



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

**ENHANCED MOBILITY MANAGEMENT ALGORITHMS
FOR CELLULAR NETWORKS**

DISSERTATION

for the Degree of

DOCTOR OF PHILOSOPHY
(Electrical Engineering)

NGUYEN MINH THANG

FEBRUARY 2021

**Enhanced Mobility Management Algorithms for Cellular
Networks**

DISSERTATION

Submitted in Partial Fulfillment
of the Requirements for the
Degree of

DOCTOR OF PHILOSOPHY
(Electrical Engineering)

at the

UNIVERSITY OF ULSAN

by

Nguyen Minh Thang
Supervisor: Professor Sungoh, Kwon
February 2021

Publication No. _____

©2021 - Nguyen Minh Thang

All rights reserved.

Enhanced Mobility Management Algorithms for Cellular
Networks

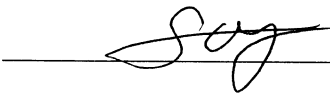
Approved by Supervisory Committee:



Prof. Sunghwan Kim, Chair



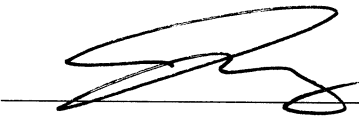
Prof. Sungoh Kwon, Supervisor



Prof. Jiho Song



Prof. Jin-Ghoo Choi



Prof. Sung-Yoon Jung

School of Electrical Engineering
University of Ulsan, South Korea

Date: February, 2021

VITA

Nguyen Minh Thang received his Bachelor of Engineering (B.E.) in Electrical Engineering from the Ho Chi Minh city University of Technology, Vietnam in 2013. From 2013 and 2014, he was employed as a Logic Design Engineer (R&D) at Arrive Technologies Inc, in Ho Chi Minh city, Vietnam. In September 2014, he started pursuing the Combined (Master's and Doctoral) degree in Electrical Engineering, under the supervision of Professor Sungoh Kwon, at the school of Electrical Engineering in the University of Ulsan (UOU), South Korea. His research interests include mobile network optimization, energy efficiency, mobility management, mmWave communications in ultra-dense networks.

*Dedicated to my grateful family
for their love and support*

ACKNOWLEDGMENTS

First and foremost, I am extremely grateful to my supervisor, Professor Sungoh Kwon, for his kindness, constant support, encouragement, enthusiasm, motivation, and immense knowledge, so that I can complete my Ph.D. degree. My association with him for six years has been a great learning experience in my life. I am also very much thankful to Professor Jiho Song for his kindness and enthusiastic guidance. His advices and comments helped me to complete the Ph.D. thesis. I am also thankfully appreciating to the other members of the committee of Ph.D. thesis defense, Professor Sunghwan Kim, Professor Jin-Ghoo Choi, and Professor Sung-Yoon Jung for their visit and valuable comments, that helped me a lot to improve the quality of this thesis.

I am grateful to the members of communication and networking (ComNet) laboratory, University of Ulsan (UoU), for their friendship, enthusiastic help and cheerfulness during my study in Korea. I send a big thank to Dr. Md Mehedi Hasan for his support to complete the Ph.D. thesis. I would also like to acknowledge the Radio Access Network S/W Research Team of ETRI, Daejeon, for their friendly collaboration in research projects.

I would like to thank the Vietnamese community at UoU for their sharing and respect, especially, I am highly grateful to Dr. Chinh Tam Le and Dr. Viet Tuan Pham for their sharing and kindness.

Last but not least, I would like to express my deep and sincere to my wife and my parents for their big love, caring, encouragement, and support, which have always stimulated me with the strength, motivation, and belief.

Nguyen Minh Thang

Ulsan, South Korea, February-2021.

ABSTRACT

Enhanced Mobility Management Algorithms for Cellular Networks

by

Nguyen Minh Thang

Supervisor: Professor Sungoh Kwon

Submitted in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy (Electrical Engineering)

February 2021

The use of mobile phones and other portable devices is continuously increasing the demand for high quality of experience (QoE) in wireless networks, such as huge data rate, and extremely low latency. To satisfy the heavily growing QoE demands, the ultra-dense network is considered a promising technique for fifth generation (5G) and beyond 5G cellular networks. Therefore, to support the data demand, as well as to increase network capacity small cells are densely deployed in present cellular networks. However, due to the low service area, the small-cell network is vulnerable to the mobility of user equipment units (UEs). During the course of movement, the wireless connections between UEs and small cells can fail frequently, such as handover failure or call drop, thus disturbing user experience. The problem will be exacerbated under ultra-dense small cell network if mobility-related parameter are not optimized. Hence, appropriate configuration and management of the network is required to enhance user quality of experience, and this thesis we studied two topics: handover optimization issues and resource management problems in cellular networks.

The first part of the thesis focuses on handover optimization for seamless mobility under ultra-dense small-cell networks. In Chapter I, in order to overcome handover failure in ultra-dense small-cell networks, we propose a low-complexity distributed mobility robustness optimization framework for small-cell networks to optimize handover parameters, such as time-to-trigger (TTT), handover offsets (A3Offset), and cell-individual offsets (CIO). The framework performs handover failure classification by exchanging message between cells in the system. The failure includes too-late handover, too-early handover, and wrong-cell handover. Due to a trade-off between too-late handover with too-early handover and wrong-cell handover, handover parameters are optimized according to the reasons for failure. Results show that the proposed algorithm improved handover performance more than baseline algorithms.

In chapter II, we analyzed handover problems to clarify when and how optimal handover parameters can be obtained. The study utilized geometry to model handover problems (such as too-late, too-early, and wrong-cell handover) and derived mathematical condition for handover failures. After that, optimal settings to avoid undesirable handover was introduced, and the trade-off between too-late handover and too-early handover was deeply investigated. We perform analyses for various aspects of wireless networks, such as impacts of interference, heterogeneous environments, mobility models, and network topology.

In chapter III, utilizing the results of chapter II, we propose a machine learning-based mobility robustness optimization framework for dynamic small-cell networks. Due to energy saving or traffic demands, small cell can be activated or deactivated, thus making the topology of wireless network become dynamic. Also, user mobility affects handover performance in a dense deployment of small cells. Taking into account the dynamics of network topology and user mobility, we apply transfer learning and reinforcement learning to optimize handover parameters. The transfer learning-based algorithm utilized the handover

analysis in chapter II to adapt the varying topology, and reinforcement learning used the transferred knowledge to optimize handover parameters with a fast convergence. The results show that the proposed framework provide significant improvement in handover performance while achieving short convergence rate under dynamic small-cell networks.

Part II of the thesis focuses on power allocation and beamforming design for multi-connectivity 5G wireless network. In chapter IV, we propose a cooperating scheme to maximize network throughput while guaranteeing user quality of experience (QoE) demands in multiple-input-multiple-output (MIMO) systems. One of the aspired-to targets of the fifth generation (5G) network is to guarantee QoE everywhere in the network. However, UEs in the edge areas are vulnerable to QoE violations, and they need dual connectivity from two nearby transmission points. Hence, with the motivation to utilize multi connectivity to satisfy the demanded QoE, our algorithm categorized UEs into two sets: single-connectivity and dual-connectivity. After classification, transmission power is allocated to maximize the network capacity while guaranteeing the minimum QoE. We show that our proposed algorithm not only satisfies all the UEs in the system but also maximizes the network capacity and outperforms benchmark algorithms.

In chapter V, we proposed a resource allocation algorithm for multi-connectivity wireless networks considering the minimum required QoE and the impact of CSI error. In practice, CSI error caused by hardware impairments or quantization of channel estimation scheme can affect the power allocation mechanism. Therefore, first, we investigate the impact of CSI error on the received signal quality at UEs in terms of signal-to-noise-plus-interference (SINR). Then, UEs are classified into two set: single-connectivity and dual-connectivity. After that, power is allocated to each set with the aims at minimizing transmit power and satisfying QoE requirement together. The results show that our algorithm can achieve 100 % satisfaction rate while minimize the transmit power.

Contents

Supervisory Committee	ii
Vita	iii
Dedication	iv
Acknowledgments	v
Abstract	vi
Table of Contents	ix
List of Figures	xii
List of Tables	xiv
Nomenclature	xv
I Part I	1
1 Mobility Robustness Optimization	3
1.1 Introduction	3
1.2 System Model and Problem Formulation	5
1.2.1 LTE Small-Cell Network and SON	5
1.2.2 Handover Procedure and Parameters	6
1.2.3 Handover Failures	7
1.2.4 Problem Formulation	9
1.3 Proposed Algorithm	10
1.3.1 Adapting TTT and Ocns	11
1.3.2 Updating Ocns and A3Offset	13
1.4 Simulation and Performance Analysis	13
1.4.1 Simulation Environment	13
1.4.2 Simulation Results	16
1.5 Closing Remarks	19
2 Analysis of Optimal Handover Parameters	21
2.1 Introduction	22
2.2 System Model	24
2.2.1 Small-cell Network	24
2.2.2 Quality Measurements of Radio Link	25

2.2.3	Handover Procedure and Control Parameters	25
2.2.4	Handover Failures and Ping-pongs	28
2.3	Analysis of Undesirable Handovers	31
2.3.1	User Mobility and Positions of Event Triggers	31
2.3.2	Modeling Undesirable Handovers	37
2.3.3	Optimal Range for Handover Offset	38
2.3.4	Discussion of The Optimal Range of Handover Offset	40
2.4	Numerical Analysis and Discussion	42
2.4.1	Simulation Environment	42
2.4.2	Impact of User Mobility	45
2.4.3	Impact of Heterogeneous Settings	47
2.4.4	Impact of Fading and Interference	48
2.4.5	Impact of Speed and TTT	48
2.4.6	Irregular Deployment of Small Cells and Non-existence of the Optimal Setting	50
2.5	Closing Remarks	55
3	Machine Learning–Based Mobility Robustness Optimization	59
3.1	Introduction	60
3.2	System Models and Problem Formulation	63
3.2.1	The SON and the Small-cell Network	63
3.2.2	Handover Procedures and Issues	65
3.2.3	Optimal Handover Parameters	71
3.2.4	Problem Formulation	72
3.3	Transfer Learning-Based MRO	73
3.3.1	Knowledge Transfer Algorithm	73
3.3.2	Environment Adaptation Algorithm	78
3.4	Numerical Results and Discussion	88
3.4.1	Simulation Environment	88
3.4.2	Handover Performance under Dynamic Network Topology	90
3.4.3	Impact of the User Mobility Model	93
3.4.4	Impact of CIO Resolution on Handover Performance	96
3.4.5	Offline Learning	98
3.5	Closing Remarks	100
II	Part II	103
4	Joint Power Allocation and Beam-forming Design for Dual-connectivity Wireless Networks	105
4.1	Introduction	106
4.2	System Model and Problem Formulation	107
4.3	Hybrid Beamforming Design	111
4.3.1	Design of the Combining and Precoding RF Matrices	111
4.3.2	Joint Design of the Baseband Combining and the Precoding Matrices	112

4.4	Power Allocation Algorithms	113
4.4.1	User Classification Algorithm	115
4.4.2	Power Allocation with Guaranteed QoE	117
4.5	Simulation Results and Analysis	121
4.6	Closing Remarks	123
5	Joint Power Allocation and Beam-forming Design for Dual-connectivity Wireless Networks Under Imperfect CSI	125
5.1	Introduction	126
5.2	System Model and Problem Formulation	129
5.2.1	System Model	130
5.2.2	Problem Formulation	133
5.3	Hybrid Beamforming Design	134
5.3.1	Design of the Combining and Precoding RF Matrices	135
5.3.2	Joint Design of the Baseband Combining and the Precoding Matrices	135
5.3.3	Impact of imperfect CSI on Beamforming and Network Performance	137
5.4	Power Allocation Algorithm Under Imperfect CSI	138
5.4.1	UE Classification Algorithm	139
5.4.2	Power Allocation for Dual-Connectivity UEs	140
5.4.3	Complexity Analysis	144
5.5	Simulation Results and Analysis	144
5.5.1	Simulation Setup	144
5.5.2	Impact of CSI error on connectivity	146
5.5.3	Impact of CSI error on network performance	147
5.5.4	Impact of QoE requirements	151
5.5.5	Impact of the number of UEs	153
5.5.6	Impact of the number of RF chains	155
5.6	Closing Remarks	155
	Publications	159
	References	161
A	Proofs in Chapter 2	175
A.1	Proof of Lemma 1	175
A.2	Proof of Theorem 1	176
A.3	Proof of Theorem 2	177
A.4	Proof of Theorem 3	178
B	Proofs in Chapter 5	181
B.1	Proof of Lemma 1	181
B.2	Proof of Lemma 2	183
B.3	Proof of Theorem 1	184

List of Figures

1.1	Event A3 and control parameters.	8
1.2	Simulation environment.	14
1.3	RLF rates ([0,10] km/h).	14
1.4	RLF classifications ([0,10] km/h).	15
1.5	Parameter optimization ([0,10] km/h).	16
1.6	RRs of various environments.	17
1.7	Ping-pong rates ([0,10] km/h).	17
2.1	Successful handover.	26
2.2	Handover failures and ping-pongs: (a) too-late handover, (b) too-early handover, and (c) a ping-pong.	29
2.3	(a) User trajectory. (b) Apollonian circles for different Δ_{01} under identical settings ($\nu_{01} = 1$). (c) Apollonian circles for individual Δ_{01} values under different settings ($\nu_{01} = 1/4$). (d) Apollonian circles for an overlaid small cell ($\nu_{01} = 1/200$).	32
2.4	A3 event borders for handover offset -3 dB in six sectors of a heterogeneous environment.	35
2.5	Regular deployment of small cells.	43
2.6	Impact of user mobility (two cells).	46
2.7	Impact of different settings (two cells).	49
2.8	$\nu_{01} = 1/200$	50
2.9	Impact of fading (two cells).	51
2.10	Impact of interference.	52
2.11	Impact of speed (four cells).	53
2.12	Impact of TTT (four cells).	54
2.13	Irregular deployment of 10 small cells and the results.	56
2.14	Non-existence of the optimal range.	57
3.1	Dynamic topologies with cells switching on and off (a) two cells are switched on, and (b) three cells are switched on.	61
3.2	A scenario with a small-cell network and its SON.	64
3.3	A successful handover.	68
3.4	Examples of dynamic small-cell network handovers.	70

3.5	Proposed transfer learning process for dynamic wireless networks.	74
3.6	A knowledge base is a database for the algorithm at a cell, including (a) TTT knowledge and (b) CIO knowledge for a neighboring cell.	87
3.7	The dynamic topology of a small-cell network.	89
3.8	At 5 km/h.	90
3.9	At 30 km/h.	92
3.10	Handover performance under a random mobility model.	94
3.11	Examples of optimal handover parameter estimation. The number indicates the optimization step.	96
3.12	Handover performance with different CIO resolutions.	97
3.13	Prior knowledge for cell 1 with a softmax probability of $\tau = 1$	98
3.14	Comparison between online and offline MRO-TL.	99
4.1	An example of a dual-connectivity wireless network.	109
4.2	Percentage of satisfied UEs at different QoE thresholds.	120
4.3	Network capacity for different QoE thresholds.	120
4.4	Adaptive DC for each time instance due to mobility.	122
5.1	An example of a dual-connectivity wireless network.	130
5.2	Percentage of dual-connectivity UEs versus QoE threshold and CSI error.	147
5.3	Network performance versus CSI Error for QoE of 50 Mbps.	148
5.4	Network performance versus CSI Error for QoE of 100 Mbps.	149
5.5	Network performance versus CSI Error for QoE of 200 Mbps.	150
5.6	Network performance with various QoE requirements and $\sigma_e^2 = 0.02$	152
5.7	Network performance with various number of UEs and $\sigma_e^2 = 0.02$	154
5.8	Network performance with various number of TP's RF chains and $\sigma_e^2 = 0.02$	156
A.1	Trajectories with $\theta = 0$	178

List of Tables

1.1	Solution to handover failures	9
2.1	Summary of undesirable handovers.	31
2.2	Summary of parameters for simulation.	44
3.1	Adaption times of the algorithms (in minutes) for a speed of 5 km/h	91
3.2	Satisfaction rates of algorithms (%) for a speed of 5 km/h	91
3.3	Adaption times (in minutes) and satisfaction rates (%) of algorithms for a speed of 30 km/h	95
4.1	Summary of parameters in this paper.	114
5.1	Summary of parameters in system model.	133
5.2	Complexity analysis.	145

Nomenclature

Notation Description

3GPP	3rd Generation Partnership Project
5G	Fifth-generation
CIO	Cell Individual Offset
EE	Energy Efficiency
HetNet	Heterogeneous Network
LTE	Long Term Evolution
MIMO	Multi-input multi-output
MLB	Mobility Load Balancing
mmWave	Millimeter wave
MRO	Mobility Robustness Optimization
QoE	Quality of Experience
RLF	Radio Link Failure
RRC	Radio Resource Control
RSRP	Reference Signal Received Power
RSRQ	Reference Signal Received Quality
SINR	Signal-to-Interference-plus-Noise-Ratio
SON	Self-organizing Network
TTT	Time to Trigger
UEs	User Equipment's

Part I

Mobility Robustness Optimization in
Small-Cell Networks.

Chapter 1

Mobility Robustness Optimization¹

In this chapter, we propose a distributed mobility robustness optimization algorithm to minimize handover failures due to radio link failures by adjusting time-to-trigger and offset parameters. According to the reason for failure, the algorithm classifies handover failure into three categories (too late, too early, and wrong cell), and simultaneously optimizes three handover parameters according to the dominant failure. Moreover, the algorithm considers handover failures to each neighboring cell and adjusts handover parameters individually. Via simulation, we show how the proposed algorithm adaptively optimizes the parameters and outperforms previous algorithms in various mobile environments.

1.1 Introduction

The small cell is one of the key technologies of the fifth-generation network, since it can share the burden of capacity and coverage demands [2], [3]. Small cells are designed to be cost-effective and power-optimized, serving limited coverage areas [2, 4], and can be installed at indoor or outdoor hyper-dense locations such as offices, campuses, stadiums,

¹The study in this chapter was published in IEEE Transactions on Vehicular Technology[1].

airports, and other large venues [2], in a planned or unplanned manner. Hence, manual configuration of such a network is difficult because of the large number of small cells and the dynamics of unplanned deployment. The self-organizing network (SON) [5] was introduced to automatically configure and optimize a wireless network by reducing manual involvement. Self-optimization aims at improving network performance by adaptively adjusting system parameters based on network status.

Among self-optimization case studies, mobility robustness optimization (MRO) aims to improve user mobility performance by optimizing handover-related parameters [5]. If the parameters are not appropriately set, when users in service move from one cell to another (referred to as handover), radio link failures (RLFs) and ping-pong can occur. As RLF causes user equipment (UE) to physically lose the radio connection, additional retransmissions or reconnections are required, which interrupts user service and wastes network resources. In contrast, under ping-pong, user service is successfully maintained while they are transferred back-and-forth between two neighboring cells within a short duration of time, called minimum time-of-state [6]. Hence, ping-pong only impacts network signaling load. Since RLFs affect the quality of service more than ping-pongs, the main objective of MRO is to minimize RLFs by optimizing the handover parameters [5].

In previous work, various MRO algorithms to minimize RLFs have been proposed [7–16]. In [7, 8], the authors introduce joint optimization algorithms for handover offset (hysteresis) and time-to-trigger (TTT). However, the hysteresis and TTT parameters globally impact all neighboring cells, so the algorithms cannot individually adapt to different neighboring cell environments. The adaptive method in [9] can locally adjust cell-individual offsets (CIOs) for specific neighboring cells, based on handover failure classification. But other handover parameters, such as TTT and hysteresis, are not of concern even though they significantly impact handover performance [10, 11]. In [10], the authors utilize an in-

interference coordination technique along with appropriate sets of handover offset and TTT. Nonetheless, the appropriate sets are fixed and chosen *a priori* via simulation. Therefore, the algorithm may not be applicable to real environments, including small-cell networks. To provide a more flexible way to adjust handover parameters, machine learning-based algorithms are proposed in [12–15]. While the algorithms in [12, 13] can tune handover offsets without TTT adjustment, the other algorithms in [14, 15] tune hysteresis and TTT, ignoring CIOs. Hence, the algorithms in [12–15] are limited in handover optimization. Also, the user speeds reported in [10, 15] are assumed to be precisely estimated, so that their algorithms are not applicable to general environments in practice.

In this chapter, we propose a distributed SON algorithm for a small-cell network to minimize the number of RLFs that can happen during handovers. To that end, we categorize handover failures according to the reasons for RLF. Handover offsets including CIOs can reshape geographic handover regions to neighboring cells and affect a handover process *spatially*. TTT can adjust handover execution time, so the parameter influences a handover process *temporally*. Hence, to minimize the number of RLFs, our algorithm jointly optimizes spatial and temporal handover parameters by considering their typical influence on the dominant failure of each neighboring cell, while the previous works only consider a part of the three parameters (TTT, hysteresis, and CIO). In addition, we discuss the impact of user speed on RLF optimization and RLF-ping-pong tradeoff.

1.2 System Model and Problem Formulation

1.2.1 LTE Small-Cell Network and SON

In this chapter, we consider a Long-Term Evolution (LTE) small-cell network with a single frequency and a SON. Small-cell base stations, called small-cell eNBs (SeNBs) in LTE,

can communicate with each other via the X2 interface. The X2 interface essentially supports the handover procedure by exchanging operational reports, parameter configuration, and radio link failure indication [17]. At each SeNB, a SON entity, known as a distributed SON [5], keeps gathering handover information and optimizing handover parameters. The handover procedure is executed when a UE moves from a serving cell to a neighboring cell. Based on measurement reports from the UE, the serving cell makes a decision of when to hand over the UE to a target cell in order to maintain service.

1.2.2 Handover Procedure and Parameters

The LTE handover procedure starts based on measurement reports from a UE to the serving SeNB that makes the handover decision. The UE periodically measures reference signal received powers (RSRPs) of all discovered SeNBs, and reports a measurement indicating that a handover should be triggered if a certain condition holds.

In this chapter, the A3 event is configured for measurement reporting [17]. The A3 event is triggered when a neighboring cell becomes better than the serving cell based on a certain offset. The moment for the UE to report measurements is right after the TTT timer has expired. The condition is described as follows:

$$M_n + O_{cn} + O_{fn} > M_p + O_{cp} + O_{fp} + O_{ff} + H_{ys}, \quad (1.1)$$

where M_p and M_n are the measured RSRP values of the serving SeNB and the neighboring SeNB, respectively. H_{ys} is the hysteresis parameter to prevent oscillation of Condition (1.1) due to fading, and O_{ff} is the handover offset for this event. To differentiate the A3 event offset from other offsets, we denote the offset for the A3 event as $A3Offset$ in this chapter. Offsets O_{cn} and O_{cp} are CIOs for the neighboring SeNB and the serving SeNB, respectively. While H_{ys} and $A3Offset$ affect handovers to all neighboring SeNBs, the CIOs can assign

different offsets to different neighbors. Of_n and Of_p are frequency-specific offsets for the neighboring SeNB and the serving SeNB, respectively. Since only intra-frequency handover is considered, and there is no specific offset for serving cells in this chapter, we set Of_p , Of_n , and Ocp to zero. Hence, we can rewrite Condition (1.1) as

$$M_n > M_p + \text{Offset}_n, \quad (1.2)$$

where $\text{Offset}_n = A3\text{Offset} + \text{Hys} - \text{Ocn}$. When (1.2) is satisfied, being as long as the given TTT, the UE sends a measurement report to the serving SeNB.

1.2.3 Handover Failures

Due to an unreliable radio link, a handover might fail unintentionally if the handover parameters are improperly configured. An RLF is detected at the physical layer when the received signal quality is too low for reliable communication. Based on the quality of the reference channel, the reliability of the current radio connection is evaluated. If timer T310 expires after the signal quality drops below threshold $Q_{\text{outofsync}}$ for N310 consecutive times, the UE triggers an RLF event, and the connection is dropped. During the process, if the signal quality is above threshold Q_{insync} for N311 consecutive measurements, the UE maintains the connection. If a radio link failure happens, the UE finds the best neighboring SeNB from the latest measurement session and sends a reestablishment request to the target SeNB to reconnect [17].

Fig. 1.1 shows an example of an RLF when the offset parameter in (1.2) and the TTT are set to $\text{Offset}_n^{(2)}$ and TTT_2 , respectively. At time t_3 , since the condition in (1.2) is met, the A3 event is triggered, and the UE prepares to send a measurement report after TTT_2 . However, the signal quality is poorer than $Q_{\text{outofsync}}$ at t_4 and remains below the threshold during N310 consecutive measurements within the T310 period. At time t_5 , the UE drops the link due to poor link quality.

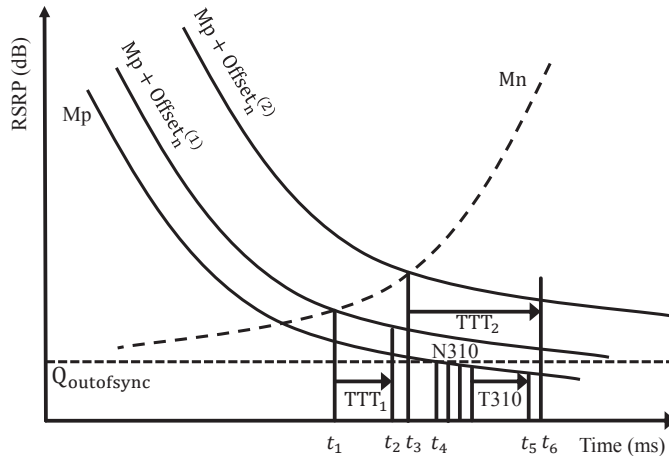


Figure 1.1: Event A3 and control parameters.

Based on the reason for the RLF, handover failures can be categorized into three types: too late, too early, and wrong cell. A too-late handover happens when a UE moves faster than the handover parameter settings allow, so that the handover procedure in the source cell starts too late. Hence, a too-late handover is detected when an RLF occurs in a serving cell before handover or during the handover procedure owing to a too-weak signal strength from the serving cell. Then the UE reconnects to a target cell that is different from the serving cell. The target cell notifies the originating cell about the RLF via the X2 interface, and the originating cell recognizes a handover that was too late.

A too-early handover happens when a UE moves into a target cell too early, and the connection is immediately lost due to poor link quality after a successful handover. Then, the victim UE reconnects to the previous cell. The serving cell itself recognizes the too-early handover, and notification between cells is unnecessary.

A wrong-cell handover involves three nearby SeNBs, and happens when a UE moves to an unexpected cell. A wrong-cell handover is detected when an RLF occurs shortly after a successful handover to the target cell, and then the UE reconnects to another cell that is

Table 1.1: Solution to handover failures

Handover failure	TTT	A3Offset	Ocn
Too late	decrease	decrease	increase
Too early	increase	increase	decrease
Wrong cell	increase	increase	decrease

neither the serving cell nor the target cell. The last connected cell informs the target cell of the first handover about the failure. The target cell can recognize that this RLF is not due to a too-late handover but to a wrong-cell handover, and notifies the originating cell.

1.2.4 Problem Formulation

Improper configuration of handover parameters brings handover failures. Too-tight handover parameters induce too-late handovers due to delay in the handover decision. By reducing the TTT or the offset in (1.2), the number of too-late handovers can be reduced. However, too-small values for the TTT or the offset can cause too-early or wrong-cell handovers, so that a longer value for the TTT, or a greater offset is required. Table 1.1 summarizes the solutions for handover failures. Because of complex radio propagation environments, all kinds of handover failures can happen in the same SeNB, and the conflict among the solutions creates a challenge. For example, a handover from SeNB A to SeNB B may be too late, whereas a handover from SeNB A to SeNB C may be too early. The problem to solve in this chapter is to minimize the total number of RLFs that happen during handovers by

adjusting the handover parameters, expressed as

$$\begin{aligned}
 & \min_{\text{TTT}, \text{Ocn}, \text{A3Offset}} && f(\text{TTT}, \text{Ocn}, \text{A3Offset}) \\
 & \text{subject to} && 0 \text{ ms} \leq \text{TTT} \leq 5120 \text{ ms}, \\
 & && -24 \text{ dB} \leq \text{Ocn} \leq 24 \text{ dB}, \\
 & && -15 \text{ dB} \leq \text{A3Offset} \leq 15 \text{ dB},
 \end{aligned}$$

where f is the number of RLFs. The value ranges for TTT, Ocn, and A3Offset follow the defined standard [18].

1.3 Proposed Algorithm

In this section, we propose a distributed SON algorithm by automatically tuning handover control parameters such as TTT, Ocn, and A3Offset. A SON entity equipped at each SeNB collects RLF-related data, and periodically optimizes handover parameters. Since CIOs (Ocn) specify neighboring cells, each SON constructs a list for CIOs of the N neighboring cells.

When an RLF happens, each SeNB classifies the failure based on RLF indication messages from neighboring SeNBs, and a distributed SON at the serving SeNB collects statistics for RLF data according to handover problems. We denote the numbers of too-late, too-early, and wrong-cell handovers from the serving SeNB to neighboring SeNB j as N_{L_j} , N_{E_j} , and N_{W_j} , respectively. Based on the RLF data, the SON entity computes the RLF rate (RR), that is, the ratio of the total number of RLFs $N_{\text{totalIRLF}}$ to the total number of handover trials $N_{\text{totalhandovertrials}}$, in order to measure handover performance. The RR is expressed as

$$\text{RR} = \frac{N_{\text{totalIRLF}}}{N_{\text{totalhandovertrials}}} = \frac{\sum_{j=1}^N (N_{L_j} + N_{E_j} + N_{W_j})}{N_{\text{totalhandovertrials}}}, \quad (1.3)$$

When the timer for updating handover parameters expires, the algorithm checks to determine if the RR is greater than a certain threshold. If the RR is less than or equal to the threshold, the algorithm resets the timer as well as the RLF-data counters, and again waits until the timer ends. If the RR is greater than the threshold, the algorithm updates the handover parameters in two steps: adapting TTT and Ocn_s, and updating Ocn_s and A3Offset.

1.3.1 Adapting TTT and Ocn_s

Since an SeNB has multiple neighboring cells in which handover environments can be different, the algorithm adjusts TTT and Ocn_s according to three types of environment: too-late, too-early or wrong-cell dominant, and mixed.

When the number of too-late handovers is dominant for all neighboring cells, i.e., $N_{L_j} \geq N_{E_j} + N_{W_j}, \forall j$, the algorithm reduces TTT to the previous value as long as it is greater than or equal to TTT_{\min} . The TTT value is standardized as an enumerated parameter [17]. When the number of too-early or wrong-cell handovers is dominant for all neighboring cells, i.e., $N_{L_j} < N_{E_j} + N_{W_j}, \forall j$, the algorithm increases TTT to the next value as long as it is less than or equal to TTT_{\max} . After the algorithm changes TTT, it then waits until the next expiration of the updating timer.

When handover failures are mixed, the algorithm adjusts the CIO with respect to each neighboring cell in the neighboring list, since TTT can affect handovers to all neighboring cells. If the number of too-late handovers is dominant during handover to neighboring cell j , i.e., $N_{L_j} \geq N_{E_j} + N_{W_j}$, the corresponding CIO (Ocn _{j}) is decreased by one step (1 dB or 2 dB, depending on the current value). Otherwise, the corresponding CIO (Ocn _{j}) is increased by one step. The actual Ocn value is obtained from a defined table [17]. Algorithm 1 summarizes the overall proposed procedure.

Algorithm 1 TTT and Ocn adaptation algorithm

```

1: Set up initial handover parameters for neighbour SeNBs
2: Update  $N_{L_j}$ ,  $N_{E_j}$ , and  $N_{W_j}$  for all  $j$ 
3: if timer for updating RR expires then
4:   Calculate RR of (1.3)
5:   if RR > Threshold then
6:     if  $N_{L_j} \geq N_{E_j} + N_{W_j}$  for all  $j$  then
7:       Decrease TTT
8:     else if  $N_{L_j} < N_{E_j} + N_{W_j}$  for all  $j$  then
9:       Increase TTT
10:    else
11:      for  $j = 1$  to  $N$  do
12:        if  $N_{L_j} \geq N_{E_j} + N_{W_j}$  then
13:          Decrease  $Ocn_j$  by one step
14:        else if  $N_{L_j} < N_{E_j} + N_{W_j}$  then
15:          Increase  $Ocn_j$  by one step
16:        end if
17:      end for
18:      Update A3Offset as in Algorithm 2
19:    end if
20:  end if
21:  Reset the timer and all counters, then return to Line 2
22: else
23:   Return to Line 2
24: end if

```

Algorithm 2 Ocn and A3Offset modification algorithm

- 1: **if** all Ocn_j are positive **then**
 - 2: $Ocn_j \leftarrow Ocn_j - Ocn_{\min}$ for all j
 - 3: $A3Offset \leftarrow A3Offset - Ocn_{\min}$
 - 4: **else if** all Ocn_j are negative **then**
 - 5: $Ocn_j \leftarrow Ocn_j - Ocn_{\max}$ for all j
 - 6: $A3Offset \leftarrow A3Offset - Ocn_{\max}$
 - 7: **end if**
-

1.3.2 Updating Ocn_s and A3Offset

After completing all CIOs, the algorithm updates all Ocn_j s and the A3Offset in order to fine-tune the offsets. The values of the Ocn_j s and A3Offset are indexed [18]. While the index for A3Offset has 0.5 dB intervals, Ocn_j has different granularity according to the range, i.e., 1-dB steps between -6 dB and 6 dB, and 2-dB steps in other ranges. For fine-tuning $Offset_n$ in (1.2), the algorithm adjusts A3Offset and the Ocn_j s while preserving $Offset_n$, so that as many Ocn_j s as possible work in the range from -6 dB to 6 dB. If the Ocn_j s are all positive, the algorithm subtracts from A3Offset and Ocn_j s the minimum of all Ocn_j s, i.e. Ocn_{\min} . If the Ocn_j s are all negative, A3Offset and the Ocn_j s are reduced by the maximum of all the Ocn_j s, i.e. Ocn_{\max} . Otherwise, A3Offset and the Ocn_j s all remain the same. Algorithm 2 summarizes the process for updating the Ocn_s and A3Offset.

1.4 Simulation and Performance Analysis

1.4.1 Simulation Environment

To verify the proposed MRO algorithm, simulations were conducted using the NS-3 network simulator with the LTE module [19]. A one-tier small-cell network composed

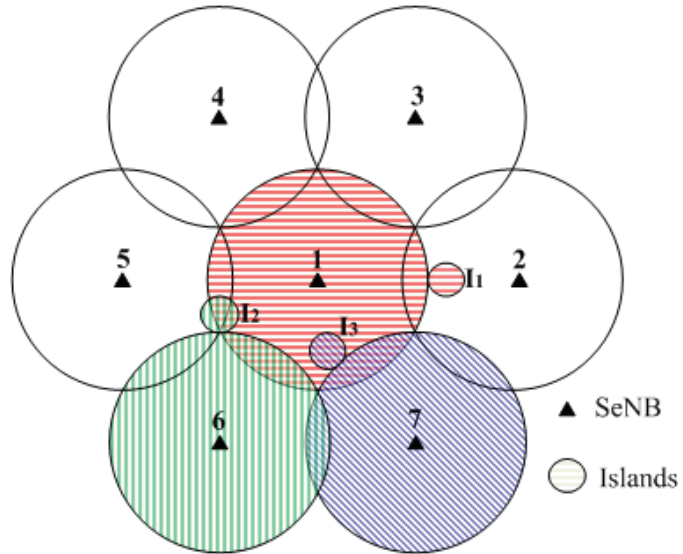
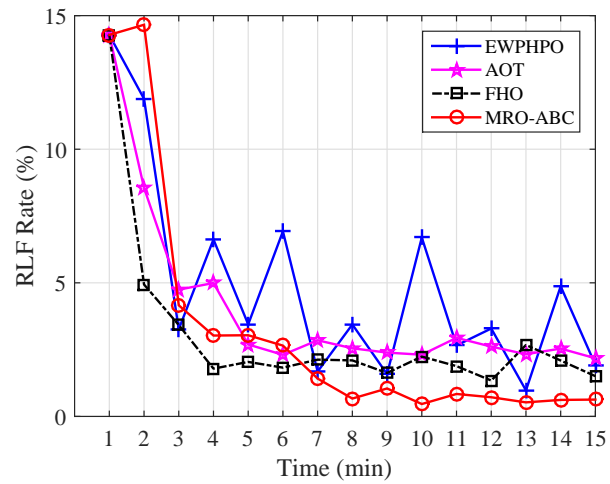


Figure 1.2: Simulation environment.

Figure 1.3: RLF rates ($[0,10]$ km/h).

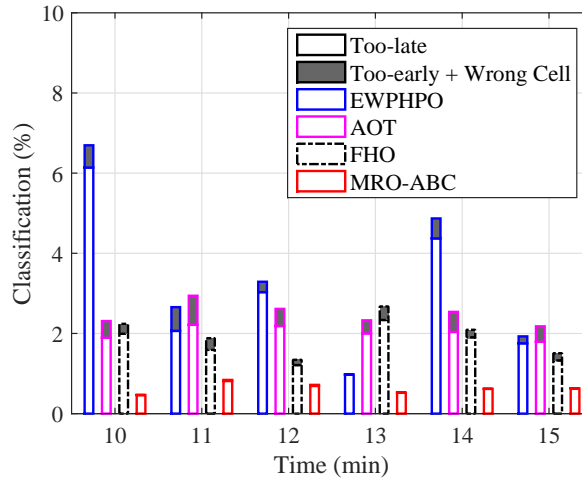


Figure 1.4: RLF classifications ($[0,10]$ km/h).

of seven SeNBs was considered and the inter-site distance was set to 30 m. The system bandwidth was set to 20 MHz. A wireless channel was modelled under a non-line-of-sight path loss model [20] and an extended typical urban multipath-fading model [21] that follows a log-normal distribution with variance of 6 dB. If the reference signal received quality (RSRQ) is below -12 dB for 500 ms, then the radio link is considered to have failed. The random waypoint (RWP) model was adopted, and the speed was selected within certain ranges. For the initial handover settings, TTT, A3Offset, and Ocnst were set as 480 ms, 3 dB, and 0 dB, respectively [10]. Additionally, TTT_{\min} and TTT_{\max} were chosen as 40 ms and 5120 ms, respectively. The threshold for activating MRO was 1% [14], which is also considered as a performance target.

To evaluate our algorithm, we considered three islands of network coverage, as shown in Fig. 1.2. In practice, due to blockages, walls, or transmission power, the coverage of cells can be irregular. Hence, we modeled such irregular coverage boundaries with multiple islands [22]. For the performance measure, we use the RR in (1.3). To compare the proposed algorithm, referred to as the MRO algorithm based on classification (MRO-ABC),

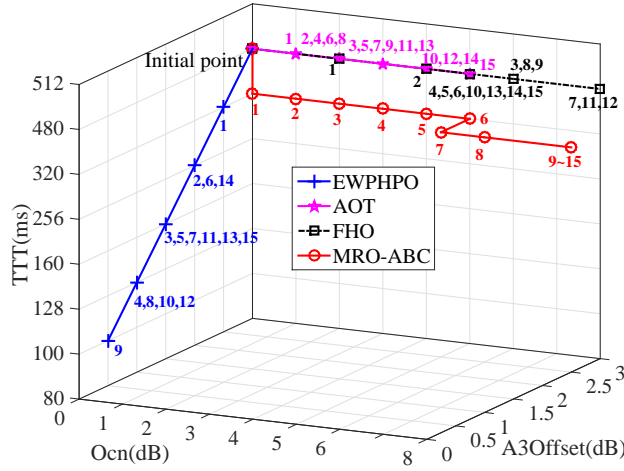


Figure 1.5: Parameter optimization ([0,10] km/h).

three previous algorithms are considered: enhanced weighted performance-based handover-parameter optimization (EWPHPO) [7], adaptive Ocn tuning (AOT) [9], and fuzzy-based handover optimization (FHO) [12].

1.4.2 Simulation Results

For the simulations, user velocities for the RWP model were randomly selected between 0 and 10 km/h to model a pedestrian environment. As shown in Figs. 1.3 and 1.4, MRO-ABC outperforms the other algorithms and smoothly converges to an RR value below the target performance (1%), while AOT and FHO converge to a higher RR level, and EWPHPO consecutively oscillates.

