rho_l^k = 1 \\ 0, \text{ if } \rho_l^k = 0. \end{cases} \quad (6.75)$$

In the following, we will prove that $\bar{R}_{Dl}^k(s_{Dl}^k, \rho_l^k)$ are concave functions for all possible cases of ρ_l^k . First, if $\rho_l^k = 1$, then $\bar{R}_{Dl}^k(s_{Dl}^k, \rho_l^k) = \log_2 \left(1 + \frac{s_{Dl}^k}{a_{kl} + b_{kl} s_{Dl}^k} \right)$ and the second partial derivative of $\bar{R}_{Dl}^k(s_{Dl}^k, \rho_l^k)$ can be expressed as

$$\frac{\partial^2 \bar{R}_{Dl}^k(s_{Dl}^k, \rho_l^k)}{\partial s_{Dl}^k{}^2} = - \frac{a_{kl} b_{kl} [a_{kl} + (b_{kl} + 1) s_{Dl}^k] + a_{kl} (b_{kl} + 1) (a_{kl} + b_{kl} s_{Dl}^k)}{(a_{kl} + b_{kl} s_{Dl}^k)^2 [a_{kl} + (b_{kl} + 1) s_{Dl}^k]^2}. \quad (6.76)$$

Since $a_{kl} > 0$ and $b_{kl} \geq 0, \forall k \in \mathcal{K}, l \in \mathcal{L}$, we have $\frac{\partial^2 \bar{R}_{Dl}^k(s_{Dl}^k, \rho_l^k)}{\partial s_{Dl}^k{}^2} \leq 0$ for all non-negative values of s_{Dl}^k . As a result, $\bar{R}_{Dl}^k(s_{Dl}^k, \rho_l^k)$ is a concave function of s_{Dl}^k for given $\rho_l^k = 1$.

We now consider the case where ρ_l^k is an undetermined variable. Since $\bar{g}_l^k(s_{Dl}^k) = \log_2 \left(1 + \frac{s_{Dl}^k}{a_{kl} + b_{kl} s_{Dl}^k} \right)$ is a concave function of variable s_{Dl}^k , the related function $g_l^k(s_{Dl}^k, \rho_l^k) = \rho_l^k \bar{g}_l^k \left(\frac{s_{Dl}^k}{\rho_l^k} \right) = \rho_l^k \log_2 \left(1 + \frac{s_{Dl}^k}{a_{kl} \rho_l^k + b_{kl} s_{Dl}^k} \right)$

is also a concave function of $(s_{Dl}^k, \rho_l^k), \forall \rho_l^k > 0$. In addition, for the given s_{Dl}^k

$$\begin{aligned} & \lim_{\rho_l^k \rightarrow 0^+} \rho_l^k \log_2 \left(1 + \frac{s_{Dl}^k}{a_{kl}\rho_l^k + b_{kl}s_{Dl}^k} \right) \\ &= \lim_{\rho_l^k \rightarrow 0^+} \rho_l^k \log_2 \left(1 + \frac{1}{a_{kl}\frac{\rho_l^k}{s_{Dl}^k} + b_{kl}} \right) = 0. \end{aligned} \quad (6.77)$$

Hence, $g_l^k(s_{Dl}^k, \rho_l^k)$ is a continuous function of ρ_l^k . According to [123], the concavity of $g_l^k(s_{Dl}^k, \rho_l^k)$ is preserved in the boundary of its domain. Therefore, $g_l^k(s_{Dl}^k, \rho_l^k)$ is a concave function of (s_{Dl}^k, ρ_l^k) . As a result, $\bar{R}_{Dl}^k(s_{Dl}^k, \rho_l^k)$ is a concave function of (s_{Dl}^k, ρ_l^k) . Finally, since $\bar{R}_{Dl}^k(s_{Dl}^k, \rho_l^k)$ is a concave function for all the cases of ρ_l^k , $g_l(\mathbf{s}_{Dl}, \boldsymbol{\rho})$ is a concave function for all $l \in \mathcal{L}$, which means that problem (6.48) is a convex optimization problem.

6.10.5 Proof of Proposition 6.4

Because problems (6.48) and (6.49) are equivalent for given node m and \mathcal{Q}_m , from the solution of problem (6.49) we can obtain the solution of problem (6.48). Assume that $(z^{(t-1)}, \mathbf{p}_D^{(t-1)}, \boldsymbol{\rho}^{(t-1)})$ and $(z^{(t)}, \mathbf{p}_D^{(t)}, \boldsymbol{\rho}^{(t)})$ are the solution of problem (6.48) in iterations $t-1$ and t respectively. In addition, let us define $\zeta_t = \min_{l \in \mathcal{L}} \frac{w_l \hat{R}_{Dl}(\mathbf{p}_D^{(t)}, \boldsymbol{\rho}^{(t)})}{P_{Dl}^{\text{total}}(\mathbf{p}_D^{(t)}, \boldsymbol{\rho}^{(t)})}$, and $\zeta_{t-1} = \min_{l \in \mathcal{L}} \frac{w_l \hat{R}_{Dl}(\mathbf{p}_D^{(t-1)}, \boldsymbol{\rho}^{(t-1)})}{P_{Dl}^{\text{total}}(\mathbf{p}_D^{(t-1)}, \boldsymbol{\rho}^{(t-1)})}$. Moreover, we have

$$\begin{aligned} & \max_{(\mathbf{p}_D, \boldsymbol{\rho}) \in \mathcal{D}} \min_{l \in \mathcal{L}} w_l \hat{R}_{Dl}(\mathbf{p}_D, \boldsymbol{\rho}) - \zeta_{t-1} P_{Dl}^{\text{total}}(\mathbf{p}_D, \boldsymbol{\rho}) \\ &= \min_{l \in \mathcal{L}} w_l \hat{R}_{Dl}(\mathbf{p}_D^{(t)}, \boldsymbol{\rho}^{(t)}) - \zeta_t P_{Dl}^{\text{total}}(\mathbf{p}_D^{(t)}, \boldsymbol{\rho}^{(t)}) \\ &\geq \min_{l \in \mathcal{L}} w_l \hat{R}_{Dl}(\mathbf{p}_D^{(t-1)}, \boldsymbol{\rho}^{(t-1)}) - \zeta_{t-1} P_{Dl}^{\text{total}}(\mathbf{p}_D^{(t-1)}, \boldsymbol{\rho}^{(t-1)}) \\ &= 0 \end{aligned} \quad (6.78)$$

where \mathcal{D} is the set of feasible solutions of problem (6.48). Therefore, $\min_{l \in \mathcal{L}} w_l \hat{R}_{Dl}(\mathbf{p}_D^{(t)}, \boldsymbol{\rho}^{(t)}) - \zeta_t P_{Dl}^{\text{total}}(\mathbf{p}_D^{(t)}, \boldsymbol{\rho}^{(t)}) \geq 0$, which means that $\zeta_t = \min_{l \in \mathcal{L}} \frac{w_l \hat{R}_{Dl}(\mathbf{p}_D^{(t)}, \boldsymbol{\rho}^{(t)})}{P_{Dl}^{\text{total}}(\mathbf{p}_D^{(t)}, \boldsymbol{\rho}^{(t)})} \geq \zeta_{t-1}$. This implies that Algorithm 3 creates a sequence of feasible solutions of problem (6.48) whose objective values monotonically increase over iterations; therefore, the algorithm converges. Assume that at convergence,

$\zeta_{t-1} = \zeta_t = \zeta^* = \min_{l \in \mathcal{L}} \frac{w_l \hat{R}_{Dl}(\mathbf{p}_D^*, \boldsymbol{\rho}^*)}{P_{Dl}^{\text{total}}(\mathbf{p}_D^*, \boldsymbol{\rho}^*)}$. Therefore, the following must hold

$$\begin{aligned}
& \max_{(\mathbf{p}_D, \boldsymbol{\rho}) \in \mathcal{D}} \min_{l \in \mathcal{L}} [w_l \hat{R}_{Dl}(\mathbf{p}_D, \boldsymbol{\rho}) - \zeta_{t-1} P_{Dl}^{\text{total}}(\mathbf{p}_D, \boldsymbol{\rho})] \\
&= \min_{l \in \mathcal{L}} [w_l \hat{R}_{Dl}(\mathbf{p}_D^{(t)}, \boldsymbol{\rho}^{(t)}) - \zeta_{t-1} P_{Dl}^{\text{total}}(\mathbf{p}^{(t)}, \boldsymbol{\rho}^{(t)})] \\
&= \min_{l \in \mathcal{L}} [w_l \hat{R}_{Dl}(\mathbf{p}_D^{(t)}, \boldsymbol{\rho}^{(t)}) - \zeta_t P_{Dl}^{\text{total}}(\mathbf{p}^{(t)}, \boldsymbol{\rho}^{(t)})] \tag{6.79} \\
&= \min_{l \in \mathcal{L}} [w_l \hat{R}_{Dl}(\mathbf{p}_D^*, \boldsymbol{\rho}^*) - \zeta^* P_{Dl}^{\text{total}}(\mathbf{p}^*, \boldsymbol{\rho}^*)] \\
&= 0.
\end{aligned}$$

Since $(\zeta^*, \mathbf{p}_D^*, \boldsymbol{\rho}^*)$ satisfies the sufficient condition of Theorem 6.1, it is the optimal solution of problem (6.47).

Chapter 7

Joint Mode Selection and Resource Allocation for Relay-based D2D Communications

The content of this chapter was published in *IEEE Communications Letters* in the following paper:

T. D. Hoang, L. B. Le, and T. Le-Ngoc, “Joint mode selection and resource allocation for relay-based D2D communications,” *IEEE Communications Letters*, vol. 21, no. 2, pp. 398–401, February 2017.

7.1 Abstract

This chapter studies the joint mode selection, resource group (RG) assignment and power allocation for D2D underlaid cellular communication systems. Our design allows D2D links to reuse the resources of cellular links, and each D2D link can operate in either direct or relay mode. We formulate a resource allocation problem which aims at maximizing the system sum rate of all D2D and cellular links while guaranteeing the required minimum rates of cellular and D2D links. To solve this problem, we first characterize the optimal power allocation when a D2D link operates in the direct or relay mode. Then, using these power allocation results, the joint mode selection

and RG assignment can be formulated as a job assignment problem whose *optimal* solution can be obtained in polynomial time. Numerical results demonstrate that the proposed design significantly outperforms conventional schemes.

7.2 Introduction

Efficient resource allocation plays a fundamental role in realizing the benefits of D2D communications in enhancing the spectral efficiency. Specifically, it is desirable to enable robust D2D communications while guaranteeing the QoS of existing cellular communications. The direct D2D communication mode may not be sufficiently reliable due to the channel fading and interferences from the cellular links [124]. Therefore, it is necessary to employ D2D relaying communications [125] for D2D links in unfavorable communication conditions to enhance their communication reliability and rates. Toward this end, adaptive mode selection where D2D communication links can choose either direct or relay communication mode can greatly improve the communications performance.

Resource allocation for D2D communications has attracted great research attention [23]. The authors in [23] studied joint mode selection and resource allocation for D2D links; nevertheless, it does not consider the relay mode. There are some other existing works dealing with the resource allocation for D2D communication with the support of relays [25]. Nonetheless, the models studied in [25] always force D2D links to communicate through relays, which may not result in the best network performance since the direct mode can be a better choice for robust D2D links. Different from the designs in [25], our model allows each D2D link to operate in either the direct mode or relay mode. In the direct mode, D2D nodes can communicate directly with each other while in the relay mode, two D2D nodes communicate with each other through the assistance of D2D relays.

In this chapter, we study the joint mode selection, resource group (RG) assignment, and power control problem for D2D underlaid cellular networks which aims at maximizing the system sum rate considering minimum rate constraints of cellular and D2D links. The resource allocation problem is formulated as an MINLP (Mixed-Integer Non-Linear Programming) problem. To solve this problem *optimally*, we first study the optimal power allocation for a given mode selection and RG assignment solution. Based on these results, the original resource allocation problem can be transformed into a RG allocation problem, which can be solved optimally by the Hungarian method. Extensive

numerical studies demonstrate that the proposed design significantly outperforms existing D2D communication schemes using fixed direct or relay mode.

7.3 System Model and Problem Formulation

7.3.1 System Model

We consider the uplink of a single macro-cell system where K cellular links in the set $\mathcal{K} = \{1, \dots, K\}$ share the same spectrum of K resource group (RG) in the set $\mathcal{N} = \{1, \dots, K\}$ with L D2D links in the set $\mathcal{L} = \{1, \dots, L\}$. We assume that cellular link $k \in \mathcal{K}$ has been pre-allocated RG $k \in \mathcal{N}$, which consists of m_k consecutive sub-channels.¹ We also assume that each D2D link reuse the resource of one RG, and each RG is assigned to at most one D2D link.

Let $\boldsymbol{\rho}$ be a matrix capturing binary resource allocation decisions of the D2D links where $[\boldsymbol{\rho}]_{kl} = \rho_{kl} = 1$ if D2D link l is assigned RG k and $\rho_{kl} = 0$, otherwise. Moreover, each D2D transmitter can communicate with its corresponding receiver via either direct or relay mode (assisted by a relay). Let $\mathbf{x} = [x_1, \dots, x_L]$ be the binary mode selection decision vector for all D2D links where $x_l = 1$ if D2D link l operates in the direct mode and $x_l = 0$, otherwise. We also assume that the relay selection for each D2D link has been pre-determined where each D2D link $l \in \mathcal{L}$ can be assisted by its assigned relay. We assume that D2D link l is supported by relay r_l in the relay set $\mathcal{R} = \{r_1, \dots, r_L\}$. Denote h_{ab}^n as the channel gain from transmitter of link or relay b to the receiver of link or relay a on RG n . ($a, b \in \mathcal{K} \cup \mathcal{L} \cup \mathcal{R}$)

We denote \mathbf{p}_C and \mathbf{p}_D are the power allocation vector of cellular and D2D links where $[\mathbf{p}_C]_k = p_{Ck}$ and $[\mathbf{p}_D]_l = p_{Dl}$ denote the transmit power of cellular link k and D2D link l , respectively. The data rate of cellular link k on its RG without any co-channel D2D links can be expressed as $R_{Ck}^{(o)} = m_k \log_2 \left(1 + p_{Ck} h_{kk}^k / \sigma^2 \right)$ where σ^2 denotes the thermal noise, and the data rate in b/s/Hz is normalized by the bandwidth of one sub-channel.

¹This is the case in uplink LTE system using SC-FDMA, each subchannel is equivalent to one resource block (RB), and RG is a group of contiguous RBs assigned to one particular cellular link.

7.3.2 D2D Communication Modes

In this work, we allow D2D nodes in each D2D link to communicate to each other using either direct or relay mode. In the direct mode, the D2D transmitter communicates directly with its D2D receiver. However, in the relay mode, we assume that the Decode and Forward (DF) relaying strategy is employed where each communication period is divided into two equal intervals corresponding to the D2D transmitter to relay (D-R) communication phase and relay to D2D receiver (R-D) communication phase.

7.3.2.1 Direct Mode

If RG k is assigned to D2D link l , the data rates of cellular link k and D2D link l are described as

$$R_{Dlk}^{(d)} = m_k \log_2 \left(1 + p_{Dl} h_{ll}^k / (\sigma^2 + p_{Ck} h_{lk}^k) \right) \quad (7.1)$$

$$R_{Ckl}^{(d)} = m_k \log_2 \left(1 + p_{Ck} h_{kk}^k / (\sigma^2 + p_{Dl} h_{kl}^k) \right). \quad (7.2)$$

7.3.2.2 Relay Mode

Let $\mathbf{p}_R = [p_{R1}, \dots, p_{RL}]$ be the power allocation vector of the relays to support their D2D links. Assume that D2D link l reuses the resource of cellular link k and cellular link k can adapt its communication rates in the two communication phases of the co-channel D2D link to achieve the link capacity (e.g., by using adaptive modulation and coding). Then, in this relay mode, the data rates of cellular link k in the first and second communications phases can be expressed as

$$R1_{Ckl}^{(r)} = 0.5 m_k \log_2 \left(1 + p_{Ck} h_{kk}^k / (\sigma^2 + p_{Dl} h_{kl}^k) \right) \quad (7.3)$$

$$R2_{Ckl}^{(r)} = 0.5 m_k \log_2 \left(1 + p_{Ck} h_{kk}^k / (\sigma^2 + p_{Rl} h_{krl}^k) \right). \quad (7.4)$$

Moreover, the data rates achieved on the D-R and R-D links in the first and second communications phases can be calculated as

$$R1_{Dlk}^{(r)} = 0.5 m_k \log_2 \left(1 + p_{Dl} h_{rl}^k / (\sigma^2 + p_{Ck} h_{rk}^k) \right) \quad (7.5)$$

and

$$R2_{Dlk}^{(r)} = 0.5m_k \log_2 \left(1 + p_{R_l} h_{lr_l}^k / (\sigma^2 + p_{Ck} h_{lk}^k) \right), \quad (7.6)$$

respectively. Finally, the data rates of cellular link k and D2D link l can be written, respectively, as

$$R_{Ckl}^{(r)} = R1_{Ckl}^{(r)} + R2_{Ckl}^{(r)} \quad (7.7)$$

$$R_{Dlk}^{(r)} = \min \left\{ R1_{Dlk}^{(r)}, R2_{Dlk}^{(r)} \right\}. \quad (7.8)$$

7.3.3 Problem Formulation

Using the above notations, the data rates of cellular link k and D2D link l can be expressed as

$$R_{Ck} = \left(1 - \sum_{l \in \mathcal{L}} \rho_{kl} \right) R_{Ck}^{(o)} + \sum_{l \in \mathcal{L}} \rho_{kl} \left[x_l R_{Ckl}^{(d)} + (1 - x_l) R_{Ckl}^{(r)} \right] \quad (7.9)$$

$$R_{Dl} = \sum_{n \in \mathcal{N}} \rho_{nl} \left[x_l R_{Dln}^{(d)} + (1 - x_l) R_{Dln}^{(r)} \right]. \quad (7.10)$$

We consider the joint mode selection, resource assignment, and power allocation problem with the following constraints. First, the minimum data rates of D2D links and cellular links in all transmission intervals must be guaranteed, i.e.,

$$R1_{Ckl}^{(r)} \geq \frac{1}{2} \rho_{kl} x_l R_{Ck}^{\min}, \quad R2_{Ckl}^{(r)} \geq \frac{1}{2} \rho_{kl} x_l R_{Ck}^{\min} \quad (7.11)$$

$$R_{Ck} \geq R_{Ck}^{\min} \quad \forall k \in \mathcal{K}, \quad R_{Dl} \geq R_{Dl}^{\min}, \quad \forall l \in \mathcal{L}. \quad (7.12)$$

Second, the transmit power of individual users must be smaller than their maximum powers, i.e.,

$$m_k p_{Ck} \leq P_{Ck}^{\max} \quad \forall k \in \mathcal{K} \quad (7.13)$$

$$\sum_{n \in \mathcal{N}} \rho_{nl} m_n p_{Dln} \leq P_{Dl}^{\max} \quad \forall l \in \mathcal{L} \quad (7.14)$$

$$\sum_{n \in \mathcal{N}} (1 - x_l) m_n p_{R_l} \leq P_{R_l}^{\max} \quad \forall r_l \in \mathcal{R}. \quad (7.15)$$

Finally, the binary constraints of mode selection and RG assignment variables can be expressed as

$$x_l \in \{0, 1\} \forall l \in \mathcal{L}, \rho_{kl} \in \{0, 1\} \forall k \in \mathcal{N} \forall l \in \mathcal{L}. \quad (7.16)$$

Our design problem aims to maximize the sum rate of all communications links which can be stated as

$$\begin{aligned} & \max_{\mathbf{p}, \boldsymbol{\rho}, \mathbf{x}} \sum_{k \in \mathcal{K}} R_{Ck} + \sum_{l \in \mathcal{L}} R_{Dl} \\ & \text{s.t. constraints (7.11) – (7.16).} \end{aligned} \quad (7.17)$$

We assume that the base station (BS) is a central controller, which collects all the necessary channel state information (CSI), perform the proposed algorithm, and then broadcasts the results to the users in the system.

7.4 Solution Approach

In the following, we propose to solve problem (7.17) *optimally* through a solution approach with three phases, namely power allocation, mode selection, and resource group (RG) allocation. First, we solve the power allocation problem for each D2D link l in either relay or direct mode if it reuses the resource of cellular link k . Then, the mode selection is implemented to determine the optimal modes of D2D links. Finally, the original problem is transformed to RG assignment problem, which can be solved *optimally* by using the Hungarian method. We present this design in the following.

7.4.1 Power allocation

Assume that D2D link l reuses the resource of cellular link k , we need to solve two power allocation problems corresponding to the direct and relay mode of D2D link l . The power allocation problem as D2D link l operates in the direct mode can be solved by using the algorithm in [23]. On the other hand, if D2D link l operates in the relay mode, we denote $\mathbf{p}_{kl} = [p_{Ck}, p_{Dl}, p_{Rl}]$, $P_c^m \triangleq P_{Ck}^{\max}/m_k$, $P_d^m \triangleq P_{Dl}^{\max}/m_k$, $P_r^m \triangleq P_{Rl}^{\max}/m_k$. Then, we have the following power allocation problem, whose

the optimal solution is developed in Section 7.5.

$$\max_{\mathbf{p}_{kl}} w_{kl}^{(r)}(\mathbf{p}_{kl}) \triangleq R_{Ckl}^{(r)} + R_{Dlk}^{(r)} \quad (7.18a)$$

$$\text{s.t. } R1_{Ckl}^{(r)} \geq \frac{1}{2}R_{Ck}^{\min}, R2_{Ckl}^{(r)} \geq \frac{1}{2}R_{Ck}^{\min}, R_{Dlk}^{(r)} \geq R_{Dl}^{\min} \quad (7.18b)$$

$$p_{Ck} \in [0, P_c^m], p_{Dl} \in [0, P_d^m], p_{Rl} \in [0, P_r^m]. \quad (7.18c)$$

7.4.2 Joint Mode Selection and RG Allocation

Denote $w_{kl}^{(d)*}$ and $w_{kl}^{(r)*}$ as the optimal total rates when D2D link l reuses the resource of cellular link k in direct and relay modes, respectively. The optimal mode selection can be determined as follows. Assume that D2D link l is assigned RG k , if $w_{kl}^{(d)*} \geq w_{kl}^{(r)*}$ then D2D link l should operate in the direct mode, otherwise, it should operate in the relay mode. As D2D link l is assigned RG k , the rate increase due to D2D resource reuse is $w_{kl}^* = \max\{w_{kl}^{(d)*}, w_{kl}^{(r)*}\} - R_{Ck}^{(o)}$. Therefore, problem (7.17) can be transformed into the following problem, which can be solved optimally by the Hungarian method [24]

$$\max_{\rho} \sum_{l \in \mathcal{L}} \sum_{k \in \mathcal{N}} w_{kl}^* \rho_{kl} \quad (7.19a)$$

$$\text{s.t. } \sum_{k \in \mathcal{N}} \rho_{kl} = 1, \forall l \in \mathcal{L}, \sum_{l \in \mathcal{L}} \rho_{kl} \leq 1, \forall k \in \mathcal{N} \quad (7.19b)$$

$$\rho_{kl} \in \{0, 1\}, \forall l \in \mathcal{L}, k \in \mathcal{N}. \quad (7.19c)$$

The complexity of the proposed design is analyzed by studying three design phases. The power allocation phase requires to solve $2KL$ power allocation problems with complexity of $O(1)$. Therefore, it has the complexity of $O(KL)$. The mode selection phase has complexity $O(KL)$ and RG assignment phase requires to implement the Hungarian algorithm with complexity of $O(\max\{K^3, L^3\})$ [24]. Therefore, the total complexity of the proposed algorithm is $O(\max\{K^3, L^3\})$.

7.5 Power Allocation for Relay Mode

In this section, we develop an algorithm to solve problem (7.18). We characterize the optimal power allocation in the following proposition.

Proposition 7.1. *If problem (7.18) is feasible then at optimality at least one node (D2D transmitter, relay, or cellular user) uses maximum transmit power and $R1_{Dlk}^{(r)} = R2_{Dlk}^{(r)}$.*

Proof. If no node uses maximum transmit power at optimality, we can increase the powers of all nodes by a scaling factor $\alpha > 1$ until one node reaches its maximum transmit power. Hence, total data rate of cellular link k and D2D link l increases, which contradicts with the assumption. Therefore, at least one node uses maximum power at optimality.