To delineate the algorithms, we plot the handover parameters in Fig. 1.5². As can be seen, MRO-ABC efficiently finds optimal parameters, because it is able to optimize three parameters together and arbitrate optimization conflicts among handover failure classes in

²EWPHPO optimizes TTT and hysteresis. Since the effect of hysteresis is identical to A3Offset when the other is fixed, as shown in (1.1), we plot EWPHPO on the TTT and A3Offset domains

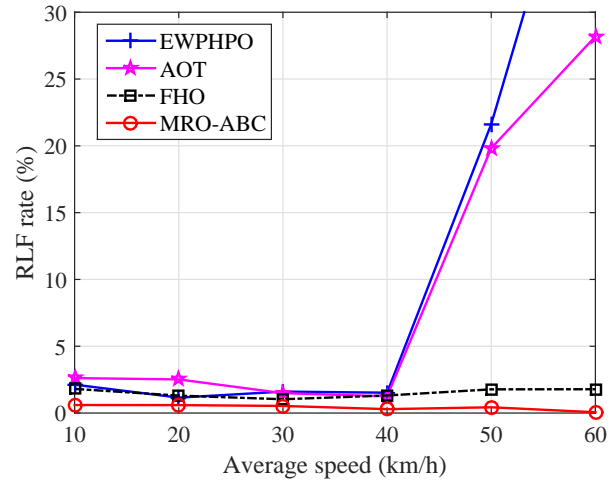


Figure 1.6: RRs of various environments.

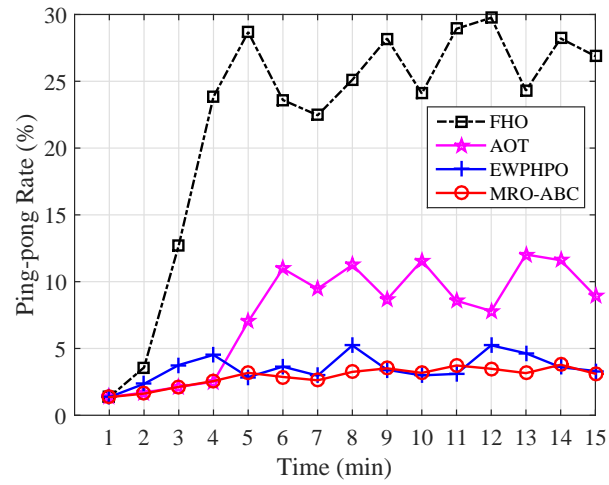


Figure 1.7: Ping-pong rates ([0,10] km/h).

a cell, while the other algorithms switch back-and-forth between parameter values, which results in the performance oscillation in Fig. 1.3. Since EWPHPPO considers TTT and hysteresis, which affect all the neighboring SeNBs, and cannot individually optimize handovers to specific neighboring SeNBs at once, the RR of EWPHPPO oscillates. FHO and AOT just attempt to spatially regulate handover regions via CIO adjustments, but they ignore TTT, which is also a critical parameter to temporally optimize handover [11], so performance is limited.

To prove the effectiveness of our algorithm in various mobile environments, we conducted simulations in the network seen in Fig. 1.2, varying the average velocities from 10 km/h to 60 km/h. After completing 15-minute simulations with 10 different random seeds, the averaged results are depicted in Fig. 1.6. The figure shows that MRO-ABC outperforms the other algorithms and maintains RRs below the target in all mobile environments, while EWPHPPO and AOT become worse as velocity increases beyond a certain point, and FHO underperforms.

There may be a tradeoff between RLF reduction and ping-pong [10], since the reduction of too-late handover bounds the increments of TTT and A3Offset (or hysteresis), which induces ping-pongs between SeNBs [23]. Hence, we compared the ping-pong rates of the algorithms, which are defined as the ratio of the total number of ping-pongs to the total number of handover trials throughout the network. Fig. 1.7 shows that the ping-pong rates of all the algorithms increase as the RRs decrease, and shows that MRO-ABC provides the lowest ping-pong rate. Except for MRO-ABC, the other algorithms have a strong trade-off between RLF reduction and ping-pong. For example, FHO performs second best in RRs, but worst in ping-pong rates. Since MRO-ABC classifies handover failures according to the reasons for RLF and cell-individually optimizes the handover parameters, MRO-ABC performs the best in both RRs and ping-pong rates.

1.5 Closing Remarks

In this chapter, a distributed MRO algorithm was proposed to improve handover performance by reducing RLFs. To that end, the proposed algorithm classifies handover failures according to the reasons for failure, and adjusts time-to-trigger, cell-individual offsets, and A3-event offset together, whereas previous algorithms simply optimize subsets of the three handover parameters. The proposed SON algorithm adaptively optimizes the handover parameters according to the dominant handover failure. Via simulation, we showed that our algorithm can effectively find the optimum handover parameters and outperforms previous algorithms. In addition, our algorithm induces the fewest ping-pongs among the considered algorithms.

This page intentionally left blank.

Chapter 2

Analysis of Optimal Handover

Parameters¹

Handover failure and ping-pongs are the common thorny issues in modern mobile networks. While handover failures caused by radio link failure (RLF) significantly reduces the reliability of network operation, ping-pongs drastically waste signaling resources. In the upcoming fifth-generation (5G) networks especially, a complex deployment of small cells can exacerbate the two problems, even though the network can be integrated with a self-organizing network (SON), which is an automation-based solution. Due to the coupling of RLFs and ping-pongs as explained in the literature, it is difficult to analyze handovers and minimize both RLFs and ping-pongs simultaneously. In this paper, we model a handover procedure with geometric elements (Apollonian circles and the straight line), and analyze handover performance. The analysis provides an optimal handover setting for minimizing both RLFs and ping-pongs together, whereas previous works only considered trade-offs between them. We show that our analysis accurately estimates the optimal setting by com-

¹The study in this chapter was published in IEEE Transactions on Wireless Communications[24].

paring it with an NS-3 simulation. From the analysis, different environments can require different optimal values: fading (as well as interference) limit the optimal values; user speed has a scaling impact; and time-to-trigger has a shifting effect.

2.1 Introduction

Trials of 5G cellular systems were demonstrated during the Winter Olympics in PyeongChang, and the first 5G mobiles can be ready for users' hands in 2020 [25]. Among crucial technologies in 5G, the small cell has been considered an approach to high network capacity and high spectrum reuse [26]. From a technical aspect, the small cell is a low-power, low-cost, and ready-to-use prototype of a base station [27]. For this reason, small cells can be deployed in a planned or unplanned manner in crowded indoor and outdoor places (such as shopping malls, stadiums, downtown streets, and so on [28]) to improve user quality of service. As a result, deployment of small cells can result in a dense and complex wireless network [26] so that traditional drive-test optimization for a cellular network is not cost-effective and applicable if reconfiguration is necessary.

To reduce administrative expenses from reconfiguring parameters, a small-cell network can be equipped with an automation-inspired solution called a self-organizing network (SON) [27,29]. The importance of a SON for small-cell networks is highlighted by its potential features, such as coverage and capacity optimization, energy efficiency, interference reduction, mobility load balancing, mobility robustness optimization (MRO), automatic neighbor relations, and random access optimization [18]. In order to ensure reliable and efficient operation of small cells for mobile subscribers, the handover procedure should be optimized by the SON so the procedure is robust in a dynamic context.

To provide seamless mobility throughout the cellular network, MRO minimizes

handover failures due to radio link failures (RLFs). Besides that, MRO should keep the number of ping-pongs as low as possible, since ping-pongs are redundant handovers in a short time, causing signaling procedure resources to be overburdened. Therefore, RLFs and ping-pongs should both be minimized. However, previous algorithms faced difficulties in finding an optimal solution for joint minimization of both RLFs and ping-pongs [9, 30, 31]. Previous works show that an optimal solution for jointly suppressing RLFs and ping-pongs is unattainable, and a compromise between them has to be reached based upon heuristic policies of network administration.

Analyses of handovers in SON-based small-cell networks have been studied in the literature [23, 32–36]. In [32], a simulation-based analysis showed the impacts of handover parameters, time-to-trigger (TTT) and handover offset, and user speed on RLFs and ping-pongs. The analysis also showed that improvement in RLFs leads to more ping-pongs. In [33], the authors showed that handover performance depends on handover parameters (TTT and handover offset), fading factors, and user mobility parameters (such as user speed and travel distance). The authors obtained a closed form of handover failure probability while neglecting ping-pongs. By contrast, in [34], the analysis was totally focused on ping-pongs, ignoring RLFs. The study showed that an improper handover offset can lead to ping-pongs, which occurs between two cells or within clusters of nearby cells. In [23], an analytical method modeled an RLF based on user mobility and handover offset without taking into account TTT. The results showed that handover offset and fading significantly affects RLFs, while ignoring ping-pongs. In [35, 36], analytical methods for RLFs and ping-pongs took into account TTT and fading, but not handover offset. The results showed that when RLFs improved, ping-pongs increased, and vice versa. However, in [1], the authors showed the possibility of jointly minimizing RLFs and ping-pongs via simulation, but without providing an analysis.

In this paper, we provide a comprehensive analysis of RLF and ping-pong minimization in SON-based small cell networks. We study the reasons for RLFs and ping-pongs, and then derive conditions for them. To that end, we model a wireless environment as geometry with a mathematical tool called the Apollonian circle. Based on the modeled geometry, we prove the existence of an *optimal* handover setting to minimize RLFs and ping-pongs together. Then, from the conditions, we lead to the optimal handover setting. Also, we investigate the impacts of various factors (such as user mobility, different settings, fading and interference, handover parameters, and cell deployment) on the optimal setting, and verify a case for non-existence of the optimal setting.

The remainder of this paper is organized as follows. Section II introduces background about undesirable handovers, RLFs, and ping-pongs. Section III presents models for undesirable handovers via the Apollonian circle, and describes the optimal handover setting. Section IV provides a numerical analysis and a discussion to verify our approximation based on various aspects.

2.2 System Model

2.2.1 Small-cell Network

In this paper, we assume that small cells operate on a dedicated spectrum to support a local area. Small cells are connected via the X2 interface to support handover procedures and to share information [18]. Also, small cells are connected to a core network via S1 interface to communicate and to receive user data [18]. The small-cell network is managed by a SON, which receives data from small cells and adjusts parameters to adapt to the environment [1].

2.2.2 Quality Measurements of Radio Link

To measure signal strength for a handover decision, reference signal receive power (RSRP) is chosen. RSRP is the average power of reference signals within a specific bandwidth [37]. While RSRP is used for handover decisions [1,30,32,35], the signal-to-interference-plus-noise ratio (SINR) includes inter-cell interference, and is used by user equipment (UE) to determine RLFs [38–41]. RLFs happen if the SINR remains below a predefined threshold, \mathcal{Q}_{out} , for a certain period of time. According to some research [38,40–42], the SINR for a UE, γ , is estimated as

$$\gamma = \frac{m_0}{\sum_{i \neq 0} m_i + N_0}, \quad (2.1)$$

where m_0 and m_i denote the received RSRP of serving Cell 0 and adjacent Cell i at the UE, respectively, and N_0 is the thermal noise.

Considering a nominated UE located at distance d_i from Cell i , the received power for such a distance can be modeled as

$$m_i = p_i \cdot G_i \cdot d_i^{-\alpha} \cdot \eta_i, \quad (2.2)$$

where p_i is the allocated transmission power of Cell i for the UE, G_i accounts for the antenna gain, α is the attenuation exponent, and η_i is the fading factor including large-scale fading and small-scale fading.

2.2.3 Handover Procedure and Control Parameters

Long Term Evolution (LTE) handover enables service switching between cells for UEs to adapt to mobility. The handover procedure starts based on measurement reports from the UEs to the serving cell that makes the handover decision. UEs periodically measure the RSRP of all discovered cells, and eventually or periodically send measurement reports to the serving cells if certain conditions hold. From the 3rd Generation Partnership Project

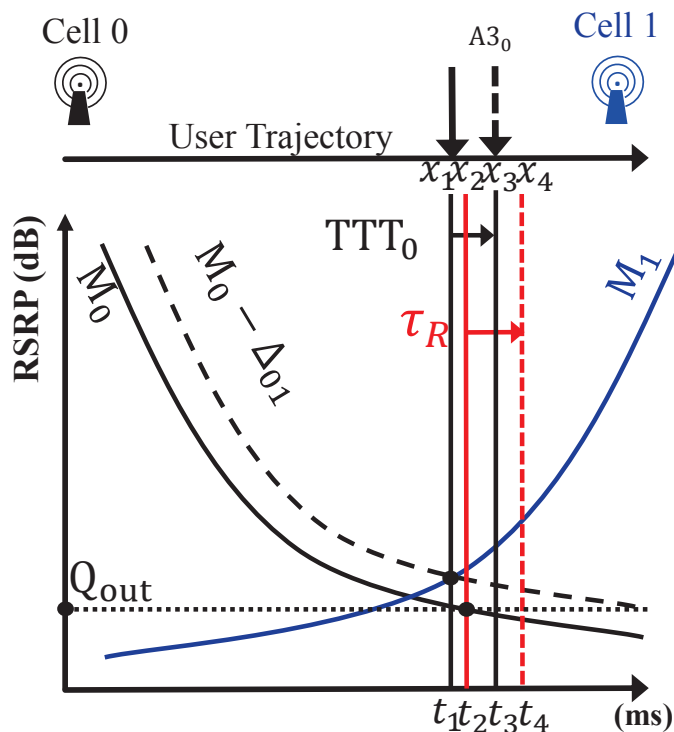


Figure 2.1: Successful handover.

(3GPP) standard, there are six intra-frequency event measurements, designated from A1 to A6 events [17]. Each event measurement is used for a specific application of a SON, like mobility management and neighbor discovery [29]. While the A1 and A2 events only consider the signal quality of the serving cell, the A4 and A6 events solely observe the signal quality of neighboring cells. The signal quality of the serving cell and neighboring cells is embodied by the A5 and A3 events, but only the A3 event assesses a relative comparison between signal quality of the serving cell and that of neighboring cells. Hence, the A3 event is used for a handover algorithm that decides which cell has the better signal quality to support moving UEs. The A3 event is triggered when the RSRP of a neighboring cell becomes better than the serving cell, based on a certain offset. UEs report measurements right after the TTT timer has expired. The condition for triggering the event on the decibel

scale is

$$M_n + O_{cn} + O_{fn} > M_p + O_{cp} + O_{fp} + O_{ff} + H_{ys}, \quad (2.3)$$

where M_p and M_n are the measured RSRP of the serving cell and the neighboring cell, respectively. H_{ys} is the hysteresis parameter to prevent oscillation of (2.3) due to fading, and O_{ff} is the A3 offset for this event. O_{cn} and O_{cp} are the cell-individual offsets (CIOs) for the neighboring cell and serving cell, respectively. O_{fp} and O_{fn} are the frequency-specific offsets for the serving cell and neighbouring cell, respectively. While H_{ys} and A3 offset affect handovers to all neighboring cells, the CIOs can assign different offsets to different neighbors. In this paper, we set the O_{fp} and O_{fn} to zero, since we consider intra-frequency handover. We rewrite (2.3) as

$$M_j + \Delta_{ij} > M_i, \quad (2.4)$$

where $\Delta_{ij} = O_{cn} - O_{cp} - O_{ff} - H_{ys}$, which is defined as the individual handover margin of serving Cell i for neighboring Cell j , M_i and M_j are the decibel-scale RSRP of Cell i and Cell j , respectively. When (2.4) is satisfied, being as long as the given TTT, the UE sends a measurement report to Cell i triggering a handover to Cell j . In this paper, the measurement report carries the RSRP of all discovered cells where the RSRP satisfies (2.4) to Cell i . Then, Cell i decides to hand over the UE to the best neighboring cell based on the reported RSRP. Fig. 2.1 shows an example of a successful handover when a UE moves from Cell 0 to Cell 1. In the figure, we denote x_i as the location of the UE from the serving cell (Cell 0) at time t_i , which is coordinated as (t_i, x_i) . When the RSRP of Cell 1 (M_1) is better than the RSRP of Cell 0 (M_0) by a value of $-\Delta_{01}$ (i.e., $M_1 > M_0 - \Delta_{01}$), the A3 event is triggered, and the timer for TTT starts counting at (t_1, x_1) for a time lapse of T_{TT0} . When the SINR of Cell 0 is below an out-of-sync indication threshold (called Q_{out}^2), a timer (called the RLF

² Q_{out} can be used as the minimum threshold of RSRP [43, 44] or SINR [38, 40]. For simplicity, we plot Q_{out} as the minimum threshold of RSRP in Fig. 2.1 and Fig. 2.2 while using Q_{out} as the minimum threshold of SINR for our analysis.

timer) starts counting for a time duration of τ_R at (t_2, x_2) ³. Once $m_1 > m_0 - \Delta_{01}$ is retained for the duration of τ_{TT0} , a handover is provoked at (t_3, x_3) . The handover to Cell 1 starts and completes successfully before a link failure. In this paper, the handover signalling delay is ignored, because the delay is much shorter than a TTT period (0 to 20 ms, as defined in [45]).

2.2.4 Handover Failures and Ping-pongs

There are two types of undesirable handover: RLF and ping-pong. While ping-pongs include successful handovers, an RLF causes failed handover attempts based on three categories of reason: too-late, too-early, and wrong-cell handovers [18]. An RLF is detected at the physical layer when the received SINR is too low for reliable communications. Based on the quality of the reference channel, the reliability of the current radio connection is evaluated. From [17], an RLF detection mechanism in the UE includes T310, N310, and N311, where T310 and N310 account for out-of-sync detection, and N311 defines the number of in-sync indications when the SINR is above an in-sync level (q_{in}). If timer T310 expires after the SINR drops below threshold q_{out} for N310 consecutive times, the link connection is dropped due to RLF. During the process, if the SINR is above threshold q_{in} for N311 consecutive measurements, the UE maintains the connection. If an RLF happens, the UE finds the best neighboring cell from the latest measurement session and sends a re-establishment request to the target cell to reconnect [17].

Fig. 2.2a shows an example of a too-late handover, which happens when the UE moves faster than the handover parameter settings allow, so the handover procedure in the source cell starts too late. When the SINR of Cell 0 is below q_{out} , the RLF timer starts counting at (t_1, x_1) for a time duration of τ_R . When m_1 is greater than m_0 by a value of

³In this paper, τ_R consists of the time duration of N310 and T310 together, where T310 is a timer measuring the time lapse when SINR is below q_{out} , and N310 defines the number of out-of-sync indications [17].

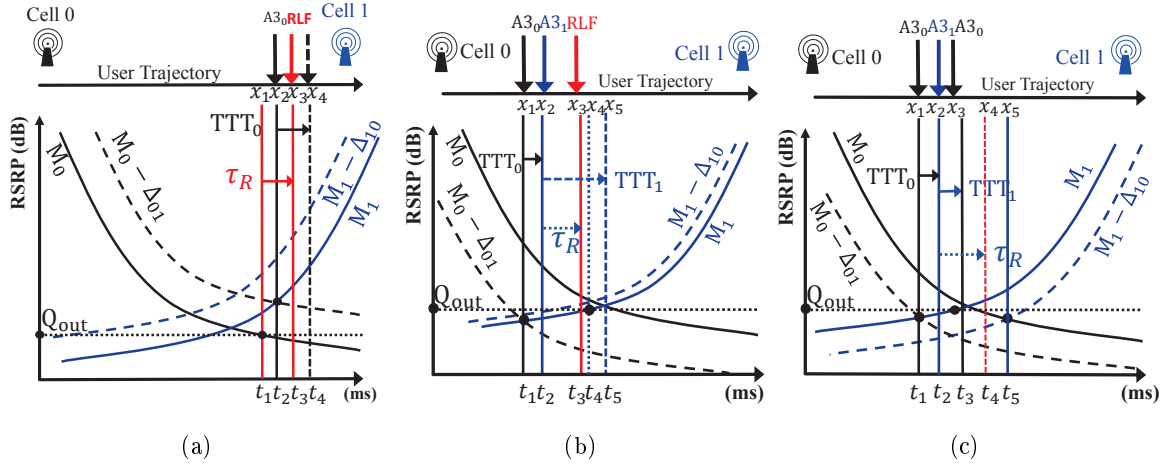


Figure 2.2: Handover failures and ping-pongs: (a) too-late handover, (b) too-early handover, and (c) a ping-pong.

$-\Delta_{01}$ (i.e., $m_1 > m_0 - \Delta_{01}$), the A3 event is triggered for Cell 1, and the timer for TTT starts counting at (t_2, x_2) for a time duration of TTT_0 . Since the SINR of Cell 0 is below q_{out} for the duration of τ_R , an RLF occurs at (t_3, x_3) before the handover to Cell 1 completes at (t_4, x_4) . This means that $x_2 + v\text{TTT}_0 \geq x_1 + v\tau_R$ (C1-1). After RLF, the UE tries to reconnect to Cell 1, which is the best neighboring cell to Cell 0. Cell 1 notifies Cell 0 about the RLF via the X2 interface, and Cell 0 recognizes a too-late handover.

A too-early handover happens when a UE joins a neighboring cell too early after a successful handover, and the connection is immediately dropped because of poor link quality, as illustrated in Fig. 2.2b. When $m_1 > m_0 - \Delta_{01}$ in (2.4), the A3 event is triggered for Cell 1, and the timer for TTT starts counting at (t_1, x_1) for a duration of TTT_0 . As long as $m_1 > m_0 - \Delta_{01}$ for the duration TTT_0 , a handover from Cell 0 to Cell 1 starts and finishes successfully at (t_2, x_2) . After a successful handover, the UE joins Cell 1. However, the SINR of Cell 1 is lower than q_{out} , so the RLF timer immediately starts counting at (t_2, x_2) for a duration of τ_R . Also, when $m_0 > m_1 - \Delta_{10}$ in (2.4), the A3 event is provoked for Cell 0, and the timer for TTT starts counting at (t_2, x_2) for a duration of TTT_1 . Once the SINR of Cell 1

is below q_{out} for the duration of τ_R , an RLF occurs at (t_3, x_3) before the SINR of Cell 1 rises above q_{out} at (t_4, x_4) . This means that $x_1 + v_{\text{TTT}_0} + v\tau_R \leq x_4$ (C2-1). The RLF occurs before the handover finishes at (t_5, x_5) , which means $\text{TTT}_1 > \tau_R$ (C2-2). After an RLF, the UE returns to Cell 0. This failure occurs shortly after a successful handover. Cell 0 itself recognizes the too-early handover based on a recent handover history. A wrong-cell handover is detected when an RLF occurs shortly after a successful handover to the target cell, and then, the UE reconnects to another cell that is neither the serving cell nor the target cell. Since a wrong-cell handover is identical to a too-early handover except for the re-connection, we merge the wrong-cell problem into the too-early problem.

In contrast to RLFs, ping-pongs maintain the link connection. However, they include multiple handovers between cells or within a cluster of cells. Fig. 2.2c shows an example of how a ping-pong is detected. Once $m_1 > m_0 - \Delta_{01}$ in (2.4), the A3 event is triggered for Cell 1, and the timer for TTT starts counting at (t_1, x_1) for a duration of TTT_0 . As long as $m_1 > m_0 - \Delta_{01}$ for the duration of TTT_0 , a handover to Cell 1 starts and completes successfully, and the UE joins Cell 1. When $m_0 > m_1 - \Delta_{10}$ in (2.4), the A3 event is triggered for Cell 0, and the timer for TTT starts counting at (t_2, x_2) for a duration of TTT_1 . Also, since the SINR of Cell 1 is below q_{out} , the RLF timer is triggered at (t_2, x_2) . As long as $m_0 > m_1 - \Delta_{10}$ for the duration of TTT_1 , a handover from Cell 1 to Cell 0 starts and finishes at (t_3, x_3) , before a failure at (t_4, x_4) . This means that $\text{TTT}_1 \leq \tau_R$ (C3-2) and that $x_1 + v_{\text{TTT}_0} + v_{\text{TTT}_1} \leq x_4$ (C3-1). It takes a very short time between two consecutive handovers in a ping-pong. This rapid pace wastes system resources such as time and signaling procedures [29]. Even though ping-pongs do not cause RLFs, regular requirement for handover optimization always includes minimizing ping-pongs as much as possible. The scenario for a ping-pong looks similar to the too-early handover problem; that is, a handover is triggered more quickly than normal, but ping-pongs do not cause RLFs.

Table 2.1: Summary of undesirable handovers.

Issues	Condition	
	Geographical condition	Timing condition
Too-late, Fig. 2.2a	(C1-1) $x_2 + v_{\text{TTT1}} \geq x_1 + v\tau_{\text{R}}$	
Too-early, Fig. 2.2b	(C2-1) $x_1 + v_{\text{TTT0}} + v\tau_{\text{R}} \leq x_4$	(C2-2) $\text{TTT1} > \tau_{\text{R}}$
Ping-pong, Fig. 2.2c	(C3-1) $x_1 + v_{\text{TTT0}} + v_{\text{TTT1}} \leq x_4$	(C3-2) $\text{TTT1} \leq \tau_{\text{R}}$

Table 2.1 summarizes the conditions for undesirable handovers.

2.3 Analysis of Undesirable Handovers

2.3.1 User Mobility and Positions of Event Triggers

The analysis of undesirable handovers is divided into two steps: between Cell 0 and Cell 1 alone, and between those two cells from among multiple cells (with both steps using general parameters). Our analysis starts with a simple scenario of two cells to explain how handover parameters affect handover failures. In practice, there exist multiple neighboring cells that have different wireless parameters such as transmit power, antenna gain, and propagation loss. Hence, we discuss the extension of our analytic result from a simple scenario to a general scenario at the end of this subsection.

In Fig. 2.3a, we depict a handover scenario of two small cells and a user moving from Cell 0 to Cell 1. Without loss of generality, we assume that the serving cell (Cell 0) is placed at the origin $O(0,0)$, and the target cell (Cell 1) is at $A(a,0)$ on the two dimensional Cartesian plane. In this figure, the dashed-line circles represent the transmission range determined by the minimum SINR, γ_{min} , to decode signals, whereas the solid-line circles illustrate the boundaries of the A3 events, according to the handover offset. We assume

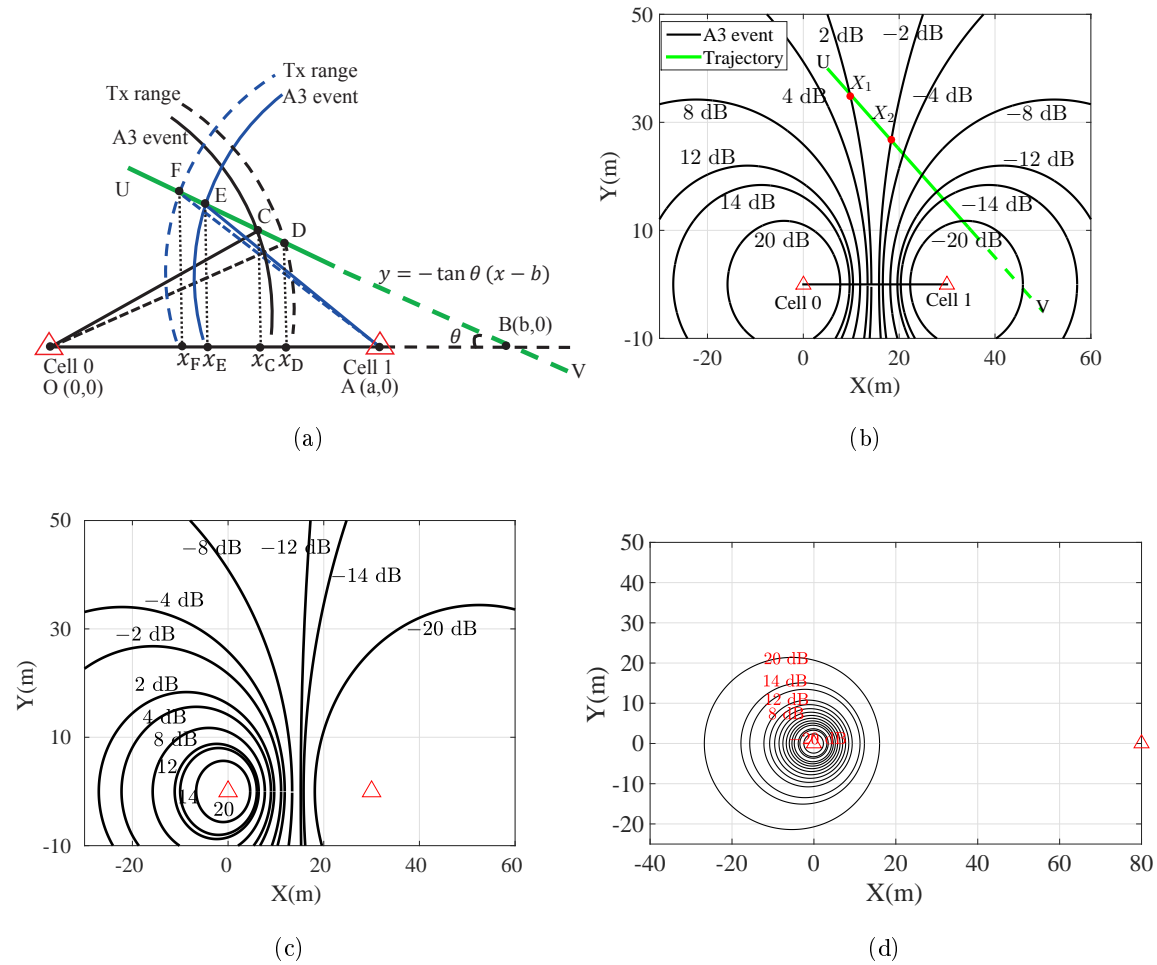


Figure 2.3: (a) User trajectory. (b) Apollonian circles for different Δ_{01} under identical settings ($\nu_{01} = 1$). (c) Apollonian circles for individual Δ_{01} values under different settings ($\nu_{01} = 1/4$). (d) Apollonian circles for an overlaid small cell ($\nu_{01} = 1/200$).

that a user moves on a linear path that is formulated as a line, ω ([35,36]), which is modeled as $\omega : y = -\tan\theta(x - b)$. The path ω intersects the x -axis at B ($b, 0$) and has a slope of $-\tan\theta$. We assume that θ is within the from range 0 to $\pi/2$, without loss of generality. Trajectory ω intersects the Apollonian circles of Cell 0 at C and D, which are the loci of the A3 event triggering (i.e., $M_0 < M_1 + \Delta_{01}$) and RLF triggering (i.e., the SINR from Cell 0, γ_0 , is less than the minimum required SINR, γ_{\min}), respectively. For Cell 1, trajectory ω also crosses the loci of the A3 event triggering (i.e., $M_1 < M_0 + \Delta_{10}$) at E and out-of-sync triggering (i.e., the SINR from Cell 1, γ_1 , is less than the minimum required SINR, γ_{\min}) at F. Next, we determine the loci C(x_C, y_C), D(x_D, y_D), E(x_E, y_E), and F(x_F, y_F).

To find C, we base it on (2.4) as $\Delta_{01} = M_0 - M_1$. Switching to the linear domain using an exponent of 10, we have

$$10^{\frac{\Delta_{01}}{10}} = \frac{m_0}{m_1} = \left(\frac{p_0}{p_1}\right) \left(\frac{G_0}{G_1}\right) \left(\frac{\eta_0}{\eta_1}\right) \left(\frac{\overline{\mathbf{OC}}}{\overline{\mathbf{AC}}}\right)^{-\alpha}. \quad (2.5)$$

After rearranging and taking an exponent of $-2/\alpha$ for both sides of (2.5), we have

$$\frac{x_C^2 + y_C^2}{(x_C - a)^2 + y_C^2} = \nu_{01}^{\frac{2}{\alpha}} 10^{\frac{-\Delta_{01}}{5\alpha}}, \quad (2.6)$$

where the ratio of Cell 0 parameters to Cell 1 parameters, ν_{01} ⁴, is given as $\nu_{01} = \frac{p_0 G_0 \eta_0}{p_1 G_1 \eta_1}$.

We rewrite (2.6) as

$$\left(x_C - \frac{k^2 a}{k^2 - 1}\right)^2 + y_C^2 = \left(\frac{ka}{k^2 - 1}\right)^2, \quad (2.7)$$

where $k = \nu_{01}^{\frac{1}{\alpha}} 10^{\frac{-\Delta_{01}}{10\alpha}}$. For an omni-directional antenna, (2.7) represents the Apollonian circle [46], where the coordinates of the center and the radius are $(k^2 a / (k^2 - 1), 0)$ and $ka / |k^2 - 1|$, respectively.

⁴The random variables for fading factors are independently initialized at positions C, D, E, and F. To address the randomness from fading factors, we compute position C, D, E, and F for each realization of the fading factors, and then, average the results.

According to the parameter settings, the Apollonian circle in (2.7) has different shapes, as shown in Fig. 2.3b and Fig. 2.3c. For simplicity, we start our analysis with identical parameters of two cells, although the analysis is applicable to a general setting.

Fig. 2.3b shows the effect of Δ_{01} on Apollonian circles with identical parameters ($\nu_{01} = 1$). When Δ_{01} decreases from -2 dB to -20 dB, leading to a rise of k , the center of the circle moves toward Cell 1. When Δ_{01} increases from 2 dB to 20 dB, causing a drop in k , the center of the circle moves toward Cell 0. When Δ_{01} equals 0 , i.e. k equals 1 , the circle becomes a straight line. The intersection of trajectory ω with the Apollonian circle is a decreasing function of Δ_{01} . Fig. 2.3b depicts two positions, X_1 and X_2 , which are intersections of a trajectory, UV , with the circles for 2 dB and -2 dB, respectively, and we have the x -coordinate of X_1 smaller than that of X_2 . This result plays a fundamental role in our analysis method. Since C belongs to ω , we find the coordinates of C via the following:

$$\begin{cases} \frac{x_C^2 + y_C^2}{(x_C - a)^2 + y_C^2} = \nu_{01}^{\frac{2}{\alpha}} 10^{\frac{-\Delta_{01}}{5\alpha}} \\ y_C = -\tan \theta (x_C - b). \end{cases} \quad (2.8)$$

Next, we find D on ω where the out-of-sync event occurs. We rearrange (2.1) as

$$\gamma_{\min} = \frac{\frac{m_0}{m_1}}{1 + \sum_{j=2}^{N-1} \frac{m_j}{m_1} + \frac{N_0}{m_1}} = \frac{m_0}{m_1} \frac{1}{\rho_0}, \quad (2.9)$$

where transmit power and antenna gain are identical for the cells, ρ_0 is the interference normalized by m_1 , which is defined as $\rho_0 = 1 + \sum_{j=2}^{N-2} m_j/m_1$. Multiplying both sides of (2.9) with ρ_0 , we have

$$\rho_0 \gamma_{\min} = \frac{m_0}{m_1} = \frac{\eta_0}{\eta_1} \left(\frac{\overline{\mathbf{OD}}}{\overline{\mathbf{AD}}} \right)^{-\alpha}. \quad (2.10)$$

Taking the exponent $-2/\alpha$ and rearranging elements for both sides of (2.10), an equivalent equation is obtained

$$\frac{x_D^2 + y_D^2}{(x_D - a)^2 + y_D^2} = \nu_{01}^{\frac{2}{\alpha}} (\rho_0 \gamma_{\min})^{\frac{-2}{\alpha}}.$$

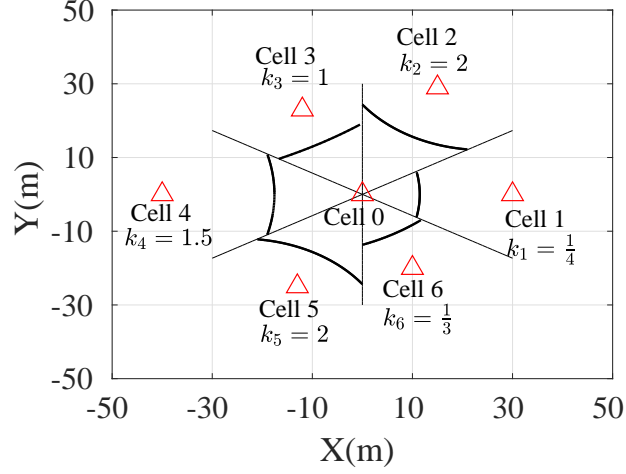


Figure 2.4: A3 event borders for handover offset -3 dB in six sectors of a heterogeneous environment.

Then, D can be found with the following:

$$\begin{cases} \frac{x_D^2 + y_D^2}{(x_D - a)^2 + y_D^2} = \nu_{01}^{\frac{2}{\alpha}} (\rho_0 \gamma_{\min})^{\frac{-2}{\alpha}} \\ y_D = -\tan \theta (x_D - b). \end{cases} \quad (2.11)$$

Similarly, E and F can be found as

$$\begin{cases} \frac{(x_E - a)^2 + y_E^2}{x_E^2 + y_E^2} = \nu_{01}^{\frac{-2}{\alpha}} 10^{\frac{-\Delta_{10}}{5\alpha}} \\ y_E = -\tan \theta (x_E - b), \end{cases} \quad (2.12)$$

$$\begin{cases} \frac{(x_F - a)^2 + y_F^2}{x_F^2 + y_F^2} = \nu_{01}^{\frac{-2}{\alpha}} (\rho_1 \gamma_{\min})^{\frac{-2}{\alpha}} \\ y_F = -\tan \theta (x_F - b), \end{cases} \quad (2.13)$$

where $\rho_1 = 1 + \sum_{j=2}^{N-2} m_j / m_0$, which is the interference normalized by m_0 .

Fig. 2.3c shows a scenario of two cells with different transmission power and antenna gains, such that $\nu_{01} = 1/4$. The handover triggering positions shift toward Cell 0, and the handover region of Cell 0 narrows due to the unbalanced settings. Similar to the identical

settings, locations C, D, E, and F can be obtained via (2.8), (2.11), (2.12), and (2.13), respectively.

In practice, there are multiple cells in wireless networks, and handovers to each neighboring cell have different wireless environments. We extend the approach to finding C, D, E, and F in a heterogeneous environment with multiple cells and different settings. Transmission power and antenna gain differ according to cell types. Also, beamforming gain can be captured in (2.2) by adopting the antenna gain as $G_i = |w_i^H \mathbf{H}_i u_i|$, where \mathbf{H}_i is the channel matrix between Cell i and the UE, u_i is a transmitting beamforming vector from Cell i to the UE, and w_i^H is the Hermitian transpose of the receiving beamforming vector of the UE toward Cell i [40,41]. If blockages and buildings interrupt a radio connection, G_i is multiplied by a factor for penetration loss [47].

Heterogeneous networks is a case of different settings, where small cells are overlaid on a macro cell with range expansion to enhance network capacity. As a macro cell has much greater transmission power and antenna gain than a small cell, the A3 event boundaries of the small cell are embedded inside the coverage of the macro cell as shown in Fig. 2.3d. Due to inappropriate handover settings, ping-pongs can occur between macro cell and small cell [32,35,36]. Also, handovers can fail due to low SINR [48]. Since the proposed analysis can model handover in general wireless settings, it is applicable to analyze handover problems in small-cell-overlaid heterogeneous networks.

Figure 2.4 visualizes a scenario of multiple cells with irregular deployment as well as different settings, where the UE moves from Cell 0 to neighboring cells. The figure shows that A3 event borders vary, sector by sector, even though the handover offset for each sector is identical. By considering them sector by sector, our analysis can be applied to a general wireless environment. In the next section, conditions for undesirable handovers are provided based on locations C, D, E, and F.

2.3.2 Modeling Undesirable Handovers

With the summary in Table 2.1 for a simple mobility, we extend to general conditions for undesirable handovers. To that end, user mobility as well as the triggering loci are projected onto the x -axis, as shown in Fig. 2.3a. For a too-late handover, handover triggering is delayed so that a UE enters deep inside a neighboring cell; hence, the link connection is disconnected due to strong interference from the neighboring cell. According to Condition (C1-1) in Table 2.1, the general condition for a too-late handover is expressed as

$$x_C + v\text{TTT}_0 \cos \theta > x_D + v\tau_R \cos \theta, \quad (2.14)$$

where the left side is moving distance during the handover, projected onto the x -axis, and the right side stands for the travelling distance that the UE experiences low SINR value ($\gamma_0 < \gamma_{\min}$).

For a too-early handover, the first handover is successful, but the second handover fails due to the RLF. Since the first handover was initialized at a spot far from the territory of the target cell, the next handover controlled by the target cell suffers tough interference from other cells. Also, TTT controlling the second handover is longer than time duration τ_R , which leads to the RLF. Based on conditions (C2-1) and (C2-2) in Table 2.1, this problem can be generally modeled as

$$\begin{cases} x_C + v\text{TTT}_0 \cos \theta + v\tau_R \cos \theta \leq x_F & (2.15) \\ \text{TTT}_1 > \tau_R. & (2.16) \end{cases}$$

In the left side of (2.15), $v\text{TTT}_0 \cos \theta$ represents the traveling distance during the first handover, and $v\tau_R \cos \theta$ represents the traveling distance before the RLF of the second handover, projected onto the x -axis. Condition (2.16) is to prevent the consequent handover from finishing earlier than the RLF.

Last, we model a ping-pong based on conditions (C3-1) and (C3-2) in Table 2.1 as follows

$$\begin{cases} x_C + v\text{TTT}_0 \cos \theta + v\text{TTT}_1 \cos \theta \leq x_E & (2.17) \\ \text{TTT}_1 < \tau_R. & (2.18) \end{cases}$$

In (2.17), $v\text{TTT}_0 \cos \theta$ and $v\text{TTT}_1 \cos \theta$, respectively, measure moving distances during the first handover (from Cell 0 to Cell 1) and the second handover (from Cell 1 to Cell 0), projected onto the x -axis. This condition enables triggering the A3 event for the second handover until its success. Condition (2.18) allows the second handover to finish before the RLF.

2.3.3 Optimal Range for Handover Offset

We define the conditions for an optimal value of Δ_{01} to minimize RLFs and ping-pongs together for a given ω as follows

$$x_C + v(\text{TTT}_0 - \tau_R) \cos \theta \leq x_D, \quad (2.19)$$

$$x_C + v(\text{TTT}_0 + \tau_R) \cos \theta > x_F, \quad (2.20)$$

$$x_C + v(\text{TTT}_0 + \text{TTT}_1) \cos \theta > x_E, \quad (2.21)$$

which are based on (2.14), (2.15), and (2.17). While Condition (2.19) is to avoid too-late handovers, conditions (2.20) and (2.21) are to avoid too-early handovers and ping-pongs, respectively. Based on Condition (2.19), we define the lower bound of Δ_{01} , called Δ_{01}^\dagger , for minimizing too-late handovers, as

$$\Delta_{01}^\dagger = \min \Delta_{01} \quad (2.22)$$

subject to (2.19),

$$O_{\min} \leq \Delta_{01} \leq O_{\max},$$

where O_{\min} and O_{\max} are the maximum and minimum values of Δ_{01} provided by the system, respectively. Therefore, if $\Delta_{01} < \Delta_{01}^\dagger$, a too-late handover occurs. Based on conditions (2.20) and (2.21), we define the upper bound, Δ_{01}^* , of Δ_{01} , for minimizing ping-pongs and too-early handovers, as

$$\begin{aligned} \Delta_{01}^* = \max \quad & \Delta_{01} \\ \text{subject to} \quad & (2.20) \text{ and } (2.21), \\ & O_{\min} \leq \Delta_{01} \leq O_{\max} \end{aligned} \quad (2.23)$$

Hence, if $\Delta_{01} > \Delta_{01}^*$, a ping-pong or a too-early handover occurs. If $\Delta_{01}^\dagger \leq \Delta_{01}^*$, the *optimal range* for minimizing RLFs and ping-pongs at the same time exists, and we denote the optimal range as $[\Delta_{01}^\dagger, \Delta_{01}^*]$. Otherwise, the optimal range does not exist (that is, MRO reduces RLFs while increasing ping-pongs, and vice versa).

Lemma 1 *The optimal range of Δ_{01} exists, i.e. $\Delta_{01}^\dagger \leq \Delta_{01}^*$, when*

$$\max \left\{ \frac{x_E - x_D}{\text{TTT}_1 + \tau_R}, \frac{x_F - x_D}{2\tau_R} \right\} \leq v \cos \theta. \quad (2.24)$$

Proof: See Appendix A.1.

Theorem 1 *When Condition (2.24) is met, the optimal range $[\Delta_{01}^\dagger, \Delta_{01}^*]$ is prolonged when speed increases, and shrinks when speed decreases.*

Proof: See Appendix A.2.

Theorem 2 *When Condition (2.24) is met, the optimal range $[\Delta_{01}^\dagger, \Delta_{01}^*]$ is shifted to the left when TTT_0 decreases, and to the right when TTT_0 increases.*

Proof: See Appendix A.3.

Since a handover offset can differently influence handover performance on individual trajectory $\omega \in \Omega$ ([49, 50]), where Ω is a set of ω , the problems to find Δ_{01}^\dagger and Δ_{01}^* for

Ω are stated as

$$\Delta_{01}^\dagger = \max_{\omega \in \Omega} \Delta_{01}^\dagger(\omega), \quad (2.25)$$

$$\Delta_{01}^* = \min_{\omega \in \Omega} \Delta_{01}^*(\omega), \quad (2.26)$$

where $\Delta_{01}^\dagger(\omega)$ and $\Delta_{01}^*(\omega)$ are given in (2.25) and (2.26) for trajectory $\omega \in \Omega$. In practice, Ω can represent a street, road, lane, or pavement where UEs usually move, and Ω is one of the vital sources of context information for handover optimization [49, 50].

2.3.4 Discussion of The Optimal Range of Handover Offset

In general, the closed forms of Δ_{01}^\dagger and Δ_{01}^* are unobtainable owing to nonlinear complex forms in (2.8), (2.11), (2.12), and (2.13). Hence, it is difficult to get any intuition into the relationships between handover parameters and handover failures. For the special case when UE moves in a straight line between the centers of Cell 0 and Cell 1 (i.e., $b = 0$ and $\theta = 0$), we can express Δ_{01}^\dagger and Δ_{01}^* as closed forms. With the closed forms, we will discuss the impact of parameters on the optimal range.

Theorem 3 *The optimal range of the handover offset is at maximum when $\theta = 0$ and $b = 0$.*

Proof: See Appendix A.4 When the UE moves from Cell 0 to Cell 1 on a trajectory with $\theta = 0$ and $b = 0$, the y -coordinate values of the position are all zero, i.e., $y_C = y_D = y_E = y_F = 0$. The x -coordinates, x_C , x_D , x_E , and x_F , are determined based on (2.8), (2.11), (2.12), and (2.13) as follows:

$$x_C = \frac{a \nu_{01}^{\frac{1}{\alpha}}}{\nu_{01}^{\frac{1}{\alpha}} + 10^{\frac{\Delta_{01}}{10\alpha}}}, \quad x_D = \frac{a \nu_{01}^{\frac{1}{\alpha}}}{\nu_{01}^{\frac{1}{\alpha}} + (\rho_0 \gamma_{\min})^{\frac{1}{\alpha}}}, \quad x_E = \frac{a \nu_{01}^{\frac{1}{\alpha}} 10^{\frac{\Delta_{10}}{10\alpha}}}{1 + \nu_{01}^{\frac{1}{\alpha}} 10^{\frac{\Delta_{10}}{10\alpha}}}, \quad x_F = \frac{a(\nu_{01} \rho_1 \gamma_{\min})^{\frac{1}{\alpha}}}{1 + (\nu_{01} \rho_1 \gamma_{\min})^{\frac{1}{\alpha}}}.$$

The lower bound of the optimal handover offset, Δ_{01}^\dagger , is found based on Condition (2.19) for too-late handover avoidance: $x_C + v_{TTT0} = x_D + v\tau_R$. After rearrangement, we obtain

Δ_{01}^\dagger as

$$\Delta_{01}^\dagger(\text{dB}) = 10 \log(\nu_{01}) - 10\alpha \log \left(\frac{1}{\frac{(\rho_0 \gamma_{\min})^{\frac{1}{\alpha}}}{\nu_{01}^{\frac{1}{\alpha}} + (\rho_0 \gamma_{\min})^{\frac{1}{\alpha}}} - \frac{v(\tau_R - \tau_{TT0})}{a}} - 1 \right). \quad (2.27)$$

The upper bound of the optimal range, Δ_{01}^* , is determined as the minimum between upper bound $\bar{\Delta}_{01}^*$ obtained from the too-early avoidance condition, and upper bound $\tilde{\Delta}_{01}^*$ acquired for the ping-pong minimization condition. In short, $\Delta_{01}^* = \min(\bar{\Delta}_{01}^*, \tilde{\Delta}_{01}^*)$. From (2.20) and by reordering, $\bar{\Delta}_{01}^*$ is found to be

$$\bar{\Delta}_{01}^*(\text{dB}) = 10 \log(\nu_{01}) - 10\alpha \log \left(\frac{1}{\frac{1}{1 + (\nu_{01} \rho_1 \gamma_{\min})^{\frac{1}{\alpha}}} + \frac{v(\tau_R + \tau_{TT0})}{a}} - 1 \right). \quad (2.28)$$

Based on Condition (2.21) for ping-pong reduction, we get $\tilde{\Delta}_{01}^*$ as

$$\tilde{\Delta}_{01}^*(\text{dB}) = 10 \log(\nu_{01}) - 10\alpha \log \left(\frac{1}{\frac{1}{1 + \nu_{01}^{\frac{1}{\alpha}} 10^{\frac{\Delta_{10}}{10\alpha}}} + \frac{v(\tau_{TT1} + \tau_{TT0})}{a}} - 1 \right). \quad (2.29)$$

It follows from (2.27), (2.28), and (2.29) that lower bound Δ_{01}^\dagger is a decreasing function of UE speed v , while upper bound Δ_{01}^* is an increasing function of v . Therefore, the optimal range broadens when v increases, whereas it narrows when v decreases, as discussed in Theorem 1. On the other hand, lower bound Δ_{01}^\dagger and upper bound Δ_{01}^* are both increasing functions of τ_{TT0} . As a result, the optimal range shifts to the right when τ_{TT0} increases, and shifts to the left when τ_{TT0} decreases, as explained in Theorem 2.

When UE speed is low compared to the inter-distance between serving Cell 0 and target Cell 1, i.e., $v/a \ll 1$, for example, and UE speed v is low in a pedestrian environment, and we approximate (2.27), (2.28), and (2.29) as

$$\Delta_{01}^\dagger(\text{dB}) \approx 10 \log(\rho_0 \gamma_{\min}), \quad (2.30)$$

$$\bar{\Delta}_{01}^*(\text{dB}) \approx -10 \log(\rho_1 \gamma_{\min}),$$

$$\tilde{\Delta}_{01}^*(\text{dB}) \approx -\Delta_{10}.$$

Then, we get Δ_{01}^* as

$$\Delta_{01}^* (\text{dB}) = \min(\bar{\Delta}_{01}^*, \tilde{\Delta}_{01}^*) = -\max(10 \log(\rho_1 \gamma_{\min}), \Delta_{10}). \quad (2.31)$$

Equations (2.30) and (2.31) reveal the impact of interference on the optimal range. When normalized interference ρ_0 and ρ_1 increase, the optimal range is tightened; and the optimal range is enlarged when ρ_0 and ρ_1 decrease.

2.4 Numerical Analysis and Discussion

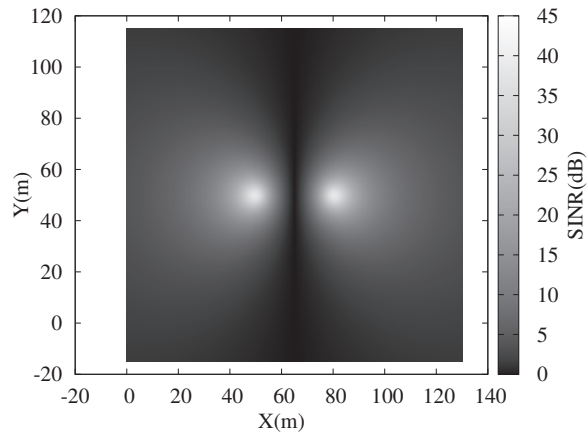
2.4.1 Simulation Environment

To measure handover performance, we considered too-late and too-early/wrong-cell handovers, and ping-pongs within a certain period of time. For Cell i , we define \mathcal{R}_{L_i} for measuring the too-late handover rate. Since a ping-pong is similar to a too-early handover, except for an RLF, we define \mathcal{R}_{E_i} for measuring too-early handover and ping-pong rates together. \mathcal{R}_{L_i} and \mathcal{R}_{E_i} are given as

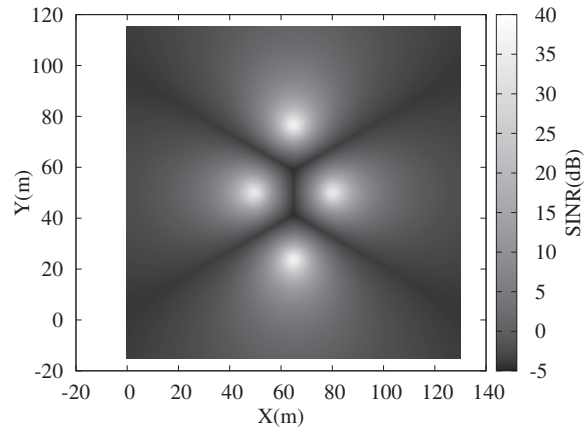
$$\mathcal{R}_{L_i} = \frac{\sum_{j=1}^N N_{L_{ij}}}{N_{\text{total}}}, \quad \text{and} \quad \mathcal{R}_{E_i} = \frac{\sum_{j=1}^N (N_{E_{ij}} + N_{PP_{ij}})}{N_{\text{total}}},$$

where $N_{L_{ij}}$, $N_{E_{ij}}$, $N_{PP_{ij}}$, and N_{total} are the numbers of too-late handovers, too-early handovers, ping-pongs, and total handover attempts from Cell i to adjacent Cell j , respectively.

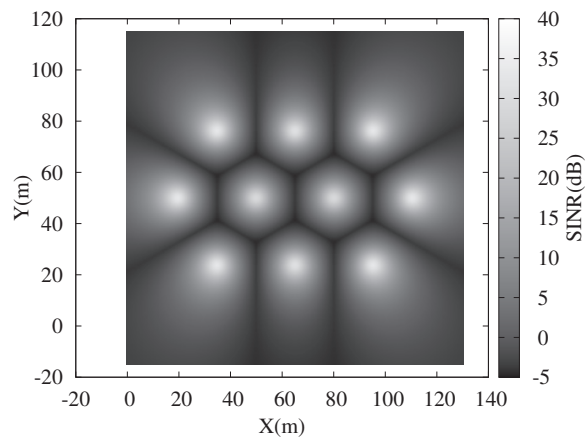
To verify the proposed model of handover problems, we compared our analysis with an NS-3 simulation [1] in various environments, taking into account the effect of user mobility, fading, inter-cell interference, handover parameters, and deployment of cells. For the propagation model, α was selected as 4.33 [20]. To model fading, a log-normal random variable together with Rayleigh fading was integrated into the propagation model, even though fast fading can be significantly filtered out by appropriate filter settings [35, 36]). This random variable has a zero mean, and standard deviation σ is 6 dB, as in a typical



(a) Two cells.



(b) Four cells.



(c) Ten cells.

Figure 2.5: Regular deployment of small cells.

Table 2.2: Summary of parameters for simulation.

Parameter	Description	value
Δ_{01}	Handover offset of Cell 0 on Cell 1 (dB)	-14 to 14
τ_{TT0}	Time-to-trigger of Cell 0 (ms)	256
Δ_{10}	Handover offset of Cell 1 on Cell 0 (dB)	2
τ_{TT1}	Time-to-trigger of Cell 1 (ms)	512
Q_{out}	Out-of-sync threshold (dB)	-4
Q_{in}	In-sync threshold (dB)	-2
τ_R	RLF detection time (ms)	500
a	Inter-site distance (m)	30
σ	Log-normal standard deviation (dB)	0, 6, 12
ω	User trajectory	
θ	Angle formed by ω x -axis (rad)	$\mathcal{U}(\pi/24, 3\pi/24)$
b	Point where ω meets x -axis (m)	$\mathcal{U}(5, 25)$
v	User speed (km/s)	5, 10, 20
α	Path loss exponent	4.33

environment [51]. An omni-directional antenna was used for all small cells. For handover settings, TTT_0 , Δ_{10} , and TTT_1 were respectively fixed at 256 ms, 2 dB, and 512 ms, while Δ_{01} varied in the range -14 to 14 dB, with granularity of 1 dB. For RLF triggering, q_{out} (i.e., γ_{min}) was set at -4 dB [52], and τ_{R} was 500 ms [1]. We set values of TTT_0 and TTT_1 according to 3GPP standard [18] and previous works [1,31]. A summary of the parameters in this section is provided in Table 2.2.

2.4.2 Impact of User Mobility

To study how user mobility affects the optimal range, we considered a linear trajectory with five different patterns (two deterministic and three random patterns) in a two-cell network with inter-site distance a of 30 m, as shown in Fig. 2.5a. The first mobility pattern is a straight line trajectory, ω , where $b = 0$ and $\theta = 0$. The second mobility pattern represents a diagonal trajectory ω , where $b = 15$ and $\theta = \pi/12$. The third mobility pattern has a deterministic b of 15, but θ is uniformly distributed between $\pi/24$ and $3\pi/24$ so that the average of θ is $\pi/12$. The fourth mobility pattern has θ fixed at $\pi/12$ and b uniformly distributed between 5 and 25 so that the average of b is 15. The last mobility pattern has randomness in b and θ together, which are uniformly distributed between 5 and 25 and between $\pi/24$ and $3\pi/24$, respectively. User speed v is 5 km/h to represent a pedestrian environment. Handover performance from Cell 0 to Cell 1 is analyzed, unless stated otherwise.

For deterministic mobility models, the analysis gives optimal ranges exactly as in the simulation, which are $[-3, 3]$ dB and $[-3, 2]$ dB, as depicted in Fig. 2.6a and Fig. 2.6b, respectively. The first optimal range is larger than the second one, because the first trajectory gives the maximum optimal range, as in Theorem 3. Regarding mobility models with randomness, comparing the results in Fig. 2.6c, Fig. 2.6d, and Fig. 2.6e to those in Fig. 2.6b, ratios \mathcal{R}_{L_0} and \mathcal{R}_{E_0} slightly increase, and thereby, the optimal ranges are narrower

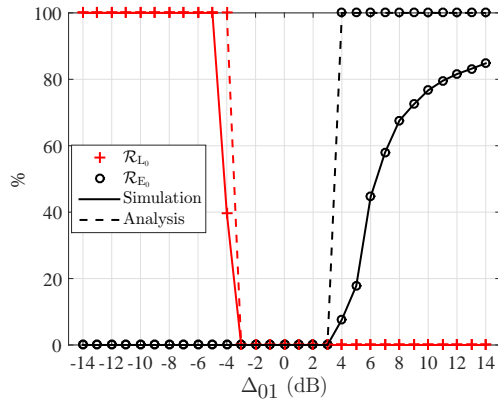
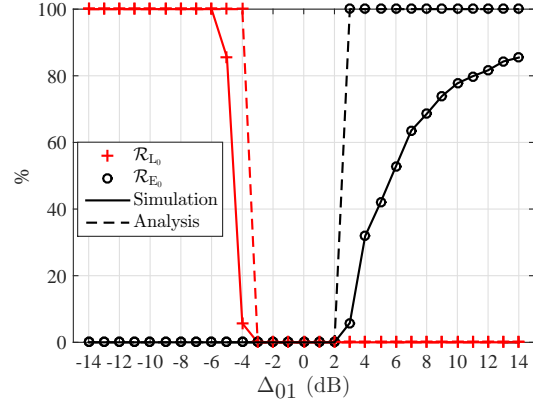
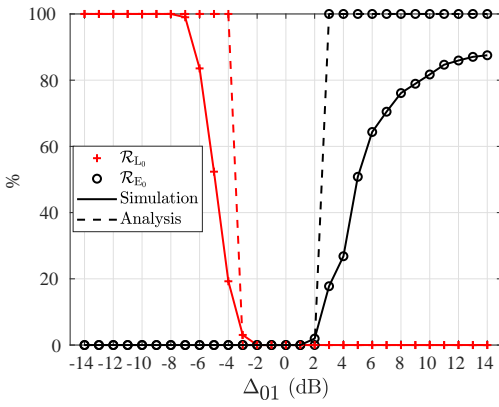
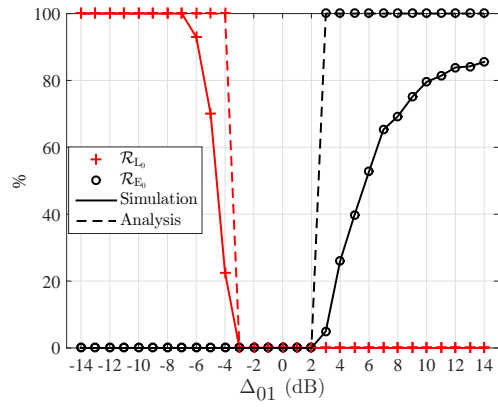
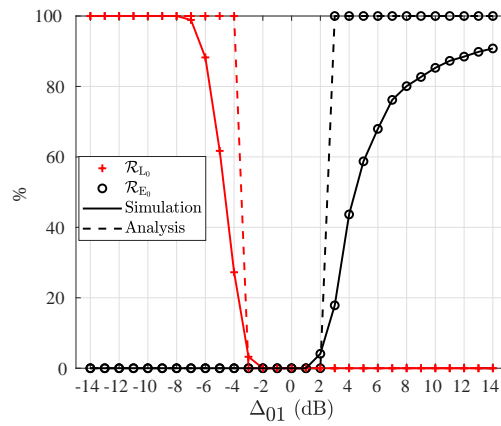
(a) $b = 0, \theta = 0$.(b) $b = 15, \theta = \pi/12$.(c) $b = 15, \theta \sim \mathcal{U}(\pi/24, 3\pi/24)$.(d) $b \sim \mathcal{U}(5, 25), \theta = \pi/12$.(e) $b \sim \mathcal{U}(5, 25), \theta \sim \mathcal{U}(\pi/24, 3\pi/24)$.

Figure 2.6: Impact of user mobility (two cells).

than in Fig. 2.6b. However, they have similar shapes. In general, the results show that the optimal range can be different due to user mobility, but our estimation matches the simulation results for deterministic models, and deviates from the simulation results with a value of handover offset resolution for random mobility models. We selected the fourth mobility pattern to model the user trajectory, since UEs tend to move in a specific direction, such as street lanes and on pavement.

2.4.3 Impact of Heterogeneous Settings

We investigated the impact of heterogeneous parameter settings of the two cells in Fig. 2.3c on handover performance. In practice, cell types can be different, such as a pico cell, a femto cell, or a macro cell, and have a different antenna gains [53]. Hence, we set the transmission power of Cell 0 and Cell 1 to 23 dBm, while the antenna gain for Cell 0 and Cell 1 was alternatively configured as 2 dBi and 8 dBi [47], to have ν_{01} of 1/4 and 4. For $\nu_{01} = 1/4$, A3 event positions of Cell 0 shift toward Cell 0 in Fig. 2.3c, compared to Fig. 2.3b for $\nu_{01} = 1$. Similarly for $\nu_{01} = 4$, A3 event positions of Cell 0 move toward Cell 1.

Fig. 2.8 shows that the optimal range for handover offset ($[-2, 1]$ dB) is narrower than that in Fig. 2.6d ($[-3, 2]$ dB) when $\nu_{01} = 1$. When $\nu_{01} = 1/4$, coverage of Cell 0 reduces, and thereby, handovers to Cell 1 have less time for success, and may fail before TTT expires. On the other hand, the optimal range in Fig 2.7b is wider than in Fig. 2.6d when $\nu_{01} = 4$. We also investigate a case of a small-cell-overlaid heterogeneous network as depicted in Fig. 2.3d. The distance between cells is 80 m. Transmission powers of Cell 0 (small cell) and Cell 1 (macro cell) are 23 dBm and 46 dBm, respectively, for which $\nu_{01} = 1/200$ being an extreme case of different settings. Fig. 2.7c shows that our analysis meets the simulation correctly. The optimal range of Δ_{01} is $[-2, 1]$ dB. In overall, our analysis can estimate the optimal range of handover offset in different wireless settings, including small-cell-overlaid networks.

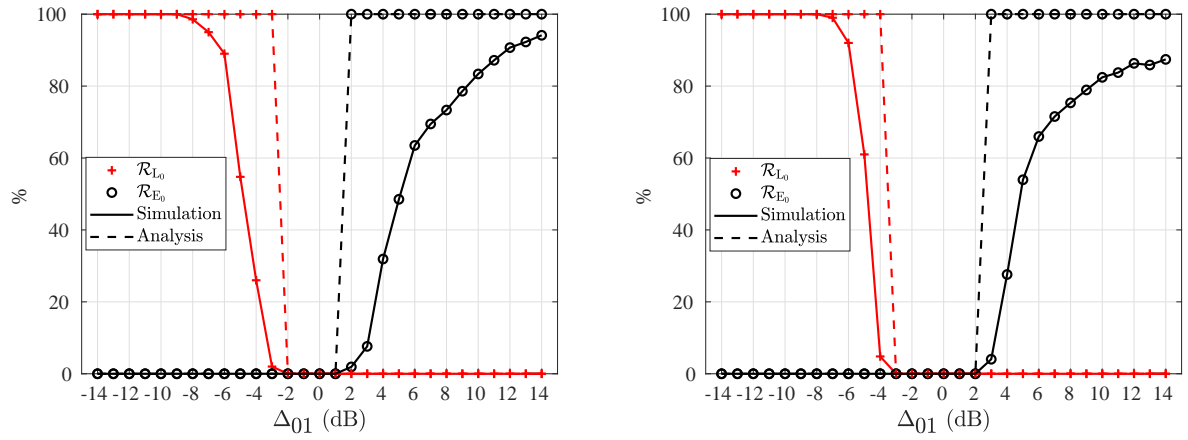
2.4.4 Impact of Fading and Interference

To analyze the effect of fading on handover performance, we kept the same scenario, except for fading factor σ . Compared to Fig. 2.6d, Fig. 2.9 shows that the optimal range narrows from $[-3, 2]$ dB to $[-2, 1]$ dB when σ increases to 12 dB, and it is extended from $[-3, 2]$ dB to $[-4, 4]$ dB when σ decreases to 0 dB. The results indicate that our analysis can estimate the optimal range with high accuracy in various fading environments. To investigate the impact of interference on handover, we increased the number of small cells to four and 10 while keeping the same environment as in Fig. 2.6d. Deployments of four small cells and 10 small cells are shown in Fig. 2.5b and Fig. 2.5c, respectively. Comparing these results to Fig. 2.6d, Fig. 2.10 shows that interference narrows the optimal range of handover offset from $[-3, 2]$ dB to $[-2, 2]$ dB, as discussed in Section III-D. Comparing simulation results and analytical results, our analysis fits the simulation well.

2.4.5 Impact of Speed and TTT

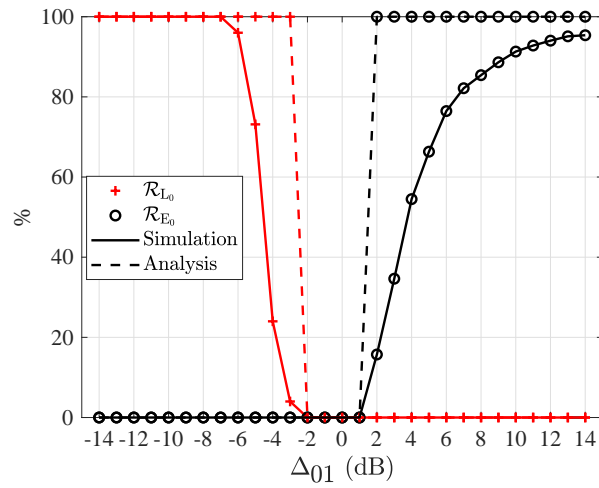
The impacts of user speed and TTT on handover performance have been studied [33, 35, 54]; however, the specific effect of speed as well as TTT on the optimal setting is outside the scope of those studies. To assess the influence of user speed on the optimal range, we used the same parameters as in Fig. 2.10a, except for user speed. Compared to Fig. 2.10a, Fig. 2.11 shows that the optimal range is extended when speed increases, and as well, our analysis accurately estimates the optimal range. When user speed increases to 10 km/h (or 20 km/h), the optimal range is extended to $[-2, 4]$ dB (or $[-5, 9]$ dB), as discussed in Theorem 1 and Section III-D.

To assess the effect of TTT on the optimal range, we chose speeds of 5 km/h and 20 km/h while varying TTT from 160 ms to 480 ms and keeping other parameters as seen in Fig. 2.10a. When TTT increases, Fig. 2.12a depicts a small shift in the optimal range for all



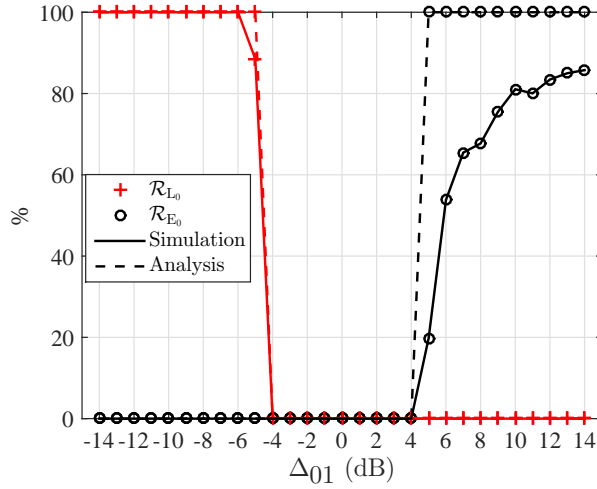
(a) $\nu_{01} = 1/4$.

(b) $\nu_{01} = 4$.



(c) $\nu_{01} = 1/200$.

Figure 2.7: Impact of different settings (two cells).

Figure 2.8: $\nu_{01} = 1/200$.

values of TTT in a low-speed environment, whereas Fig. 2.12b presents a significant shift in a high-speed environment. The optimal range is $[-4, 9]$ dB for $\tau_{TT} = 160$ ms and $[-2, 12]$ dB for $\tau_{TT} = 480$ ms. Since (2.27), (2.28), and (2.29) are affected by ν_{TTT} , the impact of TTT in a high-speed environment is more than in a low-speed environment. Also, the optimal range shifts to the right when TTT increases, and to the left when TTT decreases, as discussed in Theorem 2 and Section III-D.

2.4.6 Irregular Deployment of Small Cells and Non-existence of the Optimal Setting

Since deployment can be irregular in real wireless networks, we borrowed a field trial scenario from [55] to test the accuracy of the estimation in a practical deployment of small cells. We choose 10 cells near a shopping center, as provided in Fig. 3 of [55], and plotted the area in Fig. 2.13a. The parameters are the same as Fig. 2.10b except for the deployment of the cells. The mobility area between the two center cells was considered. Fig. 2.13b shows that our analysis provides an almost accurate estimation of the optimal

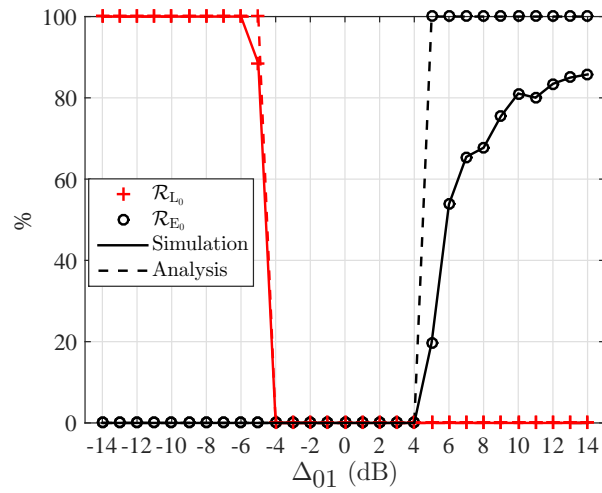
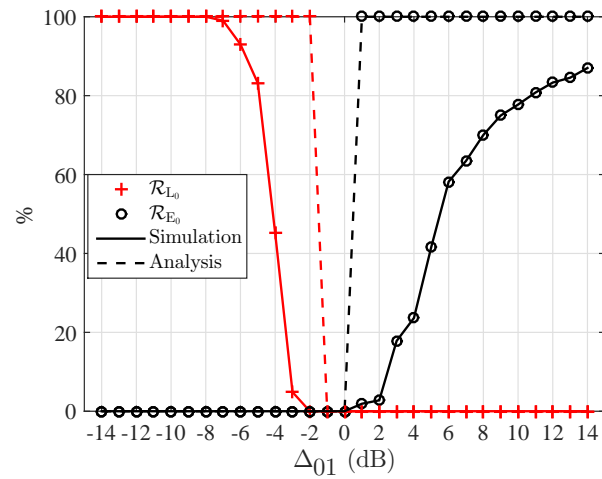
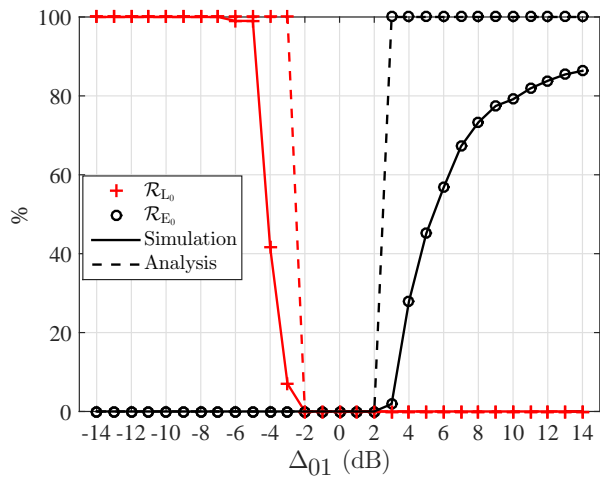
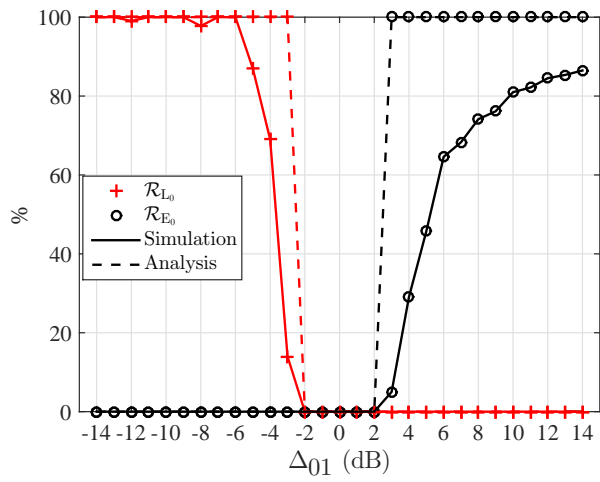
(a) $\sigma = 0$ dB.(b) $\sigma = 12$ dB.

Figure 2.9: Impact of fading (two cells).

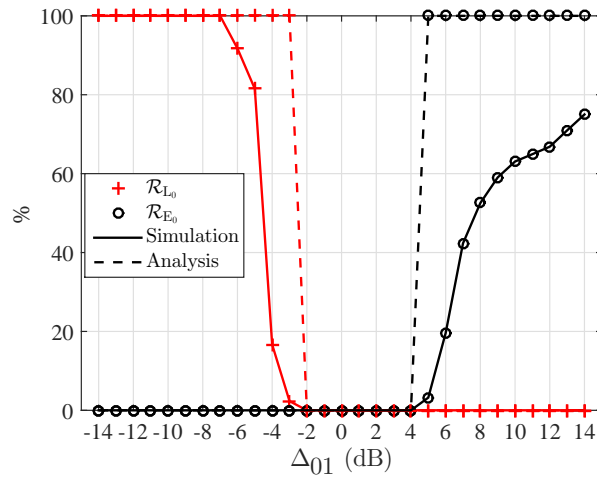


(a) Four cells.

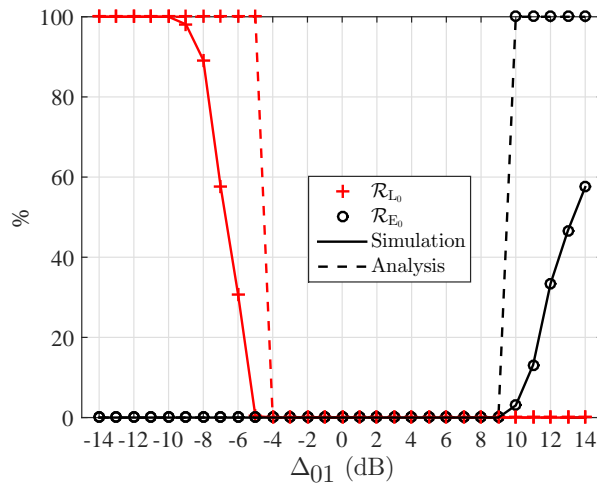


(b) 10 cells.

Figure 2.10: Impact of interference.

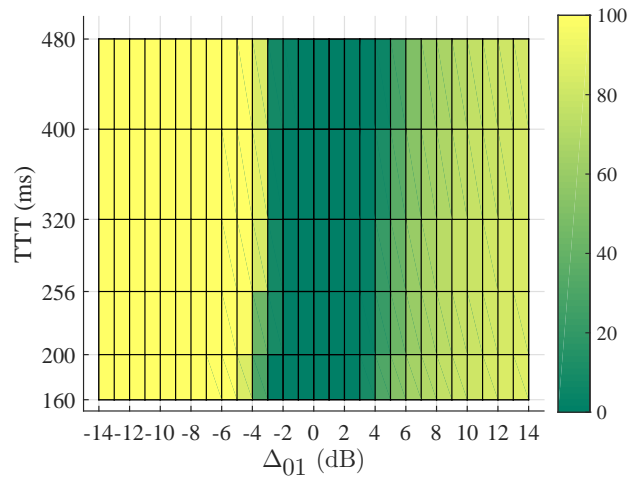


(a) 10 km/h.

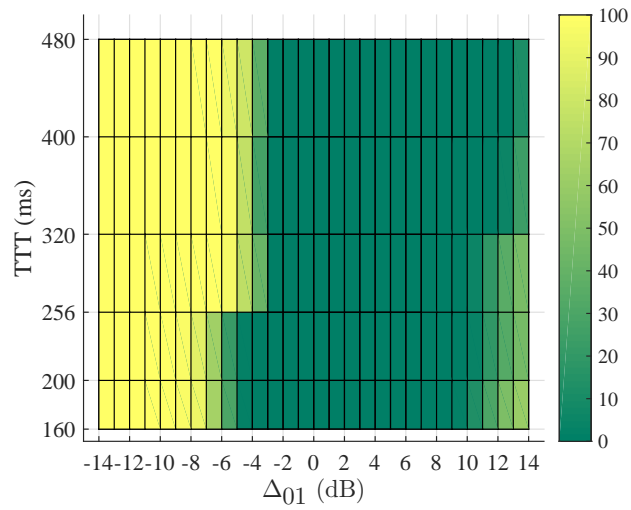


(b) 20 km/h.

Figure 2.11: Impact of speed (four cells).



(a) 5 km/h.



(b) 20 km/h.

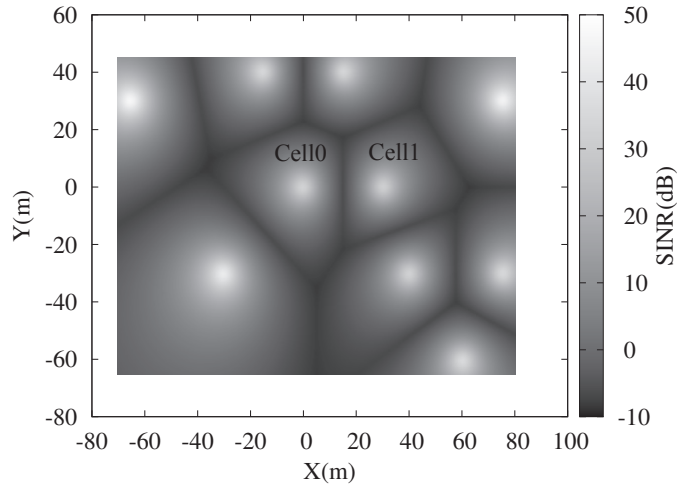
Figure 2.12: Impact of TTT (four cells).

range. The optimal range is $[-3, 2]$ dB via simulation and $[-2, 2]$ dB via analysis. Therefore, our analysis is applicable to estimating the optimal range for an irregular deployment of small cells in practice.