At optimality of problem (7.18), if $R1_{Dlk}^{(r)} > R2_{Dlk}^{(r)}$, we can decrease p_{Dl} while maintaining $R1_{Dlk}^{(r)} \geq R2_{Dlk}^{(r)}$, which leads to the increase of $R1_{Ckl}^{(r)}$. Hence, the objective value of problem (7.18) can be improved. We can prove similarly if $R1_{Dlk}^{(r)} < R2_{Dlk}^{(r)}$. Therefore, $R1_{Dlk}^{(r)} = R2_{Dlk}^{(r)}$ at optimality. \square

As problem (7.18) involves the power allocation for a relay, a D2D transmitter, and a cellular user, we can reformulate it in the following concise form where for brevity the RG index “ k ” is omitted and the subscript symbols “ c ”, “ d ”, and “ r ” denote cellular user, D2D transmitter, and relay, respectively

$$\max_{(p_c, p_d, p_r)} \zeta \tag{7.20a}$$

$$\text{s.t. } \frac{p_d h_{rd}}{\sigma^2 + p_c h_{rc}} = \frac{p_r h_{dr}}{\sigma^2 + p_c h_{dc}} \tag{7.20b}$$

$$\frac{p_c h_{cc}}{\sigma^2 + p_d h_{cd}} \geq R_c^m, \frac{p_c h_{cc}}{\sigma^2 + p_r h_{cr}} \geq R_c^m, \frac{p_r h_{dr}}{\sigma^2 + p_c h_{dc}} \geq R_d^m \tag{7.20c}$$

$$p_c \in [0, P_c^m], p_d \in [0, P_d^m], p_r \in [0, P_r^m], \tag{7.20d}$$

where

$$\zeta = \left(1 + \frac{p_c h_{cc}}{\sigma^2 + p_d h_{cd}}\right) \left(1 + \frac{p_c h_{cc}}{\sigma^2 + p_r h_{cr}}\right) \left(1 + \frac{p_r h_{dr}}{\sigma^2 + p_c h_{dc}}\right). \tag{7.21}$$

From Proposition 7.1, it can be verified that problem (7.20) can achieve its optimum if $p_c=P_c^m$, $p_d=P_d^m$, or $p_r=P_r^m$. We consider these three cases as follows.

7.5.1 Case 1: $p_c = P_c^m$

We define $\zeta_c = \zeta|_{p_c=P_c^m}$, problem (7.20) is equivalent to

$$\max_{(p_d, p_r)} \zeta_c \quad (7.22a)$$

$$\text{s.t.} \quad \frac{p_d h_{rd}}{\sigma^2 + P_c^m h_{rc}} = \frac{p_r h_{dr}}{\sigma^2 + P_c^m h_{dc}} \quad (7.22b)$$

$$\frac{P_c^m h_{cc}}{\sigma^2 + p_d h_{cd}} \geq R_c^m \quad (7.22c)$$

$$\frac{P_c^m h_{cc}}{\sigma^2 + p_r h_{cr}} \geq R_c^m \quad (7.22d)$$

$$\frac{p_r h_{dr}}{\sigma^2 + P_c^m h_{dc}} \geq R_d^m \quad (7.22e)$$

$$p_d \in [0, P_d^m], p_r \in [0, P_r^m]. \quad (7.22f)$$

Constraint (7.22b) suggests that p_d can be expressed as a linear function of p_r . Therefore, the objective function in (7.22a) can be expressed as a function of p_r , which is the ratio between one cubic polynomial and one quadratic polynomial, as

$$\zeta_c \triangleq \frac{f_c(p_r)}{g_c(p_r)} \triangleq \frac{a_{c3} p_r^3 + a_{c2} p_r^2 + a_{c1} p_r + a_{c0}}{b_{c2} p_r^2 + b_{c1} p_r + b_{c0}}. \quad (7.23)$$

Moreover, constraints (7.22c), (7.22d), (7.22e), and (7.22f) imply that $p_r \in [P_r^{\min}, P_r^{\max}]$. Hence, problem (7.22) can be transformed to

$$\max_{p_r \in [P_r^{\min}, P_r^{\max}]} \zeta_c. \quad (7.24)$$

In fact, problem (7.24) is a fractional optimization problem which can be solved *optimally* by the Dinkelbach based method [27]. Specifically, to solve problem (7.24), we need to solve the following related problem for a given parameter η_c

$$\max_{p_r \in [P_r^{\min}, P_r^{\max}]} z_c(\eta_c, p_r) \triangleq f_c(p_r) - \eta_c g_c(p_r). \quad (7.25)$$

The optimal solution of problem (7.25) can be determined by searching exhaustively over all extreme and local optimum points of the power range. We describe the procedure to determine the set of extreme and local optimum points in the following. Since $z_c(\eta_c, p_r)$ is a cubic polynomial, $\partial z_c(\eta_c, p_r)/\partial p_r$ is a quadratic polynomial of the form

$$\partial z_c(\eta_c, p_r)/\partial p_r = c_{c2}p_r^2 + c_{c1}p_r + c_{c0}. \quad (7.26)$$

Suppose that $p_{r\eta_{c1}} \leq p_{r\eta_{c2}}$ are two roots of equation

$$\frac{\partial z_c(\eta_c, p_r)}{\partial p_r} = 0. \quad (7.27)$$

Therefore, the optimal solution of problem (7.25) belongs to the following set

$$\mathcal{S}_c(\eta_c) \triangleq \{p_{r\eta_{c1}}, p_{r\eta_{c2}}, P_r^{\min}, P_r^{\max}\} \cap [P_r^{\min}, P_r^{\max}]. \quad (7.28)$$

The optimal solution of problem (7.25) can be determined by comparing the objective values at the possible solutions in $\mathcal{S}_c(\eta_c)$. Let $p_{r\eta_c}^* = \operatorname{argmax}_{p_r \in \mathcal{S}_c(\eta_c)} z_c(\eta_c, p_r)$ be the optimal solution of problem (7.25). We state the property of the optimal solution obtained from problem (7.24) in the following proposition, which is adopted from [27].

Proposition 7.2. *If problem (7.24) is feasible, $z_c(\eta_c, p_{r\eta_c}^*)$ is a decreasing function of η_c . If $z_c(\eta_c^*, p_{r\eta_c^*}^*) = 0$ then $\eta_c^* = f_c(p_{r\eta_c^*}^c)/g_c(p_{r\eta_c^*}^c)$ is the optimal objective value of problem (7.24).*

In summary, we can obtain the optimal solution of problem (7.20) by using the following algorithm.

Algorithm 7.1. Algorithm to solve problem (7.24)

- 1: Initialization: Set $\epsilon = 10^{-6}$, $\eta_c = 0$, $\eta_{ct} = -\epsilon$.
 - 2: **while** $(\eta_c - \eta_{ct}) \geq \epsilon$ **do**
 - 3: Determine $\mathcal{S}_c(\eta_c)$ according to (7.28).
 - 4: Obtain $p_{r\eta_c}^* = \operatorname{argmax}_{p_r \in [P_r^{\min}, P_r^{\max}]} z_c(\eta_c, p_r)$ and $z_c(\eta_c, p_{r\eta_c}^*)$
 - 5: $\eta_{ct} = \eta_c$, $\eta_c = \frac{f_c(p_{r\eta_c}^*)}{g_c(p_{r\eta_c}^*)}$
 - 6: **end while**
 - 7: Output η_c and $p_{r\eta_c}^*$.
-

7.5.2 Case 2: $p_d = P_d^m$ or $p_r = P_r^m$

We can see that two problems (7.20) for the two cases of $p_d = P_d^m$ and $p_r = P_r^m$ are similar in the sense that the algorithm solving the prior problem can be used to solve the latter with different parameters. Therefore, we show how to solve problem (7.20) for the case $p_r = P_r^m$ only in the following. As $p_r = P_r^m$, problem (7.20) can be transformed into the following problem, in which $\zeta_r = \zeta|_{p_r=P_r^m}$.

$$\max_{(p_c, p_d)} \zeta_r \quad (7.29a)$$

$$\text{s.t.} \quad \frac{p_d h_{rd}}{\sigma^2 + p_c h_{rc}} = \frac{P_r^m h_{dr}}{\sigma^2 + p_c h_{dc}} \quad (7.29b)$$

$$\frac{p_c h_{cc}}{\sigma^2 + p_d h_{cd}} \geq R_c^m, \quad \frac{p_c h_{cc}}{\sigma^2 + P_r^m h_{cr}} \geq R_c^m, \quad \frac{P_r^m h_{dr}}{\sigma^2 + p_c h_{dc}} \geq R_d^m \quad (7.29c)$$

$$p_c \in [0, P_c^m], p_d \in [0, P_d^m]. \quad (7.29d)$$

In problem (7.29), the first constraint of (7.29b) is equivalent to

$$p_d = \frac{P_r^m h_{dr} (\sigma^2 + p_c h_{rc})}{h_{rd} (\sigma^2 + p_c h_{dc})}. \quad (7.30)$$

Substituting this p_d to the first constraint of (7.29c) leads to the following: $d_{r_2} p_c^2 + d_{r_1} p_c + d_{r_0} \geq 0$, where $d_{r_2} > 0$, $d_{r_0} < 0$. This inequality corresponds to $p_c \in [p_{c_1}, +\infty)$, where p_{c_1} is the positive root of equation $d_{r_2} p_c^2 + d_{r_1} p_c + d_{r_0} = 0$.

In addition, the remaining constraints in (7.29c) and (7.29d) lead to $p_c \in [P_c^{\min}, P_c^{\max}]$. We define $\mathcal{D}_r \triangleq [\max\{P_c^{\min}, p_{c_1}\}, P_c^{\max}]$ as the feasible value set of p_c for problem (7.29). Substituting p_d as a function of p_c in (7.30) to the objective function of (7.29a), we can transform it to the form

$$\zeta_r \triangleq \frac{f_r(p_c)}{g_r(p_c)} \triangleq \frac{a_{r_4} p_c^4 + a_{r_3} p_c^3 + a_{r_2} p_c^2 + a_{r_1} p_c + a_{r_0}}{b_{r_2} p_c^2 + b_{r_1} p_c + b_{r_0}}, \quad (7.31)$$

which is a ratio between one quartic polynomial and one quadratic polynomial. Therefore, problem (7.29) can be transformed to

$$\max_{p_c \in \mathcal{D}_r} \zeta_r, \quad (7.32)$$

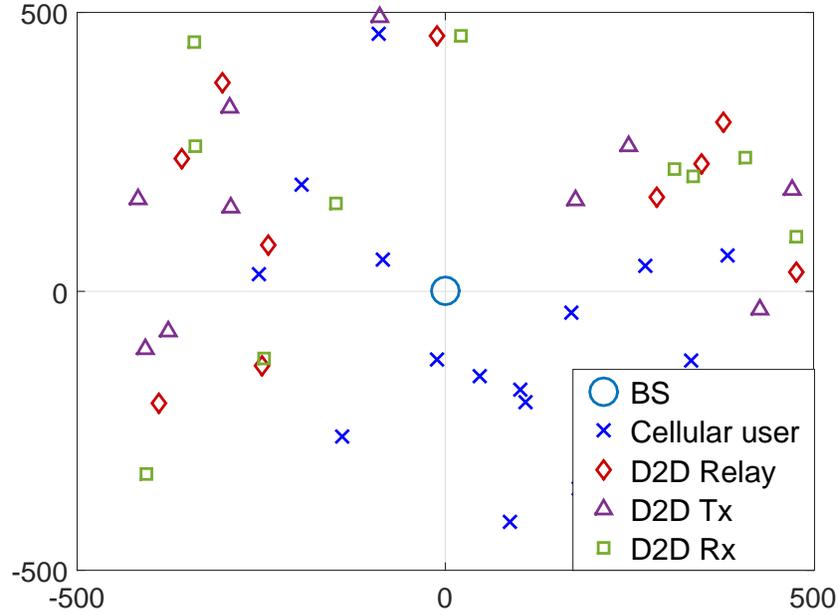


Figure 7.1 – System configuration

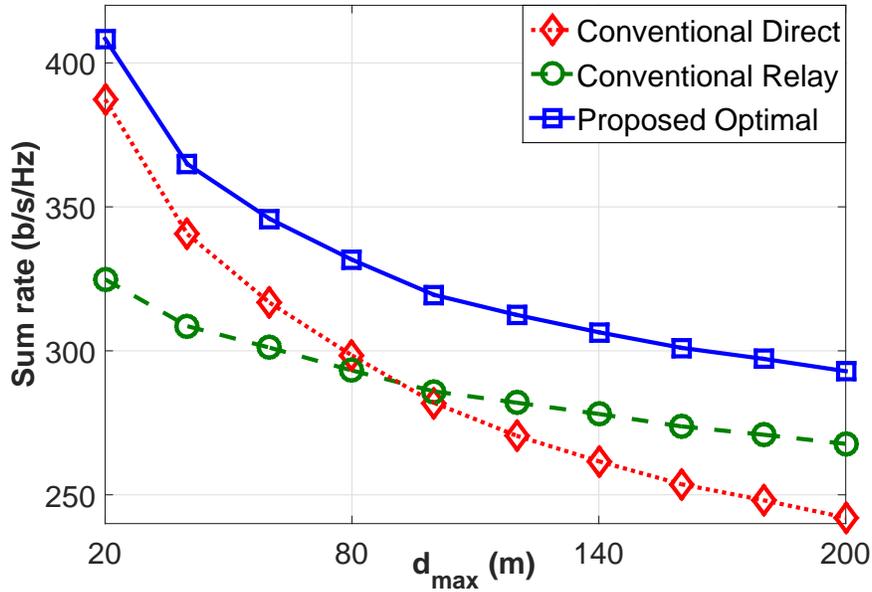
where the Dinkelbach method can be employed to solve it optimally. Nevertheless, similar to Algorithm 7.1, we need to solve the following related problem

$$\max_{p_c \in \mathcal{D}_r} z_r(\eta_r, p_c) \triangleq f_r(p_c) - \eta_r g_r(p_c). \quad (7.33)$$

It can be seen that $f_r(p_c)$ and $g_r(p_c)$ are, respectively, quartic and quadratic polynomials, $z_r(\eta_r, p_c)$ is a quartic polynomial of p_c . Hence, $\partial z_r(\eta_r, p_c)/\partial p_c$ is a cubic polynomial of the form: $\partial z_r(\eta_r, p_c)/\partial p_c = c_{r3}p_c^3 + c_{r2}p_c^2 + c_{r1}p_c + c_{r0}$, whose roots can be determined easily [126]. Assume $p_{cc1} \leq p_{cc2} \leq p_{cc3}$ are three roots of equation $\partial z_r(\eta_r, p_c)/\partial p_c = 0$ then the optimal solution of problem (7.33), $p_{c\eta_r}^*$ belongs to the set $\mathcal{S}_r(\eta_r) \triangleq \{P_c^{\min}, P_c^{\max}, p_{c1}, p_{cc1}, p_{cc2}, p_{cc3}\} \cap \mathcal{D}_r$. Hence, problem (7.29) can be solved optimally by similar procedure of Algorithm 7.1.

7.6 Numerical Results

We consider the system in Fig. 7.1 in our numerical studies where there are $L = 15$ D2D links reusing the resource of 20 cellular links, and each cellular link is allocated one subchannel, i.e., $m_k = 1 \forall k \in \mathcal{K}$. Cellular users and relays are randomly distributed in the cell area of radius 500m.

Figure 7.2 – Sum rate versus d_{\max}

Moreover, each D2D transmitter and receiver are located randomly whose distance to its relay varies within d_{\max} , where $d_{\max} = 100m$.

The channel power gain is modeled as $d^{-\alpha}\eta$ where d , α , η represent the distance, path-loss exponent, and the Rayleigh fading coefficient, respectively. We set $\alpha = 2.5$ for D-R and R-D communications, and $\alpha = 3.5$ for other links to demonstrate the relaying benefit in D2D communications. We set $\sigma^2 = 10^{-12}W$, $P_{Ck}^{\max} = P_{DI}^{\max} = 0.5W$, $R_c^{\min} = 2b/s/Hz$, and $R_d^{\min} = 5b/s/Hz$, respectively. Finally, our proposed design, which is denoted as “Proposed Optimal”, is compared with two existing schemes developed in [23] and [25] denoted as “Conventional Direct” and “Conventional Relay”, respectively.

Fig. 7.2 presents the system sum rate versus d_{\max} . As d_{\max} is small, our proposed scheme performs similarly to the “Conventional Direct” scheme and significantly outperforms the “Conventional Relay” scheme. This is because as d_{\max} is small, the optimal D2D mode is usually the direct mode. However, as d_{\max} increases, D2D links tend to operate in the relay mode more frequently since the relay D2D mode can outperform the direct D2D mode. As a result, the proposed design performs much better than the other two existing schemes thanks to the benefits of adaptively switching between the direct D2D and relay D2D modes.

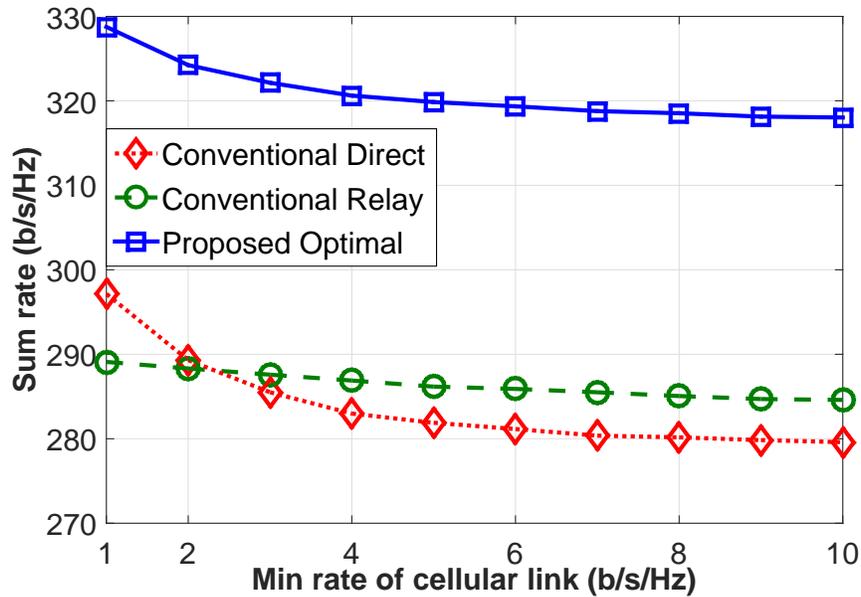


Figure 7.3 – Sum rate versus R_c^{\min}

Fig. 7.3 shows the system sum rate versus R_c^{\min} . As R_c^{\min} increases, the system sum rate decreases moderately. This is because as R_c^{\min} becomes higher, cellular links must increase their transmit powers to meet the data rate requirements. Moreover, with optimal RG assignment, D2D links which suffer from low co-channel interference from certain cellular links tend to reuse the frequency resources of these cellular links. Thus, the higher transmit powers of cellular links would not degrade the data rates of D2D links significantly. As a result, the system sum rate decreases gradually as R_c^{\min} increases.

Fig. 7.4 illustrates the variations of system sum rate with the number of D2D links L . As L increases, the system sum rates of all the schemes increase sharply. In addition, as $d_{\max} = 100m$, the sum rates achieved by the conventional relay D2D and direct D2D schemes are similar. However, our scheme performs significantly better than the other two schemes. This demonstrates the significance of optimizing the communications mode in relay-based D2D communication.

7.7 Conclusion

In this paper, we have proposed the joint optimal mode selection, RG assignment and power allocation design for D2D underlying cellular communication. Specifically, we have characterized the

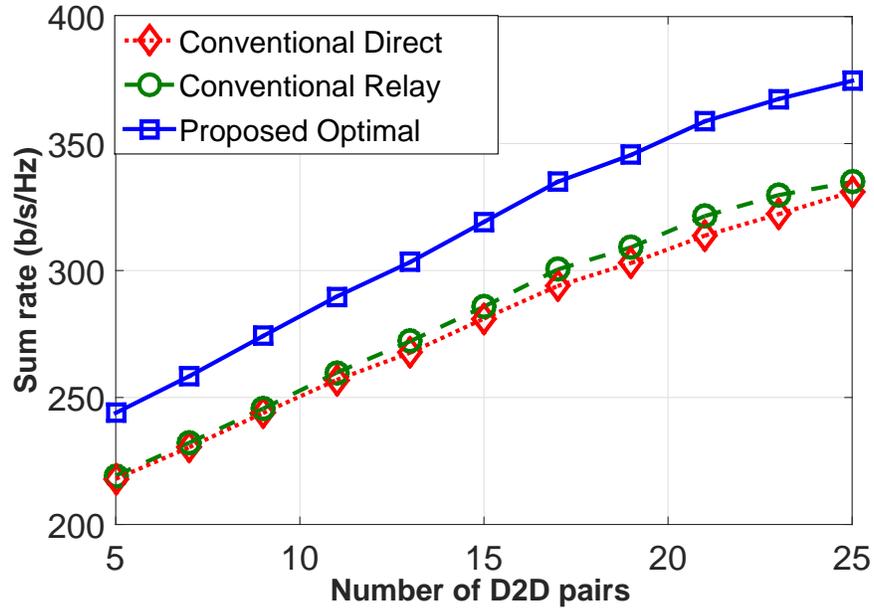


Figure 7.4 – Sum rate versus L

optimal power allocation for given mode and RG assignment solution based on which the optimal mode selection and RG assignment can be determined. Numerical studies have demonstrated that the proposed optimal design can improve the system sum rate dramatically compared to the existing D2D communication schemes.

Chapter 8

Joint Prioritized Scheduling and Resource Allocation for OFDMA-based Wireless Networks

The content of this chapter was submitted in IEEE Transactions on Wireless Communication.

8.1 Abstract

In this chapter, we study the joint prioritized link scheduling and resource allocation for OFDMA-based wireless networks which serve two classes of wireless links, namely non-prioritized (low-priority) and prioritized (high-priority) links. Our design aims to maximize the number of scheduled non-prioritized links and their sum rate while guaranteeing the minimum required rates of all active prioritized and non-prioritized links. We formulate this design as a single-stage optimization problem which simultaneously maximize number of scheduled non-prioritize links and their sum rate. It is proved that the formulated problem is NP-hard in general. We propose a Monotonic Based Optimal Approaching (MBOA) algorithm, which addresses the optimization problem by using the monotonic optimization technique and efficient rounding procedure. We prove that with slight and controllable degradation in the minimum required rates of non-prioritized links, the MBOA algorithm can schedule the maximum number of non-prioritized links. For low-complexity design,

we propose an iterative convex approximation (ICA) algorithm which sequentially performs power allocation and link removal in each iteration. We describe how the proposed algorithms can be implemented in the current cellular network standard. We conduct numerical studies considering device-to-device (D2D) communications underlaid cellular networks. Numerical results demonstrate that the proposed algorithms is applicable in the imperfect channel state information (CSI) scenarios with slight degradation in the performance. Moreover, in the perfect CSI scenarios, the proposed algorithms significantly outperform the conventional algorithms both in the number of scheduled non-prioritized links and their sum rate.

8.2 Introduction

Wireless cellular networks have seen rapidly increasing mobile traffic and growing number of connected devices. As a result, the future wireless system is expected to deliver much higher capacity and provide more efficient wireless connection management [127]. Toward this end, heterogeneous networks which support the large number of wireless connections with diverse quality of service (QoS) requirements play an important role in achieving these objectives [128]. Since radio resources are limited, joint link scheduling and resource allocation design, which determines the set of scheduled links with their allocated resources while guaranteeing their required QoS, is an important research issue. Such design enables us to achieve high system performance and required users' QoS.

In many practical network and application settings, wireless links can have different access priorities and QoS requirements. For example, in cognitive radio networks (CRNs), primary users are usually granted strictly higher access priority to the spectrum compared to secondary users. In particular, primary users should be protected from the co-channel interference due to the transmissions of primary users. Similarly, in Device-to-Device (D2D) communications underlaid cellular networks, cellular links are typically prioritized over D2D links where the cellular links must be appropriately protected from the spectrum reuse of D2D links [129]. Therefore, it is desirable that the scheduling and resource allocation design considers differentiated link priorities and efficiently manage the co-channel interference among the links.

Wireless scheduling and resource allocation have been studied extensively over the past decades. However, most existing works consider simple settings, e.g., all wireless links share a single channel

[49] while for the multiple-channel setting, existing works typically assume that each wireless link can be assigned at most one channel [50]. Moreover, scheduling and resource allocation issues have been frequently treated separately [51]. Furthermore, for the multi-channel system, differentiated access priorities of different wireless links are often ignored in the joint scheduling and resource allocation design [52]. Finally, the combinatorial nature of the scheduling problem renders the design of an optimal algorithm very challenging. To the best of our knowledge, there is no existing work proposing a joint optimal scheduling and resource allocation algorithm even in the single-channel system. Our current work aims to fill this gap in the current literature on this joint design considering multiple channels and differentiated users' QoSs.

8.2.1 Related Works

Joint power and admission control considering co-channel interference among simultaneous transmissions on a single channel has been studied by many researchers [28, 48, 49]. In [28], J. Zander *et al.* propose an iterative algorithm which sequentially performs power allocation and greedy link removal until the minimum required rates of active links are satisfied. The authors in [48, 49] consider the problem which simultaneously maximizes the number of scheduled users and minimizes the power consumption. Both [49] and [48] propose linear programming (LP) deflation algorithms which solve an LP approximation of the original problem whose solution is used to remove the “worst” link in each iteration. Although the above works study the joint admission control and resource allocation, differentiated priorities among the links are not considered and addressing a single-channel problem is obviously easier than dealing with the multiple-channel one.

Joint admission control and resource allocation has also been investigated for cognitive radio networks (CRN), where primary users (PU) are granted higher priority in spectrum access compared to secondary users (SU) [50–52]. In particular, the downlink scheduling, channel assignment, and power control problem for multi-channel CRNs is addressed in [50] where the authors propose a dynamic interference graph algorithm (DIGA) which greedily assigns a channel to the best SU in each assignment step. As an extension of [50], joint channel assignment and power control for both uplink and downlink scheduling in multi-channel CRNs is addressed in [51] where the authors propose to transform the formulated problem into a maximal weighted bipartite matching problem and develop a greedy algorithm to solve it. Nguyen *et al.* in [52] tackle the link scheduling and

resource allocation problem in CRNs by applying the coloring approach to the interference graph, which is shown to be more efficient than the algorithms in [51]. However, all above works assume that each SU can reuse only one channel from certain PU, which may limit their applicability and achievable performance.

Joint scheduling and resource allocation is also an important problem in cellular networks [13, 130]. In [130], Bashar *et al.* study a joint admission control and resource allocation problem for a heterogeneous OFDMA network which simultaneously maximizes the number of admitted high-priority users and the utility of best-effort users. Nevertheless, the authors assume that users must exploit the spectral resources orthogonally, which would reduce the system spectrum efficiency. Abdelnasser *et al.* in [13] investigate the joint admission control and resource allocation for femtocell underlaid cellular networks by proposing a two-step algorithm. In particular, the resource allocation for macro users is performed in the first step while admission control and resource allocation for femtocell users is solved in the second step. Again, this work makes an assumption that each cellular link can be assigned at most one subchannel.

Recently, D2D communication in cellular networks has emerged as an important technology for network capacity enhancement through exploiting short-range D2D communications. Joint scheduling and resource allocation for D2D-enabled communication systems are investigated in [17, 23, 53, 54]. Although these works consider both scheduling and resource allocation, scheduling design to maximize the number of admitted D2D links is not their primary design objective; instead, the scheduling result is simply the bi-product of the studied spectrum-efficiency or energy-efficiency optimization problems. Nevertheless, the subchannel assignment for the cellular links is accounted for in these works, which permit each D2D links to exploit the resource of only one cellular link.

It can be observed that none of the existing works, which are discussed above, addresses all key design aspects for the general multi-channel wireless system, i.e., the joint admission control and resource allocation with different access priorities where each wireless link can exploit multiple channels. In fact, consideration of all these aspects requires to tackle the challenging interference management problem among co-channel links with QoS constraints. Our current work fills this gap in the existing literature. Some preliminary results of this work have been published in [131].

8.2.2 Contributions and Novelty of the Current Work

In this chapter, we study the joint prioritized scheduling, subchannel assignment, and power allocation problem for multiple wireless links in the OFDMA-based wireless networks, which simultaneously (i) maximizes the number of scheduled non-prioritized links and (ii) maximizes their sum rate subject to the minimum rate requirements of prioritized and scheduled non-prioritized links. In particular, our work makes the following novel contributions.

- The scheduling and resource allocation design is formulated as a single-stage optimization problem considering QoS constraints of the prioritized and scheduled non-prioritized links. We prove that the formulated optimization problem can simultaneously maximize the number of scheduled non-prioritized links and their sum rate. For mathematical convenience, our model allows each prioritized link to utilize all subchannels; however, by introducing sufficiently high virtual interfering channel gains among the prioritized links, the prioritized links exploit the subchannels orthogonally. Therefore, the underlying joint scheduling, subchannel assignment and power allocation problem can be transformed into an equivalent joint scheduling and power allocation problem, where the subchannel assignments can be determined from the power allocation solution.
- We develop a monotonic based optimal approaching (MBOA) algorithm to solve the above problem which asymptotically¹ achieves the optimal set of scheduled non-prioritized links and their maximum sum rate. Specifically, we employ an *approximated step function* to capture the scheduling status of each non-prioritized link. Based on this function and some changes of variables, we show that the joint scheduling and resource allocation problem can be transformed into a monotonic optimization problem which can be solved efficiently by using the *polyblock approximation* algorithm. Finally, we prove that the set of scheduled non-prioritized links obtained from the MBOA algorithm is optimal with slight and controllable degradation of the minimum required rates of non-prioritized links.
- We propose another low-complexity iterative convex approximation (ICA) algorithm which sequentially performs power allocation and link removal in each iteration. Moreover, in the power allocation procedure, we propose a novel method to solve the problem by using the first Taylor approximation technique and the so-called DC optimization technique.

¹The notion of *asymptotic optimality* will be formalized in more details later.

- We describe how the proposed algorithms can be implemented to realize the studied joint prioritized scheduling, subchannel assignment and power allocation in the standardized LTE-based cellular system. Specifically, we discuss the responsibilities and operations of the user equipments (UE) and BS in the channel estimation and feedback procedure to obtain the necessary CSI at the BS for execution of the proposed algorithms. We also propose a conservative design in which the BS executes the proposed algorithms using the imperfect CSI.
- We evaluate the performance of the proposed algorithms for D2D-based cellular networks via extensive numerical studies, where the cellular and D2D links are prioritized and non-prioritized links, respectively. Numerical results show that the conservative design using the imperfect CSI results in negligible performance loss. For the perfect CSI scenarios, numerical studies demonstrate that the MBOA algorithm performs the best among all algorithms while the ICA algorithm also achieves very good performance with manageable computation complexity. Specifically, the number of scheduled D2D links and the sum rate obtained from the ICA algorithm are very close to those due to the MBOA algorithm and dramatically higher than those due to conventional algorithms.

8.3 System Model

We consider uplink communications²communications in a single-cell wireless system where K prioritized wireless links in the set $\mathcal{K} = \{1, \dots, K\}$ share the same spectrum comprising N orthogonal subchannels in the set $\mathcal{N} = \{1, \dots, N\}$ with L non-prioritized wireless links in the set $\mathcal{L} = \{K + 1, \dots, K + L\}$. Let $\mathcal{M} = \mathcal{K} \cup \mathcal{L}$ denote the set of all the links. In these notations, $K = |\mathcal{K}|$, $L = |\mathcal{L}|$, $M = |\mathcal{M}|$ and $N = |\mathcal{N}|$ denote the numbers of prioritized links, non-prioritized links, all the links, and subchannels in the system, respectively, where $|A|$ denotes the cardinality of set A . Let h_{kl}^n be the channel gain from the transmitter of link l to the receiver of link k on subchannel n . We denote p_m^n as the transmit power of link $m \in \mathcal{M}$ on subchannel n and we represent the transmit power vector of all links in the system as $\mathbf{p} = [\mathbf{p}_m]_{\forall m \in \mathcal{M}}$ where $\mathbf{p}_m = [p_m^n]_{\forall n \in \mathcal{N}}$ is the power allocation vector of link $m \in \mathcal{M}$ over the subchannels.

²This is motivated by the fact that resources in the uplink direction are not fully exploited as compared to the downlink direction. The proposed design is, however, also applicable to the downlink communications.

We assume that the prioritized links utilize subchannels orthogonally; however, the non-prioritized links are allowed to reuse all subchannels to improve the spectrum efficiency through exploiting the spatial diversity. The assignment of subchannel n to prioritized link $k \in \mathcal{K}$ is represented by a binary variable ρ_k^n , where $\rho_k^n = 1$ if subchannel n is assigned to link k , and $\rho_k^n = 0$, otherwise. We define the following subchannel assignment vectors $\boldsymbol{\rho}^n = [\rho_k^n]_{\forall k \in \mathcal{K}}$, and $\boldsymbol{\rho} = [\boldsymbol{\rho}^n]_{\forall n \in \mathcal{N}}$. Then, the signal to interference plus noise ratio (SINR) achieved by prioritized link k on subchannel $n \in \mathcal{N}$ can be expressed as³

$$\Gamma_k^n(\boldsymbol{\rho}, \mathbf{p}) = \frac{p_k^n h_{kk}^n}{\sigma_k^n + \sum_{l \in \mathcal{L}} p_l^n h_{kl}^n}, \quad (8.1)$$

where $\sum_{l \in \mathcal{L}} p_l^n h_{kl}^n$ represents the interference due to the non-prioritized links using subchannel n and σ_k^n denotes the noise power on subchannel n . Similarly, the SINR of non-prioritized link l on subchannel n can be written as

$$\Gamma_l^n(\boldsymbol{\rho}, \mathbf{p}) = \frac{p_l^n h_{ll}^n}{\sigma_l^n + \sum_{k \in \mathcal{K}} \rho_k^n p_k^n h_{lk}^n + \sum_{l' \in \mathcal{L} \setminus l} p_{l'}^n h_{ll'}^n}, \quad (8.2)$$

where $\sum_{k \in \mathcal{K}} \rho_k^n p_k^n h_{lk}^n$ and $\sum_{l' \in \mathcal{L} \setminus l} p_{l'}^n h_{ll'}^n$ are the interference from prioritized links and other non-prioritized links on subchannel n , respectively. The data rates in bits/s/Hz (i.e., normalized by the subchannel bandwidth) of prioritized link $k \in \mathcal{K}$ and non-prioritized link $l \in \mathcal{L}$ can be expressed, respectively, as

$$R_k(\boldsymbol{\rho}, \mathbf{p}) = \sum_{n \in \mathcal{N}} \rho_k^n \log_2(1 + \Gamma_k^n(\mathbf{p})) \quad (8.3)$$

$$R_l(\boldsymbol{\rho}, \mathbf{p}) = \sum_{n \in \mathcal{N}} \log_2(1 + \Gamma_l^n(\mathbf{p})). \quad (8.4)$$

8.4 Problem Formulation and Transformation

8.4.1 Problem Formulation

We would like to design the joint link scheduling, subchannel assignment, and power control for all the links considering the following design objectives and constraints: (i) the minimum required

³We assume that the transmit power of a particular prioritized link on a subchannel is zero if this subchannel is not assigned to the link.

data rates of prioritized links must be maintained, (ii) the number of scheduled non-prioritized links is maximized, and (iii) for the given set of scheduled non-prioritized links, the sum rate of the scheduled non-prioritized links is maximized. We consider the centralized design where the CSI of all links is available for the optimization. To capture the scheduling decision, we introduce a binary link scheduling vector $\mathbf{s}_{\mathcal{L}} = [s_1, \dots, s_L]^T$, where $s_l = 1$ if the non-prioritized link $l \in \mathcal{L}$ is scheduled and $s_l = 0$, otherwise. Therefore, we have the following optimization problem

$$\max_{\boldsymbol{\rho}, \mathbf{p}, \mathbf{s}_{\mathcal{L}}} \alpha \sum_{l \in \mathcal{L}} s_l + \sum_{l \in \mathcal{L}} R_l(\boldsymbol{\rho}, \mathbf{p}) \quad (8.5a)$$

$$\text{s.t.} \quad R_k(\boldsymbol{\rho}, \mathbf{p}) \geq R_k^{\min} \quad \forall k \in \mathcal{K} \quad (8.5b)$$

$$R_l(\boldsymbol{\rho}, \mathbf{p}) \geq s_l R_l^{\min} \quad \forall l \in \mathcal{L} \quad (8.5c)$$

$$\sum_{n \in \mathcal{N}} \rho_k^n p_k^n \leq P_{\max} \quad \forall k \in \mathcal{K} \quad (8.5d)$$

$$\sum_{n \in \mathcal{N}} p_l^n \leq P_{\max} \quad \forall l \in \mathcal{L} \quad (8.5e)$$

$$\sum_{k \in \mathcal{K}} \rho_k^n \leq 1 \quad \forall n \in \mathcal{N} \quad (8.5f)$$

$$\rho_k^n \in \{0, 1\} \quad \forall k \in \mathcal{K} \quad \forall n \in \mathcal{N} \quad (8.5g)$$

$$s_l \in \{0, 1\} \quad \forall l \in \mathcal{L}, \quad (8.5h)$$

where α denotes a sufficiently large number.⁴ In this problem, constraint (8.5c) requires that the minimum rate of each scheduled non-prioritized link l be satisfied. We characterize the property of this problem in the following proposition, whose the proof is given in Appendix 8.10.1.

Proposition 8.1. *By choosing $\alpha \geq LR$, where R is the maximum data rate that can be achieved by any non-prioritized link⁵, the optimal solution of problem (8.5) simultaneously maximizes the number of scheduled non-prioritized links and their sum rate.*

⁴In this formulation, we only consider individual power constraints. If certain links in \mathcal{M} correspond to downlink communications then we have to impose their total power constraint and the proposed design is still applicable.

⁵The maximum rate R can be calculated by first calculating the maximum rates of all individual links then searching for the maximum rate. Moreover, the maximum rate of a particular link can be calculated by assuming that it is allocated all subchannels and there is no co-channel interference and then using the optimal power allocation solution to calculate its maximum rate.

8.4.2 Problem Transformation

In the following, we show how to transform the joint scheduling, subchannel assignment, and power allocation problem into a joint scheduling and power allocation problem, which enables us to deal with the optimization over the binary subchannel assignment vector $\boldsymbol{\rho}$ effectively.

It was proved in [26] that if the mutual interference between two interfering links is strong enough, they should utilize the spectrum orthogonally to maximize the sum rate. Motivated by this result, the joint scheduling, subchannel assignment, and power allocation problem can be transformed into the joint scheduling and power allocation problem as follows. We allow each prioritized link to exploit all subchannels mathematically; however, the *virtual* channel gains⁶ among the prioritized links are set very high. Such setting of the high interfering channel gains will indeed force prioritized links to use the subchannels orthogonally to avoid strong co-channel interference. Specifically, by setting the channel gains among prioritized links to a sufficiently large value η , the SINR of prioritized link k on subchannel n can be expressed as

$$\bar{\Gamma}_k^n(\mathbf{p}) = \frac{p_k^n h_{kk}^n}{\sigma_k^n + \sum_{k' \in \mathcal{K} \setminus k} p_{k'}^n \eta + \sum_{l \in \mathcal{L}} p_l^n h_{kl}^n}, \quad (8.6)$$

where $\sum_{k' \in \mathcal{K} \setminus k} p_{k'}^n \eta$ represents the “virtual interference” due to other prioritized links using subchannel n . Similarly, the SINR of non-prioritized link l on subchannel n can be written as

$$\bar{\Gamma}_l^n(\mathbf{p}) = \frac{p_l^n h_{ll}^n}{\sigma_l^n + \sum_{k \in \mathcal{K}} p_k^n h_{lk}^n + \sum_{l' \in \mathcal{L} \setminus l} p_{l'}^n h_{ll'}^n}. \quad (8.7)$$

The data rates of prioritized link $k \in \mathcal{K}$ and non-prioritized link $l \in \mathcal{L}$ can be re-expressed, respectively, as

$$\bar{R}_k(\mathbf{p}) = \sum_{n \in \mathcal{N}} \log_2 \left(1 + \bar{\Gamma}_k^n(\mathbf{p}) \right) \quad (8.8)$$

$$\bar{R}_l(\boldsymbol{\rho}, \mathbf{p}) = \sum_{n \in \mathcal{N}} \log_2 \left(1 + \bar{\Gamma}_l^n(\mathbf{p}) \right). \quad (8.9)$$

⁶These virtual channel gains are not actual channel gains but they are only used for our design optimization purposes.

Therefore, problem (8.5) can be reformulated as follows:

$$\max_{\mathbf{p}, \mathbf{s}_{\mathcal{L}}} \alpha \sum_{l \in \mathcal{L}} s_l + \sum_{l \in \mathcal{L}} \bar{R}_l(\mathbf{p}) \quad (8.10a)$$

$$\text{s.t.} \quad \bar{R}_k(\mathbf{p}) \geq R_k^{\min} \quad \forall k \in \mathcal{K} \quad (8.10b)$$

$$\bar{R}_l(\mathbf{p}) \geq s_l R_l^{\min} \quad \forall l \in \mathcal{L} \quad (8.10c)$$

$$\sum_{n \in \mathcal{N}} p_k^n \leq P_{\max} \quad \forall k \in \mathcal{K} \quad (8.10d)$$

$$\sum_{n \in \mathcal{N}} p_l^n \leq P_{\max} \quad \forall l \in \mathcal{L} \quad (8.10e)$$

$$s_l \in \{0, 1\} \quad \forall l \in \mathcal{L}. \quad (8.10f)$$

The relationship between problem (8.5) and problem (8.10) is characterized in the following proposition, whose the proof is given in Appendix 8.10.2.

Proposition 8.2. *If the optimal solution of problem (8.10) satisfies the condition that prioritized links use the subchannels orthogonally, (i.e., each subchannel is assigned to at most one prioritized links), then the optimal solution of problem (8.10) is also the optimal solution of problem (8.5).*

Remark 8.1. *In general, we can set the value of η sufficiently large so that the prioritized links have no incentive to exploit the same channels. Therefore, they should use the subchannels orthogonally, i.e., each subchannel is assigned to at most one prioritized link.*

Proposition 8.2 implies that the optimum solution of (8.5) can be obtained by solving problem (8.10). Moreover, the power allocation solution obtained from problem (8.10) also determines the subchannel assignments for prioritized links as follows. If $p_k^n = 0$, the corresponding subchannel assignment $\rho_k^n = 0$, otherwise if $p_k^n > 0$ then $\rho_k^n = 1$. We develop algorithms to solve problem (8.10) in the following sections.

8.5 Monotonic Based Optimal Approaching (MBOA) Algorithm

Note that the binary nature of scheduling vector $\mathbf{s}_{\mathcal{L}}$ makes problem (8.10) difficult to solve. To overcome this challenge, we approximate a discrete variable s_l by a continuous function $q(s_l) = \frac{e^{Qs_l} - 1}{e^Q - 1}$ [132]. The approximation function has the following properties: (i) it is a continuous and

increasing function, (ii) $q(0) = 0$, (iii) and $q(1) = 1$. Moreover, as demonstrated in Fig. 8.1, the curve of $q(s_l)$ function approaches that of the step function $\delta(s_l)$ more closely as Q increases. Note that the step function $\delta(s_l)$ gives the exact representation for s_l . However, this step function is not a continuous function; therefore, it is difficult to deal with in the optimization. Using the

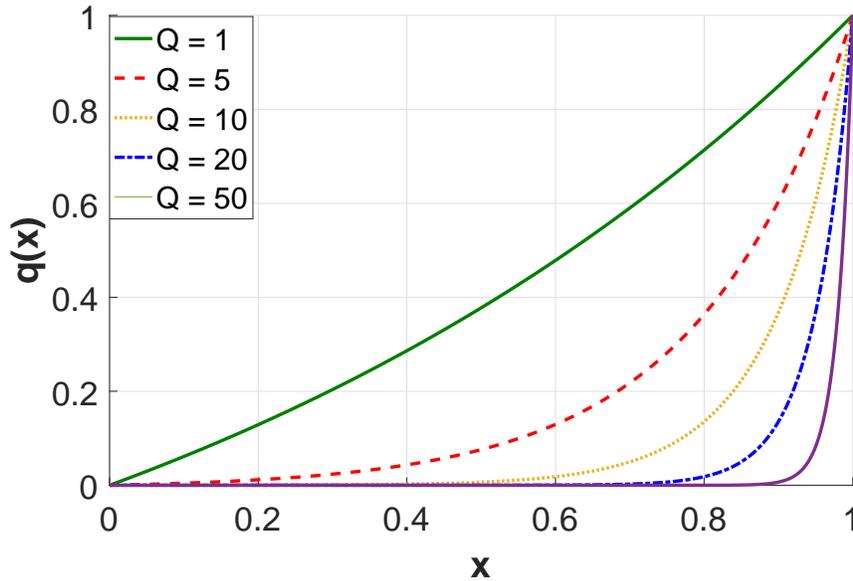


Figure 8.1 – $q(x)$ function with different values of Q

approximation function $q(\cdot)$, we arrive at the following problem

$$\begin{aligned}
 \max_{\mathbf{p}, \mathbf{s}_{\mathcal{L}}} \quad & \alpha \sum_{l \in \mathcal{L}} q(s_l) + \sum_{l \in \mathcal{L}} \bar{R}_l(\mathbf{p}) \\
 \text{s.t.} \quad & (8.10b), (8.10c), (8.10d), 8.10e \\
 & s_l \in [0, 1] \quad \forall l \in \mathcal{L}.
 \end{aligned} \tag{8.11}$$

Note that problem (8.11) is a nonconvex problem, which is, thus, difficult to solve. Fortunately, it has the hidden monotonic property. Specifically, by applying suitable changes of variables, we can transform this problem into a monotonic optimization problem so that the *polyblock approximation* technique can be employed to solve it optimally [56–58]. Toward this end, we describe some mathematical preliminaries of monotonic optimization [57], which enable us to concretely present the proposed algorithm.

8.5.1 Introduction to Monotonic Optimization

Definition 8.1 (Vector). For any two vectors $\mathbf{x}, \mathbf{y} \in R^M$, we write $\mathbf{x} \succeq \mathbf{y}$ and say that \mathbf{x} dominates \mathbf{y} if $x_i \geq y_i, \forall i = 1, \dots, M$, where x_i and y_i are the i^{th} dimension of \mathbf{x} and \mathbf{y} , respectively.

Definition 8.2 (Box). For any vertex $\mathbf{b} \in R_+^M$, the hyper rectangle $[\mathbf{0}, \mathbf{b}] = \{\mathbf{x} \in R_+^M \mid \mathbf{0} \preceq \mathbf{x} \preceq \mathbf{b}\}$ is referred as a box with vertex \mathbf{b} where R_+^M denotes the M -dimensional non-negative real domain.

Definition 8.3 (Normal set). A set $\mathcal{G} \subset R_+^M$ is called normal if, for any two points $\mathbf{x}, \mathbf{x}' \in R_+^M$ such that $\mathbf{x}' \preceq \mathbf{x}$, if $\mathbf{x} \in \mathcal{G}$, then $\mathbf{x}' \in \mathcal{G}$ too.

Definition 8.4 (Reverse normal set). A set $\mathcal{H} \subset R_+^M$ is reverse normal in box $[\mathbf{0}, \mathbf{b}]$ if $\mathbf{b} \succeq \mathbf{x}' \succeq \mathbf{x} \succeq \mathbf{0}$, then $\mathbf{x} \in \mathcal{H}$ implies $\mathbf{x}' \in \mathcal{H}$.

Definition 8.5 (Polyblock). Given any finite set of vertexes $\mathcal{V} \in R_+^M$, the union of all boxes $[\mathbf{0}, \mathbf{x}]$, $\mathbf{x} \in \mathcal{V}$, is a polyblock with vertex set \mathcal{V} .

Definition 8.6 (Monotonic Optimization). A canonical monotonic optimization problem has the following form:

$$\begin{aligned} \max_{\mathbf{x} \succeq \mathbf{0}} f(\mathbf{x}) \\ \text{s.t. } \mathbf{x} \in \mathcal{G} \cap \mathcal{H}, \end{aligned} \tag{8.12}$$

where \mathcal{G} and \mathcal{H} are nonempty normal and closed reverse normal sets, respectively, and $f(\mathbf{x})$ is an increasing function.