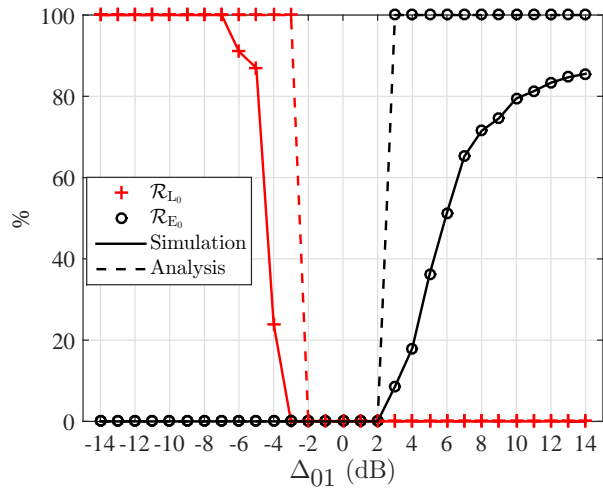
To verify the non-existence of the optimal range, we considered the scenario in Fig. 2.10a, and set Δ_{10} to 4 dB. Via Lemma 1, the condition for the existence of the optimal range is violated (i.e., $\Delta_{01}^\dagger > \Delta_{01}^*$), so there is no optimal range. As shown in Fig. 2.14, when $\Delta_{01} > -3$ dB, too-late handovers improved while ping-pongs as well as too-early handovers worsened, and vice versa. In our analysis, $\Delta_{01}^\dagger = -1$ dB and $\Delta_{01}^* = -5$ dB, which is close to the simulation. Since our analysis provides an accurate estimation, even for non-existence of the optimal range, we can apply the analysis in MRO to assess a handover setting, even if it induces no optimal solution for minimizing RLFs and ping-pongs.

2.5 Closing Remarks

In order to estimate an optimal handover setting in SON-based small-cell networks, we analyzed undesirable handovers that lead to RLFs or ping-pongs, and provided an analysis of undesirable handovers. Based on the analysis and an NS-3 simulation, we verified the existence of an optimal range for handover parameters in order to minimize RLFs and ping-pongs at the same time. The analytic estimation of the optimal range for handover parameters is highly accurate, when compared to simulation results. Furthermore, we provided a comprehensive study of handover performance in various environments that require different optimal settings. When a wireless channel experiences more fading or interference, the optimal range for handover offset shrinks. The optimal range grows when user speed increases, and shrinks when user speed decreases. The optimal range shifts to the left when TTT decreases, and to the right when TTT increases. Also, verification in a heteroge-



(a) Deployment.



(b) 5 km/h.

Figure 2.13: Irregular deployment of 10 small cells and the results.

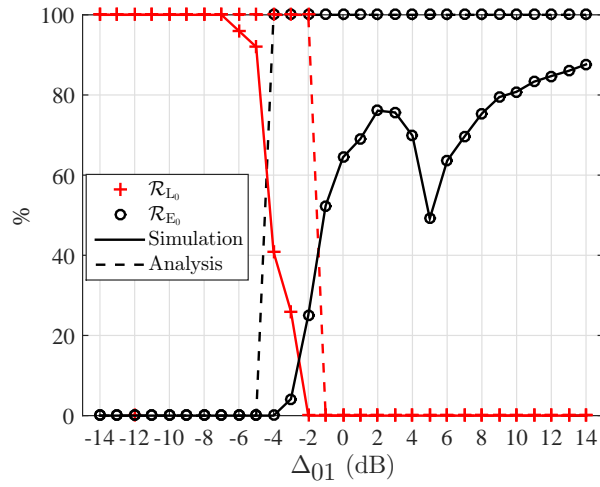


Figure 2.14: Non-existence of the optimal range.

neous scenario and a real deployment of small cells gave a high-accuracy estimation. From the analysis, future work will be developing an MRO algorithm for a SON that considers estimation of the optimal range in order to minimize RLFs and ping-pongs at the same time.

This page intentionally left blank.

Chapter 3

Machine Learning–Based Mobility Robustness Optimization Under Dynamic Cellular Networks.

In this chapter, we propose a mobility robustness optimization algorithm for dynamic small-cell networks. Due to their nature and energy-saving functionality, small cells can be arbitrarily deployed, moved, and turned on and off. As a result, the small-cell network topology varies from time to time. However, previous works have only considered dynamics due to user mobility in the given static networks. To optimize handovers under a dynamic network topology and user mobility together, we propose a machine learning–based algorithm that consists of two steps: topology adaptation and mobility adaptation. The first step obtains prior knowledge on the optimal handover settings for the given network topology and transfers that knowledge to the second step. The second step fine-tunes the handover parameters to adapt to the mobile environment based on the transferred knowledge. Via simulation, under a dynamic wireless network, the proposed algorithm reduced adaptation

time to 4.17% of the time needed by a comparative machine-based algorithm. Furthermore, the proposed algorithm improved the user satisfaction rate to 416.7% compared to the previous work.

3.1 Introduction

A small cell has been considered an essential approach to high network capacity and high spectrum reuse in order to accommodate traffic demands in the fifth-generation (5G) era [26]. Because the small cell is a low-power, low-cost, and ready-to-use prototype of a base station [27], network providers can activate their small cells in a planned or unplanned manner in crowded places (such as shopping malls, stadiums, and downtown streets [28]) to improve the user's quality of experience.

To reduce administrative expenses from frequent parameter optimization, a small-cell network is equipped with an automation-inspired solution called a self-organizing network (SON) [27,29]. The importance of a SON to small-cell networks is highlighted by its potential features, such as coverage and capacity optimization, energy efficiency, interference reduction, mobility load balancing (MLB), mobility robustness optimization (MRO), automatic neighbor relations, and random access optimization [1,56]. To ensure reliable and efficient operation of small cells, the handover procedure should be optimized by MRO to guarantee seamless mobility throughout the cellular network. MRO minimizes handover failures due to radio link failures (RLFs) and keeps the number of ping-pongs as low as possible, since RLFs disturb the user experience and ping-pongs induce resource-consuming handovers in a short time. Therefore, RLFs and ping-pongs should both be minimized.

Due to the dynamic operation of small cells, such as switching them on and off, the network topology is non-static. Small cells can be switched on and off by personal usage

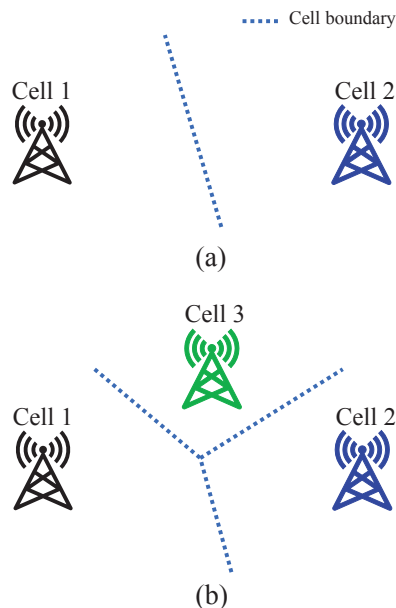


Figure 3.1: Dynamic topologies with cells switching on and off (a) two cells are switched on, and (b) three cells are switched on.

or by network operators to adapt to traffic loads and energy-saving demands [26]. Fig. 3.1 shows an example of different topologies when cells are switched on and off. Based on cell 3’s activation, the coverage of cells changes, so handover parameters should be modified. The wireless environment, such as cell boundaries, changes according to the network topology, thus affecting network optimization and performance [57]. Along with user mobility and dense deployment of small cells, the dynamic topology is a challenging issue for handover optimization.

Previous MRO algorithms have considered the adjusting of handover parameters while considering user mobility and dense deployment of small cells [1, 7, 9, 30, 38]. The authors in [7,9,38] improved handover performance by adjusting time-to-trigger (TTT) and handover offset. The algorithms are rule-based and take into account negotiation between RLFs and ping-pongs. In [30], the authors provided a more flexible optimization policy that regulates cell individual offsets (CIOs). In [1], the authors proposed an algorithm that

manipulates TTT, handover offset, and CIOs together in order to improve handover performance. That algorithm achieved an impressive handover performance without a negotiation between RLFs and ping-pongs. Machine learning techniques have also been applied in wireless networks to optimize handover parameters. In [14, 31, 58, 59], reinforcement learning (RL) was applied to adjust handover parameters. In the RL algorithm, the learning agent interacts with the outside environment to choose the optimal action based on feedback from the environment [60]. Studies in [14, 31, 58] considered a model-free RL, modeling the reward as load level (for MLB) and handover performance (for MRO), while the actions that modified handover parameters (TTT and hysteresis) were chosen according to fixed policies. In [59], the authors utilized fuzzy-based policies with RL-based algorithms to change handover offsets in order to improve handover quality. Even though previous works adjusted handover parameters while taking into account user mobility and cell deployment, they applied only to a static network topology. Therefore, optimizing handover parameters under dynamic wireless networks that have a dynamic topology as well as mobility still poses challenges to wireless network optimization.

To address the challenges in dynamic wireless networks, an optimization algorithm should have capabilities to quickly respond to network topology changes while achieving stable performance. A candidate solution is to exploit prior knowledge in optimizing different wireless networks, and to apply this experience to the new wireless network. Prior knowledge can consist of the optimal parameters from previous optimizations and/or knowledge bases of optimal parameters. Transfer learning (TL) is one of the machine learning techniques that utilize prior knowledge from previous optimizations for better performance under a new problem, because those parameters will provide better initialization instead of nothing. Applications of TL algorithms for wireless networks are found in [57, 61–63], but none of them apply the TL concept to handover optimization under dynamic wireless networks.

In this chapter, to optimize handover parameters for dynamic wireless networks, we propose an MRO algorithm that takes into account prior knowledge in optimizing different wireless environments. The algorithm includes a knowledge transfer algorithm for adapting to a dynamic topology and an adaption algorithm to suit dynamic mobility. To adapt to a dynamic topology, the TL-based algorithm gathers prior knowledge that includes the estimated optimal handover parameters, which are based on network topology, and knowledge bases of optimal parameters to adapt to dynamic on/off cells. Then, the adaptation algorithm uses the transferred knowledge to optimize three handover parameters (TTT, hysteresis, and CIO) together, in order to suit dynamic mobility. Via simulation and analysis, we verify our proposed algorithms under a dynamic topology, random mobility, and irregular deployment of small cells.

3.2 System Models and Problem Formulation

3.2.1 The SON and the Small-cell Network

In this chapter, we consider a hierarchical SON that supervises small cells, as shown in Fig. 3.2. Two types of SON are taken into account: the centralized SON (cSON) and the decentralized SON (dSON). While a cSON is located under network management for network-wide optimization, dSONs are implemented with the small cells for local adaptation. The cSON receives information from dSONs to keep track of network-wide information in order to optimize the network globally [64]. When a new cell is activated, the dSON of this cell receives initial settings from the cSON, which includes cell identification (ID), handover parameters, neighboring cell information, and other control parameters.

Due to energy saving or the arbitrary on/off manner of small cells, dSONs send notifications to the cSON before the corresponding cells are deactivated. Hence, the cSON

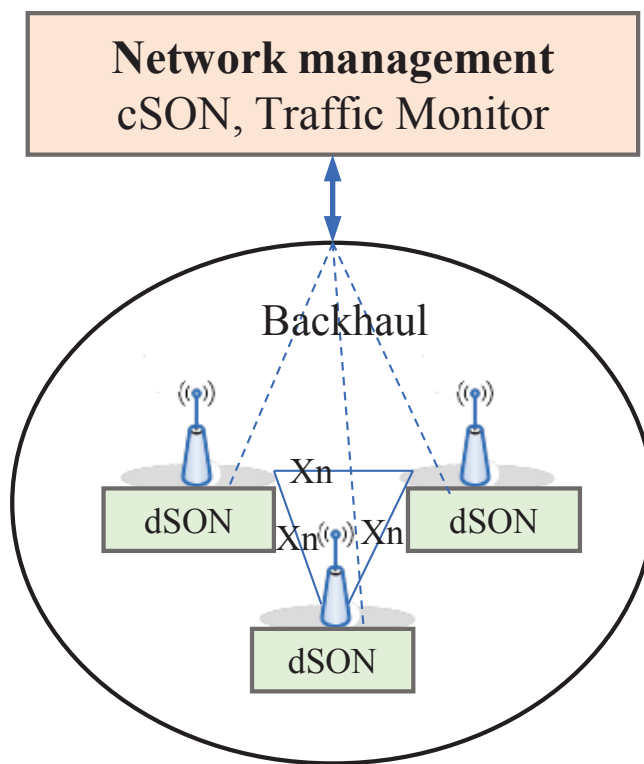


Figure 3.2: A scenario with a small-cell network and its SON.

always keeps updating the network topology in order to produce global adjustments. Furthermore, the cSON communicates with smart traffic monitoring systems [65] to monitor the environmental context, such as average user speed and traffic flows on the streets. Therefore, the environmental context is partially known and is used as input for optimization algorithms by the cSON.

For mobility management, small cells communicate with their neighbors via Xn interface to share information. Via the Xn interface, nearby small cells exchange user information, such as handover information and failure notification messages [17]. Based on the recorded data and the cSON's support, the dSON adapts to dynamic mobility by optimizing handover parameters locally.

3.2.2 Handover Procedures and Issues

Quality measurements of a radio link

Considering the nominated UEs located at distance d_i (in meters) from cell i , the received power for such a distance can be modeled as

$$m_i = p_{\text{tx}} d_i^{-\alpha} \nu G_0^{-1},$$

where p_{tx} is the transmission power of a cell, G_0 is a reference value that accounts for a fixed propagation loss, α is the attenuation exponent characterizing the level of attenuation of a specific propagation environment, and ν is the fading factor.

When a UE moves, it eventually or periodically sends signal measurements such as reference signal received power (RSRP) to the serving cell. The received measurement data are used for evaluating the quality of the wireless connection. The signal-to-interference-plus-noise ratio (SINR) is calculated based on the measurements. SINR is used by UEs to detect an RLF. The RLFs happen if the SINR remains below a predefined threshold, Q_{out} ,

for a certain period of time. Thus, the cell coverage is determined based on Q_{out} . The SINR, γ , is defined as

$$\gamma = \frac{m_i}{\sum_{j \neq i} m_j + N_0},$$

where m_i and m_j are linear values for the RSRP of the serving cell and other cells, respectively, as measured by the UE, and N_0 is thermal noise [24,37,38].

Handover procedure and control parameters

In the 3rd Generation Partnership Project (3GPP) standards, the handover procedure starts based on measurement reports from the UE to a serving cell for a handover decision. UEs periodically measure RSRPs of all discovered cells, and eventually or periodically send measurement reports to the serving cells when certain conditions hold. There are six intra-frequency event measurements, designated A1 to A6, in the 3GPP standards [17]. Each event measurement is used for a specific application of a SON, like mobility management and neighbor discovery [1,56,66].

For a handover algorithm, the A3 event is chosen because it is based on the better relative signal quality between two cells [1,24]. The A3 event is triggered when the RSRP of a neighboring cell becomes better than that of the serving cell, based on a certain offset. The moment for UEs to report measurements is right after the TTT timer has expired. The condition for triggering the event on the decibel (dB) scale is

$$M_n + O_{cn} + O_{fn} > M_p + O_{cp} + O_{fp} + O_{ff} + H_{ys}, \quad (3.1)$$

where M_p and M_n are the measured RSRP in dB of the serving cell and the neighboring cell, respectively. H_{ys} is the hysteresis parameter to prevent oscillation of Condition (3.1) due to fading, and O_{ff} is A3Offset for this event.

Offsets O_{cn} and O_{cp} are CIOs for the neighboring cell and the serving cell, re-

spectively. While the Hys and A3Offset affect handovers to all neighbors, the CIOs assign different handover offsets to each neighbor. In this chapter, we set the CIOs' Ofp and Ofn to zero, since we consider intra-frequency handover. We can rewrite Condition (3.1) as

$$m_j + \Delta_{ij} > m_i, \quad (3.2)$$

where $\Delta_{ij} = o_{cn} - o_{cp} - o_{ff} - H_{ys}$ and is the equivalent CIO of serving cell i for neighboring cell j , while $m_i = 10 \log m_i$ and $m_j = 10 \log m_j$. In practice, cell i has many neighboring cells; hence, we denote the set of neighboring cells of cell i as H_i .

When Condition (3.2) is satisfied within the TTT period at time t_3 and location x_3 , denoted as ordered pair (t_3, x_3) , the UE sends a measurement report to cell i triggering a handover from cell i to cell j , as shown in Fig. 3.3(a). The measurement report carries RSRPs of all discovered cells where the RSRP satisfies Condition (3.2) with regard to cell i . Then, cell i chooses the best neighboring cell based on the reported RSRP for a handover. However, if the SINR is lower than threshold Q_{out} for time duration τ_R , which is the timing length for RLF detection [24] that starts at (t_2, x_2) , the wireless links are corrupted leading to an RLF at time (t_4, x_4) , and the handover fails. There are two types of undesirable handover: RLFs and ping-pongs. While ping-pongs are similar to repeated successful handovers among cells, an RLF causes failed handover attempts based on one of three reasons: a too-late, a too-early, or a wrong-cell handover [18].

Undesirable handover conditions

To study conditions for undesirable handovers, we model a handover with geometry elements: a straight line for the user trajectory, and Apollonian circles for A3 event coverage and transmission ranges, as shown in Fig. 3.3(b) [24]. The trajectory is characterized by slope angle θ and x -intercept b . This trajectory meets the A3 event coverage of cell 1, the transmission range of cell 1, the A3 event coverage of cell 2, and the transmission range

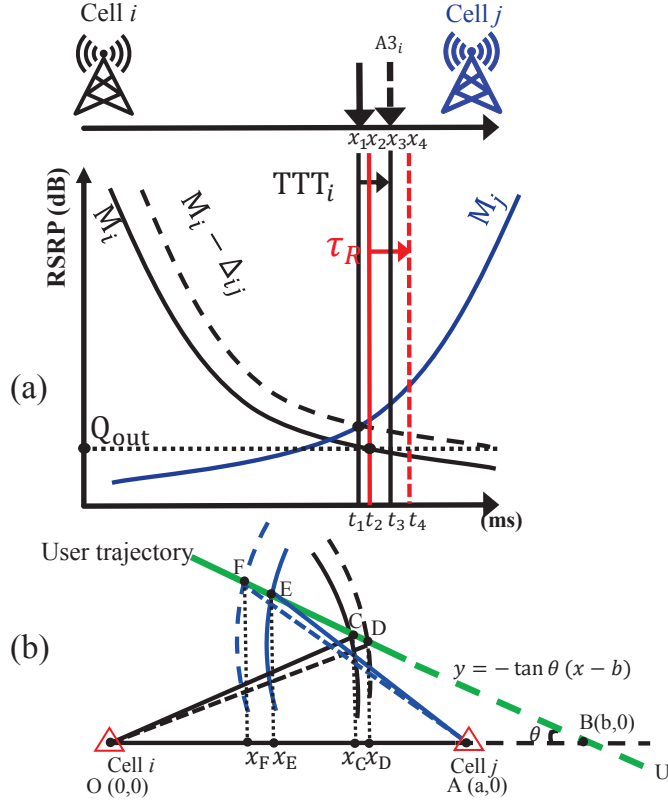


Figure 3.3: A successful handover.

of cell 2 at C, D, E, and F, respectively. The A3 event coverage for the handover of cell i to cell j is adjusted by modifying Δ_{ij} , while the transmission range is determined by the minimum SINR.

To explain conditions for undesirable handovers, we project C, D, E, and F onto the line segment between cell i and cell j , which are denoted x_C , x_D , x_E , and x_F , respectively. A too-late handover occurs when the moving depth of the UE during TTT_i , which is $x_C + vTTT_i \cos \theta$, is longer than the coverage that cell i allows for τ_R , which is $x_D + v\tau_R \cos \theta$. After an RLF, the UE tries to reconnect to cell j , which is the best neighboring cell to cell i . Cell j notifies cell i about the RLF via the Xn interface, and cell i recognizes a handover

that was too late. The condition under which a too-late handover happens is

$$x_C + v(\text{TTT}_i - \tau_R) \cos \theta > x_D \quad (\text{c1}),$$

A too-early handover happens when a UE joins target cell j too early after a successful handover, and the connection is immediately dropped because of a poor SINR ($\gamma \leq Q_{\text{out}}$). This failure occurs shortly after a successful handover from cell i to cell j , and the UE camps to cell i again because it is still within cell i coverage. Therefore, the total moving depth during TTT_j and τ_R , which is $x_C + v(\text{TTT}_j + \tau_R) \cos \theta$, is too short to enter the coverage of cell j , which is at x_F . In addition, a duration for TTT_j that is longer than τ_R causes an RLF in cell j . After the RLF, the source cell recognizes the too-early handover based on handover history. A wrong-cell handover is detected when an RLF occurs shortly after a successful handover to the target cell, and then, the UE reconnects to another cell that is neither the serving cell nor the target cell. Since a wrong-cell handover is identical to a too-early handover, except for the reconnection, we merge the wrong-cell problem into the too-early problem. The condition under which a too-early handover occurs is

$$x_C + v(\text{TTT}_j + \tau_R) \cos \theta \leq x_F \quad (\text{c2}), \text{ and } \text{TTT}_j > \tau_R,$$

where x_F is the projection of F onto the line segment between cell i and cell j , as depicted in Fig. 3.3(b).

In contrast to RLFs, a ping-pong maintains the link connection. However, it repeats the handover from cell i to cell j multiple times or even among multiple cells within a short time [66]. For example, the pattern of a ping-pong between cell i and cell j can be cell j –cell i –cell j , and so on. To cause a simple ping-pong pattern, i.e., cell j –cell i (counted as one ping-pong), the moving depth during TTT_i and TTT_j , which is $x_C + v(\text{TTT}_i + \text{TTT}_j) \cos \theta$, should not exceed the A3 event coverage of cell j , which is at x_E , to guarantee a handover back to cell j again; and TTT_j must not be greater than τ_R to avoid RLF at cell j . This

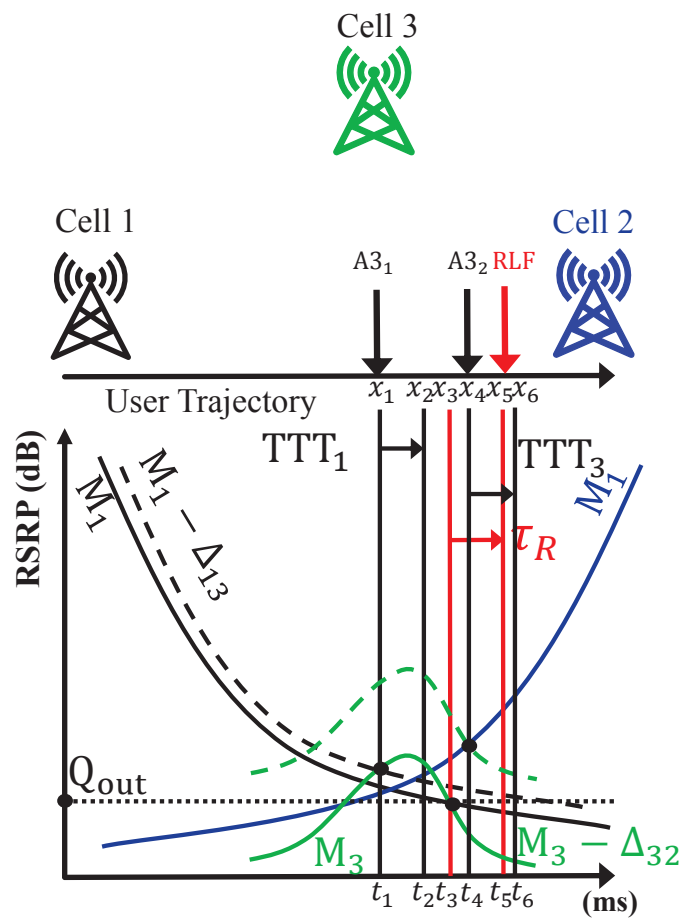


Figure 3.4: Examples of dynamic small-cell network handovers.

rapid pace of handovers wastes system resources, such as time and signaling procedures [29]. Even though ping-pongs do not cause RLFs, a regular requirement for handover optimization always includes minimizing ping-pongs as much as possible. The scenario for a ping-pong looks similar to the too-early handover problem, but ping-pongs do not cause RLFs. The condition for a ping-pong is described as

$$x_C + v(\text{TTT}_i + \text{TTT}_j) \cos \theta \leq x_E \quad (\text{C3}), \quad \text{and} \quad \text{TTT}_j \leq \tau_R,$$

where x_E is the projection of E onto the cell i -cell j line segment, as shown in Fig. 3.3(b).

3.2.3 Optimal Handover Parameters

As explained in the conditions for undesirable handovers, when handover offset Δ_{ij} increases, too-late handovers happen less often, whereas too-early handovers occur more often. When handover offset Δ_{ij} decreases, too-late handovers happen more often, whereas too-early handovers occur less often. Therefore, there is a lower bound on Δ_{ij} to prevent too-late handovers, and an upper bound on Δ_{ij} to eliminate too-early handovers.

The lower bound of Δ_{ij} , Δ_{ij}^\dagger , is obtained as follows

$$\begin{aligned} \Delta_{ij}^\dagger = \min \quad & \Delta_{ij} & (3.3) \\ \text{s.t.} \quad & x_C + v(\text{TTT}_i - \tau_R) \cos \theta \leq x_D, \\ & O_{\min} \leq \Delta_{ij} \leq O_{\max}, \end{aligned}$$

where O_{\min} and O_{\max} are the maximum and minimum values of Δ_{ij} , respectively. The

upper bound, Δ_{ij}^* , of Δ_{ij} , is obtained as follows

$$\begin{aligned} \Delta_{ij}^* = \quad & \max \quad \Delta_{ij} & (3.4) \\ \text{s.t.} \quad & x_C + v(\text{TTT}_j + \tau_R) \cos \theta > x_F, \\ & x_C + v(\text{TTT}_i + \text{TTT}_j) \cos \theta > x_E, \\ & O_{\min} \leq \Delta_{ij} \leq O_{\max}. \end{aligned}$$

The optimal range for Δ_{ij} exists when $\Delta_{ij}^\dagger \leq \Delta_{ij}^*$. The condition for the existence of the optimal range of handover parameters was explained in Lemma 1 in [24]. If the optimal range exists (i.e., $\Delta_{ij}^\dagger \leq \Delta_{ij}^*$), optimal handover parameters are obtained randomly as long as they are within the optimal range. Otherwise (i.e., $\Delta_{ij}^\dagger > \Delta_{ij}^*$), we would probably choose a value based on a predefined policy. For this chapter, we chose the average of the lower bound and the upper bound of the optimal handover offset as the approximate optimal value, which is $(\Delta_{ij}^\dagger + \Delta_{ij}^*)/2$, regardless of the existence of the optimal range.

3.2.4 Problem Formulation

The dynamic on/off switching characteristic of small cells affects handovers, since the wireless environment (e.g., cell boundaries, topology, and interference) changes accordingly. In Fig. 3.4, we show an example of a handover failure where a new cell (cell 3) is switched on for a given network with two small cells, as seen in Fig. 3.3(a). In Fig. 3.3(a), the UE should be handed over from cell 1 to cell 2 if there are only two cells. Due to the activation of cell 3, the UE is handed over to cell 3 at (t_1, x_1) . However, at (t_6, x_6) , the UE cannot be handed over successfully from cell 2 to cell 3. This failure is because handovers among cell 1, cell 2, and cell 3 are no longer optimized due to the changed topology.

We define the topology, \mathcal{T} , of a wireless network as a set of cells that are identified by cell ID as well as cell location. For instance, $\mathcal{T} = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ means that

there are 10 cells with IDs from 1 to 10 deployed in the wireless network. Also, we define a cluster of adjacent cells, \mathcal{T}_i , as a sub-topology of cells in the wireless network such that $\mathcal{T} = \cup_{i=1}^K \mathcal{T}_i$ if there are K sub-topologies. For example, $\mathcal{T}_1 = \{1, 2, 3\}$ means that sub-topology \mathcal{T}_1 contains three adjacent cells: cell 1, cell 2, and cell 3.

We formulated a handover optimization problem considering the dynamic network topology as follows:

$$\begin{aligned} \min_{\text{TTT}, \text{Ocn}, \text{A3Offset}} \quad & f(\text{TTT}, \text{Ocn}, \text{A3Offset}, \mathcal{T}) \\ \text{subject to} \quad & 0 \text{ ms} \leq \text{TTT} \leq 5120 \text{ ms}, \\ & -24 \text{ dB} \leq \text{Ocn} \leq 24 \text{ dB}, \\ & -15 \text{ dB} \leq \text{A3Offset} \leq 15 \text{ dB}, \end{aligned}$$

where f is a cost function of handover optimization, and \mathcal{T} is the network topology.

3.3 Transfer Learning-Based MRO

To minimize undesirable handovers in dynamic cellular networks, we propose a transfer learning-based MRO algorithm, as described in Fig. 3.5. The algorithm consists of two main steps. The first step, which is taken by the cSON, obtains prior knowledge of the optimal handover settings for adapting the dynamic topology, and transfers that knowledge to dSONs. The second step, taken by dSONs, fine-tunes the handover parameters to adapt to mobile environments based on the transferred knowledge.

3.3.1 Knowledge Transfer Algorithm

To deal with the dynamic topology of wireless networks, we utilize the concept of transfer learning that exploits prior knowledge in order to achieve better optimization. Transfer learning is defined as follows. Given a source domain, D^s , and the learning task

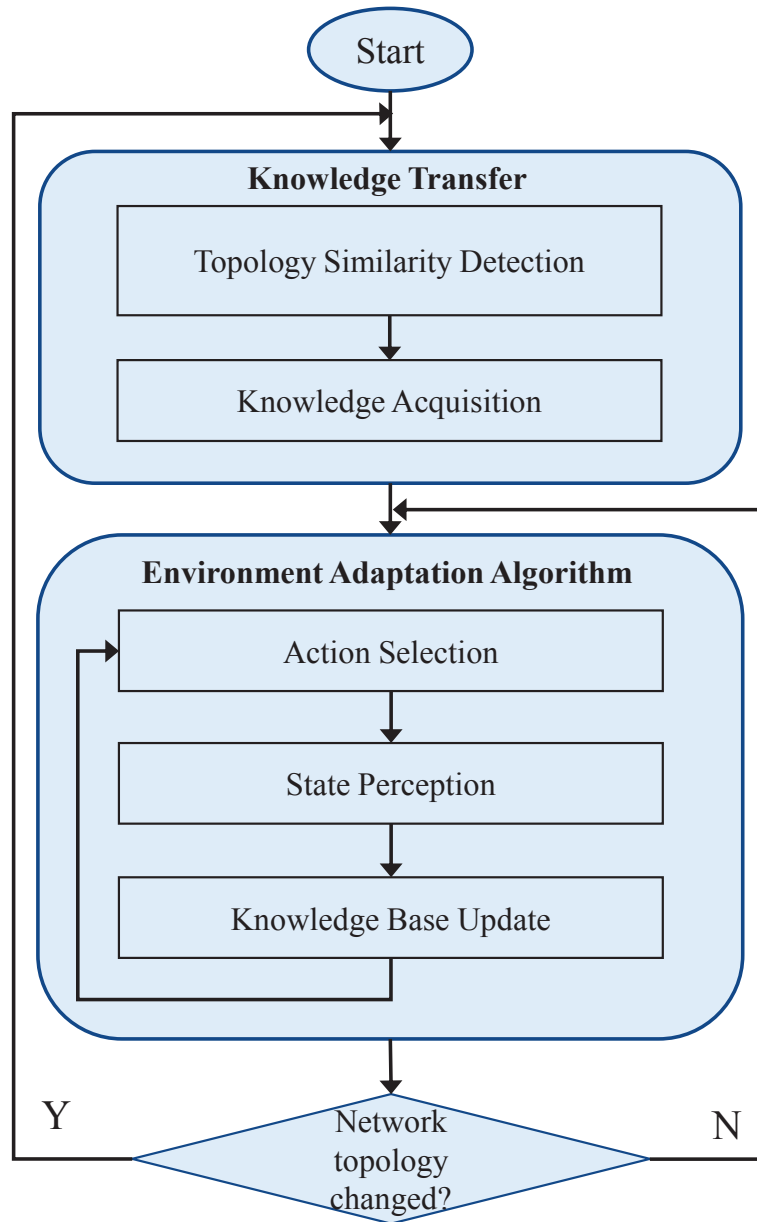


Figure 3.5: Proposed transfer learning process for dynamic wireless networks.

L^s , plus a target domain, D^t , and the learning task L^t , transfer learning aims to enhance the accuracy of the target predictive model in D^t using the knowledge in D^s and D^t in cases where $D^s \neq D^t$ and/or $L^s \neq L^t$ [67]. In our paper, the domain is network topology \mathcal{T} that changes over time, and the learning task is identical for all network topologies, which is optimization of handover parameters. A transfer learning algorithm has to address three questions: *when*, *what*, and *how* to transfer knowledge.

Topology similarity detection

Prior knowledge is transferred *when* a network topology changes due to cells switching on and off. When the network topology changes, the handover situation becomes different. While previous algorithms focused on parameter adjustments, regardless of prior knowledge, the proposed transfer-learning–based algorithm utilizes prior knowledge to improve the optimization process. At the cSON, the transfer learning algorithm extracts *features* of the old topology, \mathcal{T}^{old} , that are similar to the new topology, \mathcal{T}^{new} . Extracting the topology similarity is based on the key idea that a cell in one topology is similar to a cell in another topology if their neighboring nodes are similar.

To determine *how* similar topology \mathcal{T}^{new} is to topology \mathcal{T}^{old} , the proposed algorithm uses the information about the recent on-and-off cells that is stored in set $\mathcal{U}_{\text{ON/OFF}}$. For each cell i in $\mathcal{U}_{\text{ON/OFF}}$, the algorithm groups cell i with its neighboring cells in topology \mathcal{T}^{new} to form a sub-topology (\mathcal{T}_i). The dissimilar parts between \mathcal{T}^{old} and \mathcal{T}^{new} are sub-topologies that contain the recent on/off cells, whereas the similar parts are sub-topologies that do not have neighboring relations to the recent on/off cells.

Once the similarities and dissimilarities between \mathcal{T}^{old} and \mathcal{T}^{new} are obtained, to determine *what* knowledge should be transferred to cells in \mathcal{T}^{new} , the algorithm searches sub-topologies \mathcal{T}_i s that contain the recent on/off cells throughout the database of the topol-

Algorithm 3 Knowledge Transfer Algorithm

- 1: Select cells that were recently on and off in \mathcal{T}^{old} to make set $\mathcal{U}_{\text{ON/OFF}}$.
 - 2: **for** cell $i \in \mathcal{U}_{\text{ON/OFF}}$ **do**
 - 3: Group cell i with its neighboring cells to obtain sub-topology \mathcal{T}_i
 - 4: **if** $\mathcal{T}_i \in \mathcal{T}_{\text{DB}}$ **then**
 - 5: Transfer knowledge bases to cells in \mathcal{T}_i
 - 6: **else**
 - 7: Transfer the estimated optimal handover parameters to cells in \mathcal{T}_i (Algorithm 5)
 - 8: Store \mathcal{T}_i in \mathcal{T}_{DB}
 - 9: **end if**
 - 10: **end for**
 - 11: Keep the current handover parameters and knowledge bases for the remaining cells in \mathcal{T}^{new}
 - 12: Go to Algorithm 6
-

ogy, \mathcal{T}_{DB} . Database \mathcal{T}_{DB} contains the sub-topologies that appeared in the history and the knowledge bases of the cells in each sub-topology. If \mathcal{T}_i is in \mathcal{T}_{DB} , the algorithm transfers the knowledge bases to each cell in \mathcal{T}_i . Otherwise, the algorithm transfers the estimated optimal handover parameters to each cell in \mathcal{T}_i and stores \mathcal{T}_i into \mathcal{T}_{DB} for future queries. For similar parts between \mathcal{T}^{old} and \mathcal{T}^{new} , current parameters and knowledge bases are kept unchanged. We summarize the knowledge transfer in Algorithm 3.

Knowledge acquisition

The cSON transfers knowledge to dSONs, including knowledge bases and the estimated optimal handover parameters based on the sub-topology. The knowledge bases provide information about the optimal handover parameters that are expected to achieve the best handover performance. However, if knowledge bases are not available for a sub-topology (usually at the wireless network's deployment), the optimal handover parameters for cells in the sub-topology must be estimated. Considering two adjacent cells, cell i and cell j ($i, j \in \mathcal{T}_i$), to estimate optimal parameters for handovers by cell i to cell j , we solve optimization problems (3.3) and (3.4) to get the lower bound (Δ_{ij}^\dagger) and the upper bound (Δ_{ij}^*), respectively, of Δ_{ij} given the geometric information and cell settings.

First, positions D, E, and F are computed for a given user trajectory, (θ, b) . Then, we can find Δ_{ij}^\dagger by a trying each offset value until (C1) is satisfied. Similarly, Δ_{ij}^* is found by trying each offset value until (C2) and (C3) are met. The process is summarized in Algorithm 4.

Bound values Δ_{ij}^\dagger and Δ_{ij}^* depends on various parameters, such as $\theta, b, \Delta_{ji}, \text{TTT}_i$, and TTT_j . Hence, we express Δ_{ij}^\dagger and Δ_{ij}^* as $f_{ij}^\dagger(\theta, b, \text{TTT}_i)$ and $f_{ij}^*(\theta, b, \Delta_{ji}, \text{TTT}_i, \text{TTT}_j)$, respectively, to represent such dependencies. Since a user trajectory varies according to (θ, b) , we estimate the lower bound and the upper bound for handover offset of cell i for cell j as

follows

$$\hat{\Delta}_{ij}^\dagger = E_{\theta,b}(f_{ij}^\dagger(\theta, b, \text{TTT}_i)), \quad (3.5)$$

$$\hat{\Delta}_{ij}^* = E_{\theta,b}(f_{ij}^*(\theta, b, \text{TTT}_i, \Delta_{ji, \text{TTT}_j})), \quad (3.6)$$

where $E_{\theta,b}(X)$ is the expectation of X over two variables, θ and b , which we assume have uniform distributions [24,32]. The estimated optimal value of Δ_{ij} , $\hat{\Delta}_{ij}^{\text{opt}}$, is the average value of $\hat{\Delta}_{ij}^\dagger$ and $\hat{\Delta}_{ij}^*$. That is $\hat{\Delta}_{ij}^{\text{opt}} = (\hat{\Delta}_{ij}^* + \hat{\Delta}_{ij}^\dagger)/2$.

After obtaining the estimated optimal CIO, $\hat{\Delta}_{ij}^{\text{opt}}$, the algorithm finds $\hat{\Delta}_{ji}^{\text{opt}}$ for cell j in a process similar to the one given for $\hat{\Delta}_{ij}^{\text{opt}}$ in order to optimize handovers from cell j to cell i . The process for computing $\hat{\Delta}_{ij}^{\text{opt}}$ and $\hat{\Delta}_{ji}^{\text{opt}}$ are repeated until they converge. The results are rounded to the nearest values that follow the 3GPP standard [17]. We summarize the algorithm for estimating the optimal handover parameters for two adjacent cells in Algorithm 5. When the estimation is completed, the algorithm selects another cell pair in \mathcal{T}_i until completing all the cell pairs. Note that we change the origin of the coordinate plane to the location of the considered cell (e.g., cell i) and a neighboring cell (e.g., cell j) to estimate the optimal parameters for handovers between them.

In the next section, we provide an environment adaptation algorithm to cope with mobility and wireless environments given the transferred knowledge.

3.3.2 Environment Adaptation Algorithm

The wireless environment also varies due to UE mobility, and therefore, we propose a reinforcement learning-based algorithm for dSONs to use to adapt to the target environment given the transferred knowledge from the cSON. Our algorithm optimizes handover parameters, such as TTT and CIO, by taking into account too-late handovers, too-early handovers, and ping-pongs, to meet the required performance. Next, we are going to explain

Algorithm 4 Geometry-based Computation for the Upper Bound and Lower Bound of the CIO

- 1: Input: τ_{TTi} , Δ_{ji} , τ_{TTj} , b , θ , and locations of cells
 - 2: Output: Δ_{ij}^\dagger and Δ_{ij}^*
 - 3: Compute positions D and F with γ_{\min} , and E with Δ_{ji}
 - 4: **for** each Δ_{ij} value in a learning range **do**
 - 5: Compute position C with Δ_{ij}
 - 6: **if** (C1) is satisfied **then**
 - 7: $\Delta_{ij}^\dagger \leftarrow \Delta_{ij}$ and break
 - 8: **end if**
 - 9: **end for**
 - 10: **for** each Δ_{ij} value in a learning range **do**
 - 11: Compute position C with Δ_{ij}
 - 12: **if** (C2) & (C3) are satisfied **then**
 - 13: $\Delta_{ij}^* \leftarrow \Delta_{ij}$ and break
 - 14: **end if**
 - 15: **end for**
-

Algorithm 5 Optimal Handover Parameter Estimation

- 1: Input: τ_{TTi} , Δ_{ji} , τ_{TTj} , a range of θ , a range of b , and locations of cells
 - 2: Output: $\hat{\Delta}_{ij}^{\text{opt}}$ and $\hat{\Delta}_{ji}^{\text{opt}}$
 - 3: $\hat{\Delta}_{ji}^{\text{opt}} \leftarrow \Delta_{ji}$
 - 4: **for** each value of θ and b in the specified ranges **do**
 - 5: Compute Δ_{ij}^* and Δ_{ij}^\dagger for cell i given $\hat{\Delta}_{ji}^{\text{opt}}$ via Algorithm 4
 - 6: **end for**
 - 7: Compute $\hat{\Delta}_{ij}^\dagger$ and $\hat{\Delta}_{ij}^*$ via (3.5) and (3.6)
 - 8: $\hat{\Delta}_{ij}^{\text{opt}} \leftarrow (\hat{\Delta}_{ij}^* + \hat{\Delta}_{ij}^\dagger)/2$
 - 9: **for** each value of θ and b in the specified ranges **do**
 - 10: Compute Δ_{ji}^* and Δ_{ji}^\dagger for cell j given $\hat{\Delta}_{ij}^{\text{opt}}$ via Algorithm 4
 - 11: **end for**
 - 12: Compute $\hat{\Delta}_{ji}^\dagger$ and $\hat{\Delta}_{ji}^*$ via (3.5) and (3.6)
 - 13: $\hat{\Delta}_{ji}^{\text{opt}} \leftarrow (\hat{\Delta}_{ji}^* + \hat{\Delta}_{ji}^\dagger)/2$
 - 14: Return to Line 4 until $\hat{\Delta}_{ij}^{\text{opt}}$ and $\hat{\Delta}_{ji}^{\text{opt}}$ converged, then round them to the nearest value standardized by 3GPP.
-

the adaptation algorithm generally.

At the dSON, the environment adaptation algorithm initializes the handover parameters based on the transferred knowledge from the cSON. Then, the algorithm computes the optimization costs after the timer for handover statistics has expired. The cost of too-late handovers from cell i to cell j , \mathcal{R}_{Lij} , is given as

$$\mathcal{R}_{Lij} = \frac{N_{TLij}}{N_{totalij}}, \quad (3.7)$$

where N_{TLij} and $N_{totalij}$ are the number of too-late handovers and the total handovers from cell i to cell j , respectively. The cost of too-early handovers from cell i to cell j is defined as

$$\mathcal{R}_{Eij} = \frac{N_{PPij} + N_{TEij}}{N_{totalij}} \quad (3.8)$$

where N_{TEij} and N_{PPij} are the numbers of too-early handovers and ping-pongs from cell i to cell j .

Based on the costs, the algorithm perceives the system states following Algorithm 8, which determines directions for parameter adjustments. Then, with the computed costs and the system state, the algorithm updates knowledge bases for selecting optimal handover parameters later. The algorithm will compare the RLF rate with the target performance Th_{RR} . The RLF rate for handovers from cell i to its neighboring cells, RR_i , is given as

$$RR_i = \sum_{j \in H_i} \frac{N_{TLij} + N_{TEij}}{N_{totali}},$$

where N_{totali} is the total number of handovers from cell i . If RR_i exceeds Th_{RR} , the algorithm selects optimal actions to adjust the handover parameters of cell i in order to adapt to dynamic mobility, which follows Algorithm 7. After adjusting handover parameters, the adaptation algorithm resets the timer and flushes all the counters for handover statistics. If the network topology changes, the current optimization process switches to the knowledge transfer algorithm (Algorithm 3). We summarize the environment adaptation algorithm in Algorithm 6, and will describe it in detail after the algorithm elements definitions.

Algorithm 6 Environment Adaptation

- 1: **if** knowledge bases are available **then**
 - 2: Initialize handover parameters according to the transferred knowledge bases from cSON by (3.10)
 - 3: **else**
 - 4: Initialize handover parameters according to the transferred parameters
 - 5: **end if**
 - 6: Receive data $N_{L_{ij}}$, $N_{E_{ij}}$, and $N_{PP_{ij}}, \forall j \in H_i$
 - 7: **if** timer for receiving handover statistics has expired **then**
 - 8: Compute $\mathcal{R}_{L_{ij}}$, $\mathcal{R}_{E_{ij}}$, and \mathcal{R}_{ij} by (3.7), (3.8), and (3.9), respectively
 - 9: Measure system state $s_{ij}, \forall j \in H_i$ (Algorithm 8)
 - 10: Update knowledge bases by (3.12)
 - 11: **if** $RR_i > Th_{RR}$ **then**
 - 12: Select action $a_{ij}, \forall j \in H_i$ (Algorithm 7)
 - 13: **end if**
 - 14: Reset the timer and flush all counters of cell i , then return to Line 6
 - 15: **else**
 - 16: **if** network topology changes **then**
 - 17: Go to the knowledge transfer algorithm (Algorithm 1)
 - 18: **else**
 - 19: Return to Line 6
 - 20: **end if**
 - 21: **end if**
-

Elements of the algorithm

To adapt dynamic mobility from cell i to cell j , the environment adaptation algorithm includes four elements: cost function \mathcal{R}_{ij} ; system state $s_{ij} \in \mathcal{S}$, which is a finite set of system states; action $a_{ij} \in \mathcal{A}(s_{ij})$, which is a finite set of actions that depend on the current state, s_{ij} ; and a knowledge base $Q(a_{ij})$. In general, the algorithm chooses the optimal action in set $\mathcal{A}(s_{ij})$ for state s_{ij} to minimize the handover cost, which is analogous to reinforcement learning. System states, actions, and the cost function are defined as follows.

System state $s_{ij} \in \mathcal{S}$ is sensed through a perception stage based on the handover situation, because the objective of the algorithm is to minimize undesirable handovers. The set \mathcal{S} consists of four states ($s_{\text{cio}+}$, $s_{\text{cio}-}$, $s_{\text{ttt}+}$, and $s_{\text{ttt}-}$) to increase the CIO, decrease the CIO, increase TTT, and decrease TTT, respectively, from their current values.

Action $a_{ij} \in \mathcal{A}(s_{ij})$ adjusts the values of TTT and CIO to minimize undesirable handovers according to current state s_{ij} . Therefore, set $\mathcal{A}(s_{ij})$ varies based on state s_{ij} . An example of action values is shown in Fig. 3.6. We consider the learning range for TTT from 0 ms to 480 ms, because higher values lead to too-late handovers [1,24]. Similarly, the learning range for the CIO is between -4 dB to 4 dB.

Cost \mathcal{R}_{ij} is the total cost for a handover from cell i to cell j taking into account handover performance, including RLFs and ping-pongs, defined as

$$\mathcal{R}_{ij} = \mathcal{R}_{Lij} + \mathcal{R}_{Eij}. \quad (3.9)$$

Cost \mathcal{R}_{ij} is used to update the knowledge bases of the dSON at cell i .

The knowledge base is stored in the dSON of a cell for its neighboring cells. The knowledge base of cell i for cell j is a Q-value database, $Q(a_{ij})$, that maps a parameter value to a real value [60]. For future utilization of prior knowledge, the dSON dispatches the knowledge bases to the cSON after they are updated by the dSON. An example of a

Q-value database is shown in Fig. 3.6. There is one knowledge base for TTT and at least one knowledge base for CIO because more than one neighboring cell of a cell exists. In the next sections, we explain how the adaptation algorithm selects optimal actions, determines system states, and updates knowledge bases.

Action selection

Algorithm 7 Action Selection

- 1: Input: System state $s_{ij}, \forall j \in H_i$
 - 2: Output: Optimal action a_{ij}^*
 - 3: **if** $s_{ij} == s_{ttt+}$ **then**
 - 4: Increase TTT from the current TTT action based on softmax strategy (12)
 - 5: **else if** $s_{ij} == s_{ttt-}$ **then**
 - 6: Decrease TTT from the current TTT action based on softmax strategy (12)
 - 7: **else if** $s_{ij} == s_{cio+}$ **then**
 - 8: Increase Δ_{ij} from the current Δ_{ij} action based on softmax strategy (12)
 - 9: **else**
 - 10: Decrease Δ_{ij} from the current Δ_{ij} action based on softmax strategy (12)
 - 11: **end if**
-

At the beginning of the environment adaptation algorithm, if knowledge bases are transferred by the cSON, the optimal action a_{ij}^* of cell i for neighboring cell j is selected, based on a greedy strategy, which is

$$a_{ij}^* = \arg \min_{a_{ij}} Q(a_{ij}) \quad (3.10)$$

Otherwise, the dSON applies the estimated optimal handover parameters that were transferred by the cSON.

For later optimization, an optimal action is chosen based on a softmax strategy, given the perceived system state. A softmax strategy chooses an action based on probability distribution. An action with a higher probability is more likely to be selected because it is expected to bring better handover performance in the future. The probability of taking an action is calculated from a Gibbs (or Boltzmann) distribution, which is the popular softmax method used in reinforcement learning [60]. For neighboring cell j of cell i , each action, $a_{ij} \in \mathcal{A}(s_{ij})$, in state s_{ij} is assigned a probability of selection, $p(s_{ij}, a_{ij})$, as follows

$$p(s_{ij}, a_{ij}) = \frac{\exp\left(\frac{-Q(a_{ij})}{\tau}\right)}{\sum_{b_{ij} \in \mathcal{A}(s_{ij})} \exp\left(\frac{-Q(b_{ij})}{\tau}\right)}, \quad (3.11)$$

where τ is a positive parameter called the temperature. We apply negative Q values for the softmax method because it is preferred to a parameter with a low Q-value rather than parameters with a higher Q-value. The lower the Q-value for a parameter, the better the handover performance the parameter is expected to achieve. A higher temperature causes the actions to have a more equal probability, which encourages state space exploration. A lower temperature leads to a greater difference in the selection probability for actions, which encourages utilizing prior knowledge.

After choosing optimal actions, the algorithm waits to obtain handover optimization costs and system states from the wireless environment. Then, with the costs and states, the algorithm updates the knowledge base in (3.12), which will be explained later.

State perception

With the optimization costs, the algorithm translates them into system states. When a too-late problem is dominant for handovers to neighboring cell j (i.e., $\mathcal{R}_{L_{ij}} \geq \mathcal{R}_{E_{ij}}$), the system state for handovers to cell j is $s_{c_{io}+}$, which demands a CIO increment. When a too-early problem is dominant for handovers to cell j ($\mathcal{R}_{L_{ij}} < \mathcal{R}_{E_{ij}}$), the system state

Algorithm 8 State Perception

```

1: Input:  $\mathcal{R}_{L_{ij}}$  and  $\mathcal{R}_{E_{ij}}, \forall j \in H_i$ 
2: Output: System state  $s_{ij}, \forall j \in H_i$ 
3: for each neighboring cell  $j \in H_i$  do
4:   if  $\mathcal{R}_{L_{ij}} \geq \mathcal{R}_{E_{ij}}$  then
5:      $s_{ij} \leftarrow s_{\text{cio}+}$ 
6:   else
7:      $s_{ij} \leftarrow s_{\text{cio}-}$ 
8:   end if
9: end for
10: if  $s_{ij} == s_{\text{cio}+}, \forall j \in H_i$  then
11:    $s_{ij} \leftarrow s_{\text{ttt}-}, \forall j \in H_i$ 
12: else if  $s_{ij} == s_{\text{cio}-}, \forall j \in H_i$  then
13:    $s_{ij} \leftarrow s_{\text{ttt}+}, \forall j \in H_i$ 
14: end if

```

TTT Values (ms)	0	40	64	...	480
Q-value	0.52	0.31	0.22	...	0.05

(a)

CIO Values (dB)	−4	...	4
Q-value	0.25	...	0.41

(b)

Figure 3.6: A knowledge base is a database for the algorithm at a cell, including (a) TTT knowledge and (b) CIO knowledge for a neighboring cell.

for handovers to cell j becomes $s_{\text{cio}-}$ to decrease the CIO. If the system state for all the neighboring cells of cell i is $s_{\text{cio}+}$, the system state for each neighboring cell is $s_{\text{ttt}-}$ for TTT decrement. This is because the TTT adjustment affects handovers to all the neighboring cells of a cell. If the system state for all the neighboring cells is $s_{\text{cio}-}$, the system state for each neighboring cell is $s_{\text{ttt}+}$, which means increase TTT. The algorithm for system state perception is in Algorithm 8.

Knowledge base update

To update the knowledge base of cell i for neighboring cell j , $Q(a_{ij})$, we apply a temporal difference method that considers observed system state s_{ij} , current action a_{ij} , and optimization cost \mathcal{R}_{ij} [60]. The knowledge update is as follows

$$Q(a_{ij}) = (1 - \beta)Q(a_{ij}) + \beta(\mathcal{R}_{ij} + \lambda \max_{a'_{ij} \in A(s'_{ij})} Q(a'_{ij})), \quad (3.12)$$

where $\beta \in (0, 1]$ is the learning factor and $\lambda \in [0, 1]$ is the discount rate. A β learning factor of 1 means that the latest knowledge is considered, while the prior knowledge is ignored. A β of zero means there is no learning at all. Discount rate λ that is close to 1 increases the importance of the prior knowledge in the received cost. We show the structure of a knowledge base in Fig. 3.6, where each parameter value has its own Q-value.

3.4 Numerical Results and Discussion

3.4.1 Simulation Environment

To evaluate the potential of our algorithm under dynamic networks, we consider a network of 12 small cells, where the topology periodically changes, as shown in Fig. 3.7. In practice, small cells can be turned on and off depending on specific strategies, such as energy efficiency and load balancing [68–70]. We show the cell numbers in Fig. 3.7d for readers to easily compare different network topologies in different periods, where each period is 20 minutes long. For user mobility, we applied a Manhattan grid mobility model with different user speeds: 5 km/h and 30 km/h. The number of UEs in the simulation was 200. For the default settings, TTT, hysteresis, A3Offset, and Ocn were set to 256 ms, 3 dB, 0 dB, and 0 dB, respectively [32].

For a propagation model, signal decaying factor α and G_0 were selected at 4.33 and $10 \log 14.74$ [20]. To model the fading effect, a log-normal random variable together with a Rayleigh factor was integrated into the propagation model. The fading variable had a zero mean, and standard deviation σ was 6 dB, as in a typical environment [51]. An omni-directional antenna was used for all small cells, and transmit power p_{tx} was 23 dBm. For RLF detection, out-of-sync threshold Q_{out} was set at -4 dB, and RLF detection time τ_R was 500 ms [24]. For estimating the optimal handover parameters, we derived ranges for θ and b by uniformly sampling distributions $(0, \pi/3)$ and $(0, a)$, respectively, where a is the inter-site distance between two adjacent cells [24]. For the adaptation algorithm, β , λ , and τ were set at 0.1, 0.95, and 1, respectively [71], for utilization of the transferred knowledge.

To evaluate the proposed transfer learning-based MRO algorithm (MRO-TL), two base lines were considered: the MRO algorithm based on classification (MRO-ABC) [1],

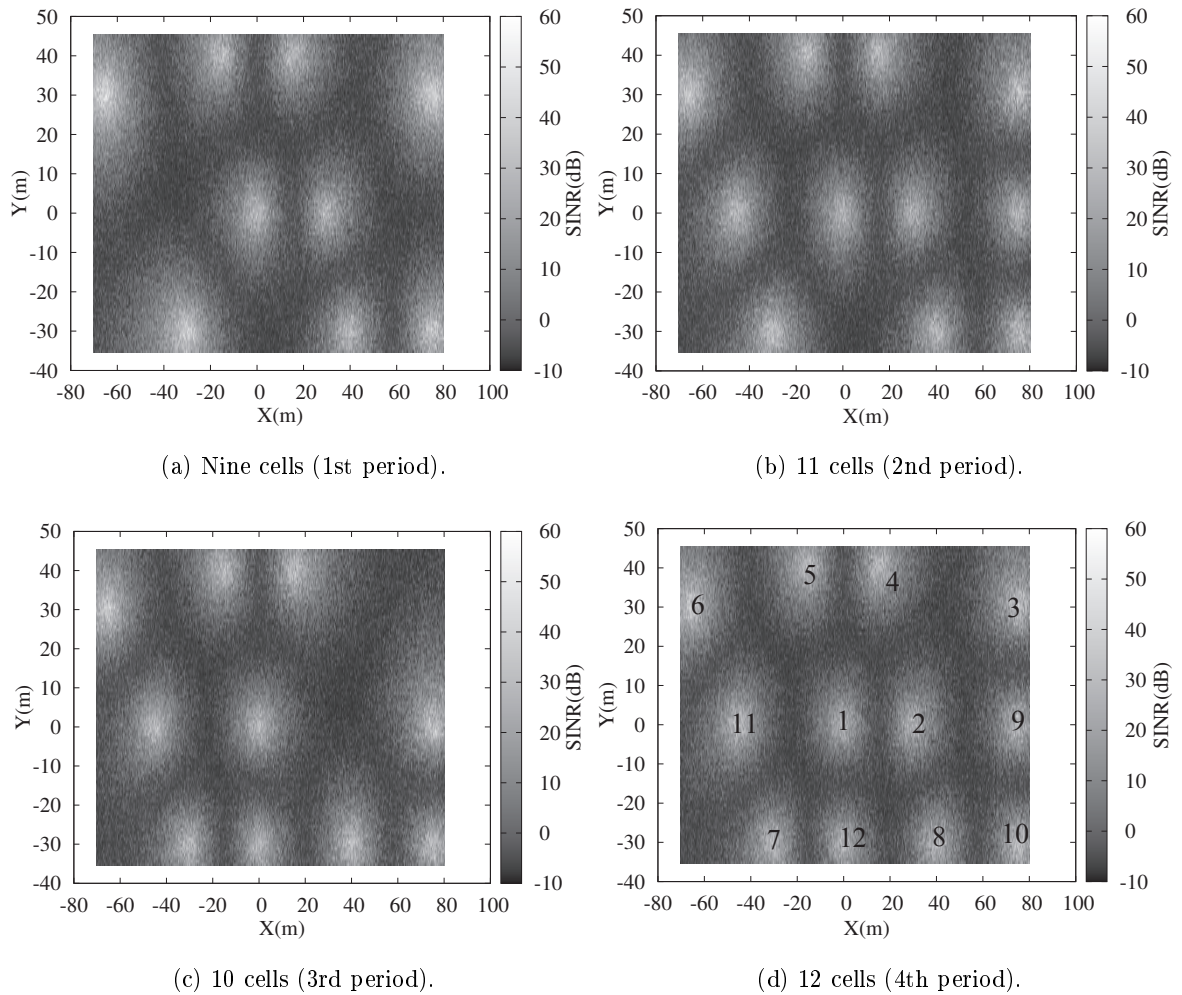
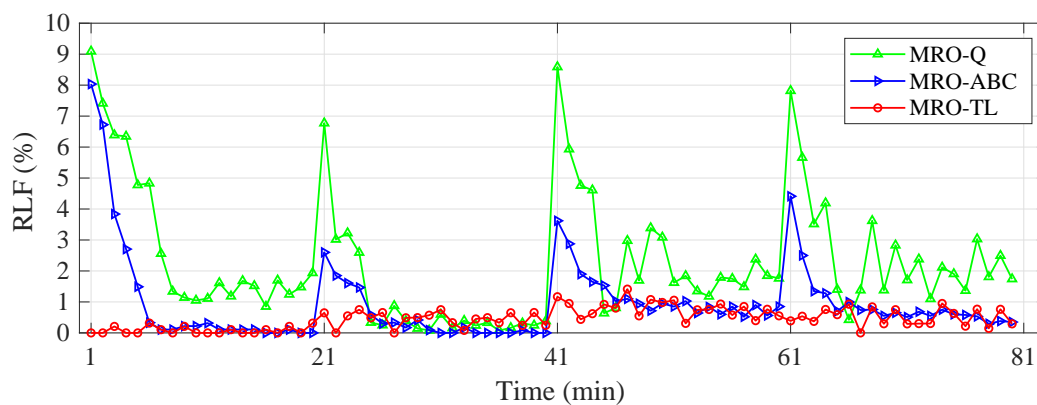
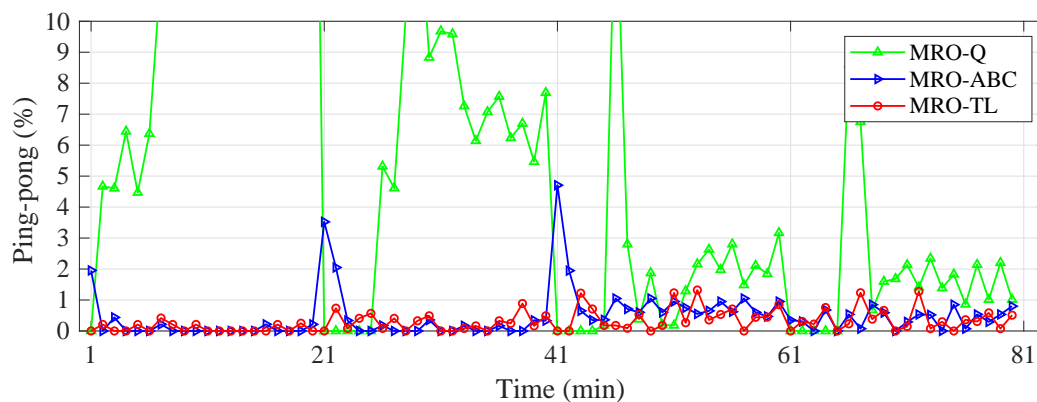


Figure 3.7: The dynamic topology of a small-cell network.



(a) RLF.



(b) Ping-pong.

Figure 3.8: At 5 km/h.

which adjusts TTT, CIO, and A3Offset to adapt mobility; and a Q-learning-based MRO (MRO-Q) [31], which is a model-free RL algorithm for adjusting TTT and A3Offset. The target performance for the considered algorithms was set at 1 % [1].

3.4.2 Handover Performance under Dynamic Network Topology

Fig. 3.8 shows the performance in terms of RLFs and ping-pongs in a 5 km/h environment. We observe that MRO-TL satisfies the target performance, improves han-

Table 3.1: Adaption times of the algorithms (in minutes) for a speed of 5 km/h

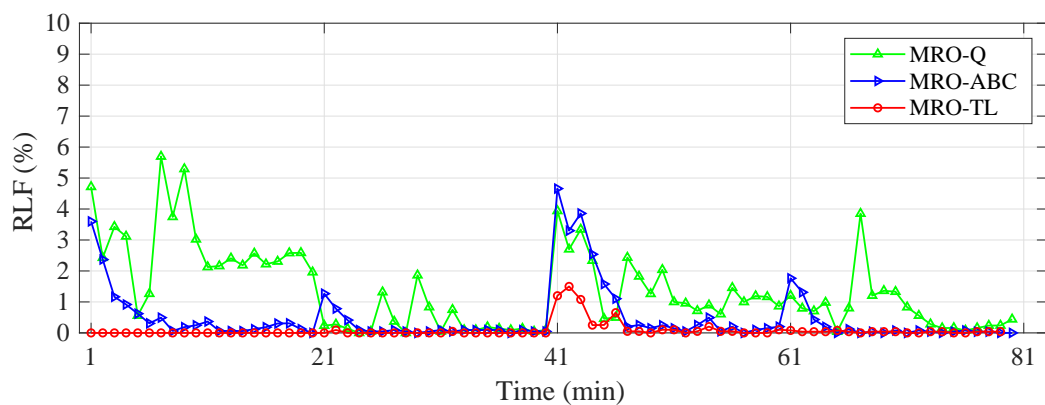
	9 cells	11 cells	10 cells	12 cells	Average	Reduction (%)
MRO-Q	10	4	5	5	6.0	100
MRO-ABC	5	4	6	4	4.75	79.17
MRO-TL	0	0	1	0	0.25	4.17

Table 3.2: Satisfaction rates of algorithms (%) for a speed of 5 km/h

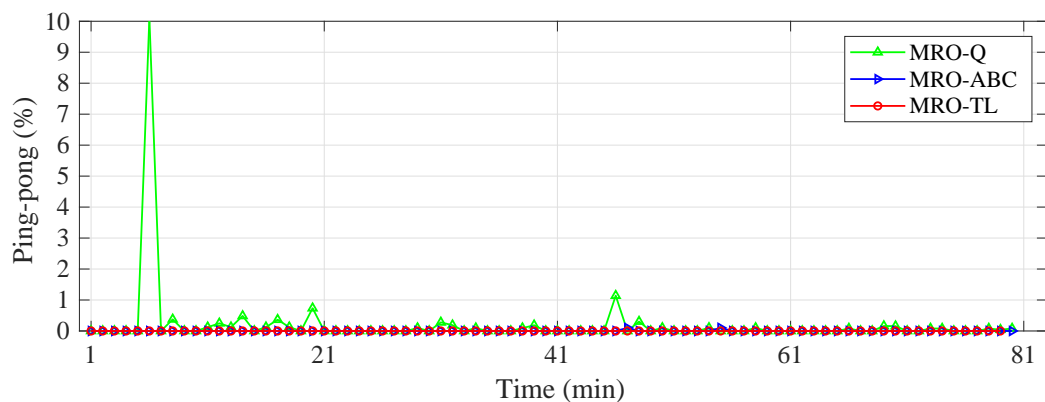
	9 cells	11 cells	10 cells	12 cells	Average	Improvement (%)
MRO-Q	0	80	5	5	22.5	100
MRO-ABC	75	80	60	75	72.5	322.2
MRO-TL	100	100	85	90	93.75	416.7

do over performance more than MRO-Q and MRO-ABC, and optimizes RLFs and ping-pongs together. In particular, MRO-TL experiences the lowest rise in RLF at the beginning of each period. This is because MRO-TL utilizes prior knowledge of the optimal handover parameters for the new topology just before the new cells are switched on and off.

To evaluate how fast the algorithms adapt to topology changes, for four topologies, we computed the average adaptation time, which is from the beginning of a topology until the first time the algorithm meets the target performance. The adaptation times of the algorithms are shown in Table 3.1. The results show that MRO-TL owns the shortest adaptation time, 0.25 minutes on average, to meet the target performance, while MRO-ABC and MRO-Q needed 4.75 minutes and six minutes on average, respectively. With respect to MRO-Q, which is based on model-free RL, MRO-TL reduced the adaptation time to 4.17% of MRO-Q, while that of MRO-ABC was 79.17% of MRO-Q. MRO-TL achieved such a significant improvement because of the optimal initialization of handover parameters



(a) RLF.



(b) Ping-pong.

Figure 3.9: At 30 km/h.

at the beginning of a new topology. Since MRO-TL achieved fast adaptation to topology changes, it is applicable to dynamic wireless networks.

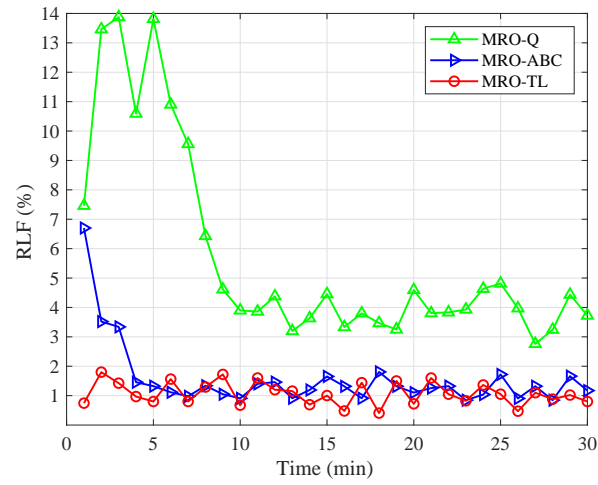
To examine how the algorithms fulfill the target handover performance, we calculated a satisfaction rate, which is the fraction of time in which handover performance is below the target performance. The results are summarized in Table 3.2. We observed that MRO-TL satisfies the handover performance requirements for all the topologies with the best satisfaction rate, 93.75% on average, while MRO-ABC and MRO-Q attained 72.5%

and 22.5%, respectively, on average. With regard to MRO-Q, MRO-TL improved the satisfaction rate by 416.7%, while MRO-ABC improved it by 322.2%. MRO-TL attained such a remarkable improvement because it fine-tunes the handover parameters around the optimal parameters based on transferred knowledge, which provides information on optimal values. Unlike MRO-TL, MRO-ABC and MRO-Q adjust the handover parameters independent of topology changes. Regarding ping-pong performance, MRO-TL and MRO-ABC were stable, while MRO-Q fluctuated. In general, the results for RLFs and ping-pongs showed that MRO-TL is a candidate for handover optimization under dynamic mobility with stable performance.

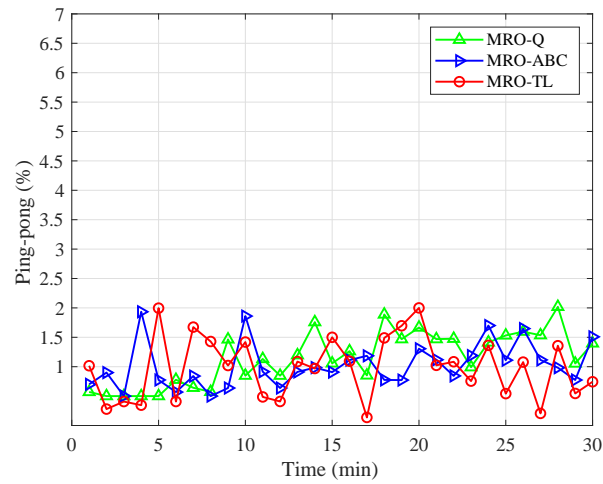
We also studied the impact of user speed to handover optimization in a dynamic topology. Fig. 3.9 visualizes the RLF and ping-pong rates of the algorithms when speed was 30 km/h. Adaptation time and satisfaction rate are presented in tables 3.3. MRO-TL owns the shortest adaptation time, 0.75 minutes on average, to meet the handover performance target, while MRO-ABC and MRO-Q spent 3.25 minutes and 2.25 minutes on average, respectively, to adapt to topology changes. For the satisfaction of UEs, MRO-TL achieved a notable satisfaction rate of 96.25%. Regarding ping-pong performance, MRO-TL was close to MRO-ABC, while MRO-Q had more fluctuations. With respect to the 5 km/h environment, the algorithms experienced fewer ping-pongs because the upper bound of the handover offset, which is to avoid too-early handovers and ping-pongs, increases when the user speed increases [24]. Overall, the results show that MRO-TL works well at a higher speed and under a dynamic network topology.

3.4.3 Impact of the User Mobility Model

To verify the impact of a random mobility model on handover performance, we chose a random waypoint (RWP) mobility model to simulate a pedestrian environment.



(a) RLF.



(b) Ping-pong.

Figure 3.10: Handover performance under a random mobility model.

Table 3.3: Adaption times (in minutes) and satisfaction rates (%) of algorithms for a speed of 30 km/h

	9 cells	11 cells	10 cells	12 cells	Average	Reduction (%)
MRO-Q	4	0	4	1	2.25	100
MRO-ABC	3	1	6	2	3	133.33
MRO-TL	0	0	3	0	0.75	33.33

	9 cells	11 cells	10 cells	12 cells	Average	Improvement (%)
MRO-Q	5	90	30	80	51.25	100
MRO-ABC	85	95	70	90	85	165.9
MRO-TL	100	100	85	100	96.25	187.8

UEs moved at a speed of 5 km/h in a scenario of 12 cells, as seen in Fig. 3.7d. Simulation time was 30 minutes long to check how the algorithms dealt with randomness in mobility.

The simulation results for RLFs and ping-pongs are depicted in Fig. 3.10. We observed that handover performance fluctuated more than in the Manhattan environment of Fig. 3.7c. This was due to the randomness of the RWP model. MRO-TL still improved handover performance more than the others, because it estimates the optimal handover parameters at the very beginning, and also adapts to the mobile environment with the estimated optimal parameters. MRO-ABC achieved the second best performance because it fine-tunes the handover parameters according to undesirable handover classifications. MRO-ABC performance was close to MRO-TL in the long run, but in the beginning, MRO-TL was better because it estimates the optimal parameters before the topology changes. MRO-TL had better performance than MRO-Q, since MRO-TL utilizes prior knowledge and fine-tunes the parameters.

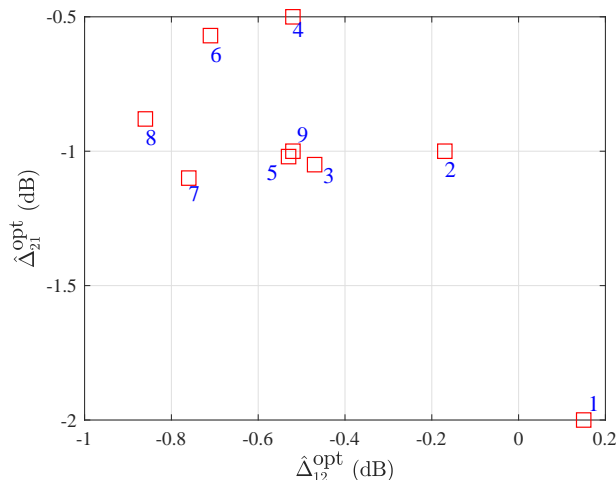
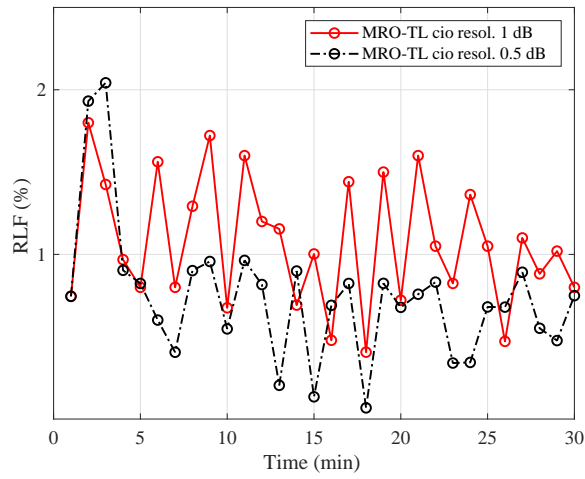


Figure 3.11: Examples of optimal handover parameter estimation. The number indicates the optimization step.

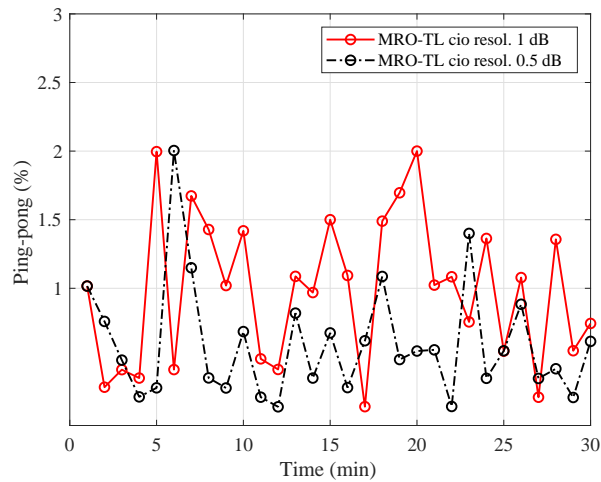
3.4.4 Impact of CIO Resolution on Handover Performance

Handover performance is affected by the resolution of CIO because the optimal handover offset is a real value instead of an integer, as standardized by the 3GPP [17]. Fig. 3.11 shows the way that MRO-TL (specifically Algorithm 5) estimates the optimal handover parameters for two adjacent cells, called cell 1 and cell 2. The inter-site distance was 30 m. T_{TT1} , Δ_{21} , and T_{TT2} were initialized at 256 ms, -2 dB, and 480 ms, respectively. Other settings, such as transmit power and antenna gain, were identical for both cells. The estimation of $\hat{\Delta}_{12}^{opt}$ and $\hat{\Delta}_{21}^{opt}$ took nine iterations to converge. The results show that $\hat{\Delta}_{12}^{opt}$ and $\hat{\Delta}_{21}^{opt}$ (in dB) were approximately 0.5 and -1 , respectively. That inspired us to evaluate handover performance for MRO-TL with a finer resolution for CIO.

To verify the impact of resolution on handover optimization, we investigated MRO-TL under the wireless scenario in Fig. 3.7d with a CIO resolution of 0.5 dB. The results are compared to those from a resolution of 1 dB, as depicted in Fig. 3.10. Fig. 3.12 shows that the resolution of 0.5 dB improved handover performance more than a resolution of 1 dB,



(a) RLF.



(b) Ping-pong.

Figure 3.12: Handover performance with different CIO resolutions.

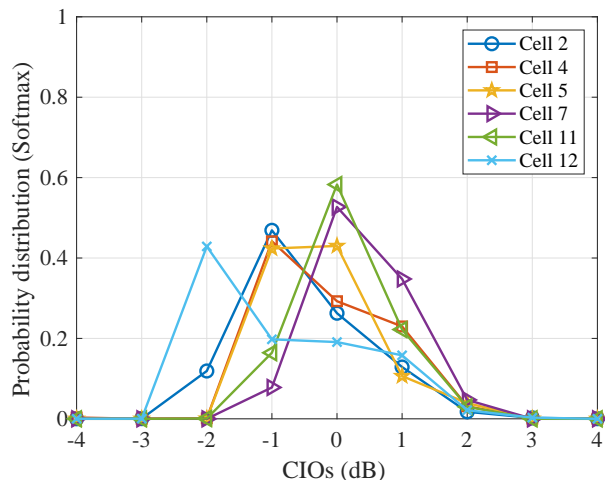


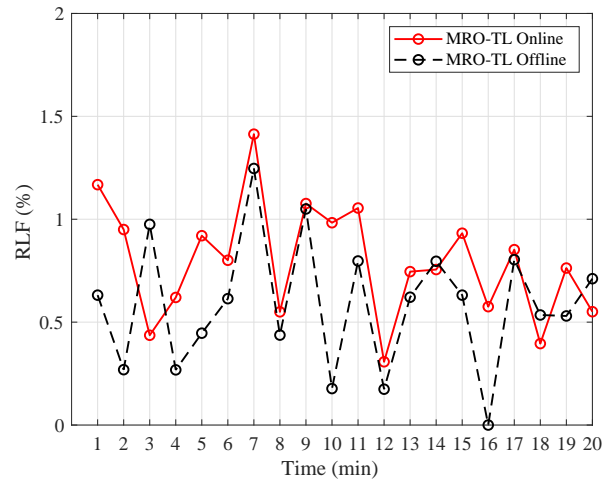
Figure 3.13: Prior knowledge for cell 1 with a softmax probability of $\tau = 1$

and the outcome of the earlier resolution fluctuated less. The reason is that the adaptation algorithm provides finer tuning of the CIO around the estimated optimal parameters, thus inducing more pleasing results.

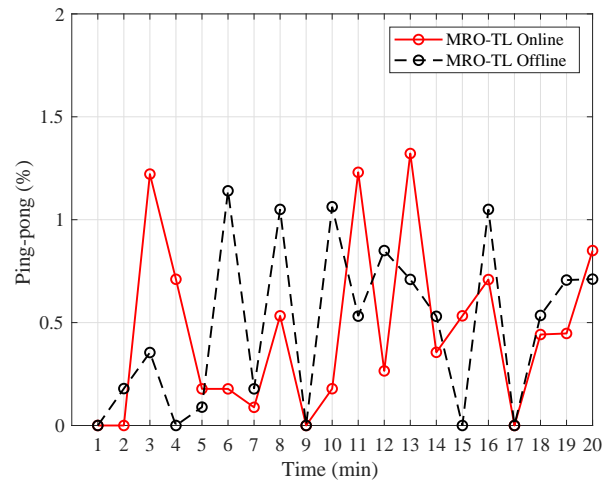
3.4.5 Offline Learning

Since MRO-TL utilizes prior knowledge to optimize handover parameters, the more prior knowledge the algorithm is provided, the better the performance it can achieve. To get more prior knowledge for MRO-TL, we kept simulating the environment of 10 cells in Fig. 3.7c for five hours and stored the knowledge bases of the dSONs in the database of the cSON. The cSON transferred the knowledge bases to the dSON at the beginning of the considered topology. We depict the knowledge bases for the optimal CIOs in the dSON of cell 1 for its neighboring cells as probability distributions in Fig. 3.13. The figure shows that the distributions of the optimal CIOs are different, cell-to-cell, which is due to the cell locations.

We compared the performance of MRO-TL in Fig. 3.7c referred to as MRO-TL



(a) RLF.



(b) Ping-pong.

Figure 3.14: Comparison between online and offline MRO-TL.