Proposition 8.3. If $g(\mathbf{x})$ and $h(\mathbf{x})$ are increasing functions, then \mathcal{G} and \mathcal{H} are normal and reverse normal sets, respectively, where

$$\mathcal{G} = \{\mathbf{x} \in R_+^M \mid g(\mathbf{x}) \leq 0\} \tag{8.13}$$

$$\mathcal{H} = \{\mathbf{x} \in R_+^M \mid h(\mathbf{x}) \geq 0\}. \tag{8.14}$$

Proposition 8.4. The intersection and union of normal sets are still normal sets.

Note that the proof of Proposition 8.3 and 8.4 are described in [56].

8.5.2 Problem Transformation

We now show that problem (8.11) can be transformed into a monotonic optimization problem. It can be seen that the objective function of problem (8.11) is increasing in $\mathbf{s}_{\mathcal{L}}$; however, it is non-increasing in \mathbf{p} . We define new variables $z_{mn} = \bar{\Gamma}_m^n(\mathbf{p}) \forall m \in \mathcal{M}, \forall n \in \mathcal{N}$, which denote the achievable SINR of link $m \in \mathcal{M}$ on subchannel $n \in \mathcal{N}$. We also define $\mathbf{z} = [z_{mn}]_{\forall m \in \mathcal{M}, \forall n \in \mathcal{N}}$.

To transform problem (8.11) into a monotonic optimization problem, we need to convert all the constraints to the forms in (8.13) and (8.14). It can be observed that constraint (8.10c) has the following form $\sum_{n \in \mathcal{N}} z_{ln} - s_l R_l^{\min} \geq 0 \forall l \in \mathcal{L}$, where the function in the left hand side is increasing in \mathbf{z} and decreasing in $\mathbf{s}_{\mathcal{L}}$. We introduce an auxiliary vector $\mathbf{t} = [t_l]_{\forall l \in \mathcal{L}}$ and $t_l \in [0, R_l^{\min}]$ then constraint (8.10c) for each non-prioritized link $l \in \mathcal{L}$ becomes equivalent to following constraints

$$\sum_{n \in \mathcal{N}} \log_2(1 + z_{ln}) + t_l \geq R_l^{\min} \quad (8.15)$$

$$t_l + s_l R_l^{\min} \leq R_l^{\min} \quad (8.16)$$

$$t_l \in [0, R_l^{\min}]. \quad (8.17)$$

It is obvious that each constraint above has the forms given in (8.13) or (8.14). Denote $\mathbf{x} = (\mathbf{t}, \mathbf{s}_{\mathcal{L}}, \mathbf{z})$ as the optimization vector which has $D = 2L + (K + L)N$ dimensions and $\mathcal{P} \triangleq \{\mathbf{p} \mid \sum_{n \in \mathcal{N}} p_m^n \leq P_{\max} \forall m \in \mathcal{M}\}$. Then, problem (8.11) can be transformed into the following one

$$\max_{\mathbf{x} \succeq \mathbf{0}} f(\mathbf{x}) \triangleq \alpha \sum_{l \in \mathcal{L}} q(s_l) + \sum_{l \in \mathcal{L}} \sum_{n \in \mathcal{N}} \log_2(1 + z_{ln}) \quad (8.18a)$$

$$\text{s.t. } s_l \leq 1 \forall l \in \mathcal{L} \quad (8.18b)$$

$$z_{mn} \leq \bar{\Gamma}_m^n(\mathbf{p}) \quad \forall m \in \mathcal{M} \forall n \in \mathcal{N} \forall \mathbf{p} \in \mathcal{P} \quad (8.18c)$$

$$t_l + s_l R_l^{\min} - R_l^{\min} \leq 0 \forall l \in \mathcal{L} \quad (8.18d)$$

$$t_l \leq R_l^{\min} \forall l \in \mathcal{L} \quad (8.18e)$$

$$\sum_{n \in \mathcal{N}} \log_2(1 + z_{kn}) - R_k^{\min} \geq 0 \forall k \in \mathcal{K} \quad (8.18f)$$

$$\sum_{n \in \mathcal{N}} \log_2(1 + z_{ln}) + t_l - R_l^{\min} \geq 0 \forall l \in \mathcal{L}. \quad (8.18g)$$

We characterize the property of problem (8.18) in the following proposition whose proof is given in Appendix 8.10.3.

Proposition 8.5. *Problem (8.18) is a monotonic optimization problem.*

8.5.3 Algorithm Design

To facilitate the description of the proposed algorithm, let \mathcal{G}_f be the normal set of problem (8.18) which contains the set of vectors \mathbf{x} satisfying (8.18b)-(8.18e). We also denote \mathcal{H}_f as the reverse normal set of problem (8.18) which contains the set of vectors \mathbf{x} satisfying (8.18f) and (8.18g). Then, $\mathcal{X} = \mathcal{G}_f \cup \mathcal{H}_f$ is the feasible region of problem (8.18). Since problem (8.18) is a monotonic optimization problem, it can be solved optimally by *polyblock outer approximation* algorithm, which is presented in Algorithm 8.1.

Since (8.18) is a monotonic optimization problem, its optimal solution lies on the boundary of \mathcal{X} . Moreover, if the feasible region of a monotonic optimization problem is a polyblock, its optimal solution belongs to the set of vertexes corresponding to that polyblock [57], which can be determined easily by examining all these vertexes. Specifically, to find the optimal solution \mathbf{x}^* of problem (8.18), Algorithm 8.1 iteratively approximates the feasible set \mathcal{X} by a polyblock containing \mathcal{X} where the size of each polyblock is reduced over iterations. In each iteration, it determines the upper-bound and lower-bound, where the upper-bound is the highest value of $f(\mathbf{x})$ over the approximated polyblock and the lower-bound is the highest feasible solution obtained in previous iterations. Finally, the algorithm terminates if the difference between the upper-bound and lower-bound is smaller than a predefined threshold.

Detailed operations of Algorithm 8.1 are described in the following. Firstly, the algorithm starts with a vertex set $\mathcal{V}^{(1)}$ which contains one infeasible vertex $\mathbf{x}^{(1)} = (\mathbf{t}^{(1)}, \mathbf{s}_{\mathcal{L}}^{(1)}, \mathbf{z}^{(1)})$ locating outside \mathcal{X} where $\mathbf{x}^{(1)}$ is determined by setting all elements at their maximum values. Specifically, we set $t_l^{(1)} = R_l^{\min} \forall l \in \mathcal{L}$, $s_l^{(1)} = 1 \forall l \in \mathcal{L}$, and $z_{mn}^{(1)} = \frac{P_{\max} h_{mm}^n}{\sigma_m^n} \forall m \in \mathcal{M} \forall n \in \mathcal{N}$.

In each iteration, we execute the following three steps. Assume that in iteration n , the considered vertex set is $\mathcal{V}^{(n)}$. In the first step, the algorithm determines the vertex which maximizes the objective function $f(\mathbf{x})$ over $\mathcal{V}^{(n)}$. We calculate $\mathbf{x}^{(n)} = \operatorname{argmax}_{\mathbf{x} \in \mathcal{V}^{(n)}} f(\mathbf{x})$, then the algorithm updates the upper-bound f_{up} as $f_{\text{up}} = f(\mathbf{x}^{(n)})$. In the second step, the algorithm finds the vertex $\Pi_{\mathcal{G}_f}(\mathbf{x}^{(n)})$

which is the projection of $\mathbf{x}^{(n)}$ on the boundary of \mathcal{G}_f . The procedure employed to determine $\Pi_{\mathcal{G}_f}(\mathbf{x}^{(n)})$ is described in Algorithm 8.2. If $\Pi_{\mathcal{G}_f}(\mathbf{x}^{(n)}) \in \mathcal{X}$ and $f(\Pi_{\mathcal{G}_f}(\mathbf{x}^{(n)})) > f_{\text{low}}$, which means that $\Pi_{\mathcal{G}_f}(\mathbf{x}^{(n)})$ is the best feasible solution until iteration n , then the algorithm updates the lower-bound as $f_{\text{low}} = f(\Pi_{\mathcal{G}_f}(\mathbf{x}^{(n)}))$.

In the third step, the algorithm updates the vertex set for the next iteration by replacing vertex $\mathbf{x}^{(n)}$ with the new vertex set $\bar{\mathcal{V}}_n = \{\bar{\mathbf{x}}_1^{(n)}, \dots, \bar{\mathbf{x}}_D^{(n)}\}$ as follows:

$$\mathcal{V}^{(n+1)} = (\mathcal{V}^{(n)} - \{\mathbf{x}^{(n)}\}) \cup \bar{\mathcal{V}}_n. \quad (8.19)$$

The additional vertex set $\bar{\mathcal{V}}_n$ consists of D vertexes in which each vertex $\bar{\mathbf{x}}_d^{(n)}$, $d \in \{1, \dots, D\}$ can be calculated as

$$\bar{\mathbf{x}}_d^{(n)} = \mathbf{x}^{(n)} - \left(\mathbf{x}^{(n)} - \Pi_{\mathcal{G}_f}(\mathbf{x}^{(n)}) \right) \odot \mathbf{e}_d, \quad (8.20)$$

where $\Pi_{\mathcal{G}_f}(\mathbf{x}^{(n)})$ is the projection of $\mathbf{x}^{(n)}$ into \mathcal{G}_f , \mathbf{e}_d is a vector with all zero elements except that the d -th element equals one, and \odot denotes Hadamard product of two vectors. Finally, the algorithm terminates if $f_{\text{up}} - f_{\text{low}} < \vartheta$ where ϑ is a predefined threshold.

Algorithm 8.2, which determines the projection of $\mathbf{x}^{(n)}$ on the boundary of \mathcal{G}_f , i.e., $\Pi_{\mathcal{G}_f}(\mathbf{x}^{(n)})$, can be described as follows. The projected vertex of $\mathbf{x}^{(n)}$ on the boundary of \mathcal{G}_f determined by Algorithm 8.2 has the following form: $\Pi_{\mathcal{G}_f}(\mathbf{x}^{(n)}) = \lambda^{(n)}(\mathbf{x}^{(n)} + \boldsymbol{\xi}) - \boldsymbol{\xi}$, which finds a vertex on the line connecting $\mathbf{x}^{(n)}$ with the point $-\boldsymbol{\xi} = -(\xi, \dots, \xi)$, and $\Pi_{\mathcal{G}_f}(\mathbf{x}^{(n)})$ is specified by $\lambda^{(n)}$ where $\xi > 0$ is a sufficiently small and positive number. In fact, the point $-\boldsymbol{\xi} = -(\xi, \dots, \xi)$ is used instead of the true origin point with all zero coordinates to guarantee the convergence of the proposed MBOA algorithm.

We must determine the $\lambda^{(n)}$ so that $\lambda^{(n)}(\mathbf{x}^{(n)} + \boldsymbol{\xi}) - \boldsymbol{\xi}$ lies inside \mathcal{G}_f but as close to the boundary of \mathcal{G}_f as possible. Such λ_n can be determined by solving the following problem

$$\begin{aligned} \lambda^{(n)} &= \max\{\lambda \mid \lambda(\mathbf{x}^{(n)} + \boldsymbol{\xi}) - \boldsymbol{\xi} \in \mathcal{G}_f\} \\ &= \max\{\lambda \mid \lambda[(\mathbf{t}^{(n)}, \mathbf{s}_{\mathcal{L}}^{(n)}, \mathbf{z}^{(n)}) + \boldsymbol{\xi}] - \boldsymbol{\xi} \in \mathcal{G}_f\} \\ &= \max \left\{ \lambda \mid \lambda \leq \min_{\substack{l \in \mathcal{L}, m \in \mathcal{M}, \\ n \in \mathcal{N}}} \left\{ \frac{1+\xi}{s_l^{(n)} + \xi}, \frac{R_l^{\min} + \xi}{t_l^{(n)} + \xi + R_l^{\min}(s_l^{(n)} + \xi)}, \frac{\bar{\Gamma}_m^n(\mathbf{p}) + \xi}{z_{mn}^{(n)} + \xi} \right\}, \mathbf{p} \in \mathcal{P} \right\}. \end{aligned} \quad (8.21)$$

Problem (8.21) is equivalent to $\lambda^{(n)} = \min\{\lambda_1^{(n)}, \lambda_2^{(n)}\}$, where

$$\lambda_1^{(n)} = \min_{l \in \mathcal{L}} \min \left\{ \frac{1+\xi}{s_l^{(n)} + \xi}, \frac{R_l^{\min} + \xi}{t_l^{(n)} + \xi + R_l^{\min}(s_l^{(n)} + \xi)} \right\} \quad (8.22)$$

$$\lambda_2^{(n)} = \min_{m \in \mathcal{M}, n \in \mathcal{N}} \max_{\mathbf{p} \in \mathcal{P}} \frac{\bar{\Gamma}_m^n(\mathbf{p}) + \xi}{z_{mn}^{(n)} + \xi}. \quad (8.23)$$

In fact, $\lambda_1^{(n)}$ captures the maximum value λ corresponding to $\mathbf{t}^{(n)}$ and $\mathbf{s}_{\mathcal{L}}^{(n)}$ while $\lambda_2^{(n)}$ captures the maximum value λ corresponding to $\mathbf{z}^{(n)}$. While $\lambda_1^{(n)}$ can be determined easily from (8.22), solving the problem on the right hand side (RHS) of (8.23) to determine $\lambda_2^{(n)}$ is more challenging. We transform this problem into the fractional program as follows. It can be seen from (8.6) and (8.7) that $\bar{\Gamma}_m^n(\mathbf{p})$ has the form $\bar{\Gamma}_m^n(\mathbf{p}) = \frac{\Gamma_{mn}^{\text{num}}(\mathbf{p})}{\Gamma_{mn}^{\text{den}}(\mathbf{p})}$, where $\Gamma_{mn}^{\text{num}}(\mathbf{p})$ and $\Gamma_{mn}^{\text{den}}(\mathbf{p})$ are linear functions of \mathbf{p} . Hence, problem in the RHS of (8.23) can be transformed into the following

$$\lambda_2^{(n)} = \min_{m \in \mathcal{M}, n \in \mathcal{N}} \max_{\mathbf{p} \in \mathcal{P}} \frac{\Gamma_{mn}^{\text{num}}(\mathbf{p}) + \xi \Gamma_{mn}^{\text{den}}(\mathbf{p})}{(z_{mn}^{(n)} + \xi) \Gamma_{mn}^{\text{den}}(\mathbf{p})}. \quad (8.24)$$

Problem in the RHS of (8.24) is a standard fractional program which can be solved *optimally* by using the Dinkelbach algorithm [133]. Specifically, to solve this problem, we consider the following problem

$$\begin{aligned} & \max_{\mathbf{p} \in \mathcal{P}} \quad \theta \\ & \text{s.t.} \quad \Gamma_{mn}^{\text{num}}(\mathbf{p}) + \xi \Gamma_{mn}^{\text{den}}(\mathbf{p}) - \lambda_{\text{temp}}(z_{mn}^{(n)} + \xi) \Gamma_{mn}^{\text{den}}(\mathbf{p}) \geq \theta, \forall m \in \mathcal{M} \forall n \in \mathcal{N}, \end{aligned} \quad (8.25)$$

which is the subtractive form of the problem in the RHS of (8.24) for given λ_{temp} . In fact, (8.25) is a linear program which can be solved efficiently by the standard linear programming methods, e.g., simplex or interior point method. To solve problem (8.24), Algorithm 8.2 alternates between solving problem (8.25) and updating λ_{temp} until θ obtained from problem (8.25) is negative.

8.5.4 Analysis of the MBOA Algorithm

Assume that $\mathbf{x}^* = (\mathbf{t}^*, \mathbf{s}_{\mathcal{L}}^*, \mathbf{z}^*)$ and \mathbf{p}^* are the optimal solution of problem (8.18) and its corresponding power allocation, respectively. Nevertheless, $\mathbf{s}_{\mathcal{L}}^*$ can be fractional which is not a feasible solution of problem (8.10). We propose to schedule the non-prioritized link l if its scheduling solution $s_l^* \geq 1 - \epsilon$. Then, the performance of the MBOA algorithm is characterized in the following theorem.

Algorithm 8.1. The MBOA Algorithm

-
- 1: **Initialization:** Set iteration index $n = 1$ and $\mathcal{V}^{(1)} = \{\mathbf{x}^{(1)}\}$. The vertex $\mathbf{x}^{(1)} = (\mathbf{t}^{(1)}, \mathbf{s}_{\mathcal{L}}^{(1)}, \mathbf{z}^{(1)})$ is determined by setting all elements at their maximum values, i.e., $t_l^{(1)} = R_l^{\min} \forall l \in \mathcal{L}$, $s_l^{(1)} = 1 \forall l \in \mathcal{L}$, and $z_{mn}^{(1)} = \frac{P_{\max} h_{mn}^n}{\sigma_m^n} \forall m \in \mathcal{M}, \forall n \in \mathcal{N}$.
 - 2: **repeat**
 - 3: **Step 1:** Determine $\mathbf{x}^{(n)} = \underset{\mathbf{x} \in \mathcal{V}^{(n)}}{\operatorname{argmax}} f(\mathbf{x})$.
Update the upper-bound $f_{\text{up}} = f(\mathbf{x}^{(n)})$.
 - 4: **Step 2:** Perform **Algorithm 8.2** to determine $\Pi_{\mathcal{G}_f}(\mathbf{x}^{(n)})$.
If $f(\Pi_{\mathcal{G}_f}(\mathbf{x}^{(n)})) > f_{\text{low}}$ and $\Pi_{\mathcal{G}_f}(\mathbf{x}^{(n)})$ satisfies (8.18f) and (8.18g), update lower-bound $f_{\text{low}} = f(\Pi_{\mathcal{G}_f}(\mathbf{x}^{(n)}))$ and current best solution $\mathbf{x}^* = \Pi_{\mathcal{G}_f}(\mathbf{x}^{(n)})$.
 - 5: **Step 3:** Determine the additional vertex set $\bar{\mathcal{V}}_n = \{\bar{\mathbf{x}}_1^{(n)}, \dots, \bar{\mathbf{x}}_D^{(n)}\}$,
where $\bar{\mathbf{x}}_d^{(n)} = \mathbf{x}^{(n)} - (\mathbf{x}^{(n)} - \Pi_{\mathcal{G}_f}(\mathbf{x}^{(n)})) \odot \mathbf{e}_d, \forall d \in \{1, \dots, D\}$.
 - 6: Update the vertex set of the next iteration $\mathcal{V}^{(n+1)} = (\mathcal{V}^{(n)} - \{\mathbf{x}^{(n)}\}) \cup \bar{\mathcal{V}}_n$.
 - 7: Increase iteration index $n = n + 1$
 - 8: **until** $f_{\text{up}} - f_{\text{low}} < \vartheta$
 - 9: Perform rounding procedure to determine the scheduling solution $\bar{\mathbf{s}}_{\mathcal{L}}$

$$\bar{s}_l = \begin{cases} 1, & \text{if } s_l^* \geq 1 - \epsilon \\ 0, & \text{otherwise.} \end{cases} \quad (8.26)$$

- 10: From \mathbf{p}^* , set the power of non-scheduled non-prioritized links to zeros to obtain $\bar{\mathbf{p}}$.
 - 11: Output $(\bar{\mathbf{s}}_{\mathcal{L}}, \bar{\mathbf{p}})$.
-

Algorithm 8.2. Calculate $\Pi_{\mathcal{G}_f}(\mathbf{x}^{(n)})$

-
- 1: Input $\mathbf{x}^{(n)}, \boldsymbol{\xi}$
 - 2: **Initialization** $\lambda_{\text{temp}} = 0$.
 - 3: **repeat**
 - 4: Solve the following problem (8.25) for given λ_{temp}

$$\begin{aligned} & \max_{\mathbf{p} \in \mathcal{P}} \quad \theta \\ & \text{s.t.} \quad \Gamma_{mn}^{\text{num}}(\mathbf{p}) + \xi \Gamma_{mn}^{\text{den}}(\mathbf{p}) - \lambda_{\text{temp}} (z_{mn}^{(n)} + \xi) \Gamma_{mn}^{\text{den}}(\mathbf{p}) \geq \theta, \forall m \in \mathcal{M}, \forall n \in \mathcal{N}. \end{aligned}$$

- 5: Update $\lambda_{\text{temp}} = \min_{m \in \mathcal{M}, n \in \mathcal{N}} \max_{\mathbf{p} \in \mathcal{P}} \frac{\Gamma_{mn}^{\text{num}}(\mathbf{p}) + \xi \Gamma_{mn}^{\text{den}}(\mathbf{p})}{(z_{mn}^{(n)} + \xi) \Gamma_{mn}^{\text{den}}(\mathbf{p})}$
 - 6: **until** $\theta < 0$
 - 7: Set $\lambda_2^{(n)} = \lambda_{\text{temp}}$ and $\mathbf{p}^* = \mathbf{p}$.
 - 8: Determine $\lambda_1^{(n)} = \min_{l \in \mathcal{L}} \left\{ \frac{1 + \xi}{s_l^{(n)} + \xi}, \frac{R_l^{\min} + \xi}{t_l^{(n)} + \xi + R_l^{\min}(s_l^{(n)} + \xi)} \right\}$
 - 9: Set $\lambda^{(n)} = \min\{\lambda_1^{(n)}, \lambda_2^{(n)}\}$
 - 10: Output $\Pi_{\mathcal{G}_f}(\mathbf{x}^{(n)}) = \lambda^{(n)}(\mathbf{x}^{(n)} + \boldsymbol{\xi}) - \boldsymbol{\xi}$.
-

Theorem 8.1. *If the minimum required data rate of each non-prioritized link can be reduced by a small and controller number ϵR_l^{\min} , by choosing $Q \geq \ln L/\epsilon$, the proposed MBOA algorithm schedules the maximum number of non-prioritized links.⁷*

⁷If we consider the studied problem where the minimum rate of non-prioritized link l equals to $(1 + \epsilon)R_l^{\min}$ then the MBOA algorithm can guarantee the required QoS of all scheduled non-prioritized links.

Denote $\bar{\mathbf{s}}_{\mathcal{L}}$ is the scheduling solution obtained from MBOA algorithm. In general, we can obtain the optimal sum rate of the scheduled links in $\bar{\mathbf{s}}_{\mathcal{L}}$ by solving the monotonic problem (8.18) for a given set of the scheduled non-prioritized links $\bar{\mathbf{s}}_{\mathcal{L}}$. Nevertheless, this approach doubles the computational complexity of the MBOA algorithm. On the other hand, as Q is sufficiently large, the values of s_l^* for unscheduled non-prioritized links $l \in \mathcal{L}$ are small, which means the rate loss due to the unscheduled non-prioritized links is negligible. Therefore, the sum rate achieved by the MBOA algorithm approaches the optimal one.

We now analyze the complexity of the MBOA algorithm in term of the number of required arithmetic operations. Since the number of iterations in the MBOA algorithm is not fixed, its complexity cannot be determined exactly. Let T be the number of iterations executed in the MBOA algorithm. In each iteration, the complexity of **Step 1** is $O(D)$ where D is the dimension of the optimization vector \mathbf{x} in problem (8.18). The complexity of **Step 2** can be determined by analyzing the complexity of Algorithm 8.2. In fact, Algorithm 8.2 requires to solve problem (8.25) iteratively by the interior point method with complexity of $O\left(m^{\frac{1}{2}}(m+n)n^2\right)$, where m is the number of inequality constraints and n is number of variables [134]. Therefore, the complexity of solving problem (8.25) for given λ_{temp} is $O(M^{3.5}N^{3.5})$. Moreover, the number of iterations in Algorithm 8.2 is $O(1)$. Thus, the complexity of the MBOA algorithm is $O(TM^{3.5}N^{3.5})$.