Online, and against MRO-TL with the transferred five-hour knowledge bases, called MRO-TL Offline. The results are depicted in Fig. 3.14 showing that MRO-TL Online and MRO-TL Offline performances were close. Particularly, MRO-TL Online and MRO-TL Offline achieved satisfaction rates at 85% (13/20) and 90% (18/20), respectively. Since MRO-TL estimates the optimal handover parameters in the very beginning of a specific topology, and fine-tunes the parameters to adapt to dynamic mobility with the transferred knowledge, the online algorithm is applicable to real-time handover optimization.

3.5 Closing Remarks

In this chapter, we proposed a transfer learning-based MRO algorithm, MRO-TL, to minimize undesirable handovers (handover failures and ping-pongs) under a dynamic network topology as well as dynamic mobility. To that end, MRO-TL has two steps: first, to adapt to the changing topology, MRO-TL estimates the optimal handover parameters for the cells in the new topology, and also transfers the knowledge bases to them, which is done by the cSON. Second, an adaptation algorithm utilizes the transferred knowledge and fine-tunes the handover parameters around the estimated optimal values to adapt to dynamic user mobility, which is done by the dSON.

We evaluated the proposed algorithms under a dynamic topology network with dynamic cell switching on/off scenarios as well as different mobility models. With optimal parameter estimation, MRO-TL adapts to dynamic wireless networks in a short time. Simulation showed that the adaptation time of MRO-TL was only 4.17% of the baseline reference algorithm. Furthermore, by utilizing transferred knowledge, MRO-TL adapted to dynamic mobility better with a stable handover performance, and satisfied the performance target at such a significantly improved rate—416.7%—compared to the baseline algorithm.

We showed that MRO-TL estimated the optimal parameters to obtain better initialization of handover parameters for the dynamic network topology, and utilized the prior knowledge as probability distributions of the optimal parameters to adapt to dynamic mobility. We also observed that the optimal CIOs were real numbers instead of integers, as in the 3GPP standards. Hence, a finer resolution for a CIO helped improve handover optimization, which was proven in the simulations. Finally, the online MRO-TL performed so closely to the offline one that the difference was negligible. Therefore, online MRO-TL is applicable to real-time handover optimization under dynamic wireless networks.

Part II

Resource Allocation in
Multi-Connectivity Wireless
Networks.

Chapter 4

Joint Power Allocation and

Beam-forming Design for

Dual-connectivity Wireless Networks¹

In this paper, we propose a cooperating scheme to maximize network throughput while guaranteeing user quality of experience (QoE) demands in multiple-input-multiple-output (MIMO) systems. One of the aspired-to targets of the fifth generation (5G) network is to boost QoE everywhere, especially in the cell-edge areas. User equipments (UEs) in the edge areas are vulnerable to QoE violations, and they need dual connectivity from two nearby transmission points. Hence, UEs are categorized into two groups: single-connectivity and dual-connectivity. After classification, transmission power is allocated to maximize the network capacity while guaranteeing the minimum QoE. By comparing performance with a single connectivity-based algorithm and a fixed multi-connectivity-based algorithm, we show that our proposed algorithm not only satisfies all the UEs in the system but also

¹The study in this chapter was published in IEEE WCNC 2020[72].

maximizes the network capacity.

4.1 Introduction

Achieving high data rates with remarkably enhanced user quality of experience (QoE) is the main driving force in the development of fifth generation (5G) wireless networks. The data rate for 5G is expected to be orders of magnitude larger than the legacy Long Term Evolution Advanced (LTE-A) system [73]. In 5G, network operators aim to offer their subscribers guaranteed QoE everywhere, which involves all the characteristics for quality of service, such as throughput, reliability (a low packet loss rate), low delay, and high availability. To improve QoE, network operators can deploy a large number of transmission points (TPs) in wireless networks [73], such as small cells and WiFi access points. In such a dense network, multi-connectivity that enables more than one TP to serve UEs is considered to be a vital solution for enhanced data rates as well as reliability [74].

Theoretical research has shown the necessity for cooperative multi-connectivity transmissions to improve QoE in multiple-input multiple-output (MIMO) networks [74, 75]. Moreover, engineering demonstrations have proven that cooperative multi-connectivity transmission concepts can provide higher data rates and increased spectral efficiency [76, 77]. The concepts are based on appropriate beamforming algorithms so that signals transmitted from coordinating TPs are mutually suppressed to eliminate interference, or are constructively added to increase the data rate.

Cooperative multi-connectivity transmission has been studied [78–84]. In [78], a block diagonalization technique was applied to remove inter-user interference in cooperative MIMO networks. In [79], the authors proposed a sub-optimal user-selection algorithm under a zero-forcing beamforming mechanism in multi-cell MIMO systems. To maximize network

capacity under a per-base station power constraint, a joint transmission algorithm based on a block diagonalization precoding technique was developed [80,81]. In [82], the authors utilized a game theory-based approach for cooperation between TPs, and to maximize the network sum rate. Even though they attempted to maximize capacity, guaranteed QoE was ignored.

To improve QoE in LTE-A networks, Shen et al. [83] provided joint transmission to support cell-edge UEs while limiting interference among cooperating base stations. In [84], the authors studied cooperation over transmissions in WLAN networks to improve QoE in the systems. Though they adopted a multi-connectivity concept to improve QoE, capacity maximization was excluded. Furthermore, previous work utilized all TPs in a system to support all UEs; however, the number of TPs needed for each UE can be different due to various the channel conditions of the UEs.

In this paper, we propose a solution for maximizing network capacity while taking into account a minimum QoE in cooperating MIMO systems. To maximize network capacity and fulfill the minimum QoE, we apply a beamforming design to eliminate inter-user interference, and implement an adaptive dual-connectivity algorithm to guarantee QoE at the edge areas. Based on the channel conditions and the minimum QoE, we propose a UE classification algorithm to identify whether UEs need dual connectivity to fulfill QoE. After classification, we introduce power allocation to meet the QoE for UEs, while also maximizing network capacity.

4.2 System Model and Problem Formulation

We consider a coordinating MIMO network, where multiple TPs can serve UEs simultaneously via a joint transmission scheme. The scheme enables data to be available at

multiple TPs so that data can be delivered to UEs in the same time-frequency resource to improve throughput [85]. We denote a set of multi-connectivity UEs as \mathcal{U} , and the set is served by a set of coordinating TPs denoted as \mathcal{J} . In practice, there can be several multi-connectivity UE sets and corresponding TP sets, as shown in Fig. 4.1. For example, \mathcal{U}_1 and \mathcal{U}_2 are served by \mathcal{J}_1 and \mathcal{J}_2 , respectively. There are N_{TP} TPs in the system.

The MIMO system operates at millimeter wave (mmWave) frequencies, in which TPs and UEs are equipped with a fully-connected hybrid beamforming architecture [86]. For beamforming, a TP has N_a antennas and N_{rf} radio frequency (RF) chains. Each UE owns M_a antennas and M_{rf} RF chains. A single data stream is used by each of the UEs.

Assuming channels experience block fading, the input-output expression of user $u \in \mathcal{U}$ is generally given by

$$y_u = \mathbf{W}_{\text{bb}u}^H \mathbf{W}_{\text{rf}u}^H \sum_{j \in \mathcal{J}} \left(\mathbf{H}_{j,u} \mathbf{F}_{\text{rf}j} \mathbf{F}_{\text{bb}j,u} s_u + \sum_{v \in \mathcal{U}, v \neq u} \mathbf{H}_{j,u} \mathbf{F}_{\text{rf}j} \mathbf{F}_{\text{bb}j,v} s_{j,v} + \mathbf{n}_u \right), \quad (4.1)$$

where \mathbf{s}_u is the transmitting baseband signal for UE u with $\mathbb{E}[\mathbf{s}_u \mathbf{s}_u^H] = 1$ ($\mathbb{E}[x]$ denotes the expectation of x , and x^H is the conjugate transpose vector of x). $\mathbf{H}_{j,u} \in \mathbb{C}^{M_a \times N_a}$ is the channel matrix of UE u from TP j , and $\mathbf{n}_u \in \mathcal{CN}(0, \sigma^2)$ is additive complex Gaussian noise and is identically and independently distributed. $\mathbf{F}_{\text{bb}j}$ is an $N_{\text{rf}} \times K$ baseband precoding matrix of TP j for serving K users. $\mathbf{F}_{\text{bb}j,u}$ is an $N_{\text{rf}} \times 1$ baseband precoding vector of TP j for UE u . $\mathbf{F}_{\text{rf}j}$ is the baseband precoding matrix of TP j . $\mathbf{W}_{\text{bb}j,u}$ is the $M_{\text{rf}} \times 1$ baseband combining vector of UE u for TP j , and $\mathbf{W}_{\text{rf}u}$ is the $M_a \times M_{\text{rf}}$ RF combining matrix.

For expressing a multi-connectivity transmission, we rewrite (5.1) as

$$y_u = \mathbf{W}_{\text{bb}u}^H \tilde{\mathbf{H}}_u \mathbf{F}_{\text{bb}u} s_u + \sum_{v \in \mathcal{U}, v \neq u} \mathbf{W}_{\text{bb}u}^H \tilde{\mathbf{H}}_u \mathbf{F}_{\text{bb}v} s_v + \mathbf{W}_{\text{bb}u}^H \mathbf{W}_{\text{rf}u}^H n_u, \quad (4.2)$$

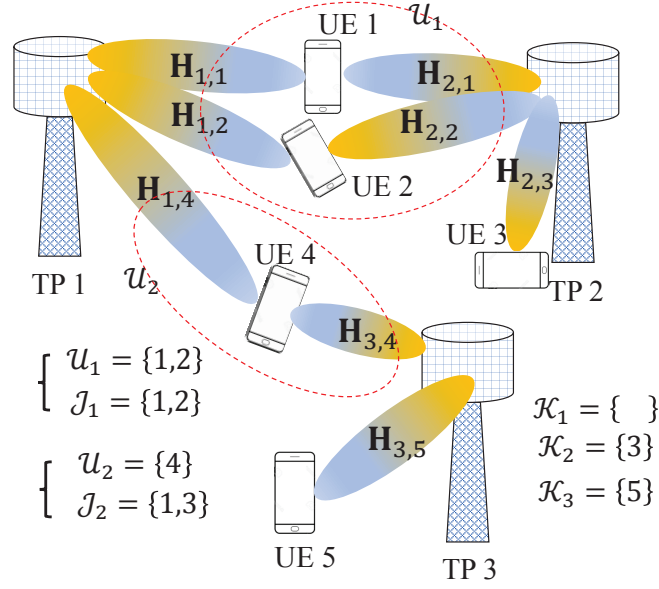


Figure 4.1: An example of a dual-connectivity wireless network.

where $\check{\mathbf{H}}_{j,u} \in \mathbb{C}^{M_{\text{rf}} \times N_{\text{rf}}}$ is the effective channel matrix and is given as

$$\check{\mathbf{H}}_{j,u} = \mathbf{W}_{\text{rf}_u}^H \mathbf{H}_{j,u} \mathbf{F}_{\text{rf}_j}, \quad (4.3)$$

and $\tilde{\mathbf{H}}_u = [\check{\mathbf{H}}_{1,u}, \dots, \check{\mathbf{H}}_{|\mathcal{J}|,u}]$. The operand $|\cdot|$ means the cardinality of a set. The achievable rate for UE u is

$$R_u = \log_2 \left(1 + D_u^{-1} \mathbf{W}_{\text{bb}_u}^H \tilde{\mathbf{H}}_u \mathbf{F}_{\text{bb}_u} Q_u \mathbf{F}_{\text{bb}_u}^H \tilde{\mathbf{H}}_u^H \mathbf{W}_{\text{bb}_u} \right), \quad (4.4)$$

where Q_u is the power loading coefficient for UE u , and D_u is given by

$$\begin{aligned} D_u &= \sum_{v \in \mathcal{U}, v \neq u} \mathbf{W}_{\text{bb}_u}^H \tilde{\mathbf{H}}_u \mathbf{F}_{\text{bb}_v} P_v \mathbf{F}_{\text{bb}_v}^H \tilde{\mathbf{H}}_u^H \mathbf{W}_{\text{bb}_u} \\ &\quad + \mathbf{W}_{\text{bb}_u}^H \mathbf{W}_{\text{rf}_u}^H \mathbf{W}_{\text{rf}_u} \mathbf{W}_{\text{bb}_u} \sigma^2. \end{aligned} \quad (4.5)$$

For modeling the wireless channel, we use the cluster-based mmWave channel model to simulate the limited scattering feature of the mmWave channel [86, 87]. The mmWave

channel for UE k from TP j , which is $\mathbf{H}_{j,k}$, is given as

$$\mathbf{H}_{j,k} = \sqrt{\frac{N_a M_a}{\mu_{j,k} N_C N_P}} \sum_{m=1}^{N_C} \sum_{n=1}^{N_P} \alpha_{m,n}^{j,k} \mathbf{a}_{UE}^{j,k}(\theta_{m,n}^{j,k}) \mathbf{a}_{TP}^{j,k}(\phi_{m,n}^{j,k}),$$

where $\mu_{j,k}$ is the large-scale path loss including shadowing from TP j to UE k [88], N_C is the number of clusters. N_P is the number of radio paths scattering from one cluster, $\alpha_{m,n}^{j,k}$ is the complex gain of radio path n in cluster m , and $\mathbf{a}_{UE}^{j,k}(\theta_{m,n}^{j,k})$ and $\mathbf{a}_{TP}^{j,k}(\phi_{m,n}^{j,k})$ are the receiving and transmitting ray-like radio paths at the azimuth angles of $\theta_{m,n}^{j,k}$ and $\phi_{m,n}^{j,k}$, respectively; $\theta_{m,n}^{j,k}$ and $\phi_{m,n}^{j,k}$ are the angle of arrival and departure, respectively, and are uniformly distributed from 0 to 2π . Assuming a uniform linear array antenna layout is considered, each radio path can be characterized by an array response vector as

$$\mathbf{a}(\theta) = \frac{1}{\sqrt{L_a}} \left[1, e^{j2\pi \frac{d}{\lambda} \sin \theta}, \dots, e^{j(L_a-1)2\pi \frac{d}{\lambda} \sin \theta} \right]^T, \quad (4.6)$$

where L_a is the number of antenna elements, λ is the wavelength, and d stands for the inter-element distance. In this paper, we consider half-lambda spacing, e.g. $d/\lambda = 1/2$.

We assume that TPs can share channel state information (CSI) and user data through an inter-cell connection link with extremely low-latency optical fibers [77]; thus, the synchronization delay is negligible. Moreover, CSI is assumed to be recorded at both transmitters and receivers.

To maximize the total capacity while considering the minimum QoE, we propose an optimization problem as follows

$$\begin{aligned} \text{(P1)} : \quad & \max_{\{\mathbf{F}_{bb}, \mathbf{F}_{rf}, \mathbf{W}_{bb}, \mathbf{W}_{rf}\}} \sum_{u \in \mathcal{U}} R_u \\ \text{s.t.} \quad & \sum_{u \in \mathcal{U}} \text{Tr} \left(\mathbf{F}_{bbj,u} \mathbf{F}_{rfj} Q_u \mathbf{F}_{rfj}^H \mathbf{F}_{bbj,u}^H \right) \leq P_{\max}, \quad j \in \mathcal{J} \\ & R_u \geq R_{\min}, \end{aligned}$$

where P_{\max} and R_{\min} are the maximum transmission power and the minimum required throughput, respectively. The problem considers multi-connectivity transmission, which

includes all TPs, to satisfy UEs in the system. Since wireless connections can experience high propagation loss, we consider dual connectivity to leverage the efficiency of power allocation.

To solve optimization problem (P1), in Section III, we first work on the hybrid beamforming design for multi-connectivity transmission. After that, we propose a UE classification algorithm and power allocation algorithms in Section IV.

4.3 Hybrid Beamforming Design

This section provides a hybrid beamforming design to obtain a large antenna, gain as well as to suppress the inter-user interference for multi-cell transmissions.

4.3.1 Design of the Combining and Precoding RF Matrices

The objective of the RF design is to compensate for higher than expected channel losses by facilitating a large beamforming gain. First, the combining RF matrix of each of the UEs is maximized by solving the following problem [86]

$$\max_{\mathbf{W}_{\text{rf}u} \in \mathcal{W}} \sum_{j \in \mathcal{J}} \left\| (\mathbf{W}_{\text{rf}u})^H \mathbf{H}_{j,u} \right\|_1^2$$

where \mathcal{W} is a discrete fourier transform matrix of dimension $M_a \times M_a$, in which the spatial frequency is $2\pi d/\lambda \sin \theta$. Given $\mathbf{W}_{\text{rf}u}$, we define matrix $\hat{\mathbf{H}}_j$ as $\hat{\mathbf{H}}_j = [\hat{\mathbf{H}}_{j,1}^T, \dots, \hat{\mathbf{H}}_{j,|\mathcal{U}|}^T]^T$, where $\hat{\mathbf{H}}_{j,u} = \mathbf{W}_{\text{rf}u}^H \mathbf{H}_{j,u}$. Then, $\mathbf{F}_{\text{rf}j}$ is computed based on an equal gain transmission algorithm [86]

$$\mathbf{F}_{\text{rf}j} = \frac{1}{\sqrt{N_a}} e^{i\xi_{j,u}},$$

where $\xi_{j,u}$ is the phase of the (j, u) -th element of $\hat{\mathbf{H}}_j^H$.

4.3.2 Joint Design of the Baseband Combining and the Precoding Matrices

To suppress the inter-user interference on dual-connectivity downlink, we apply the block diagonalization (BD) technique to obtain the baseband combining and precoding matrices, \mathbf{W}_{bb} and \mathbf{F}_{bb} . To leverage the low-dimensional BD processing, we consider the effective channel matrices. Via (5.9), we aim at designing \mathbf{F}_{bb} such that $\bar{\mathbf{H}}_u \mathbf{F}_{\text{bb}v} = 0, \forall u \neq v$. We define a concatenating matrix for UE u as

$$\bar{\mathbf{H}}_u = \left[\tilde{\mathbf{H}}_1^T, \dots, \tilde{\mathbf{H}}_{u-1}^T, \tilde{\mathbf{H}}_{u+1}^T, \dots, \tilde{\mathbf{H}}_{|\mathcal{U}|}^T \right]^T. \quad (4.7)$$

The fundamentals of BD are to obtain the $\mathbf{F}_{\text{bb}u}$ that lies in the null space of $\bar{\mathbf{H}}_u$. By singular value decomposition (SVD), we can obtain $\mathbf{F}_{\text{bb}u}$. The SVD of $\bar{\mathbf{H}}_u$ is

$$\bar{\mathbf{H}}_u = \bar{\mathbf{U}}_u \bar{\Sigma}_u \left[\bar{\mathbf{V}}_u^1, \bar{\mathbf{V}}_u^0 \right]^H, \quad (4.8)$$

where $\bar{\mathbf{V}}_u^1 \in \mathcal{C}^{2|\mathcal{U}|M_{\text{rf}} \times (|\mathcal{U}|-1)M_{\text{rf}}}$ contains the first $(|\mathcal{U}|-1)M_{\text{rf}}$ right singular vectors of $\bar{\mathbf{H}}_u$, and $\bar{\mathbf{V}}_u^0$ includes the last $2|\mathcal{U}|M_{\text{rf}} - (|\mathcal{U}|-1)M_{\text{rf}}$ right singular vectors that form the null basis of the null space of $\bar{\mathbf{H}}_u$. The rank of $\bar{\mathbf{H}}_u$ implicitly satisfies the condition $2|\mathcal{U}|M_{\text{rf}} > (|\mathcal{U}|-1)M_{\text{rf}}$.

Since $\mathbf{F}_{\text{bb}u}$ lies on the null space of $\bar{\mathbf{H}}_u$, we can rewrite (4.2) as

$$y_u = \mathbf{W}_{\text{bb}u}^H \tilde{\mathbf{H}}_u \mathbf{F}_{\text{bb}u} s_u + \mathbf{W}_{\text{bb}u}^H \mathbf{W}_{\text{rf}u}^H n_u.$$

To achieve the optimal achievable rate, we perform SVD as follows

$$\tilde{\mathbf{H}}_u \bar{\mathbf{V}}_u^0 = \hat{\mathbf{U}}_u \hat{\Sigma}_u \hat{\mathbf{V}}_u^H. \quad (4.9)$$

Then, the optimal precoding and combining baseband matrices of $\mathbf{F}_{\text{bb}u}$ and $\mathbf{W}_{\text{bb}u}$ are $\bar{\mathbf{V}}_u^0 \hat{\mathbf{V}}_u^{(1)}$ and $\hat{\mathbf{U}}_u^{(1)}$, respectively, where $A^{(1)}$ denotes the first column of matrix A. The overall precoding matrix for dual-connectivity UEs in TP j is given as $\mathbf{F}_{\text{DC}j} = [\mathbf{F}_{\text{rf}j} \mathbf{F}_{\text{bb}j,1}, \dots, \mathbf{F}_{\text{rf}j} \mathbf{F}_{\text{bb}j,|\mathcal{U}|}]$.

With the obtained precoding and combining baseband matrices, the inter-user interference is eliminated so that we can rewrite the achievable data rate (5.8) as

$$R_u = \log_2 \left(1 + \sigma^{-2} (\hat{\Sigma}_u^*)^2 Q_u \right),$$

where $\hat{\Sigma}_u^*$ is the first, and also the largest, singular value in diagonal singular matrix $\hat{\Sigma}_u$.

Since UEs can be served by a single TP, we can apply the BD technique for single-connectivity transmission. Assuming that UEs select the best TP based on reference signal received power (RSRP), we define the single-connectivity set served by TP j as \mathcal{K}_j . For UE $k \in \mathcal{K}_j$, we apply (5.10), (5.11), and (5.12) for concatenating $\check{\mathbf{H}}_{j,k}$ in (5.5). Then, we can get $\mathbf{F}_{\text{bb},j,k}$, $\mathbf{W}_{\text{bb},j}$, and $\hat{\Sigma}_{j,k}^*$, which are, respectively, the baseband precoding vector of TP j for UE k , the baseband combining vector of UE k for TP j , and the largest singular value of the channel between UE k and TP j . For single-connectivity UEs, the inter-cell interference can be negligible because of the propagation loss of the mmWave channel. We summarize the notations used in this paper in Table 4.1. In the next section, we provide UE classification and a power allocation algorithm, which are based on the obtained singular values.

4.4 Power Allocation Algorithms

This section provides the proposed power allocation algorithm that can guarantee QoE via dual connectivity while maximizing the network capacity. According to channel conditions and QoE, single-connectivity transmission is enough for some UEs, while dual connectivity is essential for others. Hence, we first identify which UEs need dual connectivity. After the classification, we allocate the power to fulfill QoE demands of the dual-connectivity UEs, and the remaining power is for maximizing the sum rate of the single-connectivity UEs. The main algorithm is described in Algorithm 9

Table 4.1: Summary of parameters in this paper.

Parameter	Description
$\mathbf{H}_{j,u}$	Channel matrix between UE u and TP j
$\mathbf{F}_{\text{bb}j}$	Baseband precoding matrix of TP j
$\mathbf{F}_{\text{rf}j}$	RF precoding matrix of TP j
$\mathbf{F}_{j,u}$	Overall precoding matrix of TP j for UE u
$\mathbf{W}_{\text{bb}u}$	Baseband combining matrix of UE u
$\mathbf{W}_{\text{rf}u}$	RF combining matrix of UE u
$\check{\mathbf{H}}_{j,u}$	Effective channel matrix between UE u and TP j
$\bar{\mathbf{H}}_u$	Concatenation of effective matrix for BD technique
$g_{j,k}$	Chanel quality between single-connectivity UE k and TP j
g_u	Chanel quality received by dual-connectivity UE u
$P_{j,k}^{\min}$	Minimum power of TP j in order for UE k to satisfy the QoE demand
$P_{\text{DC}j}^{\max}$	Maximum power of TP j for dual connectivity
\mathcal{U}_n	The n -th dual-connectivity UE set
$\mathcal{U}_{\text{total}}$	The dual-connectivity set in the system
\mathcal{U}^\dagger	The optimal selected dual-connectivity set in the system
\mathcal{J}_n	The n -th TP set supporting \mathcal{U}_n
\mathcal{K}_j	The single-connectivity UE set of TP j
$\check{\mathcal{K}}_j$	The optimal single-connectivity set of TP j

Algorithm 9 Adaptive Dual-Connectivity Power Allocation Algorithm

- 1: Classify UEs into single-connectivity UEs and dual-connectivity UEs (Algorithm 10)
- 2: Allocate power to satisfy as many dual-connectivity UEs as possible (Algorithm 11)
- 3: Allocate the remaining power to maximize network capacity by solving (P3)

4.4.1 User Classification Algorithm

This algorithm classifies UEs based on the idea of a water-filling algorithm that maximizes the network capacity [78]. The algorithm examines whether single-connectivity transmission is enough for all UEs to meet the required QoE. If single-connectivity transmission can guarantee QoE for all UEs, power is allocated to all UEs by solving (P3).

For each TP j , the algorithm sorts single-connectivity UE list \mathcal{K}_j to obtain a sorted list, $\check{\mathcal{K}}_j$, with respect to their channel quality, $\bar{\Sigma}_{j,k}$ ($k = 1, \dots, |\mathcal{K}_j|$). To identify which UEs cannot be satisfied by single-connectivity transmissions, we compute the minimum power for UE $\check{k} \in \check{\mathcal{K}}_j$ to fulfill QoE under single-connectivity transmission, denoted as $P_{j,\check{k}}^{\min}$, given as

$$P_{j,\check{k}}^{\min} = \|\mathbf{F}_{j,\check{k}}\|^2 Q_{j,\check{k}}^{\min}, \quad (4.10)$$

where $Q_{j,\check{k}}^{\min} = (2^{R_{\check{k}}^{\min}} - 1)/(g_{j,\check{k}}^2)$, which is the minimum power loading coefficient, and $g_{j,\check{k}} = \sigma^{-2}(\hat{\Sigma}_{j,\check{k}}^*)^2$, which is the channel gain of single-connectivity transmission for UE \check{k} from TP j , and $\|\cdot\|$ denotes a norm-2 operand. Since $P_{j,\check{k}}^{\min}$ is reciprocal to $g_{j,\check{k}}$, the last UE in $\check{\mathcal{K}}_j$ needs the largest amount of power to meet QoE.

Due to the channel condition, we need to find the optimal single-connectivity list for TP j , because all UEs are not satisfied by single connections. To get the optimal single-connectivity list for TP j , the algorithm iteratively removes the last UE of $\check{\mathcal{K}}_j$ such that $\sum_{\check{k}=1}^{|\check{\mathcal{K}}_j|} P_{j,\check{k}}^{\min} \leq P_{\max}$. For UEs that cannot be satisfied by single connections, the algorithm stores them in a dual-connectivity UE set, called $\mathcal{U}_{\text{total}}$.

Each dual-connectivity UE $u \in \mathcal{U}_{\text{total}}$ chooses the second TP for dual connectivity, which has the second-largest RSRP in the system. Because dual-connectivity UEs can be served by different pairs of TPs, we group dual-connectivity UEs by pairs of TPs. For example in Fig. 4.1, dual-connectivity UE set \mathcal{U}_1 including UE 1 and UE 2 are served by TP

set \mathcal{J}_1 consisting of TP 1 and TP 2. Likewise, dual-connectivity UE set \mathcal{U}_2 including only UE 4 is served by TP set \mathcal{J}_2 consisting of TP 1 and TP 3. We summarize the algorithm for user classification in Algorithm 10. In the next subsection, we propose a power allocation algorithm for dual-connectivity UEs in order to meet QoE while maximizing the sum rate of single-connectivity UEs.

Algorithm 10 User Classification Algorithm

- 1: Input: $\mathbf{F}_{j,k}$, $\bar{\Sigma}_{j,k}$, $j = 1, \dots, N_{\text{TP}}$, $k = 1, \dots, |\mathcal{K}_j|$
 - 2: Output: \mathcal{U}_n s and corresponding \mathcal{J}_n s, $\check{\mathcal{K}}_j$, $j = 1, \dots, N_{\text{TP}}$
 - 3: **for** each TP j **do**
 - 4: Sort single-connectivity UE list \mathcal{K}_j in ascending order based on the single-connectivity channel quality $\bar{\Sigma}_{j,k}$ ($k = 1, \dots, |\mathcal{K}_j|$). Achieve sorted list $\check{\mathcal{K}}_j$
 - 5: Calculate $P_{j,\check{k}}^{\min}$ ($\check{k} = 1, \dots, |\check{\mathcal{K}}_j|$) via (4.10)
 - 6: **while** $\sum_{k=1}^{|\check{\mathcal{K}}_j|} P_{j,\check{k}}^{\min} > P_{\max}$ **do**
 - 7: Remove the last UE in sorted list $\check{\mathcal{K}}_j$ and put the UE in dual-connectivity UE set $\mathcal{U}_{\text{total}}$
 - 8: **end while**
 - 9: **end for**
 - 10: **for** each UE in dual-connectivity UE set \mathcal{U} **do**
 - 11: Find the second-best TP based on RSRP
 - 12: **end for**
 - 13: In $\mathcal{U}_{\text{total}}$, group dual-connectivity UEs that have common TPs in order to get \mathcal{U}_n s and the corresponding \mathcal{J}_n s.
-

4.4.2 Power Allocation with Guaranteed QoE

To meet the required QoE while maximizing the network capacity, we allocate power to dual-connectivity UEs to meet QoE with a power constraint, and the remaining power is for maximizing network capacity with single-connectivity UEs. First, we compute the sufficient power for each of the dual-connectivity UEs based on their channel quality. For each TP j that serves dual-connectivity UE u , the minimum power assigned by TP j for dual-connectivity UE u is given as

$$P_{j,u}^{\min} = \|\mathbf{F}_{j,u}\|^2 Q_u^{\min} \quad (4.11)$$

where $Q_u^{\min} = (2^{R_u^{\min}} - 1)/g_u^2$, and $g_u = \sigma^{-2}(\hat{\Sigma}_u^*)^2$, which is the overall channel gain of dual connectivity to UE u from the corresponding TP set.

Since all dual-connectivity UEs may not be served at once due to the power limitation, we choose an optimal subset of dual-connectivity UEs from $\mathcal{U}_{\text{total}}$, called \mathcal{U}^\dagger . To that end, we solve the following optimization problem to get \mathcal{U}^\dagger

$$\begin{aligned} \text{(P2)} : \max & \quad |\mathcal{U}^\dagger| \\ \text{s.t.} & \quad \sum_{u \in \mathcal{U}^\dagger} \|\mathbf{F}_{j,u}\|^2 Q_u^{\min} \leq P_{\text{DC}_j}^{\max}, \quad j = 1, \dots, N_{\text{TP}} \\ & \quad Q_u \geq 0, \quad u \in \mathcal{U}^\dagger. \end{aligned}$$

We note that $\|\mathbf{F}_{j,u}\|^2 Q_u^{\min} = 0$ if TP j does not serve UE u .

We obtain \mathcal{U}^\dagger as follows. We initialize \mathcal{U}^\dagger such that $\mathcal{U}^\dagger = \mathcal{U}$. For each TP j , the maximum remaining power $P_{\text{DC}_j}^{\max}$ for each TP j available for satisfying dual-connectivity UEs is computed as

$$P_{\text{DC}_j}^{\max} = P_{\text{max}} - \sum_{k=1}^{|\tilde{\mathcal{K}}_j|} P_{j,k}^{\min}. \quad (4.12)$$

Algorithm 11 Power Allocation Algorithm for Dual Connectivity

- 1: Input: \mathbf{F}_{SC_j} , $\bar{\Sigma}_{j,k}$, $j = 1, \dots, N_{\text{TP}}$, $k = 1, \dots, |\mathcal{K}_j|$
 - 2: Output: \mathcal{U}^\dagger and P_{DC_j} , $j = 1, \dots, N_{\text{TP}}$
 - 3: **for** each UE $u \in \mathcal{U}$ **do**
 - 4: Compute $P_{j,u}^{\min}$ by (5.23) for each TP j serving UE u
 - 5: **end for**
 - 6: Set $\mathcal{U}^\dagger = \mathcal{U}$
 - 7: **for** each TP j in the system **do**
 - 8: Compute $P_{\text{DC}_j}^{\max}$ via (5.24)
 - 9: Sort dual-connectivity UEs that are served by TP j based on the minimum power
 - 10: Compute P_{DC_j} via (4.13)
 - 11: **while** $P_{\text{DC}_j} > P_{\text{DC}_j}^{\max}$ **do**
 - 12: Remove the last UE of the sorted list for TP j and also remove it from \mathcal{U}^\dagger
 - 13: **end while**
 - 14: **end for**
-

Then, a list of the dual-connectivity UEs served by TP j is sorted in ascending order based on $P_{j,u}^{\min}$. After the sorting process, the algorithm iteratively removes the last UE from the sorted list, and extracts it from \mathcal{U}^\dagger until the power for dual connectivity does not exceed $P_{\text{DC}j}^{\max}$, i.e., $\sum_{u \in \mathcal{U}^\dagger} \|\mathbf{F}_{j,u}\|^2 Q_u^{\min} \leq P_{\text{DC}j}^{\max}$.

After \mathcal{U}^\dagger has been achieved, the total power spent for dual connectivity for each TP j , $P_{\text{DC}j}$, is given by

$$P_{\text{DC}j} = \sum_{u \in \mathcal{U}^\dagger} \|\mathbf{F}_{j,u}\|^2 Q_u^{\min}. \quad (4.13)$$

The procedure to acquire \mathcal{U}^\dagger is summarized in Algorithm 11.

After satisfying dual-connectivity UEs, we maximize the sum rate for single-connectivity UEs with the remaining power at each TP j , $P_{\text{SC}j}^{\max}$, given as

$$P_{\text{SC}j}^{\max} = P_{\text{max}} - P_{\text{DC}j}.$$

For each TP j , we allocate the power for single-connectivity UEs according to the following optimization problem

$$\begin{aligned} (\text{P3}) : \max & \sum_{\check{k}=1}^{|\check{\mathcal{K}}_j|} R_{\check{k}} \\ \text{s.t.} & \sum_{\check{k}=1}^{|\check{\mathcal{K}}_j|} \|\mathbf{F}_{j,\check{k}}\|^2 Q_{j,\check{k}} \leq P_{\text{SC}j}^{\max}, \\ & Q_{j,\check{k}} \geq 0, \quad j = 1, \dots, |\check{\mathcal{K}}_j|, \\ & R_{\check{k}} \geq R_{\check{k}}^{\min}. \end{aligned}$$

Since (P3) is a convex problem, and the constraints are feasible, it is solved by a water-filling approach [78].

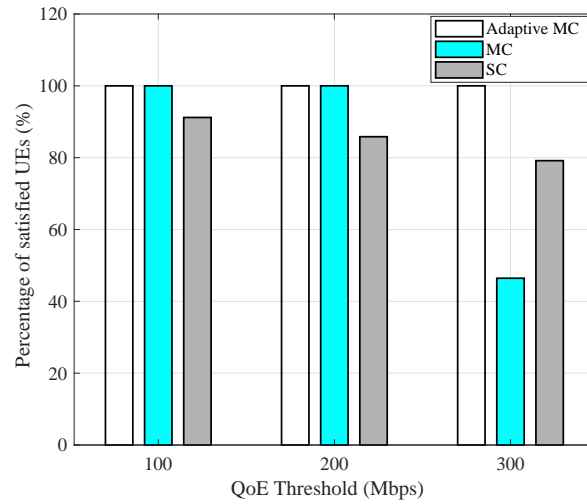


Figure 4.2: Percentage of satisfied UEs at different QoE thresholds.

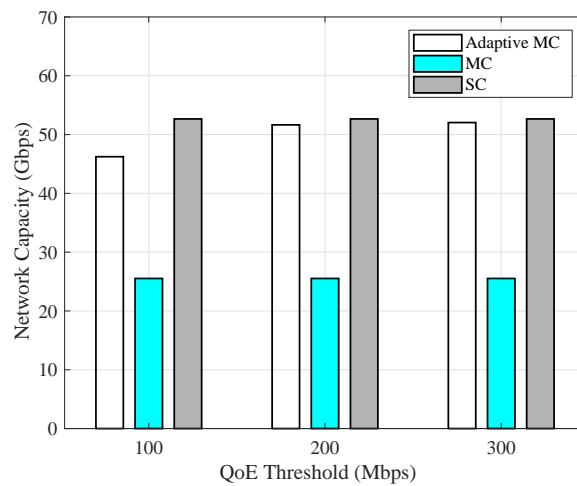


Figure 4.3: Network capacity for different QoE thresholds.

4.5 Simulation Results and Analysis

In this section, we evaluate the performance of the proposed algorithm in terms of network capacity and the percentage of satisfied UEs using Matlab. We compared the proposed adaptive dual-connectivity algorithm, denoted as Adaptive DC, with a previous algorithm that considers single-connectivity transmission [86], called the SC algorithm, and an algorithm that encompasses multiple TPs for joint transmissions [89], called the MC algorithm.

The simulation environment includes seven TPs, i.e $N_{\text{TP}} = 7$, each of which serves 12 UEs [85] in the beginning. TPs are deployed regularly at an inter-site distance of 50 m [85]. The maximum transmission power of a TP is 30 dBm. The system operates on a carrier frequency of 28 GHz with a 100 MHz bandwidth [85]. The noise figure at a receiver is set at 10 dB. Hybrid beamforming is commonly used for all considered algorithms. For beamforming parameters, $N_{\text{a}} = 256$, $N_{\text{rf}} = 48$, $M_{\text{a}} = 8$, and $M_{\text{rf}} = 4$, as in [85]. For channel realization, we set N_{C} and N_{P} at 3 and 2, respectively. The UE mobility is modeled as a random walk at 3 km/h. We set identical QoE thresholds for all UEs in the system.

Fig. 5.2 depicts the percentage of satisfied UEs achieved by our proposed algorithm and in two previous works by setting three QoE thresholds: 100 Mbps, 200 Mbps, and 300 Mbps. As we can observe, the adaptive DC algorithm can satisfy all the UEs, while the SC algorithm can only fulfill QoE for a portion of the UEs. This is because the SC algorithm neglects the minimum required QoE, and lacks coordination between TPs in order to support cell-edge UEs. The MC algorithm can satisfy UEs for 100 Mbps and 200 Mbps QoE requirements; however, the algorithm can guarantee the minimum QoE for just under 50% of the UEs when the demand is 300 Mbps. Since the MC algorithm requires all TPs in the system to support UEs, power is inefficiently allocated when some UEs are too far from

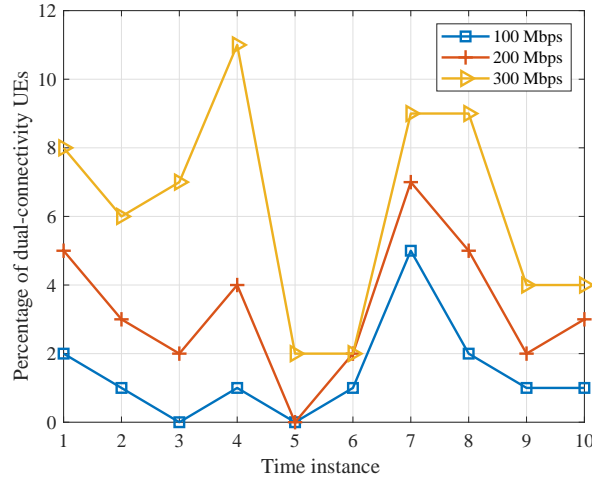


Figure 4.4: Adaptive DC for each time instance due to mobility.

a TP for efficient communication. We provide a solution for adaptive dual connectivity, and therefore, we have better performance.

Fig. 5.3 illustrates the system capacity from the three algorithms. The results show that the adaptive DC algorithm obtains the second-best performance but it is close to the SC algorithm. Meanwhile, the MC algorithm provides the lowest performance. This is because our algorithm can maximize the sum rate of single-connectivity UEs that are near the center of the TPs, and can also fulfill the QoE requirements of the UEs via dual connectivity. The MC algorithm pushes all TPs in the system to support a UE, which is inefficient due to high propagation loss, so it induces the worst performance.

Fig. 5.4 shows that the portion of dual-connectivity UEs varies over time due to mobility. We chose snapshots from 10 different times. The results show that our algorithm can adaptively classify UEs for dual connectivity in order to fulfill the required QoE. When the demanded QoE is higher the percentage of dual-connectivity UEs increases because more power is needed to meet the QoE requirements.