8.6 Iterative Convex Approximation Algorithm

The MBOA algorithm can achieve an asymptotically optimal solution but it has high computational complexity. Therefore, we propose an iterative convex approximation (ICA) algorithm, which is more efficient in computational complexity and achieves satisfactory performance. The ICA algorithm solves problem (8.10) by sequentially performing power allocation and link removal which is described in Algorithm 8.3. In particular, the power allocation solution is obtained by solving the

following problem

$$\begin{aligned}
& \max_{\mathbf{p}, \mathbf{s}_{\mathcal{L}_i}} \alpha \sum_{l \in \mathcal{L}_i} s_l + \sum_{l \in \mathcal{L}} \bar{R}_l(\mathbf{p}) \\
& \text{s.t.} \quad \bar{R}_k(\mathbf{p}) \geq R_k^{\min} \quad \forall k \in \mathcal{K} \\
& \quad \quad \bar{R}_l(\mathbf{p}) \geq s_l R_l^{\min} \quad \forall l \in \mathcal{L}_i \\
& \quad \quad \sum_{n \in \mathcal{N}} p_k^n \leq P_{\max} \quad \forall k \in \mathcal{K} \\
& \quad \quad \sum_{n \in \mathcal{N}} p_l^n \leq P_{\max} \quad \forall l \in \mathcal{L}_i \\
& \quad \quad s_l \in [0, 1] \quad \forall l \in \mathcal{L}_i,
\end{aligned} \tag{8.27}$$

where $\mathcal{L}_i \subset \mathcal{L}$ is the set of considered non-prioritized links. Denote problem (8.27) associated with the link set \mathcal{L}_i as $\mathcal{P}(\mathcal{L}_i)$ where problem $\mathcal{P}(\mathcal{L}_i)$ is obtained from problem (8.10) by changing the set of considered non-prioritized links \mathcal{L} to a smaller set $\mathcal{L}_i \subset \mathcal{L}$, relaxing the scheduling decision variables, and approximating $\delta(s_l)$ by a linear function s_l .

Algorithm 8.3. ITERATIVE CONVEX APPROXIMATION ALGORITHM

- 1: $\mathcal{L}_1 = \mathcal{L}$, $\mathbf{p}^{(0)} = \mathbf{0}$.
 - 2: Set iteration index $i = 0$
 - 3: **repeat**
 - 4: $i = i + 1$
 - 5: Perform power allocation by solving problem $\mathcal{P}(\mathcal{L}_i)$ to obtain $\mathbf{p}^{(i)}$ and $\mathbf{s}_{\mathcal{L}_i}^*$.
 - 6: Run the link removal procedure to update \mathcal{L}_i
 - 7: **until** Required rates of all links in \mathcal{L}_i are satisfied.
 - 8: Output \mathcal{L}_i and $\mathbf{p}^* = \mathbf{p}^{(i)}$.
-

Detailed operations of Algorithm 3 can be described as follows. In iteration i , let \mathcal{L}_i and $\mathbf{p}^{(i)}$ be the set of non-prioritized links to be scheduled and the initial power allocation. Initially, we set $\mathcal{L}_i = \mathcal{L}$ and $\mathbf{p}^{(i)} = \mathbf{p}^{(0)}$. Algorithm 3 performs power allocation (line 5) and link removal procedure (line 6) sequentially until the minimum required rates of all non-prioritized links in \mathcal{L}_i are satisfied. The power allocation and link removal procedures are described in the following.

8.6.1 Power Allocation

We propose an algorithm to solve problem $\mathcal{P}(\mathcal{L}_i)$, which is non-convex and hence difficult to solve. In the following, we propose an algorithm to solve it suboptimally and efficiently by employing a convex approximation approach. For convex approximation, we express the data rate of link

$m \in \mathcal{K} \cup \mathcal{L}_i$ in the form $\bar{R}_m(\mathbf{p}) = f_m(\mathbf{p}) - g_m(\mathbf{p})$, where

$$f_m(\mathbf{p}) = \sum_{n \in \mathcal{N}} \log_2 \left(\sigma_m^n + p_m^n h_{mm}^n + \sum_{k \in \mathcal{K} \setminus m} p_k^n \eta + \sum_{l \in \mathcal{L}_i} p_l^n h_{ml}^n \right) \quad (8.28)$$

$$g_m(\mathbf{p}) = \sum_{n \in \mathcal{N}} \log_2 \left(\sigma_m^n + \sum_{k \in \mathcal{K} \setminus m} p_k^n \eta + \sum_{l \in \mathcal{L}_i} p_l^n h_{ml}^n \right), \quad (8.29)$$

if $m \in \mathcal{K}$, and

$$f_m(\mathbf{p}) = \sum_{n \in \mathcal{N}} \log_2 \left(\sigma_m^n + \sum_{k \in \mathcal{K}} p_k^n h_{mk}^n + \sum_{l' \in \mathcal{L}_i} p_{l'}^n h_{ll'}^n \right) \quad (8.30)$$

$$g_m(\mathbf{p}) = \sum_{n \in \mathcal{N}} \log_2 \left(\sigma_m^n + \sum_{k \in \mathcal{K}} p_k^n h_{mk}^n + \sum_{l' \in \mathcal{L}_i \setminus l} p_{l'}^n h_{ll'}^n \right), \quad (8.31)$$

if $m \in \mathcal{L}_i$. It can be verified that $f_m(\mathbf{p})$ and $g_m(\mathbf{p})$ are concave functions of \mathbf{p} . Hence, $\bar{R}_m(\mathbf{p})$ has the DC (difference of two concave functions) structure. We can approximate $g_m(\mathbf{p})$ by the first order Taylor approximation as follows [73]:

$$g_m(\mathbf{p}) \leq \bar{g}_m(\mathbf{p}, \mathbf{p}^t) = g_m(\mathbf{p}^t) + \nabla g_m^T(\mathbf{p}^t)(\mathbf{p} - \mathbf{p}^t), \quad (8.32)$$

where $\nabla g_m(\mathbf{p}^t)$ is a subgradient vector of $g_m(\mathbf{p})$ at $\mathbf{p} = \mathbf{p}^t$. Using this approximation, we can obtain the rate lower-bound of link m as

$$\bar{R}_m(\mathbf{p}) \geq \hat{R}_m(\mathbf{p}, \mathbf{p}^t) \quad \forall m \in \mathcal{K} \cup \mathcal{L}_i \quad \forall \mathbf{p}^t, \quad (8.33)$$

where $\hat{R}_m(\mathbf{p}, \mathbf{p}^t) = f_m(\mathbf{p}) - g_m(\mathbf{p}^t) - \nabla g_m^T(\mathbf{p}^t)(\mathbf{p} - \mathbf{p}^t)$. Using this lower-bound for all rate terms in problem $\mathcal{P}(\mathcal{L}_i)$, we obtain an approximated problem of $\mathcal{P}(\mathcal{L}_i)$ at initial power \mathbf{p}^t denoted as $\mathcal{AP}(\mathcal{L}_i, \mathbf{p}^t)$ as follows:

$$\begin{aligned} & \max_{\mathbf{p}, \mathbf{s}^{\mathcal{L}_i}} \alpha \sum_{l \in \mathcal{L}_i} s_l + \sum_{l \in \mathcal{L}_i} \hat{R}_l(\mathbf{p}, \mathbf{p}^t) \\ & \text{s.t.} \quad \hat{R}_k(\mathbf{p}, \mathbf{p}^t) \geq R_k^{\min} \quad \forall k \in \mathcal{K} \\ & \quad \hat{R}_l(\mathbf{p}, \mathbf{p}^t) \geq s_l R_l^{\min} \quad \forall l \in \mathcal{L}_i \\ & \quad \sum_{n \in \mathcal{N}_k} p_k^n \leq P_{\max} \quad \forall k \in \mathcal{K} \\ & \quad \sum_{n \in \mathcal{N}} p_l^n \leq P_{\max} \quad \forall l \in \mathcal{L}_i \\ & \quad s_l \in [0, 1] \quad \forall l \in \mathcal{L}_i. \end{aligned} \quad (8.34)$$

It can be verified that this problem is convex, which can be solved optimally by standard convex optimization techniques, e.g., interior point method [118]. We now propose Algorithm 4 to solve problem $\mathcal{P}(\mathcal{L}_i)$, which is used in line 5 of Algorithm 3. Particularly, it first solves problem $\mathcal{AP}(\mathcal{L}_i, \mathbf{p}^t)$ with the initial power allocation \mathbf{p}^t obtained from the previous iteration. In each iteration t , it solves problem $\mathcal{AP}(\mathcal{L}_i, \mathbf{p}^{t-1})$ to obtain \mathbf{p}^t ; then, it calculates the subgradient of $g_m(\mathbf{p})$ at \mathbf{p}^t . These two steps are performed iteratively until convergence.

Algorithm 8.4. Solve Problem $\mathcal{P}(\mathcal{L}_i)$

- 1: Initial power allocation \mathbf{p}^0 .
 - 2: Calculate $\nabla g_m^T(\mathbf{p}^0), \forall m \in \mathcal{K} \cup \mathcal{L}_i$ and set $t = 0$.
 - 3: **repeat**
 - 4: $t = t + 1$
 - 5: Solve problem $\mathcal{AP}(\mathcal{L}_i, \mathbf{p}^t)$ to obtain \mathbf{p}^t .
 - 6: Update $\nabla g_m^T(\mathbf{p}^t), \forall m \in \mathcal{K} \cup \mathcal{L}_i$.
 - 7: **until** Convergence.
-

8.6.2 Link Removal Procedure and Computational Complexity Analysis

The link removal procedure is executed as follows. We initially start with the largest set of links where all links are scheduled. Then, in each iteration i , we run the ICA algorithm to obtain the link scheduling vector $\mathbf{s}_{\mathcal{L}_i}^*$ and power allocation vector \mathbf{p}^* . Then, the set of scheduled links is determined based on $\mathbf{s}_{\mathcal{L}_i}^*$ as follows. If we have $\min_{l \in \mathcal{L}_i} s_l^* = 1$, then minimum rate requirements of all links in \mathcal{L}_i are satisfied so we stop the algorithm. Otherwise, link $l^* = \operatorname{argmin}_{l \in \mathcal{L}_i} s_l^*$ will be removed from the set \mathcal{L}_i , and we perform the next iteration of ICA algorithm with the updated set \mathcal{L}_i .

The ICA algorithm requires to solve problem $\mathcal{P}(\mathcal{L}_i)$ by Algorithm 4 and perform link removal sequentially in each iteration. The number of iterations required by this algorithm has complexity of $O(L)$. Algorithm 4 solves problem (8.34) iteratively by interior point method with complexity of $O(N^3 M^{3.5})$ [134]. In addition, number of iterations required in Algorithm 4 has complexity of $O(1)$. Therefore, the complexity of the ICA algorithm is $O(LN^3 M^{3.5})$.

8.7 Practical Implementation

In general, to perform the resource allocation algorithm, the BS has to collect the CSI of all links. The BS can obtain CSI information by channel estimation and feedback procedure implemented in

each user. The CSI estimation and feedback period of each user depends on the variation of its channel state and buffer status. In the following, we describe the current 3GPP standard of CSI estimation and feedback for D2D communication in cellular network and discuss how our algorithms can be executed [135, 136].

8.7.1 Channel state information estimation and feedback phase

Since the accuracy of CSI estimation and feedback affects greatly the performance of the resource allocation algorithm, it is investigated intensively in nowadays 3GPP standard. Recently, release 14 of 3GPP standard allows the BS to perform scheduling and resource allocation effectively due to the careful design in CSI estimation and feedback procedure [135].

To request the channel quality index (CQI) feedback from the user equipment (UE), the BS can tell the UE to send CSI report by setting CSI Request field in Downlink Control Information (DCI) Format of Physical Downlink Control Channel (PDCCH). As the UE can decode the request from the BS, it can feedback the required information to the BS in the Physical Downlink Shared Channel (PDSCH). This procedure can be executed either periodically or aperiodically depending on the configuration of the BS [135, 137].

In order to support the BS in performing the scheduling and resource allocation procedure, the UEs involved in D2D communication are required to perform sensing continuously up to 1000 subframe before its transmission [137]. The sensed information is broadcasted to the other UEs and BS via Sidelink Control Information (SCI) message in Physical Sidelink Control Channel (PSSCH) [138]. Each UE is required to decode the SCI messages of other UEs to evaluate the cross-interference between it and the other UEs. This information, then can be transmit to the BS via the SCI messages [137]. Therefore, from the above procedures, the BS can collect the necessary information of channel gains and cross-interferences to perform scheduling and resource allocation algorithm.

8.7.2 Algorithm execution phase

Upon receiving the necessary CSI information of all links in the system, the BS can perform designed scheduling and resource allocation algorithm. In practical wireless scenarios, the CSI available at the BS is not perfect due to quantization, limited training time, channel estimation error, and CSI

feedback delay. Therefore, performing proposed algorithm based on the obtained CSI might not guarantee the quality of service (QoS) of all links in the system. Nevertheless, we can execute the conservative design, which can guarantee the QoS of all links in the worst scenario, to solve the scheduling and resource allocation problem [139–141]. The conservative design for imperfect CSI scenario is described in the following.

Let \hat{h}_{kl}^n be the channel gain from the transmitter of link l to the receiver of link k on subchannel n , which is available at the BS. Note that h_{kl}^n represent the exact channel gain that are unknown to the BS due to the uncertainty. Let δ_{kl}^n and Δ_{kl}^n be the estimation error and maximum estimation error, respectively, of channel gain from the transmitter of link l to the receiver of link k on subchannel n .

Therefore, the SINR of prioritized link k and non-prioritized link l on subchannel n in equation (8.6) and (8.7) can be reformulated, respectively, as

$$\bar{\Gamma}_k^n(\mathbf{p}) = \frac{p_k^n (\hat{h}_{kk}^n + \delta_{kk}^n)}{\sigma_k^n + \sum_{k' \in \mathcal{K} \setminus k} p_{k'}^n \eta + \sum_{l \in \mathcal{L}} p_l^n (\hat{h}_{kl}^n + \delta_{kl}^n)} \quad (8.35)$$

$$\bar{\Gamma}_l^n(\mathbf{p}) = \frac{p_l^n (\hat{h}_{ll}^n + \delta_{ll}^n)}{\sigma_l^n + \sum_{k \in \mathcal{K}} p_k^n (\hat{h}_{lk}^n + \delta_{lk}^n) + \sum_{l' \in \mathcal{L} \setminus l} p_{l'}^n (\hat{h}_{ll'}^n + \delta_{ll'}^n)}. \quad (8.36)$$

Then, the minimum SINR of prioritized link k and non-prioritized link l are expressed, respectively, as

$$\hat{\Gamma}_k^n(\mathbf{p}) = \frac{p_k^n (\hat{h}_{kk}^n - \Delta_{kk}^n)}{\sigma_k^n + \sum_{k' \in \mathcal{K} \setminus k} p_{k'}^n \eta + \sum_{l \in \mathcal{L}} p_l^n (\hat{h}_{kl}^n + \Delta_{kl}^n)} \quad (8.37)$$

$$\hat{\Gamma}_l^n(\mathbf{p}) = \frac{p_l^n (\hat{h}_{ll}^n - \Delta_{ll}^n)}{\sigma_l^n + \sum_{k \in \mathcal{K}} p_k^n (\hat{h}_{lk}^n + \Delta_{lk}^n) + \sum_{l' \in \mathcal{L} \setminus l} p_{l'}^n (\hat{h}_{ll'}^n + \Delta_{ll'}^n)}. \quad (8.38)$$

Note that above equations are obtained by considering the worst case SINR expression where the numerator and denominator achieve its smallest and largest, respectively. Hence, the minimum data rates of prioritized link $k \in \mathcal{K}$ and non-prioritized link $l \in \mathcal{L}$ can be re-expressed, respectively,

as

$$\hat{R}_k(\mathbf{p}) = \sum_{n \in \mathcal{N}} \log_2 \left(1 + \hat{\Gamma}_k^n(\mathbf{p}) \right) \quad (8.39)$$

$$\hat{R}_l(\boldsymbol{\rho}, \mathbf{p}) = \sum_{n \in \mathcal{N}} \log_2 \left(1 + \hat{\Gamma}_l^n(\mathbf{p}) \right). \quad (8.40)$$

Therefore, we can formulate the following scheduling and resource allocation problem

$$\max_{\mathbf{p}, \mathbf{s}_{\mathcal{L}}} \alpha \sum_{l \in \mathcal{L}} \delta(s_l) + \sum_{l \in \mathcal{L}} \hat{R}_l(\mathbf{p}) \quad (8.41a)$$

$$\text{s.t.} \quad \hat{R}_k(\mathbf{p}) \geq R_k^{\min} \quad \forall k \in \mathcal{K} \quad (8.41b)$$

$$\hat{R}_l(\mathbf{p}) \geq s_l R_l^{\min} \quad \forall l \in \mathcal{L} \quad (8.41c)$$

$$\sum_{n \in \mathcal{N}} p_k^n \leq P_{\max} \quad \forall k \in \mathcal{K} \quad (8.41d)$$

$$\sum_{n \in \mathcal{N}} p_l^n \leq P_{\max} \quad \forall l \in \mathcal{L} \quad (8.41e)$$

$$s_l \in \{0, 1\} \quad \forall l \in \mathcal{L}. \quad (8.41f)$$

We then can perform the proposed algorithms, i.e., MBOA and ICA algorithms, to solve problem (8.41) for the imperfect CSI scenario.

8.8 Numerical Results

We evaluate the performance of our proposed algorithms for wireless cellular networks supporting D2D communications. We conduct numerical studies for two different network settings, namely the sparse and dense networks, as illustrated in Fig. 8.2 and 8.3, respectively. The two networks are different in the numbers of cellular links, D2D links, and subchannels. We set $K=4$, $L=5$, $N=5$ in the sparse network, and $K=5$, $L=10$, $N=10$ in the dense network. The performance of the ICA algorithm is investigated in the dense network while the performance evaluation of both MBOA and ICA algorithms are studied in the sparse network.

In particular, we consider the uplink resource allocation in a single cell system with coverage radius $R = 500\text{m}$. There are K prioritized cellular links sharing the spectrum resource of N subchannels with L non-prioritized D2D links. In addition, the subchannel power gain is modeled

as $h_{kl}^n = d_{kl}^{-3} \kappa$, where κ represents the Rayleigh fading following the exponential distribution with the mean value of 1. We set the noise power in each subchannel to 10^{-12} W and the maximum transmit power of each link as 0.5W. The maximum distance of each D2D link $d_{\max} = 100$ m, and the minimum rates of each cellular and scheduled D2D link are $R_c^{\min} = 3$ bps/Hz and $R_d^{\min} = 5$ bps/Hz, respectively unless stated otherwise. All numerical results are obtained by averaging over 1000 random realizations of user locations and channel gains.

We first evaluate the performance of the proposed algorithms in the imperfect CSI scenarios. Specifically, we study the scenarios where CSI of the direct links are perfect, i.e., $\Delta_{mm}^n = 0$, $\forall m \in \mathcal{M}$, $\forall n \in \mathcal{N}$, and the CSI of the cross-interference links are imperfect, i.e., $\Delta_{kl}^n \leq h_{kl}^n \Delta_{\max}$, $\forall k, l \in \mathcal{M}, k \neq l$. The performance results of the proposed algorithms for imperfect CSI scenarios are obtained by conducting the conservative design described in Section 8.7.2. We compare the performance of the proposed algorithms with different values of Δ_{\max} , i.e., $\Delta_{\max} = 0\%$ for perfect CSI, and $\Delta_{\max} = 10\%$, $\Delta_{\max} = 30\%$, $\Delta_{\max} = 50\%$ for imperfect CSI, in different network settings.

We then evaluate the performance of the proposed algorithms and compare them with the ‘‘Conv. SE’’ and ‘‘Conv. Scheduling’’ algorithms, which are adopted from [28]. The ‘‘Conv. SE’’ algorithm begins by using the ICA algorithm to solve problem (8.10) without the QoS constraints of the D2D links and with $\mathcal{L}_a^* = \mathcal{L}$. Then, if the QoS of all D2D links in \mathcal{L}_a^* are not satisfied, the D2D link with the lowest data rate is removed from \mathcal{L}_a^* . This procedure is performed iteratively until the QoS requirements of all D2D links in \mathcal{L}_a^* are satisfied. The ‘‘Conv. Scheduling’’ algorithm directly deals with the joint scheduling and resource allocation problem within a single step. It first solves problem (8.27) using the convex approximation approach, then all D2D links which do not satisfy their QoS constraints are removed from the set of D2D links \mathcal{L} to be scheduled.

8.8.1 Numerical Studies in the Imperfect CSI scenarios

Figs. 8.4 and 8.5 illustrate the number of scheduled D2D links and their sum rate versus R_d^{\min} for different values of Δ_{\max} in MBOA algorithm. As expected, both number of scheduled D2D links and their sum rate are higher as Δ_{\max} is smaller, and MBOA algorithm with perfect CSI information performs the best in both performance metrics. It is remarkable that the number of scheduled D2D links and their sum rate in MBOA algorithm with $\Delta_{\max} = 50\%$ are always greater than 95% those of the MBOA algorithm with perfect CSI information. It is because the MBOA algorithm is

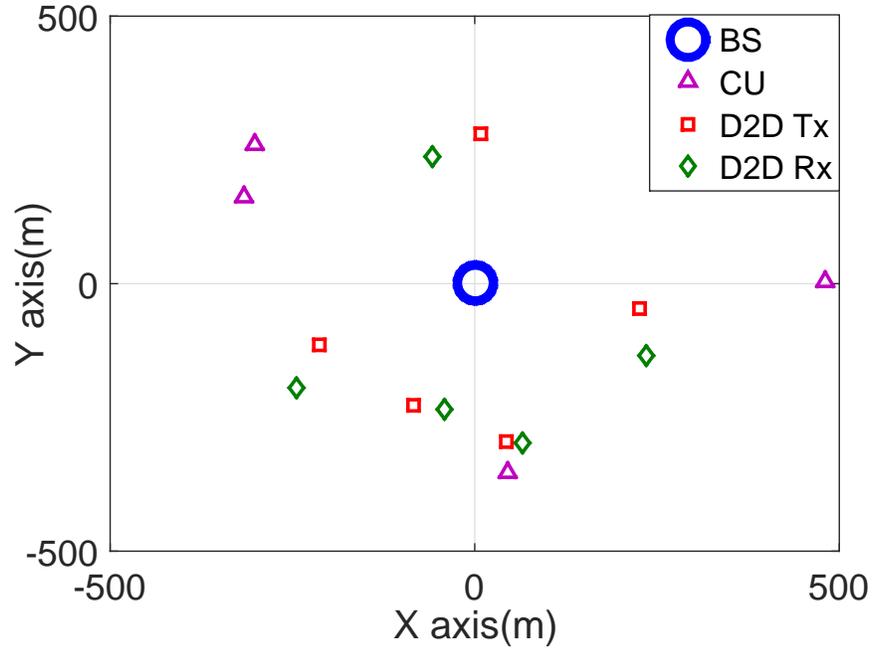


Figure 8.2 – Number of admitted D2D links versus number of cellular links K

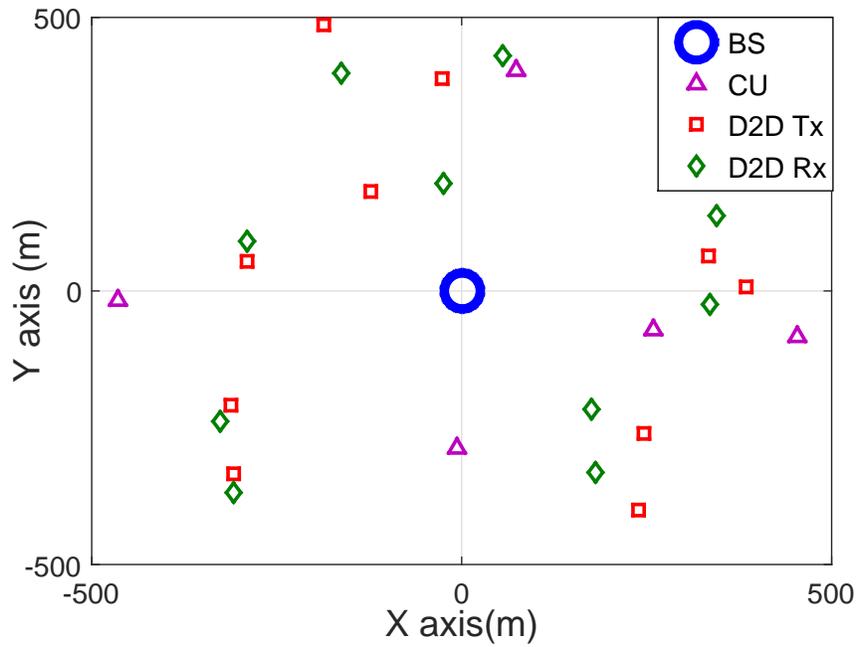


Figure 8.3 – Sum rate of D2D links versus number of cellular links K

very robust in scheduling and resource allocation. It targets to allocate the subchannels with high channel gain and low interference to the users. Therefore, in most cases, the user will suffer less cross-interference from the other users. Hence, even with large values of Δ_{max} , the performance lost of the MBOA algorithm is negligible.

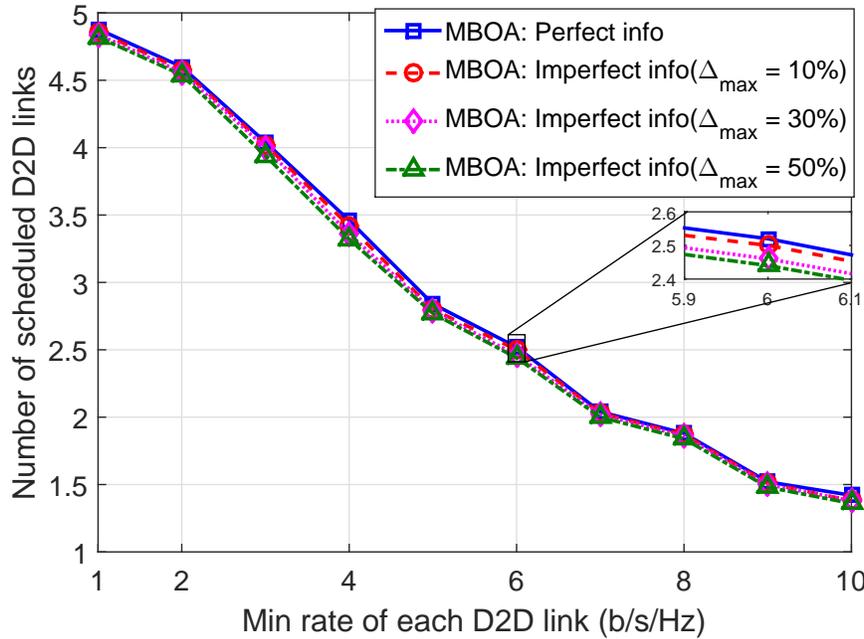


Figure 8.4 – Number of admitted D2D links versus minimum rate of D2D link R_d^{\min}

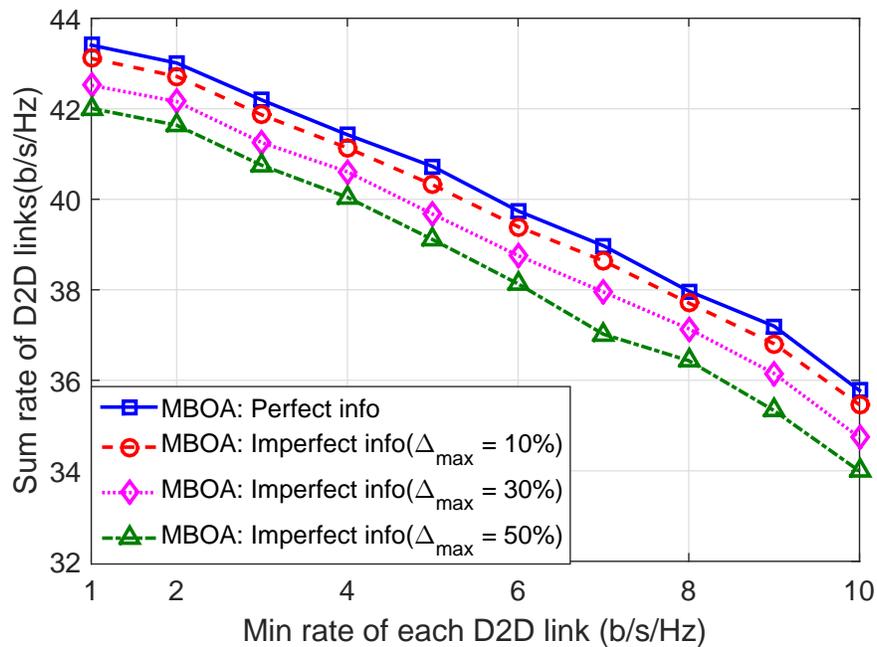


Figure 8.5 – Sum rate of D2D links versus minimum rate of D2D link R_d^{\min}

Figs. 8.6 and 8.7 present the number of scheduled D2D links and their sum rate versus the number of D2D links L for different values of Δ_{\max} in ICA algorithm. As L is higher, both the number of scheduled D2D links and their sum rate become larger for all values of Δ_{\max} . It is noticeable that the performance lost due to the CSI error in ICA algorithm is always smaller than

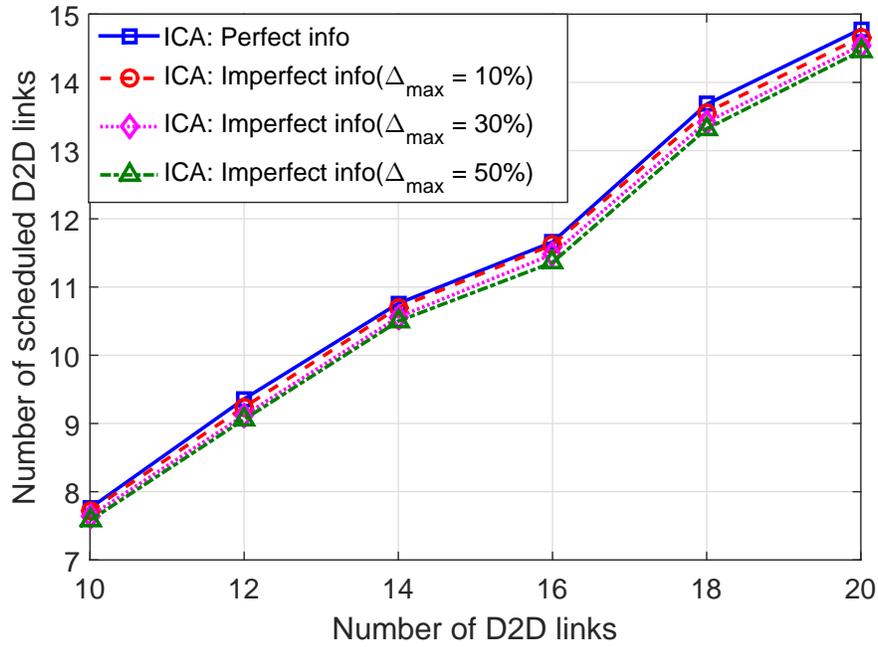


Figure 8.6 – Number of admitted D2D links versus K

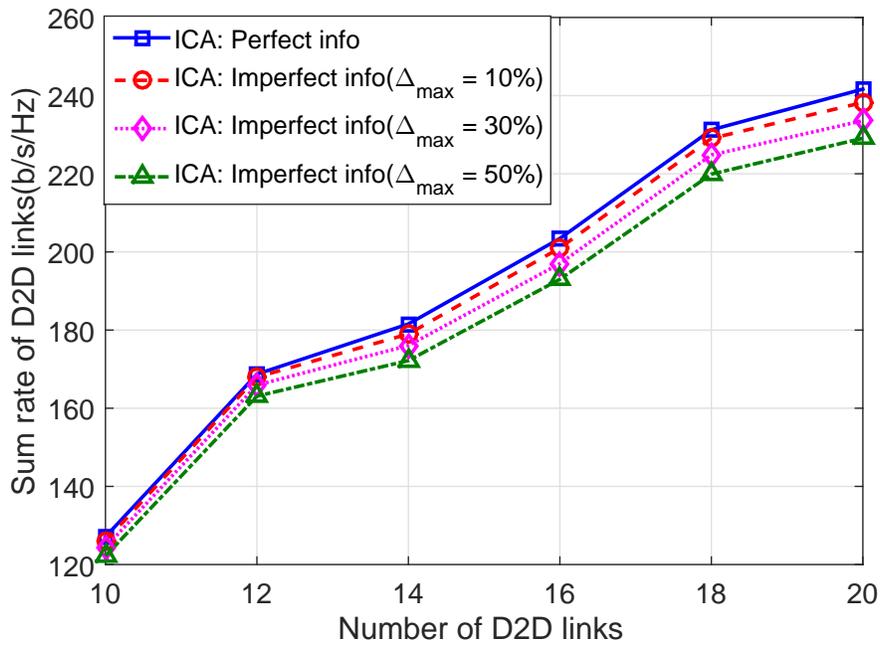


Figure 8.7 – Sum rate of D2D links versus number of cellular links K

10%. Moreover, as number of D2D links is larger, performance lost due to the error in CSI increases, and the performance lost in number of scheduled D2D links is always smaller than that of sum rate. It is because as number of D2D links increases, each link will suffer more interference from the other links. Therefore, the total error in evaluating interference of each link is amplified. Consequently,

the performance lost is higher. Finally, in our proposed the first priority is maximizing number of scheduled D2D links. Therefore, the performance lost in number of scheduled D2D links is minimize.

8.8.2 Numerical Studies in the Perfect CSI scenarios

Figs. 8.8 and 8.9 show the number of scheduled D2D links and their sum rate versus the cellular links' minimum required rate R_c^{\min} , respectively. As R_c^{\min} increases, the system has to assign more resources to the cellular links; therefore, a smaller number of D2D links and a smaller sum rate of the scheduled D2D links can be achieved, which can be observed for all algorithms. These results confirm that the MBOA algorithm performs the best in term of both number of scheduled D2D links and their sum rate. This is because the MBOA algorithm can schedule the maximum number of D2D links as stated in Theorem 8.1.

It is notable that the ICA algorithm performs very well where the number of scheduled D2D links and sum rate of the D2D links due to this algorithm are close to those of the MBOA algorithm. This is because the ICA algorithm targets at maximizing both metrics simultaneously. The performance gap between the MBOA and the ICA algorithms is mostly due to the sub-optimality of the power allocation procedure employed in the ICA algorithm. Moreover, it can be seen that the ICA algorithm outperforms the two conventional algorithms in term of number of scheduled D2D links and their sum rate. These results confirm the efficiency of the proposed algorithms, which can simultaneously optimize the two considered design objectives.

Figs. 8.10 and 8.11 illustrate the number of scheduled D2D links and their sum rate versus maximum distance of each D2D link d_{\max} . Unsurprisingly, the number of scheduled D2D links and their sum rate decrease as d_{\max} increases. This is because as d_{\max} becomes higher, channel gain of more D2D links are worse. Therefore, more D2D links cannot satisfy their QoS requirements and are removed from the set of potentially scheduled D2D links, which results in the reduction in the number of scheduled D2D links and their sum rate. As d_{\max} is small, the number of scheduled D2D links due to the "Conv. SE" algorithm is much smaller than that of other algorithms, and this gap is reduced as d_{\max} is higher.

Even the "Conv. SE" algorithm can schedule a smaller number of D2D links than other algorithms, the sum rate of scheduled D2D links from this algorithm is close to that of the ICA

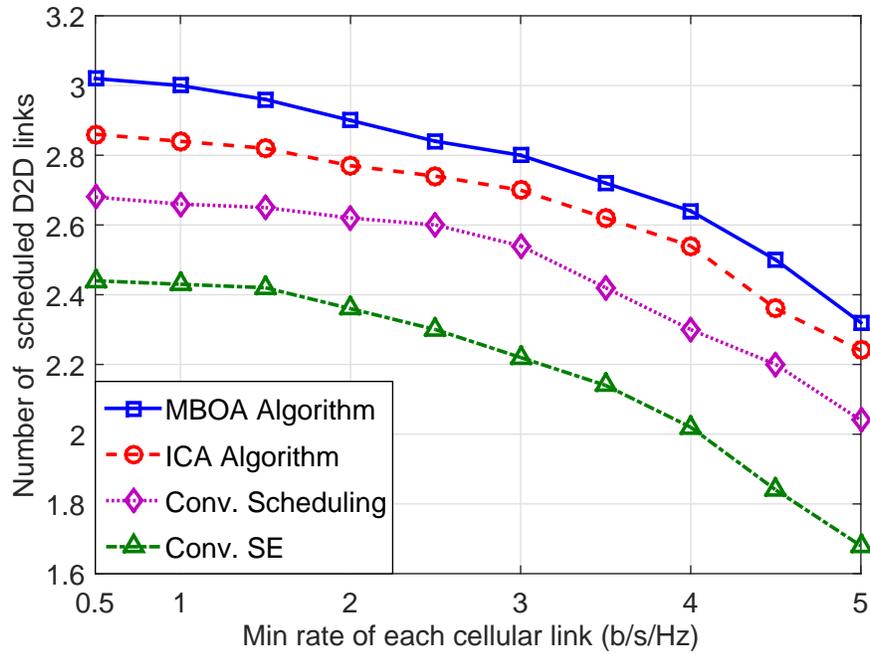


Figure 8.8 – Number of admitted D2D links versus R_c^{\min}

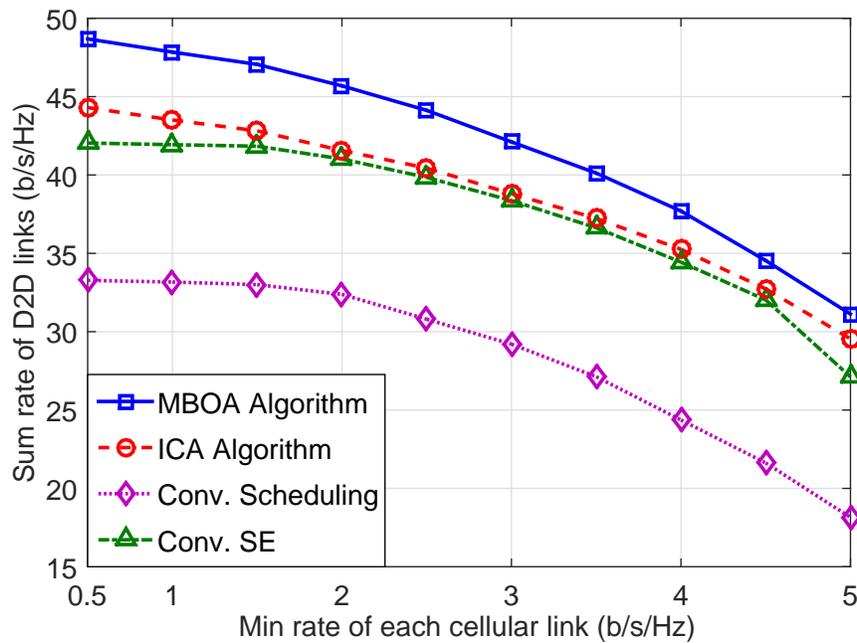


Figure 8.9 – Sum rate of D2D links versus minimum rate of cellular link R_c^{\min}

algorithm, slightly smaller than that of the MBOA algorithm, and much higher than the sum rate due to the “Conv. Scheduling” algorithm. This is because the “Conv. SE” algorithm aims at maximizing the sum rate of the scheduled D2D links while all other algorithms attempt to schedule as many D2D links as possible. Therefore, these algorithms can schedule more D2D links as d_{\max} is

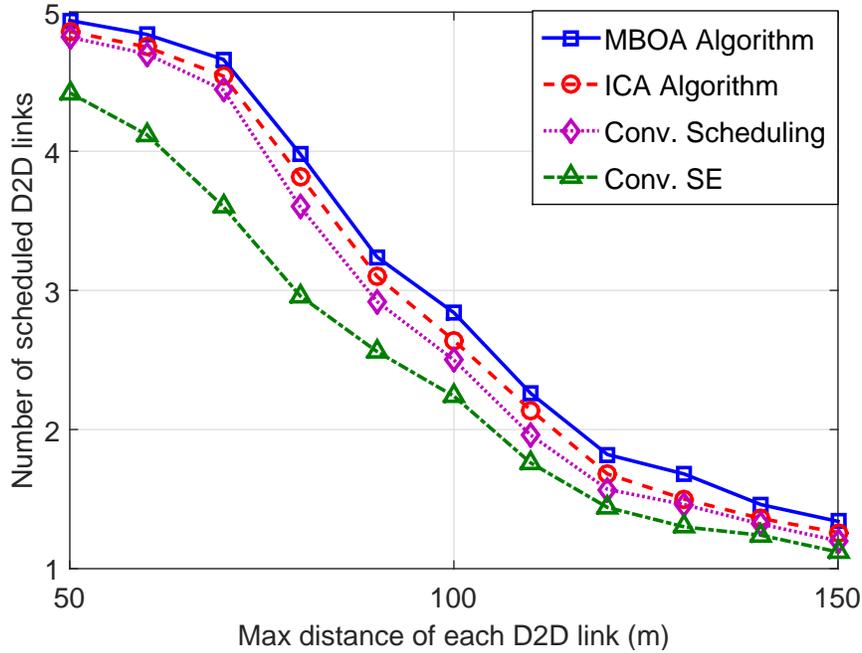


Figure 8.10 – Number of admitted D2D links versus d_{max}

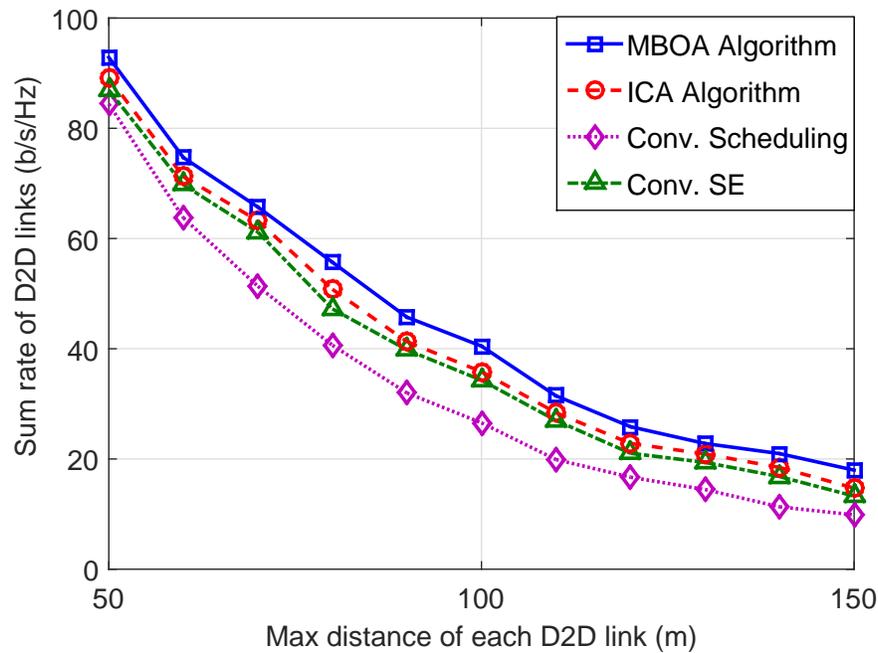


Figure 8.11 – Sum rate of D2D links versus maximum distance of each D2D link d_{max}

small. Moreover, as d_{max} increases, the numbers of scheduled D2D links due to MBOA, ICA, and "Conv. Scheduling" algorithms decrease significantly and the gaps in this performance metric between these three algorithms and the "Conv. SE" algorithm become smaller. Overall, the proposed algorithms outperform the two conventional algorithms in both performance metrics.

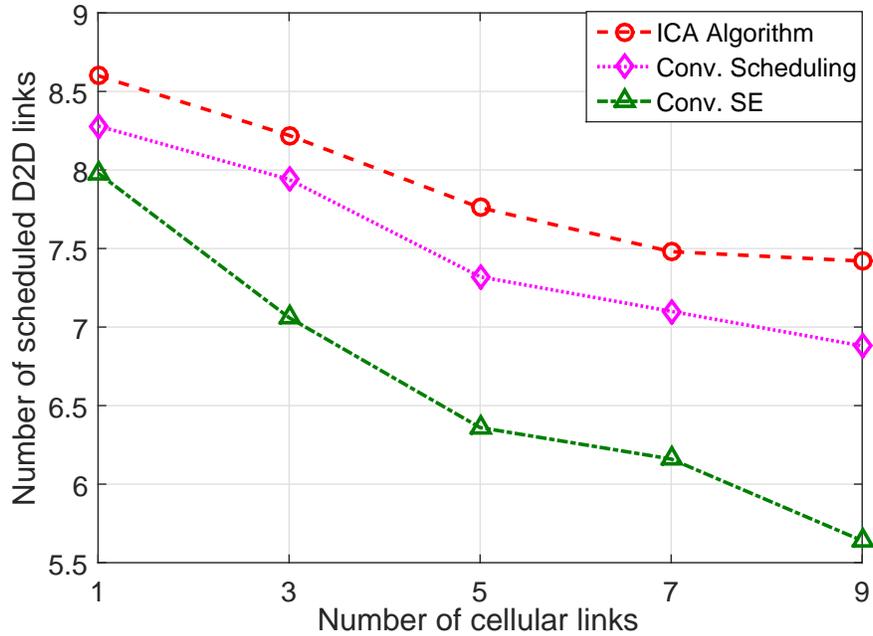


Figure 8.12 – Number of admitted D2D links versus number of cellular links K

Figs. 8.12 and 8.13 demonstrate the number of scheduled D2D links and their sum rate versus the number of cellular links K , respectively. Both number of scheduled D2D links and their sum rate decreases as K increases and the ICA algorithm performs the best in both performance metrics. This is because as K increases, more radio resources must be assigned to the cellular links which results in less radio resources available for D2D links. Accordingly, the number of scheduled D2D links and their sum rate become smaller accordingly. Moreover, the “Conv. Scheduling” algorithm can schedule more D2D links compared to the “Conv. SE” algorithm; however, the sum rate obtained from the “Conv. Scheduling” algorithm is smaller than that due to the “Conv. SE” algorithm.

Interestingly, as K becomes higher, the performance gaps between the three algorithms in both figures become larger where the proposed ICA algorithm always outperforms the two conventional algorithms. These results show that the ICA algorithm is efficient in maximizing both the number of scheduled D2D links and their sum rate. Moreover, with larger K , the available radio resources for D2D links become more limited, and the scheduling and resource allocation tasks become more challenging. This explains the increasing performance gaps between the ICA algorithm and the conventional algorithms with higher value of K .

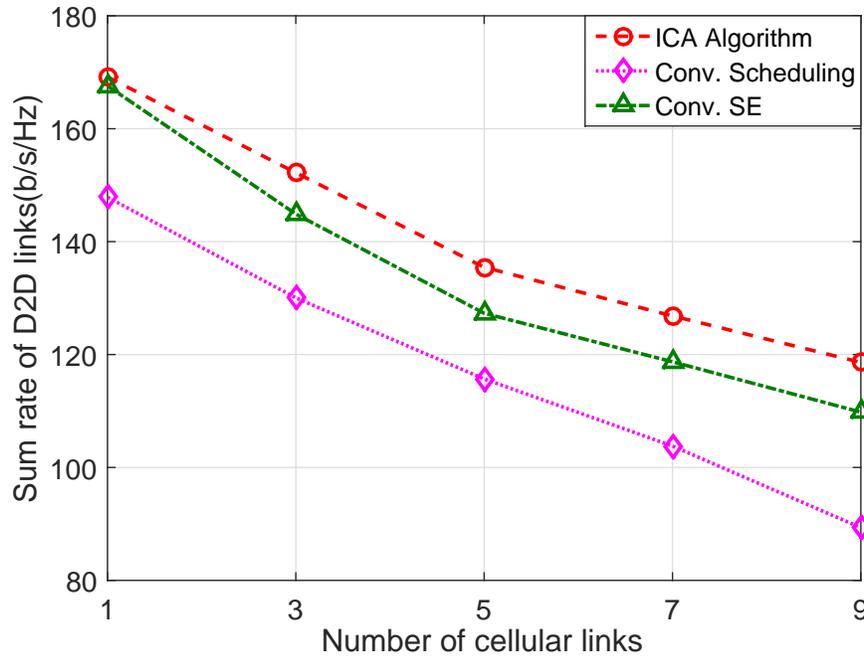


Figure 8.13 – Sum rate of D2D links versus number of cellular links K

8.9 Conclusion

We have studied the joint prioritized scheduling and resource allocation for OFDMA-based wireless networks, which aims to maximize the number of scheduled non-prioritized links and their sum rate simultaneously. Specifically, we have formulated a single-stage optimization problem, which simultaneously maximize number of non-prioritized links scheduled and their sum rate. The hardness of this problem is studied. We have proposed the MBOA algorithm and a low-complexity iterative convex approximation (ICA) algorithm to solve it. We then describe the possible implementation of the proposed algorithm in the current D2D communication in cellular network. Numerical studies have confirmed that the proposed algorithms can be applicable in the CSI error scenarios with negligible performance lost. It also illustrates that our proposed algorithms significantly outperform the “Conv. SE” and “Conv. Scheduling” algorithms.