4.6 Closing Remarks

In this paper, we proposed a power allocation algorithm with hybrid beamforming to maximize the network capacity, along with satisfying the minimum QoE. The beamforming design was applied to reduce interference and to maximize network throughput. In cases where single connections cannot guarantee the minimum QoE for all UEs, depending on the wireless environment, for a given beamforming, UEs are classified into two groups. According to the UE categories, power is allocated to maximize the network capacity while satisfying the minimum QoE. Simulations show that our proposed algorithm dynamically classifies UE groups based on QoE and the channel conditions of the UEs, and maximizes network capacity while satisfying the QoE for UEs.

This page intentionally left blank.

Chapter 5

Joint Power Allocation and Beam-forming Design for Dual-connectivity Wireless Networks Under Imperfect CSI

Multi-connectivity is a promising solution for current wireless technologies and future networks to fulfill ever increasing demands of quality of experience (QoE). User data will be transmitted from multiple source using massive multi-output multi-input (MIMO) to support subscribers, which can not be happy due to a low QoE. In multi-connectivity wireless networks, the tradeoff between energy efficiency and spectrum efficiency poses a challenge on to the radio access design, such as power allocation and beamforming. Furthermore, performance of MIMO systems can be majorly degraded because of the accuracy of the channel estimation, which is known as channel state information (CSI). In this work, we propose a joint power allocation and beamforming design considering the impact of imperfect

CSI to guarantee user quality of experience (QoE) demands while minimizing the total transmission power in multi-connectivity multiple-input-multiple-output (MIMO) systems. Then, the design took into account dual connectivity of 5G new radio (NR) standard. First, the effect of imperfect CSI is analyzed and estimated by a closed form in term of signal-to-interference-plus-noise (SINR) ratio. Then a two-stage algorithm will categorize UEs into two groups: single-connectivity and dual-connectivity. After classification, transmission power is allocated to guarantee the demanded QoE while minimizing the transmission power. By comparing performance with a single connectivity-based algorithm and a fixed multi-connectivity-based algorithm, we show that our proposed algorithm not only satisfies all the UEs in the system but also consumes less transmission power.

5.1 Introduction

Guaranteeing high quality of experience (QoE) while spending power efficiently is one of the main driving force for the sustainable development of fifth generation wireless networks (5G) and beyond 5G [73,90]. However, the trade-off between energy efficiency and high quality of service has been challenging the optimization of the wireless systems [91–93]. To guarantee QoE every where in a wireless network, a dense deployment of transmission points (TPs) and a large number of antenna will be a potential solution [73,92]. In such dense networks, a cooperative multi-connectivity communication scheme will enable more than one TP to support UEs to enhance QoE [74,94].

Theoretical research has shown the necessity for cooperative multi-connectivity transmissions to improve QoE in multiple-input multiple-output (MIMO) networks [74, 75]. Moreover, engineering demonstrations have proven that cooperative multi-connectivity transmission concepts can provide higher data rates and increased spectral efficiency [76,

77,95]. The concepts are based on beamforming design so that signals transmitted from coordinating TPs are mutually suppressed to eliminate interference, or are constructively added to increase the data rate. However, beamforming gain is strongly affected by the channel state information (CSI), which is obtained by channel estimation scheme.

Assuming wireless communications is under perfect CSI, cooperative multi-connectivity transmission has been studied [78–84,96]. In [78], a block diagonalization technique was applied to remove inter-user interference in cooperative MIMO networks. In [79], the authors proposed a sub-optimal user-selection algorithm under a zero-forcing beamforming mechanism in multi-cell MIMO systems. To maximize network capacity under a per-base station power constraint, a joint transmission algorithm based on a block diagonalization precoding technique was developed [80,81]. In [96], complexity of the block diagonalization technique was investigated and a low-complexity precoding algorithm was developed. In [82], the authors utilized a game theory-based approach for cooperation between TPs, and to maximize the network sum rate. To improve QoE in LTE-A networks, Shen et al. [83] provided joint transmission to support cell-edge UEs while limiting interference among cooperating base stations. In [84], the authors studied cooperation over transmissions in WLAN networks to improve QoE in the systems. These previous works attempted to maximize capacity without considering minimum QoE requirement and energy efficiency.

Considering energy efficiency in multi-connectivity MIMO networks, joint beamforming design and power allocation were studied [97–102]. In [97], an optimization problem for maximizing the energy efficiency was developed by designing a hybrid beamforming architecture based on a statistical model for the transmitted signal. In [98], a transmission point (TP) selection scheme and the optimal power allocation algorithm were proposed, which aimed at maximizing the total energy efficiency subject to a per-user spectral efficiency constraint and a power constraint. In [99], the security aspect of multi-connectivity

MIMO network under a pilot spoofing attack was studied. The problem of minimizing the power consumption subject to security constraints was also considered. In [100], the author provided a usercentric architectures operating at millimeter wave frequencies, considering a training-based channel estimation phase, and the downlink and uplink data transmissions, hybrid beamforming, and resource allocation strategies maximizing the global energy efficiency. In [101], provide the a comprehensive analysis of multi connectivity wireless network under different degrees of cooperation among the TPs. In [102], a power saving strategy with load balancing was developed to minimize total power consumption in multi-connectivity wireless networks. However, from these researches, joint power allocation and beamforming design for guaranteed QoE with imperfect CSI in multi-connectivity MIMO network is still left as an open issue.

Considering imperfect CSI in joint power allocation and beamforming design, [103–106]. In [103], a robust coordinated beamforming for mitigating the inter-cell interference under imperfect CSI was studied. The objective of research is to minimize sum transmission power of base stations while considering worst-case SINR of UEs. In [104], the authors developed a power control technique with zero-forcing precoding to improve the energy efficiency considering the backhaul power consumption. In [105], the authors investigated a coordinated beamforming with error channel estimation via a uplink training scheme, and then devised a max-min fairness power control algorithm for energy efficiency. In [106], a full-duplex mmWave multi-connectivity system was investigated, a robust hybrid beamforming in an imperfect CSI scenario was studied, and sum rate maximization problem was considered. Nevertheless, guaranteeing minimum QoE requirement was still absent from these researches.

In this paper, we propose a solution for guaranteeing QoE under imperfect CSI acquisition while minimizing the total power power in cooperative multi-connectivity MIMO systems. We applied block diagonalization technique for beamforming design aiming at

minimizing inter-user interference. Due to the impact of imperfect CSI on interference aspect of downlink transmission, the channel quality is degraded and we estimated that impact in terms of signal-to-noise-and-interference (SINR) metric. We formulated the optimization problem using the estimated SINR. To solve the problem, we proposed a two-stage adaptive dual-connectivity algorithm aiming at utilizing dual connectivity, which is standardized by 3GPP. The first stage of the proposed algorithm classifies UEs into two different sets based on the estimated SINR and identifies condition when UEs need more than one connectivity to be satisfied. The second stage allocate power allocation to each UE meet the required QoE while also minimizing the transmission power. For the performance evaluation, we verify the proposed algorithm with various aspects of the wireless networks, such as different QoE requirement, level of CSI error, network load, and antenna settings.

Notations: In this paper, matrices is denoted by bold-face upper-case letters. $\mathbf{0}$, \mathbf{I} denote an identity matrix and an zero matrix, respectively, with appropriate dimensions. \mathbf{A}^T , \mathbf{A}^\dagger , and $\text{Tr}(\mathbf{A})$ are the transpose, conjugate transpose, and trace of a matrix \mathbf{A} , respectively. $\mathbf{A} \geq 0$ means that \mathbf{A} is postive semi-define matrix. $\|\mathbf{A}\|_l$ is the l -norm of matrix \mathbf{A} . $\text{Diag}(a_1, \dots, a_N)$ denotes a square block-diagonal matrix with diagonal elements a_1, \dots, a_N . The operand $|\cdot|$ means the cardinality of a set. $\bar{\delta}_{mn}$ means that $\bar{\delta}_{mn} = 1$ if $m \neq n$, and $\bar{\delta}_{mn} = 0$ if $m = n$. The operand \odot is the Hadamard product. $(\mathbf{A})_{mn}$ denotes the element at row m and column n of matrix \mathbf{A} . $(\mathbf{A})_{:m}$ is the column m of matrix \mathbf{A} . $\text{vec}(\mathbf{A})$ is the vectorization of matrix \mathbf{A} .

5.2 System Model and Problem Formulation

In this section, we provide the system model and then formulate the power consumption minimization problem.

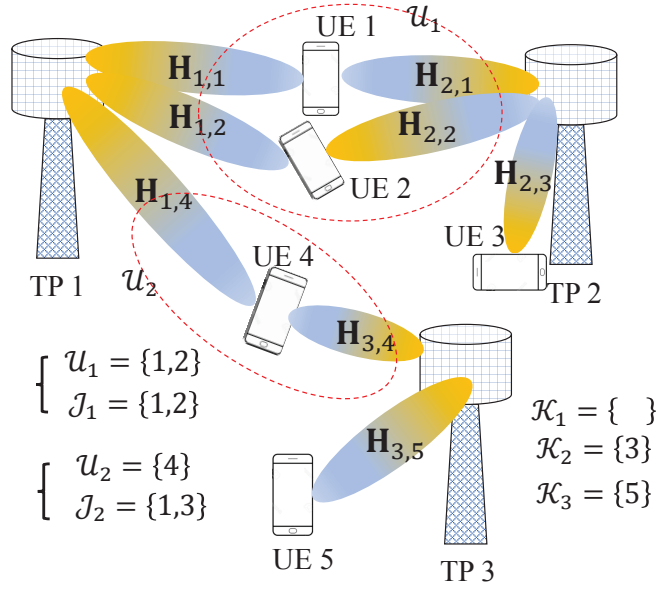


Figure 5.1: An example of a dual-connectivity wireless network.

5.2.1 System Model

We consider a coordinating MIMO network, where multiple TPs can serve UEs simultaneously via a joint transmission scheme. The scheme enables data to be available at multiple TPs so that data can be delivered to UEs in the same time-frequency resource to improve throughput [85]. We denote a set of multi-connectivity UEs as \mathcal{U} , and the set is served by a set of coordinating TPs denoted as \mathcal{J} . In practice, there can be several multi-connectivity UE sets and corresponding TP sets, as shown in Fig. 5.1. For example, \mathcal{U}_1 and \mathcal{U}_2 are served by \mathcal{J}_1 and \mathcal{J}_2 , respectively. There are N_{TP} TPs in the system.

The MIMO system operates at millimeter wave (mmWave) frequencies, in which TPs and UEs are equipped with a fully-connected hybrid beamforming architecture [86]. For beamforming, a TP has N_a antennas and N_{rf} radio frequency (RF) chains. Each UE owns M_a antennas and M_{rf} RF chains. A single data stream is used by each of the UEs.

Assuming channels experience block fading, the down-link input–output expression

of UE $u \in \mathcal{U}$ that is expected by TPs is expressed as

$$y_u = \mathbf{W}_{\text{bb}u}^H \mathbf{W}_{\text{rf}u}^H \sum_{j \in \mathcal{J}} \left(\mathbf{H}_{ju} \mathbf{F}_{\text{rf}j} \mathbf{F}_{\text{bb}ju} \mathbf{Q}_u^{\frac{1}{2}} s_u + \sum_{v \in \mathcal{U}, v \neq u} \mathbf{H}_{ju} \mathbf{F}_{\text{rf}j} \mathbf{F}_{\text{bb}jv} \mathbf{Q}_v^{\frac{1}{2}} s_{jv} + \mathbf{n}_u \right), \quad (5.1)$$

where \mathbf{s}_u is the transmitting baseband signal for UE u with $\mathbb{E}[\mathbf{s}_u \mathbf{s}_u^H] = 1$ ($\mathbb{E}[x]$ denotes the expectation of x , and x^H is the conjugate transpose vector of x). $\mathbf{H}_{ju} \in \mathbb{C}^{M_a \times N_a}$ is the downlink channel matrix of UE u observed by TP j , and $\mathbf{n}_u \in \mathcal{CN}(0, \sigma^2)$ is additive complex Gaussian noise and is identically and independently distributed. $\mathbf{F}_{\text{bb}j}$ is an $N_{\text{rf}} \times K$ baseband precoding matrix of TP j for serving K users. $\mathbf{F}_{\text{bb}ju}$ is an $N_{\text{rf}} \times 1$ baseband precoding vector of TP j for UE u . $\mathbf{F}_{\text{rf}j}$ is the baseband precoding matrix of TP j . $\mathbf{W}_{\text{bb}ju}$ is the $M_{\text{rf}} \times 1$ baseband combining vector of UE u for TP j , and $\mathbf{W}_{\text{rf}u}$ is the $M_a \times M_{\text{rf}}$ RF combining matrix.

To model the wireless channel, we use the cluster-based mmWave channel model to simulate the limited scattering feature of the mmWave channel [86, 87]. The downlink mmWave channel vector for UE u from TP j , \mathbf{H}_{ju} , is given by

$$\mathbf{H}_{ju} = \sqrt{\frac{N_a M_a}{N_C N_P}} \sum_{m=1}^{N_C} \sum_{n=1}^{N_P} \alpha_{jumn} \mathbf{a}_{\text{UE}}^{ju}(\theta_{jumn}) \mathbf{a}_{\text{TP}}^{ju}(\phi_{jumn}), \quad (5.2)$$

where N_C is the number of clusters. N_P is the number of radio paths scattering from one cluster, α_{jumn} is the complex gain of radio path n in cluster m , and $\mathbf{a}_{\text{UE}}^{ju}(\theta_{jumn})$ and $\mathbf{a}_{\text{TP}}^{ju}(\phi_{jumn})$ are the receiving and transmitting ray-like radio paths at the azimuth angles of θ_{jumn} and ϕ_{jumn} , respectively; θ_{jumn} and ϕ_{jumn} are the angle of arrival and departure, respectively, and are uniformly distributed from 0 to 2π . Assuming a uniform linear array antenna layout is considered, each radio path can be characterized by an array response vector as

$$\mathbf{a}(\theta) = \frac{1}{\sqrt{L_a}} \left[1, e^{j2\pi \frac{d}{\lambda} \sin \theta}, \dots, e^{j(L_a-1)2\pi \frac{d}{\lambda} \sin \theta} \right]^T, \quad (5.3)$$

where L_a is the number of antenna elements, λ is the wavelength, and d stands for the inter-element distance. In this paper, we consider half-lambda spacing, e.g. $d/\lambda = 1/2$.

For expressing a multi-connectivity transmission, we rewrite the signal at (5.1) as

$$\begin{aligned} y_u &= \mathbf{W}_{\text{bb}u}^H \sum_{j \in \mathcal{J}} \left(\tilde{\mathbf{H}}_{ju} \mathbf{F}_{\text{bb}ju} \mathbf{Q}_u^{\frac{1}{2}} s_u + \sum_{v \in \mathcal{U}, v \neq u} \tilde{\mathbf{H}}_{ju} \mathbf{F}_{\text{bb}jv} \mathbf{Q}_v^{\frac{1}{2}} s_{jv} + \mathbf{n}_u \right) \\ &= \mathbf{W}_{\text{bb}u}^H \left(\tilde{\mathbf{H}}_u \mathbf{F}_{\text{bb}u} \mathbf{Q}_u^{\frac{1}{2}} s_u + \sum_{v \in \mathcal{U}, v \neq u} \tilde{\mathbf{H}}_u \mathbf{F}_{\text{bb}v} \mathbf{Q}_v^{\frac{1}{2}} s_v \right) + \mathbf{W}_{\text{bb}u}^H \mathbf{W}_{\text{rf}u}^H n_u \end{aligned} \quad (5.4)$$

where $\tilde{\mathbf{H}}_{ju} \in \mathbb{C}^{M_{\text{rf}} \times N_{\text{rf}}}$ is the effective channel matrix and is given as

$$\tilde{\mathbf{H}}_{ju} = \mathbf{W}_{\text{rf}u}^H \mathbf{H}_{ju} \mathbf{F}_{\text{rf}j}, \quad (5.5)$$

and $\tilde{\mathbf{H}}_u = [\tilde{\mathbf{H}}_{1u}, \dots, \tilde{\mathbf{H}}_{|\mathcal{J}|u}] \in \mathbb{C}^{M_{\text{rf}} \times |\mathcal{J}| N_{\text{rf}}}$ is the concatenating matrix for effective channel matrix of all the TPs for UE u .

We assume that TPs can share channel state information (CSI) and user data through an inter-cell connection link, and user perfectly estimate the channel. Since the quality of the CSI feedback from UEs to TPs can be affected by quantization at the RF transceivers, the real effective channel witnessed by UE u can be different to that at TP j . We model the relation between the effective channel witnessed by UE u , $\tilde{\mathbf{H}}'_{ju}$, and that observed by TP j , $\tilde{\mathbf{H}}_{ju}$, as follow

$$\tilde{\mathbf{H}}'_{ju} = \tilde{\mathbf{H}}_{ju} + \Delta \tilde{\mathbf{H}}_{ju}, \quad (5.6)$$

where $\Delta \tilde{\mathbf{H}}_{ju} \in \mathcal{CN}(0, \sigma_e^2 \mathbf{I}_{M_{\text{rf}} \times N_{\text{rf}}})$.

In our paper, some basic assumptions are summarized as follows

- For channel model, the direction of angle of arrival/departure are uniformly distributed in $[0, 2\pi]$.
- UEs can perfectly estimate the DL channel using channel estimation algorithm; however the feedback CSI to TPs has error, which distribution is known.

Table 5.1: Summary of parameters in system model.

Parameter	Description
\mathbf{H}_{ju}	Channel matrix between UE u and TP j
\mathbf{F}_{bbj}	Baseband precoding matrix of TP j
\mathbf{F}_{rfj}	RF precoding matrix of TP j
\mathbf{F}_{ju}	Overall precoding matrix of TP j for UE u
\mathbf{W}_{bbu}	Baseband combining matrix of UE u
\mathbf{W}_{rfu}	RF combining matrix of UE u
$\tilde{\mathbf{H}}_{ju}$	Effective channel matrix between UE u and TP j
$\tilde{\mathbf{H}}_u^{\text{BD}}$	Concatenation of effective matrix for BD technique

- Precoding matrices are designed based on the channel observed by TPs, $\tilde{\mathbf{H}}_{ju}$.

5.2.2 Problem Formulation

Under imperfect CSI acquisition, the decoded signal at UE u is actually as

$$y'_u = \mathbf{W}_{bbu}^H \left(\tilde{\mathbf{H}}'_{ju} \mathbf{F}_{bbu} \mathbf{Q}_u^{\frac{1}{2}} s_u + \sum_{v \in \mathcal{U}, v \neq u} \tilde{\mathbf{H}}'_{ju} \mathbf{F}_{bbv} \mathbf{Q}_v^{\frac{1}{2}} s_v \right) + \mathbf{W}_{bbu}^H \mathbf{W}_{rfu}^H n_u, \quad (5.7)$$

The achievable data rate at UE u under imperfect CSI is expressed as

$$R'_u = \log_2 \left(1 + D_u^{-1} \mathbf{W}_{bbu}^H \tilde{\mathbf{H}}'_u \mathbf{F}_{bbu} \mathbf{Q}_u \mathbf{F}_{bbu}^H \tilde{\mathbf{H}}_u'^H \mathbf{W}_{bbu} \right) \quad (5.8)$$

where Q_u is the power loading coefficient for UE u , and D_u is given by

$$D_u = \sum_{v \in \mathcal{U}, v \neq u} \mathbf{W}_{\text{bb}u}^H \tilde{\mathbf{H}}_u' \mathbf{F}_{\text{bb}v} Q_v \mathbf{F}_{\text{bb}v}^H \tilde{\mathbf{H}}_u'^H \mathbf{W}_{\text{bb}u} + \mathbf{W}_{\text{bb}u}^H \mathbf{W}_{\text{rf}u}^H \mathbf{W}_{\text{rf}u} \mathbf{W}_{\text{bb}u} \sigma_n^2. \quad (5.9)$$

Since the error is unknown except for its error distribution, our main objective is to minimize the minimum power consumption while guaranteeing the expected achievable data rate meeting the minimum required QoE and power constraint. For that end, We will address the following optimization

$$\begin{aligned} \text{(P1)} : \quad & \min_{\{\mathbf{F}_{\text{bb}ju}, \mathbf{F}_{\text{rf}j}, \mathbf{W}_{\text{bb}u}, \mathbf{W}_{\text{rf}u}\}} \sum_{j \in \mathcal{J}} \sum_{u \in \mathcal{U}} \text{Tr} \left(\mathbf{F}_{\text{bb}ju} \mathbf{F}_{\text{rf}j} Q_u \mathbf{F}_{\text{rf}j}^H \mathbf{F}_{\text{bb}ju}^H \right) \\ \text{s.t.} \quad & \text{E} [R_u'] \geq R_{\min}, \\ & \sum_{u \in \mathcal{U}} \text{Tr} \left(\mathbf{F}_{\text{bb}ju} \mathbf{F}_{\text{rf}j} Q_u \mathbf{F}_{\text{rf}j}^H \mathbf{F}_{\text{bb}ju}^H \right) \leq P_{\max}, \forall j \in \mathcal{J}, \end{aligned}$$

where R_{\min} and P_{\max} is the minimum required throughput and the maximum power, respectively. The problem considers multi-connectivity transmission, which includes all TPs, to satisfy UEs in the system. However, to leverage the 5G NR feature, we consider dual connectivity and single connectivity together.

To solve optimization problem (P1), in Section III, we first work on the hybrid beamforming design for multi-connectivity transmission while considering perfect acquisition of CSI. After that, we analyze the impact of imperfect CSI on the beamforming design and formulate the expected SINR at UEs. With the expected SINR, we propose a UE classification algorithm and power allocation algorithms to satisfy QoE requirement while minimizing power consumption in Section IV.

5.3 Hybrid Beamforming Design

This section provides a hybrid beamforming design for obtaining a large antenna gain and suppress the inter-user interference for multi-cell transmissions while considering

perfect CSI.

5.3.1 Design of the Combining and Precoding RF Matrices

The objective of the RF design is to compensate for higher than expected channel losses by facilitating a large beamforming gain. First, the combining RF matrix of each UE is maximized by solving the following problem [86]

$$\max_{\mathbf{W}_{\text{rf}_u} \in \mathcal{W}} \sum_{j \in \mathcal{J}} \left\| (\mathbf{W}_{\text{rf}_u})^H \mathbf{H}_{ju} \right\|_1^2,$$

where \mathcal{W} is a discrete fourier transform matrix of dimension $M_a \times M_a$, in which the spatial frequency is $2\pi d/\lambda \sin \theta$. Given \mathbf{W}_{rf_u} , we define a temporary matrix $\mathbf{H}_j^{\text{temp}} = [\mathbf{H}_{j1}^{\text{temp}T}, \dots, \mathbf{H}_{j|\mathcal{U}|}^{\text{temp}T}]^T$, where $\mathbf{H}_{ju}^{\text{temp}} = \mathbf{W}_{\text{rf}_u} \mathbf{H}_{ju}$. Then, \mathbf{F}_{rf_j} is computed based on an equal gain transmission algorithm [86]

$$\mathbf{F}_{\text{rf}_j} = \frac{1}{\sqrt{N_a}} e^{i\xi_{ju}},$$

where ξ_{ju} is the phase of the (ju) -th element of $\mathbf{H}_j^{\text{temp}H}$.

5.3.2 Joint Design of the Baseband Combining and the Precoding Matrices

To suppress the inter-user interference on dual-connectivity downlink under perfect CSI, we apply the block diagonalization (BD) technique to obtain the precoding matrices, \mathbf{F}_{bb} , at TPs. To leverage the low-dimensional BD processing, we consider the effective channel matrices. Via (5.9), we aim at designing \mathbf{F}_{bb} such that $\tilde{\mathbf{H}}_u \mathbf{F}_{\text{bb}_v} = 0, \forall u \neq v$. We define a concatenating matrix for BD of the channel of UE u as

$$\tilde{\mathbf{H}}_u^{\text{BD}} = [\tilde{\mathbf{H}}_1^T, \dots, \tilde{\mathbf{H}}_{u-1}^T, \tilde{\mathbf{H}}_{u+1}^T, \dots, \tilde{\mathbf{H}}_{|\mathcal{U}|}^T]^T. \quad (5.10)$$

The fundamentals of BD are to obtain the \mathbf{F}_{bb_u} that transform the channel matrix of UE u , $\tilde{\mathbf{H}}_u$, on to the null space of the hyper-plan containing the channel matrices of the other UEs, $\tilde{\mathbf{H}}_u^{\text{BD}}$. By the singular value decomposition (SVD), we can obtain \mathbf{F}_{bb_u} as follows. We define the SVD of $\tilde{\mathbf{H}}_u^{\text{BD}}$ as

$$\tilde{\mathbf{H}}_u^{\text{BD}} = \bar{\mathbf{U}}_u^1 \bar{\boldsymbol{\Sigma}}_u^1 \bar{\mathbf{V}}_u^{1H} + \bar{\mathbf{U}}_u^0 \bar{\boldsymbol{\Sigma}}_u^0 \bar{\mathbf{V}}_u^{0H}, \quad (5.11)$$

where $\bar{\mathbf{V}}_u^1 \in \mathcal{C}^{2|\mathcal{U}|M_{\text{rf}} \times (|\mathcal{U}|-1)M_{\text{rf}}}$ contains the first $(|\mathcal{U}|-1)M_{\text{rf}}$ right singular vectors of $\tilde{\mathbf{H}}_u^{\text{BD}}$, and $\bar{\mathbf{V}}_u^0$ includes the last $2|\mathcal{U}|M_{\text{rf}} - (|\mathcal{U}|-1)M_{\text{rf}}$ right singular vectors that form the null basis of the null space of $\tilde{\mathbf{H}}_u^{\text{BD}}$. The rank of $\tilde{\mathbf{H}}_u^{\text{BD}}$ implicitly satisfies the condition $2|\mathcal{U}|M_{\text{rf}} > (|\mathcal{U}|-1)M_{\text{rf}}$.

By utilizing $\bar{\mathbf{V}}_u^0$ as part of precoding matrix, the inter-user interference is eliminated. So, we can rewrite (5.4) as

$$y_u = \mathbf{W}_{\text{bb}_u}^H \tilde{\mathbf{H}}_u \mathbf{F}_{\text{bb}_u} \mathbf{Q}_u^{\frac{1}{2}} s_u + \mathbf{W}_{\text{bb}_u}^H \mathbf{W}_{\text{rf}_u}^H n_u.$$

To design the combining and full precoding matrices, we perform SVD as follows

$$\tilde{\mathbf{H}}_u \bar{\mathbf{V}}_u^0 = \hat{\mathbf{U}}_u^1 \hat{\boldsymbol{\Sigma}}_u^1 \hat{\mathbf{V}}_u^{1H} + \hat{\mathbf{U}}_u^0 \hat{\boldsymbol{\Sigma}}_u^0 \hat{\mathbf{V}}_u^{0H}. \quad (5.12)$$

The optimal precoding and combining baseband matrices of \mathbf{F}_{bb_u} and \mathbf{W}_{bb_u} are $\bar{\mathbf{V}}_u^0 \left(\hat{\mathbf{V}}_u^1 \right)_{:1}$ and $\left(\hat{\mathbf{U}}_u^1 \right)_{:1}$, respectively, since we consider single stream transmission scheme. With the obtained precoding and combining baseband matrices, the inter-user interference is eliminated. So, the decoded signal in (5.4) is simplified as

$$y_u = \left(\hat{\boldsymbol{\Sigma}}_u^1 \right)_{11} \mathbf{Q}_u^{\frac{1}{2}} s_u + \mathbf{W}_{\text{bb}_u}^H \mathbf{W}_{\text{rf}_u}^H n_u,$$

where $\left(\hat{\boldsymbol{\Sigma}}_u^1 \right)_{11}$ is the first, and also the largest, singular value in diagonal matrix of non-zero singular $\hat{\boldsymbol{\Sigma}}_u^1$. Thus, the achievable data rate in (5.8) under perfect CSI is rewritten as

$$R_u = \log_2 \left(1 + \frac{\left(\hat{\boldsymbol{\Sigma}}_u^1 \right)_{11}^2 Q_u}{\sigma_n^2} \right).$$

Since UEs can be served by single TP, we can apply the BD technique for single-connectivity transmission. UEs can select the best TP based on reference signal received power (RSRP). We define the single-connectivity set served by TP j as $\mathcal{U}_j^{\text{SC}}$. For UE $u \in \mathcal{U}_j^{\text{SC}}$, we apply the similar procedure (5.10), (5.11), and (5.12) by concatenating $\tilde{\mathbf{H}}_{ju}$ in (5.5) for TP j only. Then, we can get $\mathbf{F}_{\text{bb}ju}$, $\mathbf{W}_{\text{bb}j}$, and $\left(\hat{\boldsymbol{\Sigma}}_{ju}^1\right)_{11}$, which are, respectively, the baseband precoding vector of TP j for UE k , the baseband combining vector of UE u for TP j , and the largest singular value of the channel between UE u and TP j . In the next section, we provide an analysis for the impact of imperfect CSI on beamforming and network performance.

5.3.3 Impact of imperfect CSI on Beamforming and Network Performance

Because the CSI is imperfect, the inter-user interference remains. The decoded signal in (5.7) is equivalent to the following

$$y'_u = \mathbf{W}_{\text{bb}u}^H \left(\tilde{\mathbf{H}}_u + \Delta \tilde{\mathbf{H}}_u \right) \bar{\mathbf{V}}_u^0 \left(\hat{\mathbf{V}}_u^1 \right)_{:1} Q_u^{\frac{1}{2}} s_u + \mathbf{W}_{\text{bb}u}^H \sum_{v \in \mathcal{U}, v \neq u} \left(\tilde{\mathbf{H}}_u + \Delta \tilde{\mathbf{H}}_u \right) \bar{\mathbf{V}}_v^0 \left(\hat{\mathbf{V}}_v^1 \right)_{:1} Q_v^{\frac{1}{2}} s_v + \mathbf{W}_{\text{bb}u}^H \mathbf{W}_{\text{rf}u}^H n_u. \quad (5.13)$$

To analyze the impact of imperfect CSI on beamforming design, we express the SVD of $\tilde{\mathbf{H}}'_u \bar{\mathbf{V}}_u^0$ as

$$\tilde{\mathbf{H}}'_{ju} \bar{\mathbf{V}}_u^0 = \left(\tilde{\mathbf{H}}_u + \Delta \tilde{\mathbf{H}}_u \right) \bar{\mathbf{V}}_u^0 = \left(\hat{\mathbf{U}}_u^1 + \Delta \hat{\mathbf{U}}_u^1 \right) \left(\hat{\boldsymbol{\Sigma}}_u^1 + \Delta \hat{\boldsymbol{\Sigma}}_u^1 \right) \left(\hat{\mathbf{V}}_u^{1H} + \Delta \hat{\mathbf{V}}_u^{1H} \right), \quad (5.14)$$

where $\Delta \hat{\mathbf{U}}_u^1$, $\Delta \hat{\boldsymbol{\Sigma}}_u^1$, and $\Delta \hat{\mathbf{V}}_u^1$ are the errors in the combiner vector, singular matrix, and precoding vector, respectively.

Lemma 1 *The error in precoding for UE u , $\Delta \hat{\mathbf{V}}_u^1$, is expressed as*

$$\Delta \hat{\mathbf{V}}_u^1 \approx \hat{\mathbf{V}}_u^1 \mathbf{T}_u + \hat{\mathbf{V}}_u^0 \bar{\mathbf{V}}_u^{0H} \Delta \tilde{\mathbf{H}}_u^H \hat{\mathbf{U}}_u^1 \hat{\boldsymbol{\Sigma}}_u^{1^{-1}}, \quad (5.15)$$

where \mathbf{T}_u is computed as

$$\mathbf{T}_u = \Psi_u \odot \left(\widehat{\mathbf{V}}_u^{1H} \overline{\mathbf{V}}_u^{0H} \Delta \widetilde{\mathbf{H}}_u^H \widehat{\mathbf{U}}_u^1 \widehat{\Sigma}_u^1 + \widehat{\Sigma}_u^1 \widehat{\mathbf{U}}_u^{1H} \Delta \widetilde{\mathbf{H}}_u^H \overline{\mathbf{V}}_u^0 \widehat{\mathbf{V}}_u^1 \right), \quad (5.16)$$

and the elements of matrix Ψ_u are given as

$$(\Psi_u)_{mn} = \left(\left(\widehat{\Sigma}_u^1 \right)_{mm}^2 - \left(\widehat{\Sigma}_u^1 \right)_{nn}^2 \right)^{-1} \delta_{mn}. \quad (5.17)$$

Proof: See Appendix B.1.

Lemma 2 The error singular matrix of UE u , $\Delta \widehat{\Sigma}_u^1$, has the diagonal elements

$$\left(\Delta \widehat{\Sigma}_u^1 \right)_{mm} = \frac{1}{2} \left(\left(\widehat{\mathbf{U}}_u^1 \right)_{:m}^H \Delta \widetilde{\mathbf{H}}_u \overline{\mathbf{V}}_u^0 \left(\widehat{\mathbf{V}}_u^1 \right)_{:m} + \left(\widehat{\mathbf{V}}_u^1 \right)_{:m}^H \overline{\mathbf{V}}_u^{0H} \Delta \widetilde{\mathbf{H}}_u^H \left(\widehat{\mathbf{U}}_u^1 \right)_{:m} \right).$$

Proof: See Appendix B.2.

Theorem 1 The expected SINR at UE u , $E[\Gamma_u]$, is approximated as

$$E[\Gamma_u] \approx \frac{\left(\left(\widehat{\Sigma}_u^1 \right)_{11}^2 + \frac{1}{2} \sigma_e^2 \right) Q_u}{\sigma_e^2 \sum_{v \in \mathcal{U}, v \neq u} Q_v + \sigma_n^2}. \quad (5.18)$$

Proof: See Appendix B.3.

The expected achievable data rate at UE u under imperfect CSI is given as

$$E[R_u'] \approx \log_2 \left(1 + \frac{\left(\left(\widehat{\Sigma}_u^1 \right)_{11}^2 + \frac{1}{2} \sigma_e^2 \right) Q_u}{\sigma_e^2 \sum_{v \in \mathcal{U}, v \neq u} Q_v + \sigma_n^2} \right).$$

5.4 Power Allocation Algorithm Under Imperfect CSI

This section describes the proposed algorithm under imperfect CSI for guaranteeing QoE via dual connectivity while minimizing the power consumption. The algorithm is shown in Algorithm 12. First, UEs are classified based on which UEs need dual connectivity to be satisfied, based on channel quality. Then, with the classified UEs, power is allocated to fulfill QoE demands of the dual-connectivity UEs, and the remaining power is for satisfying the single-connectivity UEs.

Algorithm 12 Adaptive Dual-Connectivity Power Allocation Algorithm

- 1: Classify UEs into single-connectivity UEs and dual-connectivity UEs (Algorithm 2)
 - 2: Allocate power to satisfy as many dual-connectivity UEs as possible (Algorithm 3)
 - 3: Allocate the remaining power to satisfy single-connectivity UEs
-

5.4.1 UE Classification Algorithm

This algorithm classifies UEs based on the minimum power to meet QoE requirement, which is determined by (5.18). The algorithm examines whether single-connectivity is enough for all UEs to meet the required QoE. If single-connectivity transmission can guarantee QoE for all UEs, power is immediately allocated to all UEs by solving (5.20).

For each TP j , the problem is to compute the minimum power for satisfying single-connectivity UE of list $\mathcal{U}_j^{\text{SC}}$, such that the following condition is fulfilled

$$E[\Gamma_{ju}] \geq \Gamma_u^{\min} \quad \forall u \in \mathcal{U}_j^{\text{SC}}. \quad (5.19)$$

If inter-user interference is eliminated, the problem is solved trivially [72]. Due to the existence of inter-user interference, this problem becomes nontrivial and is solved by applying the Pareto optimal approach [107]. The optimal power loading vector of TP j for satisfying UE in $\mathcal{U}_j^{\text{SC}}$, $\mathbf{Q}_j^{\text{SC}*} = (Q_{j1}^{\text{SC}*}, \dots, Q_{j|\mathcal{K}_j|}^{\text{SC}*})^T$, is given as follows

$$\mathbf{Q}_j^{\text{SC}*} = (\mathbf{I} - \mathbf{G}_j^{\text{SC}})^{-1} \mathbf{B}_j^{\text{SC}} \quad (5.20)$$

where \mathbf{G}_j^{SC} is a matrix with

$$(\mathbf{G}_j^{\text{SC}})_{kl} = \begin{cases} 0 & l = k \\ \frac{\sigma_e^2 \Gamma_k^{\min}}{(\hat{\Sigma}_{jk}^1)_{11}^2 + \frac{\sigma_e^2}{2}} & l \neq k \end{cases} \quad (5.21)$$

and $\mathbf{B}_j^{\text{SC}} = (b_{j1}, \dots, b_{j|\mathcal{U}_j^{\text{SC}}|})^T$ is a vector with

$$b_{jk} = \frac{\sigma_n^2 \Gamma_k^{\min}}{(\hat{\Sigma}_{jk}^1)_{11}^2 + \frac{\sigma_e^2}{2}}.$$

Due to the channel condition and power limitation, all UEs may not be satisfied by single connections. So, we need to find the optimal single-connectivity list for TP j . To get the optimal single-connectivity list for TP j , the algorithm sorts the UEs in ascending order according to the minimum required power, and obtained sorted list $\mathcal{U}_j^{\text{SC}*}$. Then, the algorithm iteratively removes the last UE in $\mathcal{U}_j^{\text{SC}*}$, such that the minimum total required power does not exceed P_{\max}

$$\sum_{u=1}^{|\mathcal{U}_j^{\text{SC}*}|} P_{ju}^{\text{SC}*} = \sum_{u=1}^{|\mathcal{U}_j^{\text{SC}*}|} \|\mathbf{F}_{ju}\|_2^2 Q_{ju}^{\text{SC}*} \leq P_{\max}.$$

For UEs that cannot be satisfied by single connections, the algorithm stores them in a dual-connectivity UE set, denoted as $\mathcal{U}_{\text{total}}^{\text{DC}}$. The next step is categorizing the dual-connectivity UE according to pairs of TPs because dual-connectivity UEs can be served by different pairs of TPs.

Each dual-connectivity UE $u \in \mathcal{U}_{\text{total}}^{\text{DC}}$ chooses the second TP for dual connectivity, which has the second-largest RSRP in the system. Then, dual-connectivity UEs are grouped according pairs of TPs, as depicted in Fig. 5.1. Dual-connectivity UE set $\mathcal{U}_{\{1,2\}}^{\text{DC}}$ including UE 1 and UE 2 are served by TP set $\{1,2\}$ consisting of TP 1 and TP 2. Likewise, dual-connectivity UE set $\mathcal{U}_{\{1,3\}}^{\text{DC}}$ including only UE 4 is served by TP set $\{1,3\}$. We summarize the algorithm for user classification in Algorithm 13. In the next subsection, we propose a power allocation algorithm for dual-connectivity UEs in order to meet the required QoE.