8.10 Appendices

8.10.1 Proof of Proposition 8.1

Let $(\boldsymbol{\rho}^*, \mathbf{p}^*, \mathbf{s}_{\mathcal{L}}^*)$ be the optimal solution of problem (8.5). We prove Proposition 8.1 by contradiction as follows. Assume that $(\boldsymbol{\rho}^*, \mathbf{p}^*, \mathbf{s}_{\mathcal{L}}^*)$ does not schedule the maximum number of non-prioritized links, then there exists a feasible solution $(\boldsymbol{\rho}', \mathbf{p}', \mathbf{s}'_{\mathcal{L}})$ which schedules more links than that in the optimal solution. Because $\mathbf{s}_{\mathcal{L}}$ is a binary vector, we have $\sum_{l \in \mathcal{L}} s'_l \geq \sum_{l \in \mathcal{L}} s_l^* + 1$. We then have the following inequalities

$$\begin{aligned}
& \alpha \sum_{l \in \mathcal{L}} s'_l + \sum_{l \in \mathcal{L}} R_l(\boldsymbol{\rho}', \mathbf{p}') \\
& \geq \alpha \sum_{l \in \mathcal{L}} s'_l \geq \alpha (\sum_{l \in \mathcal{L}} s_l^* + 1) \\
& \geq \alpha \sum_{l \in \mathcal{L}} s_l^* + LR \geq \alpha \sum_{l \in \mathcal{L}} s_l^* + \sum_{l \in \mathcal{L}} R_l(\boldsymbol{\rho}^*, \mathbf{p}^*).
\end{aligned} \tag{8.42}$$

From the above inequalities, the new feasible solution $(\boldsymbol{\rho}', \mathbf{p}', \mathbf{s}'_{\mathcal{L}})$ achieves higher objective value than that due to the optimal solution which contradicts with the assumption. Therefore, problem (8.5) admits the maximum number of non-prioritized links.

Assume that there is a feasible solution $(\boldsymbol{\rho}', \mathbf{p}', \mathbf{s}'_{\mathcal{L}})$ which schedules the maximum number of non-prioritized links and returns a higher sum rate than that due to $(\boldsymbol{\rho}^*, \mathbf{p}^*, \mathbf{s}_{\mathcal{L}}^*)$. Then, we have

$$\alpha \sum_{l \in \mathcal{L}} s'_l + \sum_{l \in \mathcal{L}} R_l(\boldsymbol{\rho}', \mathbf{p}') \geq \alpha \sum_{l \in \mathcal{L}} s_l^* + \sum_{l \in \mathcal{L}} R_l(\boldsymbol{\rho}^*, \mathbf{p}^*). \tag{8.43}$$

Therefore, $(\boldsymbol{\rho}^*, \mathbf{p}^*, \mathbf{s}_{\mathcal{L}}^*)$ is not the optimal solution of problem (8.5) which contradicts with the assumption. Hence, the optimal solution of problem (8.5) simultaneously maximizes the number of scheduled non-prioritized links and their sum rate.

8.10.2 Proof of Proposition 8.2

Assume that $(\boldsymbol{\rho}^*, \mathbf{p}^*, \mathbf{s}_{\mathcal{L}}^*)$ is an optimal solution of problem (8.5). Hence, $(\mathbf{p}^*, \mathbf{s}_{\mathcal{L}}^*)$ is a feasible solution of problem (8.10). Now we assume $(\mathbf{p}^*, \mathbf{s}_{\mathcal{L}}^*)$ is an optimal solution of problem (8.10) and it satisfies the condition that each subchannel is assigned to at most one prioritized link. Therefore, we propose

the following subchannel assignment vector $\boldsymbol{\rho}^*$

$$\rho_k^{n^*} = \begin{cases} 1, & \text{if } p_k^{n^*} > 0 \\ 0, & \text{otherwise.} \end{cases} \quad (8.44)$$

We have $(\boldsymbol{\rho}^*, \mathbf{p}^*, \mathbf{s}_{\mathcal{L}}^*)$ as a feasible solution of problem (8.5). Therefore, $(\mathbf{p}^*, \mathbf{s}_{\mathcal{L}}^*)$ is an optimal solution of problem (8.10) iff $(\boldsymbol{\rho}^*, \mathbf{p}^*, \mathbf{s}_{\mathcal{L}}^*)$ is an optimal solution of problem (8.5), where $\boldsymbol{\rho}^*$ is determined by (8.44), which means that the optimal solution of problem (8.10) is also the optimal solution of problem (8.5).

8.10.3 Proof of Proposition 8.5

We denote $\mathcal{G}_f = \{\mathbf{x} \mid \mathbf{x} \text{ satisfies constraints (8.18b)–(8.18e)}\}$ and $\mathcal{H}_f = \{\mathbf{x} \mid \mathbf{x} \text{ satisfies constraints (8.18f)–(8.18g)}\}$. Problem (8.18) can be written as follows:

$$\begin{aligned} & \max_{\mathbf{x} \succeq \mathbf{0}} f(\mathbf{x}) \\ & \text{s.t. } \mathbf{x} \in \mathcal{G}_f \cap \mathcal{H}_f. \end{aligned} \quad (8.45)$$

According to Propositions 8.3 and 8.4, \mathcal{G}_f and \mathcal{H}_f are normal and reverse normal sets, respectively. Moreover, $f(\mathbf{x})$ is an increasing function. Therefore, as stated in Definition 8.6, (8.45) is a monotonic optimization problem, which means that (8.18) is a monotonic optimization problem.

8.10.4 Proof of Theorem 8.1

Let $(\mathbf{t}^*, \mathbf{s}_{\mathcal{L}}^*, \mathbf{z}^*)$ and \mathbf{p}^* be the optimal solution of problem (8.18) and its corresponding power allocation vector obtained from the MBOA algorithm, respectively. If $\mathbf{s}_{\mathcal{L}}^*$ is integer, Theorem 1 is proved. If $\mathbf{s}_{\mathcal{L}}^*$ is fractional, we denote $\mathbf{s}_{\mathcal{L}}^0$ and \mathbf{p}^0 as the optimal scheduling solution and its corresponding power allocation, respectively, of the original scheduling and resource allocation

problem $(\mathbf{s}_{\mathcal{L}}^0 \neq \mathbf{s}_{\mathcal{L}}^*)$. We calculate $\mathbf{z}^0 = [z_{mn}^0]_{\forall m \in \mathcal{M}, \forall n \in \mathcal{N}}$ and $\mathbf{t}^0 = [t_l^0]_{\forall l \in \mathcal{L}}$ as

$$z_{mn}^0 = \bar{\Gamma}_m^n(\mathbf{p}^0), \quad (8.46)$$

$$t_l^0 = \begin{cases} 0, & \text{if } s_l^0 = 1 \\ R_l^{\min}, & \text{otherwise.} \end{cases} \quad (8.47)$$

Since $(\mathbf{t}^0, \mathbf{s}_{\mathcal{L}}^0, \mathbf{z}^0)$ is a feasible solution of problem (8.18), we have

$$\alpha \sum_{l \in \mathcal{L}} q(s_l^*) + \sum_{n \in \mathcal{N}, l \in \mathcal{L}} z_{ln}^* > \alpha \sum_{l \in \mathcal{L}} q(s_l^0) + \sum_{n \in \mathcal{N}, l \in \mathcal{L}} z_{ln}^0. \quad (8.48)$$

Since we can choose α sufficiently large, z_{ln}^* and z_{ln}^0 are bounded. We have

$$\sum_{l \in \mathcal{L}} q(s_l^*) \geq \sum_{l \in \mathcal{L}} q(s_l^0). \quad (8.49)$$

Let $\bar{\mathbf{s}}_{\mathcal{L}} = [\bar{s}_l]_{\forall l \in \mathcal{L}}$ be the scheduling solution of the MBOA algorithm after applying the rounding procedure, which is determined as follows:

$$\bar{s}_l = \begin{cases} 1, & \text{if } s_l^* \geq 1 - \epsilon \\ 0, & \text{otherwise.} \end{cases} \quad (8.50)$$

Therefore, we arrive at the following inequality

$$\sum_{l \in \mathcal{L}} q(\bar{s}_l) + L \frac{e^{Q(1-\epsilon)} - 1}{e^Q - 1} \geq \sum_{l \in \mathcal{L}} q(s_l^*). \quad (8.51)$$

Since $Q > \ln L/\epsilon$, we have $L \frac{e^{Q(1-\epsilon)} - 1}{e^Q - 1} < 1$. Therefore, we have

$$\sum_{l \in \mathcal{L}} q(\bar{s}_l) + 1 > \sum_{l \in \mathcal{L}} q(s_l^*) \geq \sum_{l \in \mathcal{L}} q(s_l^0). \quad (8.52)$$

As $q(\bar{s}_l)$ and $q(s_l^0)$ are integers $\forall l \in \mathcal{L}$, we arrive at

$$\sum_{l \in \mathcal{L}} q(\bar{s}_l) \geq \sum_{l \in \mathcal{L}} q(s_l^0), \quad (8.53)$$

which means the MBOA algorithm can schedule the maximum number of non-prioritized links.

Chapter 9

Conclusions and Future Research Directions

In this chapter, we summarize our research contributions and discuss some potential directions for further research.

9.1 Major Research Contributions

In the first contribution [96, 97, 124], we have investigated a spectrum-efficient resource allocation for D2D communication in a cellular network. Specifically, we have presented the subchannel assignment and power control problem which aims at maximizing the weighted system sum-rate while guaranteeing the minimum required rates of individual cellular and D2D links. We have derived an optimal power allocation for a given subchannel assignment which has been used to transform the original problem to a subchannel assignment problem. Then, we have developed an optimal BnB algorithm and an iterative graph-based algorithm to solve the subchannel problem. These proposed algorithms have been shown to perform significantly better than the other state-of-the-art design in the literature.

In the second contribution [19, 27, 77, 115], we have studied a general energy-efficient resource allocation problem for D2D communication. In particular, we have formulated a subchannel as-

signment and power allocation problem which targets to maximize the minimum weighted EE of the D2D links while maintaining the minimum data rates of individual cellular links. We have then characterized the optimal power allocation solution for a cellular link and we have exploited this result to transform the original resource allocation problem into the resource allocation problem for only D2D links. Moreover, we have employed the max-min fractional programming technique to iteratively transform the resource allocation problem to a Mixed Integer Nonlinear Programming (MINLP) problem, which can be solved by dual decomposition based and BnB algorithms. Finally, we have described the centralized and distributed implementation with limited message passing to execute the proposed algorithms.

In the third contribution [144], we have designed the joint mode selection, subchannel assignment, and power control problem for relay-based D2D communication in cellular networks which aims at maximizing the system sum rate considering the minimum rate constraints of cellular and D2D links. Specifically, we have formulated the mode selection and resource allocation problem as an MINLP (Mixed-Integer Non-Linear Programming) problem. We have investigated the optimal power allocation for a given mode selection and subchannel allocation based on which the original resource allocation problem can be transformed into a resource allocation problem, which can be solved optimally by the Hungarian method. Finally, we have shown that the proposed design significantly outperforms existing D2D communication schemes.

In the final contribution [131, 145], we have addressed the joint scheduling and resource allocation design for D2D communication in cellular network where cellular links are more prioritized in comparison with D2D links. To solve the underlying problem, we have proposed a monotonic-based algorithm which asymptotically achieves the optimal solution. To design a low-complexity algorithm, we have developed a convex approximation algorithm which sequentially performs power allocation and link removal in each iteration. Finally, we have described how the proposed algorithms can be implemented in the future cellular network system.

9.2 Further Research Directions

Our research work in this dissertation focuses on the resource allocation designs for D2D communication in cellular networks for significant enhancements of different performance measures such

as SE and EE of the future wireless cellular networks. The following research directions are of importance and deserve further investigation.

9.2.1 Multi-cell D2D Communication

Generally, a cellular network consists of multiple cells sharing the available spectrum resource. Moreover, when D2D communication is enabled in each cell, each communication link is suffered from co-channel interference from not only the links in the same cell but also the links in the other cells. Efficient resource allocation for multi-cell D2D communication would require the tight coordination of different cells in the system, which can be realized by using backhaul links connecting the BSs. Nevertheless, information exchanges over backhaul links to enable such coordination might experience long transmission delay, which can degrade the performance of the employed resource allocation algorithm. Hence, one potential future research direction is to develop a distributed algorithm for D2D communication in the multi-cell wireless system, in which individual BSs execute properly designed resource allocation and interference mitigation operations based on the observed system states and received information from their neighboring BSs.

9.2.2 MIMO based D2D Communication

MIMO communication is the dominant technology for air interface of current and future wireless communication networks. In the MIMO based wireless system, besides subchannel assignment and power allocation, one has to optimize the beamforming vectors to achieve the best system utility. Such beamforming design can be exploited to mitigate the multi-user and co-channel interference as well. Hence, joint consideration of subchannel assignment, power allocation, and beamforming for D2D communication is an important research direction for our future research.

9.2.3 Full-Duplex based D2D Communication

For the research study conducted in this dissertation, we have only considered the half-duplex (HD) communication setting, in which each node could not transmit and receive data at the same time over the same frequency band. Recently, full-duplex (FD) communication technology, in which each node can receive and transmit a message at the same time over the same frequency

band, has been proved to be feasible and considered for future wireless systems. Theoretically, FD communication can allow us to significantly enhancing the spectral efficiency without requiring more radio spectrum. Additionally, the communication latency can be reduced thanks to the FD technology since feedback of certain signaling information such as CSI can be realized during the data transmission. Therefore, development of effective resource allocation algorithms to support FD D2D communication is another potential research direction.

9.2.4 Robust Resource Allocation for D2D Communication in Imperfect CSI Scenarios

In general, the performance of a resource allocation algorithm depends on the quality of the CSI, which is used to perform the resource allocation. Nevertheless, the CSI available at the BS is usually imperfect due to quantization error, limited training time, channel estimation error, and CSI feedback delay. The conservative design proposed in Chapter 8 may not exploit the maximum potential gain of the system since it is based on the worst-case design. Development of robust resource allocation algorithms which use chance constraints to capture the imperfect CSI will be considered in our future work.

9.2.5 Scheduling and Resource Allocation for Vehicular-to-Vehicular (V2V) Communication

Overall, D2D communication involves the communications between two nearby wireless devices, which can applied to the V2V communication setting if each vehicle is equipped with a communication transceiver. Enabling V2V communication in the road can enable to support many emerging applications for the intelligent transportation system. However, one has to design an appropriate mechanism which schedules the transmissions and allocates radio resources for these V2V communications. This design can be more challenging to tackle in comparison with other communication scenarios because V2V applications such as vehicle platooning, advanced driving, and remote driving typically have very stringent requirements. Hence, autonomous scheduling and resource allocation design for V2V communication is an interesting direction for our future research.

9.3 List of Publications

9.3.1 Journals

- [J4]. T. D. Hoang, and L. B. Le “Joint scheduling and resource allocation for OFDMA-based prioritized wireless networks,” *IEEE Transaction on Wireless Communication*, under minor revision.
- [J3]. T. D. Hoang, L. B. Le, and T. Le-Ngoc, “Joint mode selection and resource allocation for relay-based D2D communications,” *IEEE Communications Letters*, vol. 21, no. 2, pp. 398–401, February 2017.
- [J2]. T. D. Hoang, L. B. Le, and T. Le-Ngoc, “Resource allocation for D2D underlaid cellular networks using graph-based approach,” *IEEE Transaction on Wireless Communication*, vol. 15, no. 10, pp. 7099–7113, October 2016.
- [J1]. T. D. Hoang, L. B. Le, and T. Le-Ngoc, “Energy-efficient resource allocation for D2D communications in cellular networks,” *IEEE Transaction on Vehicular Technology*, vol. 65, no. 9, pp. 6972–6986, September 2016.

9.3.2 Conferences

- [C7]. T. D. Hoang, and L. B. Le, “Content caching for heterogeneous small-cell networks with QoS-aware BS association,” *Proc. IEEE VTC’2017*, October 2017.
- [C6]. T. D. Hoang, L. B. Le, and T. Le-Ngoc, “Joint prioritized link scheduling and resource allocation for OFDMA-based wireless networks,” *Proc. IEEE ICC’2016*, June 2016.
- [C5]. T. D. Hoang, L. B. Le, and T. Le-Ngoc, “Dual decomposition method for energy-efficient resource allocation in D2D communications underlying cellular networks,” *Proc. IEEE GLOBECOM’2015*, December 2015.
- [C4]. T. D. Hoang, L. B. Le, and T. Le-Ngoc, “Radio resource management for optimizing energy efficiency of D2D Communications in cellular networks,” *Proc. IEEE PIMRC’2015*, September 2015.

- [C3]. T. D. Hoang, L. B. Le, and T. Le-Ngoc, "Energy-Efficient resource allocation for D2D communication in cellular networks ," in *Proc. IEEE ICC'2015*, June 2015.
- [C2]. T. D. Hoang, L. B. Le, and T. Le-Ngoc, "Resource allocation for D2D communications under proportional fairness," in *Proc. IEEE GLOBECOM'2014*, December 2014.
- [C1]. T. D. Hoang, L. B. Le, and T. Le-Ngoc, "Joint subchannel and power allocation for D2D communications in cellular networks," in *Proc. IEEE WCNC'2014*, April 2014.

Références

- [1] D. Wu, J. Wang, R. Q. Hu, Y. Cai, and L. Zhou, “Energy-efficient resource sharing for mobile Device-to-Device multimedia communications,” *IEEE Trans. Veh. Technol.*, vol. 63, pp. 2093–2103, June 2014.
- [2] D. Wu, Y. Cai, R. Q. Hu, and Y. Qian, “Dynamic distributed resource sharing for mobile D2D communications,” *IEEE Trans. Wireless Commun.*, vol. 14, pp. 5417–5429, October 2015.
- [3] Ericson, *5G radio access, Research and Vision*. White Paper, 2013.
- [4] M. S. Corson, R. Laroia, J. Li, V. Park, T. Richardson, and G. Tsirtsis, “Toward proximity-aware internetworking,” *IEEE Commun. Mag.*, vol. 17, pp. 26–33, December 2010.
- [5] 3GPP, “3rd generation partnership project; Technical specification group radio access network; Study on LTE Device to Device proximity services; Radio aspects (release 12),” in *TR 36.843 V12.0.1*, 2014.
- [6] 3GPP, “3rd generation partnership project; Technical specification group services and system aspects; Proximity-based services (Prose); stage 3 (release 13),” *TS 23.303 V13.1.1*, September 2015.
- [7] 3GPP, “3rd generation partnership project; Technical specification group services and system aspects; Proximity-based services (Prose); stage 2 (release 14),” *TS 23.303 V14.1.0*, December 2016.
- [8] 3GPP, “3rd generation partnership project; Technical specification group services and system aspects; Proximity-based services (Prose); stage 2 (release 12),” *TS 23.303 V12.0.0*, June 2014.
- [9] 3GPP, “3rd generation partnership project; Technical specification group services and system aspects; Service requirements for machine-type communications (MTC),” *TS 22.368 V13.2.0*, September 2016.
- [10] A. Asadi, Q. Wang, and V. Mancuso, “A survey on Device-to-Device communication in cellular networks,” *IEEE Communications Surveys Tutorials*, vol. 16, pp. 1801–1819, April 2014.
- [11] L. Chen, Y. Huang, F. Xie, Y. Gao, L. Chu, H. He, Y. Li, F. Liang, and Y. Yuan, “Mobile relay in LTE-Advanced systems,” *IEEE Commun. Mag.*, vol. 51, pp. 144–151, November 2013.
- [12] R. B. Yehuda, K. Bendel, A. Freund, and D. Rawitz, “Local ratio: A unified framework for approximation algorithms. In memoriam: Shimon Even 1935–2004,” *ACM Comput. Surv.*, vol. 36, pp. 422–463, December 2004.

- [13] A. Abdelnasser, E. Hossain, and D. I. Kim, "Tier-aware resource allocation in OFDMA macrocell-small cell networks," *IEEE Trans. Commun.*, vol. 63, pp. 695–711, March 2015.
- [14] M. Ebrahimi, M. A. Maddah-Ali, and A. K. Khandani, "Power allocation and asymptotic achievable sum-rates in single-hop wireless networks," in *Proc. IEEE CISS'06*, March 2006.
- [15] C. Roos, T. Terlaky, and J. Vial, *Interior point methods for linear optimization*. Press: Springer, 2005.
- [16] X. Ma, J. Liu, and H. Jiang, "Resource allocation for heterogeneous applications with D2D communication underlaying cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 34, pp. 15–26, January 2016.
- [17] Y. Gu, Y. Zhang, M. Pan, and Z. Han, "Matching and cheating in Device to Device communications underlying cellular networks," *IEEE Journal on Selected Areas in Communications*, vol. 33, pp. 2156–2166, Oct 2015.
- [18] C. Xiong, G. Y. Li, S. Zhang, Y. Chen, and S. Xu, "Energy-efficient resource allocation in OFDMA networks," *IEEE Trans. Wireless Commun.*, vol. 60, pp. 3767–3778, December 2012.
- [19] T. D. Hoang, L. B. Le, and T. Le-Ngoc, "Energy-efficient resource allocation for D2D communications in cellular networks," in *Proc. IEEE ICC'2015*, June 2015.
- [20] J. P. G. Crouzeix and J. A. Ferland, "Algorithms for generalized fractional programming," *Mathematical Programming*, vol. 52, pp. 191–207, 1991.
- [21] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2009.
- [22] Z. Zhou, M. Dong, K. Ota, J. Wu, and T. Sato, "Energy efficiency and Spectral efficiency tradeoff in Device-to-Device (D2D) communications," *IEEE Commun. Lett.*, vol. 3, pp. 485–488, October 2014.
- [23] G. Yu, L. Xu, D. Feng, R. Yin, G. Y. Li, and Y. Jiang, "Joint mode selection and resource allocation for Device-to-Device communications," *IEEE Trans. Commun.*, vol. 62, pp. 3814–3824, November 2014.
- [24] D. Jungnickel, "Chapter "Weighted matchings," in *Graphs, Networks and Algorithms*, Heidelberg, vol. 5, pp. 419–456, Berlin: Springer, 2008.
- [25] T. Kim and M. Dong, "An iterative Hungarian method to joint relay selection and resource allocation for D2D communications," *IEEE Wireless Commun. Lett.*, vol. 3, pp. 625–628, December 2014.
- [26] S. Hayashi and Z. q. Luo, "Spectrum management for interference-limited multiuser communication systems," *IEEE Trans. Information Theory*, vol. 55, pp. 1153–1175, March 2009.
- [27] T. D. Hoang, L. B. Le, and T. Le-Ngoc, "Energy-efficient resource allocation for D2D communications in cellular networks," *IEEE Trans. on Veh. Tech.*, vol. 65, pp. 6972–6986, September 2016.
- [28] M. Andersin, Z. Rosberg, and J. Zander, "Gradual removals in cellular PCS with constrained power control and noise," *ACM/Baltzer Wireless Networks J*, vol. 2, no. 1, pp. 27–43, 1996.

- [29] T. S. Rappaport, *Wireless Communications: Principles and Practice*. Upper Saddle River, NJ: Prentice-Hall, 1996.
- [30] Z. q. Luo and S. Zhang, "Dynamic spectrum management: Complexity and duality," *IEEE J. Sel. Topics Signal Process*, vol. 2, pp. 57–73, February 2008.
- [31] C. H. Yu, K. Doppler, C. B. Ribeiro, and O. Tirkkonen, "Resource sharing optimization for Device-to-Device communication underlaying cellular networks," *IEEE Trans. Wireless Commun.*, vol. 10, pp. 2752–2763, August 2011.
- [32] J. Wang, D. Zhu, C. Zhao, J. C. F. Li, and M. Lei, "Resource sharing of underlaying Device-to-Device and uplink cellular communications," *IEEE Commun. Letters*, vol. 17, pp. 1148–1151, June 2013.
- [33] L. B. Le, "Fair resource allocation for Device-to-Device communications in wireless cellular networks," in *Proc. IEEE Globecom'2012*, pp. 5451–5456, December 2012.
- [34] D. Feng, L. Lu, Y. W. Yi, G. Y. Li, G. Geng, and S. Li, "Device-to-Device communications underlaying cellular networks," *IEEE Trans. Commun.*, vol. 61, pp. 3541–3551, August 2013.
- [35] F. Malandrino, Z. Limani, C. Casetti, and C. F. Chiasserini, "Interference-aware downlink and uplink resource allocation in Hetnets with D2D support," *IEEE Trans. Wireless Commun.*, vol. 14, pp. 2729–2741, January 2015.
- [36] D. Lee, K. Choi, W. Jeon, and D. Jeong, "Two-stage semi-distributed resource management for Device-to-Device communication in cellular networks," *IEEE Trans. Wireless Commun.*, vol. 13, pp. 1908–1920, April 2014.
- [37] M. Zulhasnine, C. Huang, and A. Srinivasan, "Efficient resource allocation for Device-to-Device communication underlaying LTE network," in *Proc. IEEE WiMob'2010*, pp. 368–375, October 2010.
- [38] Z. Zhou, M. Dong, K. Ota, J. Wu, and T. Sato, "Energy efficiency and spectral efficiency tradeoff in Device-to-Device (D2D) communications," *IEEE Commun. Letters*, vol. 3, pp. 485–488, October 2014.
- [39] D. Wu, J. Wang, R. Q. Hu, Y. Cai, and L. Zhou, "Energy-efficient resource sharing for mobile Device-to-Device multimedia communications," *IEEE Trans. Veh. Technol.*, vol. 63, pp. 2093–2103, June 2014.
- [40] W. O. Oduola, X. Li, L. Qian, and Z. Han, "Power control for Device-to-Device communications as an underlay to cellular system," in *Proc. IEEE ICC'2014*, pp. 5257–5262, June 2014.
- [41] C. Gao, X. Sheng, J. Tang, W. Zhang, S. Zou, and G. Mohsen, "Joint mode selection, channel allocation and power assignment for green Device-to-Device communications," in *Proc. IEEE ICC'2014*, pp. 178–183, June 2014.
- [42] W. Zhao and S. Wang, "Resource allocation for Device-to-Device communication underlaying cellular networks: An alternating optimization method," *IEEE Commun. Letter*, vol. PP, pp. 1–1, June 2015.

- [43] Y. Pei and Y.-C. Liang, "Resource allocation for Device-to-Device communications overlaying two-way cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, pp. 3611–3621, July 2013.
- [44] M. Hassan and E. Hossain, "Distributed resource allocation for relay-aided Device-to-Device communication: A message passing approach," *IEEE Trans. Wireless Commun.*, vol. 13, pp. 6326–6341, November 2014.
- [45] T. Kim and M. Dong, "An iterative Hungarian method to joint relay selection and resource allocation for D2D communications," *IEEE Wireless Commun. Lett.*, vol. 3, pp. 625–628, December 2014.
- [46] L. Wei, R. Q. Hu, Y. Qian, and G. Wu, "Energy-efficiency and spectrum-efficiency of multi-hop Device-to-Device communications underlaying cellular networks," *IEEE Trans. Vehi. Tech.*, 2015.
- [47] 3GPP, "3rd generation partnership project; technical specification group services and system aspects; service requirements for the evolved packet system (EPS) (Release 13)," in *TS 22.278 V13.2.0*, 2014.
- [48] I. Mitliagkas, N. D. Sidiropoulos, and A. Swami, "Joint power and admission control for ad-hoc and cognitive underlay networks: Convex approximation and distributed implementation," *IEEE Trans. Wireless Commun.*, vol. 10, pp. 4110–4121, December 2011.
- [49] Y. f. Liu, Y. h. Dai, and Z. q. Luo, "Joint power and admission control via linear programming deflation," *IEEE Trans. Signal Process.*, vol. 61, pp. 1327–1339, March 2013.
- [50] A. T. Hoang and Y. c. Liang, "Downlink channel assignment and power control for cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 3106–3117, August 2008.
- [51] A. T. Hoang, Y. c. Liang, and M. H. Islam, "Power control and channel allocation in cognitive radio networks with primary users' cooperation," *IEEE Trans. Mobile Comput.*, vol. 9, pp. 348–361, March 2010.
- [52] M. V. Nguyen and H. S. Lee, "Effective scheduling in infrastructure-based cognitive radio networks," *IEEE Trans. Mobile Comput.*, vol. 10, pp. 853–868, June 2011.
- [53] D. Feng, L. Lu, Y. W. Yi, G. Y. Li, G. Geng, and S. Li, "Device-to-Device communications underlaying cellular networks," *IEEE Trans. Commun.*, vol. 61, pp. 3541–3551, August 2013.
- [54] W. Zhao and S. Wang, "Resource sharing scheme for Device-to-Device communication underlaying cellular networks," *IEEE Trans. Commun.*, vol. 63, pp. 4838–4848, December 2015.
- [55] H. W. Kuhn, "The Hungarian method for the assignment problem," *Naval Res. Logistics Quart.*, vol. 2, pp. 83–97, March 1955.
- [56] H. Tuy, "Normal sets, polyblocks and monotonic optimization," *Vietnam J. Math.*, vol. 27, no. 4, pp. 277–300, 1999.
- [57] H. Tuy, "Monotonic optimization: Problem and solution approaches," *SIAM J. Optim.*, vol. 11, no. 2, pp. 464–494, 2000.
- [58] Y. J. Zhang, L. Qian, and J. Huang, "Monotonic optimization in communication and networking systems," *Found. Trends Netw.*, vol. 7, no. 1, pp. 1–75, 2012.

- [59] E. Yaacoub and Z. Dawy, "A survey on uplink resource allocation in OFDMA wireless networks," *IEEE Communications Surveys Tutorials*, vol. 14, pp. 322–337, May 2012.
- [60] L. Gao and S. Cui, "Efficient subcarrier, power, and rate allocation with fairness consideration for OFDMA uplink," *IEEE Transactions on Wireless Communications*, vol. 7, pp. 1507–1511, May 2008.
- [61] C. Y. Ng and C. W. Sung, "Low complexity subcarrier and power allocation for utility maximization in uplink OFDMA systems," *IEEE Transactions on Wireless Communications*, vol. 7, pp. 1667–1675, May 2008.
- [62] J. Huang, V. G. Subramanian, R. Agrawal, and R. Berry, "Joint scheduling and resource allocation in uplink OFDM systems for broadband wireless access networks," *IEEE Journal on Selected Areas in Communications*, vol. 27, pp. 226–234, February 2009.
- [63] S. Sadr, A. Anpalagan, and K. Raahemifar, "Radio resource allocation algorithms for the downlink of multiuser OFDM communication systems," *IEEE Communications Surveys Tutorials*, vol. 11, pp. 92–106, Third Quarter 2009.
- [64] J. Huang, V. G. Subramanian, R. Agrawal, and R. A. Berry, "Downlink scheduling and resource allocation for OFDM systems," *IEEE Transactions on Wireless Communications*, vol. 8, pp. 288–296, Jan 2009.
- [65] L. Venturino, N. Prasad, and X. Wang, "Coordinated scheduling and power allocation in downlink multicell OFDMA networks," *IEEE Transactions on Vehicular Technology*, vol. 58, pp. 2835–2848, July 2009.
- [66] T. Wang and L. Vandendorpe, "Iterative resource allocation for maximizing weighted sum min-rate in downlink cellular OFDMA systems," *IEEE Transactions on Signal Processing*, vol. 59, pp. 223–234, Jan 2011.
- [67] K. Son, S. Lee, Y. Yi, and S. Chong, "REFIM: A practical interference management in heterogeneous wireless access networks," *IEEE Journal on Selected Areas in Communications*, vol. 29, pp. 1260–1272, June 2011.
- [68] D. T. Ngo, S. Khakurel, and T. Le-Ngoc, "Joint subchannel assignment and power allocation for OFDMA femtocell networks," *IEEE Transactions on Wireless Communications*, vol. 13, pp. 342–355, January 2014.
- [69] R. Y. Chang, Z. Tao, J. Zhang, and C. C. J. Kuo, "Multicell OFDMA downlink resource allocation using a graphic framework," *IEEE Transactions on Vehicular Technology*, vol. 58, pp. 3494–3507, Sept 2009.
- [70] S. Lin and H. Tian, "Clustering based interference management for QoS guarantees in OFDMA femtocell," in *2013 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 649–654, April 2013.
- [71] S. J. Kim and I. Cho, "Graph-based dynamic channel assignment scheme for femtocell networks," *IEEE Communications Letters*, vol. 17, pp. 1718–1721, September 2013.
- [72] A. N. Zaki and A. O. Fapojuwo, "Optimal and efficient graph-based resource allocation algorithms for multiservice frame-based OFDMA networks," *IEEE Transactions on Mobile Computing*, vol. 10, pp. 1175–1186, Aug 2011.

- [73] H. H. Kha, H. D. Tuan, and H. H. Nguyen, "Fast global optimal power allocation in wireless networks by local D.C. programming," *IEEE Transactions on Wireless Communications*, vol. 11, pp. 510–515, February 2012.
- [74] J. Papandriopoulos and J. S. Evans, "Scale: A low-complexity distributed protocol for spectrum balancing in multiuser DSL networks," *IEEE Transactions on Information Theory*, vol. 55, pp. 3711–3724, August 2009.
- [75] M. Chiang, C. W. Tan, D. P. Palomar, D. O'Neill, and D. Julian, "Power control by geometric programming," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 2640–2651, July 2007.
- [76] X. Lin, J. G. Andrews, A. Ghosh, and R. Ratasuk, "An overview on 3GPP Device-to-Device proximity services," *IEEE Commun. Mag.*, vol. 52, pp. 40–48, April 2014.
- [77] T. D. Hoang, L. B. Le, and T. Le-Ngoc, "Dual decomposition method for energy-efficient resource allocation in D2D communications underlying cellular networks," in *in Proc. IEEE GLOBECOM'2015*, December 2015.
- [78] 3GPP, "3rd generation partnership project; technical specification group services and system aspects; service requirements for the evolved packet system (EPS) (release 13),." December 2014.
- [79] Z. Q. Luo and S. Zhang, "Dynamic spectrum management: Complexity and duality," *IEEE J. Sel. Topics Signal Process*, vol. 2, pp. 57–73, February 2008.
- [80] C. H. Yu, K. Doppler, C. B. Ribeiro, and O. Tirkkonen, "Resource sharing optimization for Device-to-Device communication underlying cellular networks," *IEEE Trans. Wireless Commun*, vol. 10, pp. 2752–2763, August 2011.
- [81] H. Song, J. Y. Ryu, W. Choi, and R. Schober, "Joint power and rate control for Device-to-Device communications in cellular systems," *IEEE Trans. Wireless Commun*, vol. 14, pp. 5750–5762, October 2015.
- [82] D. Feng, G. Yu, C. Xiong, Y. Yuan-Wu, G. Y. Li, G. Feng, and S. Li, "Mode switching for energy-efficient Device-to-Device communications in cellular networks," *IEEE Trans. Wireless Commun*, vol. 14, pp. 6993–7003, December 2015.
- [83] M. Azam, M. Ahamad, M. Naeem, M. Iqbal, A. S. Khwaja, A. Anpalagan, and S. Quaser, "Joint admission control, mode selection and power allocation in D2D communication systems," *IEEE Trans. Veh. Tech*, vol. 65, pp. 7322–7333, October 2016.
- [84] L. Wang and G. L. Stuber, "Pairing for resource sharing in cellular Device-to-Device underlays," *IEEE Network*, vol. 30, pp. 122–128, March 2016.
- [85] Y. Ren, F. Liu, Z. Liu, C. Wang, and Y. Ji, "Power control in D2D-based vehicular communication networks," *IEEE Trans. Veh. Tech*, vol. 64, pp. 5547–5562, December 2015.
- [86] L. A.-Kanj, H. V. Poor, and Z. Dawy, "Optimal cellular offloading via D2D communication networks with fairness constraints," *IEEE Trans. Wireless Commun*, vol. 13, pp. 4628–4643, August 2014.
- [87] F. Malandrino, Z. Limani, C. Casetti, and C. F. Chiasserini, "Interference-aware downlink and uplink resource allocation in hetnets with D2D support," *IEEE Trans. Wireless Commun*, vol. 14, pp. 2729–2741, January 2015.

- [88] S. Maghsudi and S. Stanczak, "Hybrid centralized-distributed resource allocation for Device-to-Device communication underlaying cellular networks," *IEEE Trans. Veh. Tech.*, vol. 65, pp. 2481–2495, April 2016.
- [89] R. Yin, C. Zhong, G. Yu, Z. Zhang, K. K. Wong, and X. Chen, "Joint spectrum and power allocation for D2D communications underlaying cellular networks," *IEEE Trans. Veh. Tech.*, vol. 65, pp. 2182–2195, April 2016.
- [90] H. Zhang, L. Song, and Z. Han, "Radio resource allocation for Device-to-Device underlay communication using hypergraph theory," *IEEE Trans. Wireless Commun.*, vol. 15, pp. 4852–4861, July 2016.
- [91] R. Zhang, X. Cheng, L. Yang, and B. Jiao, "Interference graph based resource allocation (INGRA) for D2D communications underlaying cellular networks," *IEEE Trans. Veh. Technol.*, vol. 64, pp. 3844–3850, August 2015.
- [92] C. Xu, L. Song, Z. Han, Q. Zhao, X. Wang, X. Cheng, and B. Jiao, "Efficiency resource allocation for Device-to-Device underlay communication systems: A reverse iterative combinatorial auction based approach," *IEEE J. Sel. Areas Commun.*, vol. 31, pp. 348–358, September 2013.
- [93] Q. Ye, M. Al-Shalash, C. Caramanis, and J. G. Andrews, "Distributed resource allocation in Device-to-Device enhanced cellular networks," *IEEE Trans. Commun.*, vol. 63, pp. 441–454, February 2015.
- [94] R. Yin, G. Yu, H. Zhang, Z. Zhang, and G. Y. Li, "Pricing-based interference coordination for D2D communications in cellular networks," *IEEE Trans. Wireless Commun.*, vol. 14, pp. 1519–1532, March 2015.
- [95] Y. Zhao, Y. Li, Y. Cao, T. Jiang, and N. Ge, "Social-aware resource allocation for Device-to-Device communications underlaying cellular networks," *IEEE Trans. Wireless Commun.*, vol. 14, pp. 6621–6634, December 2015.
- [96] T. D. Hoang, L. B. Le, and T. Le-Ngoc, "Joint subchannel and power allocation for D2D communications in cellular networks," in *Proc. IEEE WCNC'2014*, April 2014.
- [97] T. D. Hoang, L. B. Le, and T. Le-Ngoc, "Resource allocation for D2D communications under proportional fairness," in *in Proc. IEEE Globecom'2014*, December 2014.
- [98] D. Gale and L. S. Shapley, "College admissions and the stability of marriage," *Amer. Math. Monthly*, vol. 69, pp. 9–15, January 1962.
- [99] L. C. Lau, R. Ravi, and M. Singh, *Iterative methods in combinatorial optimization*. Cambridge: Univ. Press, 2011.
- [100] K. Doppler, M. Rinne, C. Wijting, C. B. Ribeiro, and K. Hugl, "Device-to-Device communication as an underlay to LTE-Advanced networks," *IEEE Commun. Mag.*, vol. 47, pp. 42–49, December 2009.
- [101] M. S. Corson, R. Laroia, J. Li, V. Park, T. Richardson, and G. Tsirtsis, "Toward proximity-aware internetworking," *IEEE Wireless Commun.*, vol. 17, pp. 26–33, December 2010.
- [102] G. Fodor, E. Dahlman, G. Mildh, S. Parkvall, N. Reider, G. Miklos, and Z. Turanyi, "Design aspects of network assisted Device-to-Device communications," *IEEE Commun. Mag.*, vol. 50, pp. 170–177, March 2012.

- [103] E. Hossain, L. B. Le, and D. Niyato, *Radio resource management in Multi-Tier Cellular Wireless Networks*. New York, NY, USA: Wiley, 2013.
- [104] Z. Hasan, H. Boostanimehr, and V. K. Bhargava, "Green cellular networks: A survey, some research issues and challenges," *IEEE Commun. Surveys & Tutorials*, vol. 13, pp. 524–540, November 2011.
- [105] G. Y. Li, Z. Xu, C. Xiong, C. Yang, S. Zhang, Y. Chen, and S. Xu, "Energy-efficient wireless communications: Tutorial, survey, and open issues," *IEEE Wireless Commun*, vol. 18, pp. 28–35, December 2011.
- [106] V. N. Ha and L. B. Le, "Fair resource allocation for OFDMA femtocell networks with macrocell protection," *IEEE Trans. Veh. Technol.*, vol. 63, no. 3, pp. 1388–1401, 2013.
- [107] C. Xiong, G. Y. Li, S. Zhang, Y. Chen, and S. Xu, "Energy-efficient resource allocation in OFDMA networks," *IEEE Trans. Commun*, vol. 60, pp. 3767–3778, December 2012.
- [108] C. Isheden, Z. Chong, E. Jorswieck, and G. Fettweis, "Framework for link-level energy efficiency optimization with informed transmitter," *IEEE Trans. Wireless Commun*, vol. 11, pp. 2946–2957, August 2012.
- [109] S. Khakurel, L. Musavian, and T. Le-Ngoc, "Energy-efficient resource and power allocation for uplink multi-user ofdma systems," in *Proc. IEEE PIMRC'2012*, pp. 357–361, September 2012.
- [110] M. Ismail, A. Gamage, and W. Z. X. Shen, "Energy efficient uplink resource allocation in a heterogeneous wireless medium," in *Proc. IEEE ICC'2014*, pp. 5275–5280, June 2014.
- [111] C. H. Yu, O. Tirkkonen, K. Doppler, and C. Ribeiro, "On the performance of Device-to-Device underlay communication with simple power control," in *Proc. IEEE VTC'2009*, pp. 1–5, April 2009.
- [112] J. Wang, D. Zhu, C. Zhao, G. Y. Li, and M. Lei, "Resource sharing of underlying Device-to-Device and uplink cellular communications," *IEEE Commun. Lett.*, vol. 17, pp. 1148–1151, June 2013.
- [113] L. B. Le, "Fair resource allocation for Device-to-Device communications in wireless cellular networks," in *Proc. IEEE Globecom'2012*, pp. 5451–5456, December 2012.
- [114] F. Wang, C. Xu, L. Song, Z. Han, and B. Zhang, "Energy-efficient radio resource and power allocation for Device-to-Device communication underlying cellular networks," in *Proc. IEEE WCSP'2012*, October 2012.
- [115] T. D. Hoang, L. B. Le, and T. Le-Ngoc, "Radio resource management for optimizing energy efficiency of D2D communications in cellular networks," in *Proc. IEEE PIMRC'2015*, September 2015.
- [116] W. O. Oduola, X. Li, L. Qian, and Z. Han, "Power control for Device-to-Device communications as an underlay to cellular system," in *Proc. IEEE ICC'2014*, pp. 5257–5262, June 2014.
- [117] C. Gao, X. Sheng, J. Tang, W. Zhang, S. Zou, and G. Mohsen, "Joint mode selection, channel allocation and power assignment for green Device-to-Device communications," in *Proc. IEEE ICC'2014*, pp. 178–183, June 2014.

- [118] D. P. Bertsekas, *Nonlinear programming*. Cambridge, MA, USA: MIT Press, 2nd ed., 2008.
- [119] Y. Xu, J. Wang, Q. Wu, A. Anpalagan, and Y. d. Yao, "Opportunistic spectrum access in cognitive radio networks: Global optimization using local interaction games," *IEEE J. Sel. Topics Signal Process*, vol. 6, pp. 180–194, April 2012.
- [120] Y. Xu, J. Wang, Q. Wu, A. Anpalagan, and Y. d. Yao, "Opportunistic spectrum access in unknown dynamic environment: A game-theoretic stochastic learning solution," *IEEE Trans. Wireless Commun*, vol. 11, pp. 1380–1391, April 2012.
- [121] A. H. Land and A. G. Doig, "An automatic method of solving discrete programming problems," *Econometrica: Journal of the Econometric Society*, vol. 28, pp. 497–520, July 1960.
- [122] A. Nemirovski, "Interior point polynomial methods in convex programming," in *Lecture notes, Georgia Inst. of Techno., Atlanta, GA., USA*, ch. 10, p. 169, Springer Berlin Heidelberg, 2004.
- [123] J. b. Hiriart-Urruty and C. Lemarechal, *Fundamental of Convex Analysis*. Heigelbergeg: Springer Verlag, 2001.
- [124] T. D. Hoang, L. B. Le, and T. Le-Ngoc, "Resource allocation for D2D underlaid cellular networks using graph-based approach," *IEEE Trans. on Wireless Communication*, vol. 15, pp. 7099–7113, October 2016.
- [125] H. Nishiyama, M. Ito, and N. Kato, "Relay-by-smartphone: realizing multihop Device-to-Device communications," *IEEE Commun. Mag.*, vol. 52, pp. 56–65, April 2014.
- [126] G. C. Holmes, "The use of hyperbolic cosines in solving cubic polynomials," *The Math. Gazette*, vol. 86, pp. 473–477, November 2002.
- [127] Ericson, "5G radio access, capabilities and technologies".
- [128] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag*, vol. 52, pp. 74–80, February 2014.
- [129] 3GPP , "3rd generation partnership project; Technical specification group services and system aspects; Service requirements for the evolved packet system (EPS) (release 13)," *TS 22.278 V13.2.0*, December 2016.
- [130] S. Bashar and Z. Ding, "Admission control and resource allocation in a heterogeneous ofdma wireless network," *IEEE Trans. Wireless Commun*, vol. 8, pp. 4200–4210, August 2009.
- [131] T. D. Hoang, L. B. Le, and T. Le-Ngoc, "Joint prioritized link scheduling and resource allocation for OFDMA-based wireless networks," in *Proc. IEEE ICC'2016*, May 2016.
- [132] V. N. Ha, L. B. Le and D. N. Dao, "Coordinated multipoint transmission design for Cloud-RANs with limited fronthaul capacity constraints," *IEEE Trans. Vehi. Tech.*, vol. 65, no. 9, pp. 7432–7447, September 2016.
- [133] W. Dinkelbach, "On nonlinear fractional programming," *Management Science*, vol. 13, pp. 492–498, March 1967.
- [134] A. Nemirovski, "Interior point polynomial methods in convex programming," *Lecture notes, Georgia Inst*, vol. 10, 2004.

- [135] 3GPP, “Evolved universal terrestrial radio access (E-UTRA); radio resource control (RRC) protocol specification,” *TS 36.331 v14.0.0*, December 2016.
- [136] 3GPP, “Evolved universal terrestrial radio access (E-UTRA); medium access control (MAC) protocol specification,” *TS 36.321 v14.0.0*, December 2016.
- [137] 3GPP, “Evolved universal terrestrial radio access (E-UTRA); physical layer procedures,” *TS 36.213 v14.0.0*, December 2016.
- [138] 3GPP, “Evolved universal terrestrial radio access (E-UTRA); multiplexing and channel coding,” *TS 36.212 v14.0.0*, December 2016.
- [139] N. Zorba and A. I. Neira, “Robust power allocation schemes for multibeam opportunistic transmission strategies under quality of service constraints,” *IEEE J. Sel. Areas Commun.*, vol. 26, pp. 1025–1034, August 2008.
- [140] E. Bjornson, G. Zheng, M. Bengtsson, and B. Ottersten, “Robust monotonic optimization framework for multicell MISO systems,” *IEEE Trans. Signal Process.*, vol. 60, pp. 2508–2523, May 2012.
- [141] C. Shen, T. Chang, L. Wang, Z. Qui, and C. Chi, “Distributed robust multicell coordinated beamforming with imperfect CSI: An admm approach,” *IEEE Trans. Signal Process.*, vol. 60, pp. 2988–3003, June 2012.
- [142] L. Wolsey, *Integer Programming*. New York: John Wiley and Sons, 1998.
- [143] A. Kulik and H. Shachnai, “There is no EPTAS for two-dimensional knapsack,” *Inf. Process. Lett.*, vol. 110, no. 16, pp. 707–710, 2010.
- [144] T. D. Hoang, L. B. Le, and T. Le-Ngoc, “Joint mode selection and resource allocation for relay-based D2D communications,” *IEEE Communications Letters*, vol. 21, pp. 398–401, February 2017.
- [145] T. D. Hoang and L. B. Le, “Joint prioritized scheduling and resource allocation for ofdma-based wireless networks,” *IEEE Trans. on Wireless Commun.*, Submitted.