5.4.2 Power Allocation for Dual-Connectivity UEs

To meet the required QoE while minimizing the power consumption, we allocate power to dual-connectivity UEs to meet QoE with a power constraint, and the remaining power is for satisfying network capacity with single-connectivity UEs. First, we compute the sufficient power for each of the dual-connectivity UEs based on their channel quality and

Algorithm 13 User Classification Algorithm

- 1: Input: $\mathbf{F}_{ju}, \bar{\mathbf{S}}_{ju}, j = 1, \dots, N_{\text{TP}}, u = 1, \dots, |\mathcal{U}_j^{\text{SC}}|$
 - 2: Output: $\mathcal{U}_n^{\text{DC}}$ s and corresponding $\mathcal{J}_n^{\text{DC}}$ s, $\mathcal{U}_j^{\text{SC}*}, j = 1, \dots, N_{\text{TP}}$
 - 3: **for** each TP j **do**
 - 4: Compute the optimal power for satisfying UEs with single-connectivity scheme, \mathbf{Q}_j^* , by (5.20)
 - 5: Sort single-connectivity UE list $\mathcal{U}_j^{\text{SC}}$ in ascending order based on minimum needed power, \mathbf{Q}_{ju}^* , achieving sorted list $\mathcal{U}_j^{\text{SC}*}$
 - 6: **while** $\sum_{u=1}^{|\mathcal{U}_j^{\text{SC}*}|} \|\mathbf{F}_{ju}\|_2^2 Q_{ju}^{\text{SC}*} > P_{\text{max}}$ **do**
 - 7: Remove the last UE in sorted list $\mathcal{U}_j^{\text{SC}*}$ and store it in dual-connectivity UE set $\mathcal{U}_{\text{total}}^{\text{DC}}$
 - 8: **end while**
 - 9: **end for**
 - 10: **for** each UE in dual-connectivity UE set $\mathcal{U}_{\text{total}}^{\text{DC}}$ **do**
 - 11: Find the second-best TP based on RSRP
 - 12: **end for**
 - 13: In $\mathcal{U}_{\text{total}}^{\text{DC}}$, group dual-connectivity UEs that have common TPs in order to get \mathcal{U}_n s and the corresponding \mathcal{J}_n s.
-

QoE requirement. The optimal power loading vector for satisfying dual-connectivity UEs served by TP i and TP j , $\mathbf{Q}_{\{ij\}}^{\text{DC}^*} = \left(Q_{\{ij\}1}^{\text{DC}^*}, \dots, Q_{\{ij\}|\mathcal{U}_{\{ij\}}^{\text{DC}}|}^{\text{DC}^*} \right)^T$, is given as

$$\mathbf{Q}_{\{ij\}}^{\text{DC}^*} = \left(\mathbf{I} - \mathbf{G}_{\{ij\}}^{\text{DC}} \right)^{-1} \mathbf{B}_{\{ij\}}^{\text{DC}}. \quad (5.22)$$

where \mathbf{G}_j^{SC} is a matrix with

$$\left(\mathbf{G}_{\{ij\}}^{\text{DC}} \right)_{kl} = \begin{cases} 0 & l = k \\ \frac{\sigma_e^2 \Gamma_k^{\min}}{\left(\widehat{\Sigma}_k^1 \right)_{11} + \frac{\sigma_e^2}{2}} & l \neq k \end{cases},$$

and $\mathbf{B}_{\{ij\}}^{\text{DC}} = \left(b_{\{ij\}1}, \dots, b_{\{ij\}|\mathcal{U}_{\{ij\}}^{\text{DC}}|} \right)^T$ is a vector with

$$b_{\{ij\}k} = \frac{\sigma_n^2 \Gamma_k^{\min}}{\left(\widehat{\Sigma}_{\{ij\}k}^1 \right)_{11} + \frac{\sigma_e^2}{2}}.$$

Then, for each TP in set $\{i, j\}$ that serves dual-connectivity UE u , the minimum power assigned by TP j for dual-connectivity UE u is

$$P_{ju}^{\text{DC}^*} = \|\mathbf{F}_{ju}\|_2^2 Q_{\{ij\}u}^{\text{DC}^*}. \quad (5.23)$$

Since all dual-connectivity UEs may not be served at once due to the power limitation, we choose an optimal subset of dual-connectivity UEs from $\mathcal{U}_{\text{total}}^{\text{DC}}$, called $\mathcal{U}_{\text{total}}^{\text{DC}^*}$. The objective is to satisfy as many as dual-connectivity UEs as possible. To obtain $\mathcal{U}_{\text{total}}^{\text{DC}^*}$, we solve the following optimization problem

$$\begin{aligned} \text{(P2)} : \max & \quad \left| \mathcal{U}_{\text{total}}^{\text{DC}^*} \right| \\ \text{s.t.} & \quad \sum_{u \in \mathcal{U}_{\text{total}}^{\text{DC}^*}} \|\mathbf{F}_{ju}\|_2^2 Q_{\{ij\}u}^{\text{DC}^*} \leq P_{\text{DC}j}^{\max}, \quad \forall i, j \in \mathcal{J} \\ & \quad Q_u \geq 0, \quad \forall u \in \mathcal{U}_{\text{total}}^{\text{DC}^*}, \end{aligned}$$

where $P_{\text{DC}j}^{\max}$ is the maximum power for dual-connectivity transmission at TP j , which is calculated as

$$P_{ju}^{\text{DC}^{\max}} = P_{\max} - \sum_{u=1}^{|\mathcal{U}_j^{\text{SC}^*}|} P_{ju}^{\text{SC}^*}. \quad (5.24)$$

We note that $P_{ju}^{\text{DC}^*} = \|\mathbf{F}_{ju}\|^2 Q_{\{ij\}u}^{\text{DC}^*} = 0$ if TP i or TP j does not serve UE u .

Algorithm 14 Power Allocation Algorithm for Dual Connectivity UEs

- 1: **for** each UE $u \in \mathcal{U}$ **do**
 - 2: Compute $P_{ju}^{\text{DC}^*}$ by (5.23) for each TP j serving UE u
 - 3: **end for**
 - 4: Set $\mathcal{U}_{\text{total}}^{\text{DC}^*} = \mathcal{U}_{\text{total}}^{\text{DC}}$
 - 5: **for** each TP j in the system **do**
 - 6: Compute $P_{ju}^{\text{DC}^{\max}}$ via (5.24)
 - 7: Sort dual-connectivity UEs that are served by TP j based on their minimum required power for dual connectivity, $P_{ju}^{\text{DC}^*}$
 - 8: Compute the total power for connectivity, $P_{\text{DC}j}$
 - 9: **while** $P_{\text{DC}j} > P_{\text{DC}j}^{\max}$ **do**
 - 10: Remove the last UE of the sorted list UE served by TP j and also remove the UE from $\mathcal{U}_{\text{total}}^{\text{DC}^*}$
 - 11: **end while**
 - 12: **end for**
-

We obtain $\mathcal{U}_{\text{total}}^{\text{DC}^*}$ as follows. Initially, $\mathcal{U}_{\text{total}}^{\text{DC}^*} = \mathcal{U}_{\text{total}}^{\text{DC}}$; then, a list of the dual-connectivity UEs served by TP j is sorted in ascending order based on $P_{ju}^{\text{DC}^*}$. After that, the algorithm iteratively removes the last UE from the sorted list, and also extracts it from $\mathcal{U}_{\text{total}}^{\text{DC}^*}$ until the total power for dual connectivity at TP j does not exceed $P_{\text{DC}j}^{\max}$, which is

$$P_{\text{DC}j} = \sum_{u \in \mathcal{U}_{\text{total}}^{\text{DC}^*}} P_{ju}^{\text{DC}^*} \leq P_{\text{DC}j}^{\max} \quad (5.25)$$

The process is applied to all TPs that support dual-connectivity UEs.

After $\mathcal{U}_{\text{total}}^{\text{DC}^*}$ has been achieved, dual-connectivity UEs in $\mathcal{U}_{\text{total}}^{\text{DC}^*}$ will be satisfied by allocating the minimum required power from the corresponding pairs of TPs. The total

power spent for dual connectivity at TP j , P_{DC_j} , is computed as

$$P_{\text{DC}_j} = \sum_{u \in \mathcal{U}_{\text{total}}^{\text{DC}^*}} P_{ju}^{\text{DC}^*}. \quad (5.26)$$

The algorithm for obtaining $\mathcal{U}_{\text{total}}^{\text{DC}^*}$ is summarized in Algorithm 14. The remaining power after allocating power for dual-connectivity UEs will be allocated to single-connectivity UE at each TP based on (5.20).

5.4.3 Complexity Analysis

In this section, we provide a complexity analysis for the algorithm. Specifically, the computational complexity of the problems are given in the following table.

5.5 Simulation Results and Analysis

In this section, we show the numerical results to evaluate the performance of the proposed adaptive-dual connectivity (ADC) power control. For comparison, the base lines are the single-connectivity (SC) scheme [86] and dual-connectivity scheme (DC), in which all UEs are supported by dual connectivity.

5.5.1 Simulation Setup

The simulation environment includes seven TPs, i.e. $N_{\text{TP}} = 7$. TPs are deployed regularly at an inter-site distance of 50 m [85]. The maximum transmission power of a TP is 30 dBm. The system operates on a carrier frequency of 28 GHz with a 100 MHz bandwidth [85]. The noise figure at a receiver is 10 dB. Hybrid beamforming is all used for all considered algorithms. For beamforming parameters, $N_{\text{a}} = 256$, $N_{\text{rf}} = 48$, $M_{\text{a}} = 8$, and $M_{\text{rf}} = 4$, as provided in [85]. For channel realization in (5.2), we set N_{C} and N_{P} as 3 and 2, respectively. Large-scale path loss for the wireless channel between TP j and UE u is given

Table 5.2: Complexity analysis.

Elements	Complexity
\mathbf{W}_{rf}	$\mathcal{O}(M_a^2)$
\mathbf{F}_{rf}	$\mathcal{O}(M_a N_a M_{\text{rf}})$
$\bar{\mathbf{V}}_u^0$	$\mathcal{O}\left(\mathcal{U}_j^{\text{SC}} \left(\mathcal{U}_j^{\text{SC}} - 1\right)^2 M_{\text{rf}}^3\right)$ with $u \in \mathcal{U}_j^{\text{SC}}$ $\mathcal{O}\left(2 \mathcal{U}_j^{\text{SC}} \left(\mathcal{U}_j^{\text{DC}} - 1\right)^2 M_{\text{rf}}^3\right)$ with $u \in \mathcal{U}^{\text{DC}}$
$\mathbf{F}_{\text{bb}u}$	$\mathcal{O}\left(N_{\text{rf}} M_{\text{rf}}^2\right)$ with $u \in \mathcal{U}^{\text{SC}}$ $\mathcal{O}\left(2N_{\text{rf}} M_{\text{rf}}^2\right)$ with $u \in \mathcal{U}^{\text{DC}}$
\mathbf{Q}_j^{SC}	$\mathcal{O}\left(\mathcal{U}_j^{\text{SC}} ^3\right)$
\mathbf{Q}_j^{DC}	$\mathcal{O}\left(\mathcal{U}_j^{\text{DC}} ^3\right)$

as $32.4 + 20 * \log_{10}(d) + 20\log_{10}(f_c) + \Delta_{\text{shadow}}$ [88], where d is the Euclidean distance (in metre) between the TP and the UE, $\Delta_{\text{shadow}} = 2$ dB [88] is the shadowing factor, and f_c is the carrier frequency, which is set to 28 GHz. Thus, the full channel model is The UE mobility is modeled as a random walk at 3 km/h [72]. Each simulation was performed 100 times and the results were averaged. Measurement metrics are the total transmission power and the percentage of satisfied UEs, which is the ratio between the number of UEs receiving at least the demanded QoE and the total of UEs in the network. QoE is is measured by the actual throughput at each UE, and the throughput is calculated based on Shannon capacity. For all UEs, we set identical QoE threshold, and the value is varied (e.g., 100 Mbps, 200 Mbps, 300 Mbps, and 400 Mbps) for performance evaluation.

5.5.2 Impact of CSI error on connectivity

To verify the adaptation to varying channel quality of our proposed algorithm, we changed the variance of CSI error value, σ_e^2 , in (5.6) from 0.01 to 0.1 [108]. Fig. 5.2 provides the percentage of dual-connectivity UEs under different error variance values. As shown in the figure, the number of connectivity tends to increase when a CSI error rises. This trend is explained by (5.18) showing that interference can increase due to the CSI error. And, to meet the QoE requirement under varying channel quality, UEs may need more connectivity supported by the network. The proposed algorithm classifies the UEs based on the expected SINR and UE QoE requirements to provide necessary resource. Additionally, we also witness an increment of dual-connectivity UEs when QoE demand is higher. Due to varying channel quality and power limitation, one wireless connection may not be sufficient to satisfy some UEs, hence, dual connectivity should be involved.

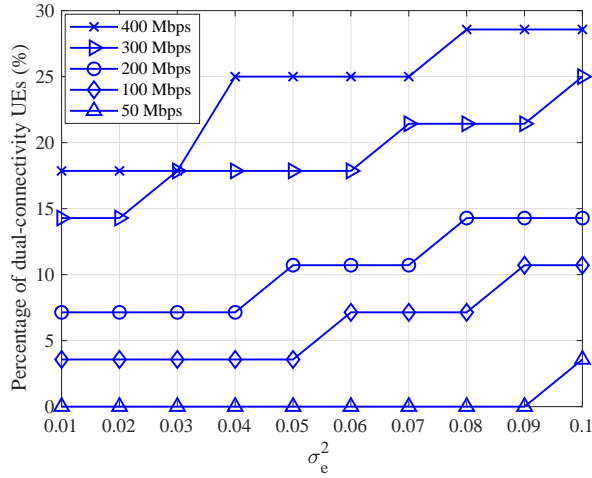
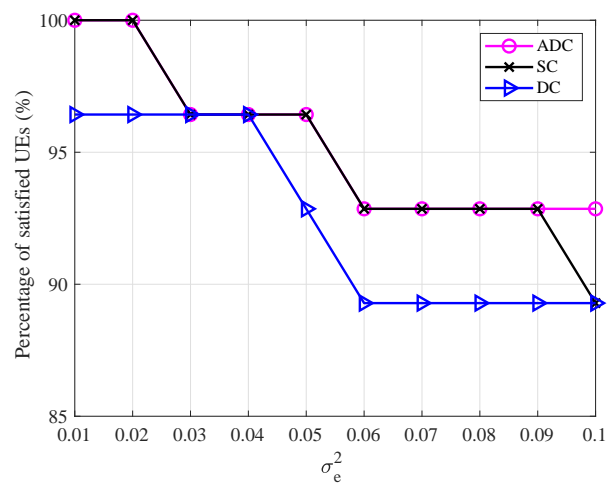


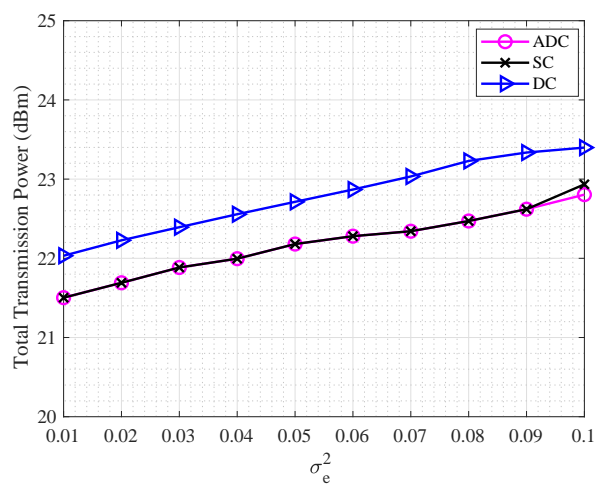
Figure 5.2: Percentage of dual-connectivity UEs versus QoE threshold and CSI error.

5.5.3 Impact of CSI error on network performance

In this section, we evaluated the effect of CSI error on UE satisfaction rate and the total spent power of the system. With the QoE target of 50 Mbps, the results are plotted in Fig. 5.3a and 5.3b. In Fig. 5.3a, ADC satisfies all UEs when CSI error, σ_e^2 , is smaller than 0.02. When σ_e^2 reaches 0.1, ADC satisfies around 93% UEs. The SC algorithm performs similar to ADC except for when σ_e^2 increases from 0.09 to 0.1, where SC falls behind ADC to around 90%. This is because ADC found some UEs that needs more than one connection due to the increasing interference and changing channel quality. This situation is shown in Fig. 5.2 for QoE of 50 Mbps that an increment of σ_e^2 increases dual-connectivity UEs. The DC algorithm satisfies less UEs than ADC and SC algorithm, however, it consumes more power than the ADC and SC, as shown in Fig. 5.3b. This is because UE only needs dual connectivity to meet the required QoE depending on its channel condition [105]. Since ADC leverages the channel quality to classify the UEs, it efficiently allocates power to meet the required QoE.

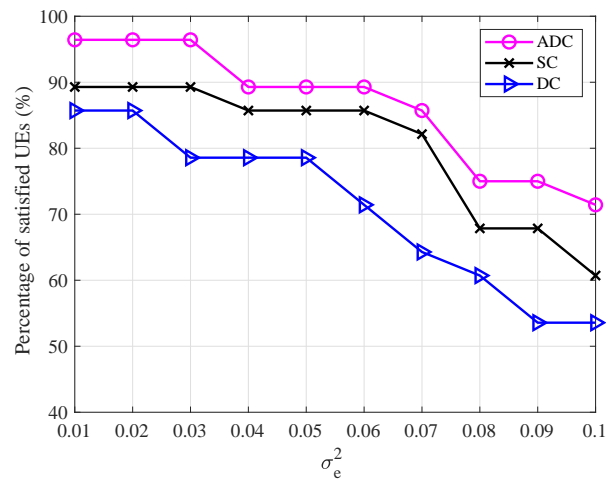


(a) Percentage of satisfied UEs.

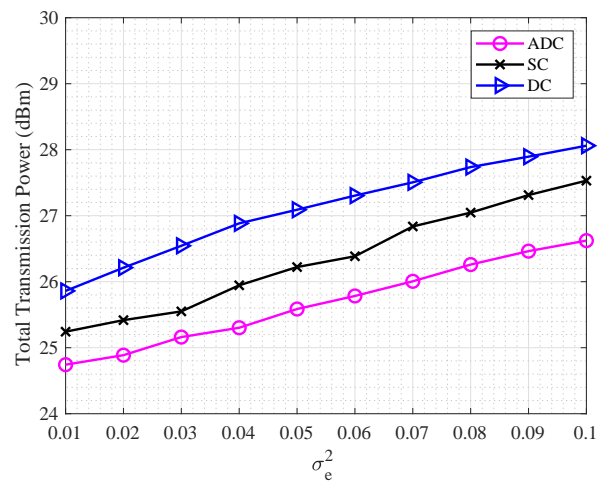


(b) Total transmission power .

Figure 5.3: Network performance versus CSI Error for QoE of 50 Mbps.

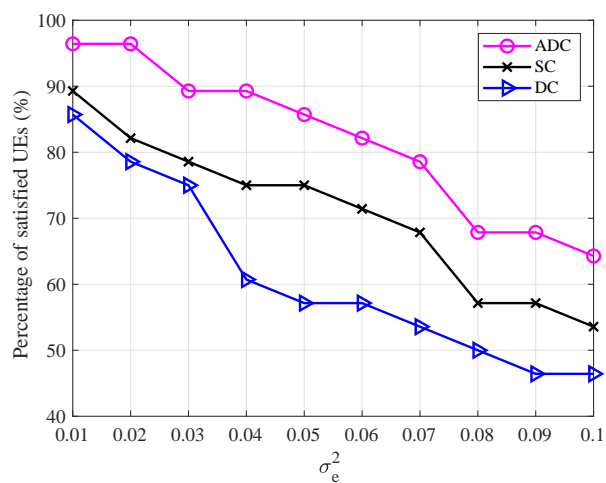


(a) Percentage of satisfied UEs.

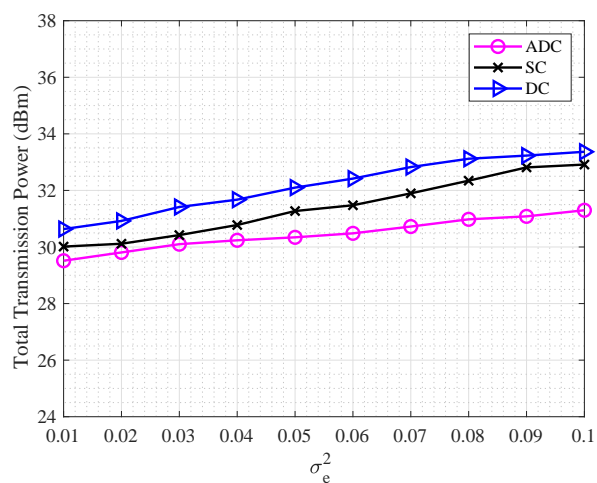


(b) Total transmission power .

Figure 5.4: Network performance versus CSI Error for QoE of 100 Mbps.



(a) Percentage of satisfied UEs.



(b) Total transmission power .

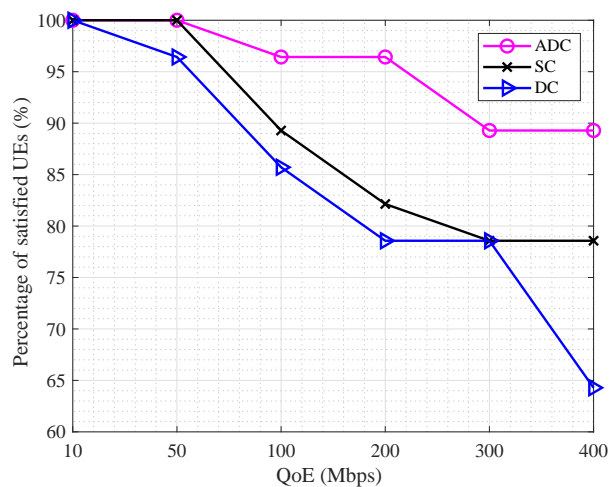
Figure 5.5: Network performance versus CSI Error for QoE of 200 Mbps.

We investigate the impact of the CSI error with higher QoE targets, such as 100 Mbps, as shown in Fig. 5.4. It is shown that the ADC algorithm satisfies more UEs while spending less power than its counterparts. For QoE of 100 Mbps, at a low CSI error level ($\sigma_e^2 = 0.01$), the ADC algorithm can satisfy around 96% UEs while the SC and DC algorithms just accomplish around 90% and 86%, respectively. Regarding the total transmission power, the ADC algorithm spends only around 24.7 dBm, however, ADC and SC algorithms use around 27.5 and 28 dBm, respectively. When σ_e^2 rises to 0.1, ADC algorithm can satisfy up to approximately 72% UEs, which outperforms SC (around 61%) and DC (around 54%) algorithms. In context of power consumption, ADC needs only around 26.6 dBm, which is greener than SC's (around 27.5 dBm) and DC's (around 28 dBm).

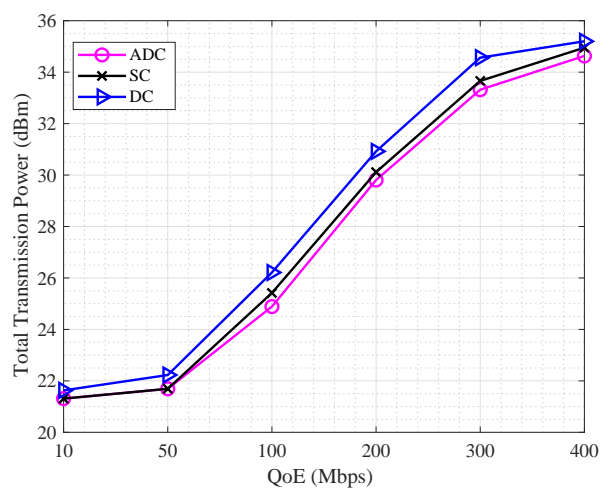
Increasing QoE demand to 200 Mbps, we plot the results in Fig. 5.5. It is observed that ADC algorithm improved the UE satisfaction rate more than the others, even though the CSI error affects the performance more, as compared to the results of 100 Mbps QoE target. ADC algorithm can afford QoE demand of around 96% UEs when $\sigma_e^2 \leq 0.02$, and around 64% when σ_e^2 increases to 0.1. Looking at Fig. 5.2, to guarantee the requirement, ADC algorithm chose around 7% UEs and 14% for dual connectivity when $\sigma_e^2 = 0.01$ and $\sigma_e^2 = 0.1$, respectively. Thus, the performance of ADC algorithm is better than the baseline algorithms. Furthermore, by effectively selecting UE for dual connectivity rather than fixed-connectivity scheme, ADC spent less power, around 0.5 dB and 1.1 dB than SC and DC algorithms, respectively, when $\sigma_e^2 = 0.01$; and, around 1.6 dB and 2.1 dB than SC and DC algorithms, respectively, when $\sigma_e^2 = 0.1$.

5.5.4 Impact of QoE requirements

In this section, we scrutinized the effectiveness of the algorithms with various QoE targets, where we set σ_e^2 to 0.02. Fig. 5.6a and 5.6b presents the UE satisfaction rate



(a) Percentage of satisfied UEs.



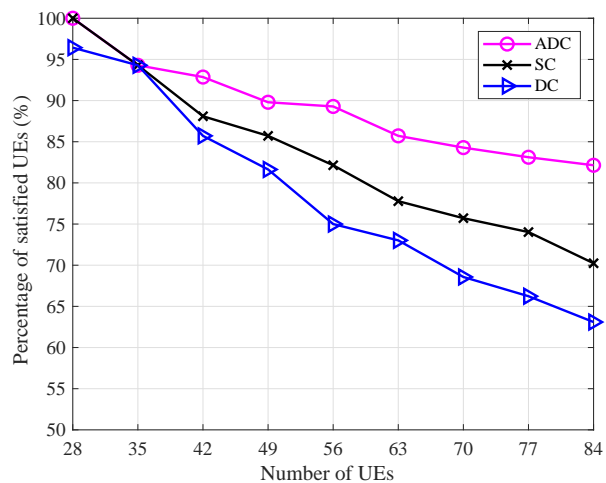
(b) Total transmission power .

Figure 5.6: Network performance with various QoE requirements and $\sigma_e^2 = 0.02$.

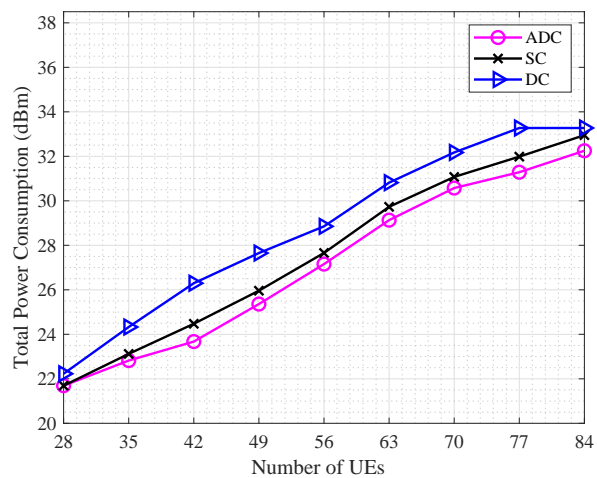
and total transmission power, respectively. We can observe from Fig. 5.6a that ADC algorithm satisfies more UEs than SC and DC algorithms. Starting from 50 Mbps, where ADC and SC algorithms satisfy all the UEs and DC algorithm pleases around 96% UEs, the difference between ADC and SC/DC significantly enlarges when QoE goal extends beyond 100 Mbps. When it reaches 400 Mbps, ADC can still offer around 89% UEs the QoE requirements, however SC and DC can only afford the requirement for around 77% and 64% UEs, respectively. In Fig. 5.6b, although the total transmission power grows for all the algorithms to fulfill increasing QoE demand, ADC algorithm still spends less power than SC and DC algorithm. By utilizing dual connectivity efficiently based on channel quality and estimating the impact of CSI error for power allocation, ADC algorithm can improve the UE satisfaction rate while being greener than the benchmark algorithms.

5.5.5 Impact of the number of UEs

In this section, we clarified the impact of network load in terms of the number of UEs on the performance of three algorithms. The number of UEs are varied as 28, 32, 36, 40, 44, and 48 while the $\sigma_e^2 = 0.02$ and the QoE requirement is set to 50 Mbps. Percentage of satisfied UEs and the total transmission power are depicted in Fig. 5.7a and Fig. 5.7b, respectively. One can observe from Fig. 5.7a that UE satisfaction rate decreases when the network load increases. This is caused by the growing interference due to imperfect CSI. However, the performance of the proposed algorithm is better than others. Starting from 28 UEs, where ADC, SC, and DC satisfies 100%, 100%, and around 96%, respectively. When the number of UEs rises to 84, ADC algorithm can supply around 82% of UEs with matching QoE requirement, while SC and DC algorithms obtain only around 70% and 63%, respectively. Fig. 5.7b shows that the energy consumption grows in keeping with the growth of the number of UEs to provide a pleased service. Since ADC algorithm utilized



(a) Percentage of satisfied UEs.



(b) Total transmission power .

Figure 5.7: Network performance with various number of UEs and $\sigma_e^2 = 0.02$.

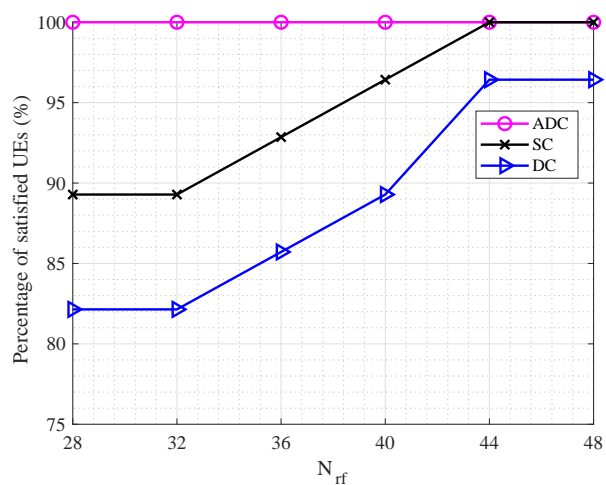
dual connectivity efficiently, it spends less energy than SC and DC algorithms.

5.5.6 Impact of the number of RF chains

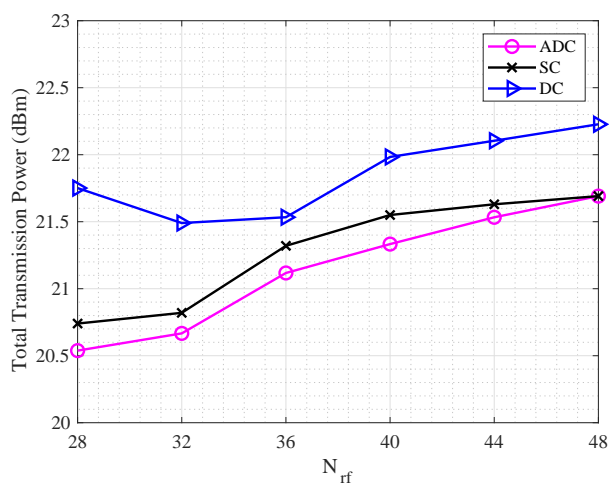
Here, we examined the impact of the number of RF chain elements at TP, N_{rf} , on the UE satisfaction rate and transmission power. We changed the number of RC chains at TP, such as 28, 32, 36, 40, 44, and 48, while setting CSI error, QoE requirement, and the number of UEs to 0.02, 50 Mbps, and 28 respectively. Fig. 5.8a shows that ADC algorithm can satisfy all of UEs for all the RF chain settings. Meanwhile, performance of SC and DC algorithms are only improved by increasing the number of RF chains (from around 90% when $N_{\text{rf}} = 28$ to 100% when $N_{\text{rf}} = 44$ for SC algorithm, and from around 82% when $N_{\text{rf}} = 28$ to around 96% when $N_{\text{rf}} = 48$ for DC algorithm). This is because increasing the RF chains can improve the spectral efficiency [109, 110]. For transmission power, we observed from Fig. 5.8b that the total transmission power tends to increase with the number of RC chains since more RC chains will consume more power. However, the ADC algorithm still saved more transmission power than the benchmarks.

5.6 Closing Remarks

In this work, we proposed a power allocation algorithm with hybrid beamforming to guarantee the required QoE while minimizing transmission power. A cooperative beamforming design was applied and the impact of imperfect CSI was analyzed in terms of estimated SINR. Based on the estimated SINR, the proposed algorithm follows two stages, UE classifications and power allocation. In cases where single connections cannot guarantee the minimum QoE for all UEs, depending on the wireless environment, for a given beamforming, UEs are classified into two groups. After that, according to the UE categories, minimum power is allocated to satisfy the demanded QoE. Simulations show that our pro-



(a) Percentage of satisfied UEs.



(b) Total transmission power .

Figure 5.8: Network performance with various number of TP's RF chains and $\sigma_e^2 = 0.02$.

posed algorithm dynamically classifies UE groups based on QoE and the channel conditions of the UEs, and satisfy the UE while minimizing transmission power. Furthermore, various aspects of a wireless network such as QoE requirements, accuracy of CSI, network load, and RF settings are investigated.

This page intentionally left blank.

Publications

International journal

- [1] M. T. Nguyen and S. Kwon, “Geometry-Based Analysis of Optimal Handover Parameters for Self-Organizing Networks,” in *IEEE Transactions on Wireless Communications*, , vol. 19, no. 4, pp. 2670-2683, April 2020, doi: 10.1109/TWC.2020.2967668
- [2] M. T. Nguyen, S. Kwon and H. Kim, “Mobility Robustness Optimization for Handover Failure Reduction in LTE Small-Cell Networks,” in *IEEE Transactions on Vehicular Technology*, vol. 67, no. 5, pp. 4672-4676, May 2018, doi: 10.1109/TVT.2017.2787602.

Journal papers under Preparation

- [3] M. T. Nguyen and S. Kwon, “Transfer Learning for Handover Optimization in Self-Organizing Networks.”
- [4] M. T. Nguyen, S. Kwon, and J. Song, “Joint Power Allocation and Beam-forming Design for Dual-connectivity Wireless Networks.”

International conference

- [5] M. T. Nguyen, S. Kwon and Y. H. Cho, “Estimation of Measurement Errors in 3GPP Filtering Models”, *2020 International Conference on Information Networking (ICOIN)*, Barcelona, Spain, 2020, pp. 510-512, doi: 10.1109/ICOIN48656.2020.9016431.
- [6] M. T. Nguyen, J. Song, S. Kwon and K. Kim, “Joint Power Allocation and Beamforming Design for Dual-connectivity Wireless Networks”, *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, Seoul, Korea (South), 2020, pp. 1-6, doi: 10.1109/WCNC45663.2020.9120728.

Domestic conference

- [7] M. T. Nguyen, S. Kwon and S. Kim, “Capacity Maximization for CoMP-based Networks with Per-User Guaranteed Throughput”, *Proceedings of Symposium of the Korean Institute of communications and Information Sciences*, 1457-1458, 2019
- [8] L. H. Nguyen, M. T. Nguyen and S. Kwon, “Adaptive Reduced Power Subframe in LTE Heterogeneous Networks”, *Proceedings of Symposium of the Korean Institute of communications and Information Sciences*, 73-74, 2018

References

- [1] M. T. Nguyen, S. Kwon, and H. Kim, “Mobility robustness optimization for handover failure reduction in LTE small-cell networks,” *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 4672–4676, May 2018.
- [2] Qualcomm (Feb. 2017), “Small cells: enhancing coverage, capacity and experiences with shared/unlicensed spectrum.” [Online]. Available: <https://www.qualcomm.com/news/onq/2017/02/22/small-cells-enhancing-coverage-capacity-and-experiences-sharedunlicensed>
- [3] E. Hossain and M. Hasan, “IEEE instrum. meas. mag.” *IEEE Instrumentation and Measurement Magazine*, vol. 18, no. 3, pp. 11–21, Jun. 2015.
- [4] J. Hoydis, M. Kobayashi, and M. Debbah, “Green small-cell networks,” *IEEE Veh. Technol. Mag.*, vol. 6, no. 1, pp. 37–43, Mar. 2011.
- [5] M. Peng, D. Liang, Y. Wei, J. Li, , and H. H. Chen, “Self-configuration and self-optimization in LTE-advanced heterogeneous networks,” *IEEE Commun. Mag.*, vol. 51, pp. 36–45, May 2013.
- [6] 3GPP, *TR 36.839 Mobility enhancements in heterogeneous networks*, 2013.
- [7] I. M. Bălan, B. Sas, T. Jansen, I. Moerman, K. Spaey, and P. Demeester, “An en-

- hanced weighted performance-based handover parameter optimization algorithm for LTE networks,” *EURASIP J. Wireless Com. Network.*, vol. 2011, pp. 98, Sep. 2011.
- [8] T. Jansen, I. Balan, J. Turk, I. Moerman, and T. Kurner, “Handover parameter optimization in LTE self-organizing networks,” in *Proc. IEEE VTC 2010-Fall*, Sep. 2010, pp. 1–5.
- [9] K. Kitagawa, T. Komine, T. Yamamoto, and S. Konishi, “A handover optimization algorithm with mobility robustness for LTE systems,” in *Proc. IEEE PIMRC*, Sep. 2011, pp. 1647–1651.
- [10] D. Lopez-Perez, I. Guvenc, and X. Chu, “Mobility management challenges in 3GPP heterogeneous networks,” *IEEE Commun. Mag.*, vol. 50, pp. 70–78, Dec. 2012.
- [11] —, “Theoretical analysis of handover failure and ping-pong rates for heterogeneous networks,” in *Proc. IEEE ICC*, Jun. 2012, pp. 6774–6779.
- [12] P. Muñoz, R. Barco, and I. de la Bandera, “On the potential of handover parameter optimization for self-organizing networks,” *IEEE Trans. Veh. Technol.*, vol. 62, pp. 1895–1905, Jun. 2013.
- [13] J. Wu, J. Liu, Z. Huang, and S. Zheng, “Dynamic fuzzy q-learning for handover parameters optimization in 5g multi-tier networks,” in *Proc. IEEE WCSP*, Oct. 2015, pp. 1–5.
- [14] A. Klein, N. P. Kuruvatti, J. Schneider, and H. D. Schotten, “Fuzzy q-learning for mobility robustness optimization in wireless networks,” in *Proc. IEEE GLOBECOM Workshops*, Dec. 2013, pp. 76–81.
- [15] S. S. Mwanje, L. C. Schmelz, and A. Mitschele-Thiel, “Cognitive cellular networks: a q-

- learning framework for self-organizing networks,” *IEEE Trans. Netw. Service Manag.*, vol. 13, pp. 85–98, Mar. 2016.
- [16] C. H. M. de Lima, M. Bennis, and M. Latva-aho, “Modeling and analysis of handover failure probability in small cell networks,” in *Proc. IEEE INFOCOM Workshops*, Apr. 2014, pp. 736–741.
- [17] 3GPP, *TS 36.331 Universal mobile telecommunications system, radio resource control protocol specification*, 2017.
- [18] ———, *TS 36.300 LTE; E-UTRAN; Overall description; Stage 2*, 2018.
- [19] NS3 (Feb. 2015), “LTE-EPC network simulator (LENA).” [Online]. Available: <http://networks.cttc.es/mobile-networks/software-tools/lena/>
- [20] 3GPP, *TS 36.814 Further advancements for E-UTRA physical layer aspects*, 2017.
- [21] ———, *TS 36.104 Base station radio transmission and reception*, 2017.
- [22] ———, *TS 37.320 Radio measurement collection for minimization of drive tests (MDT): overall description Stage 2*, 2017.
- [23] Z. Xiao, T. Li, W. Cheng, , and D. Wang, “Apollonius circles based outbound handover in macro-small wireless cellular networks,” in *Proc. IEEE GLOBECOM*, Dec. 2016, pp. 1–6.
- [24] M. T. Nguyen and S. Kwon, “Geometry-based analysis of optimal handover parameters for self-organizing networks,” *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, pp. 2670–2683, Apr. 2020.
- [25] A. Bleicher, “5G goes for the gold,” *IEEE Spectrum*, vol. 55, no. 1, pp. 32–33, Jan. 2018.

-
- [26] T. S. Rappaport, Y. Xing, G. R. MacCartney, A. F. Molisch, E. Mellios, and J. Zhang, "Overview of millimeter wave communications for fifth-generation (5G) wireless networks-with a focus on propagation models," *IEEE Trans. Antennas Propag.*, vol. 65, no. 12, pp. 6213–6230, Dec. 2017.
- [27] A. J. Fehske, I. Viering, J. Voigt, C. Sartori, S. Redana, and G. P. Fettweis, "Small-cell self-organizing wireless networks," *Proceedings of the IEEE*, vol. 102, no. 3, pp. 334–350, Mar. 2014.
- [28] R. Webb, T. Wehmeier, and K. Dyer, "Small cells 2012 integration and optimisation," *Mobile Europe, Tech. Rep.*, 2012.
- [29] 3GPP, *TR 36.902 Self-configuring and self-optimizing network use cases and solutions*, 2011.
- [30] P. Muñoz, R. Barco, and I. de la Bandera, "On the potential of handover parameter optimization for self-organizing networks," *IEEE Trans. Veh. Technol.*, vol. 62, pp. 1895–1905, Jun. 2013.
- [31] S. S. Mwanje, L. C. Schmelz, and A. Mitschele-Thiel, "Cognitive cellular networks: a q-learning framework for self-organizing networks," *IEEE Trans. Netw. Service Manag.*, vol. 13, pp. 85–98, Mar. 2016.
- [32] D. Lopez-Perez, I. Guvenc, and X. Chu, "Mobility management challenges in 3GPP heterogeneous networks," *IEEE Commun. Mag.*, vol. 50, pp. 70–78, Dec. 2012.
- [33] C. H. M. de Lima, M. Bennis, and M. Latva-aho, "Modeling and analysis of handover failure probability in small cell networks," in *Proc. IEEE INFOCOM Workshops*, Apr. 2014, pp. 736–741.

-
- [34] O. C. Iacobaiea, B. Sayrac, S. B. Jemaa, and P. Bianchi, "On mobility parameter configurations that can lead to chained handovers," *IEEE Trans. Commun.*, vol. 64, no. 12, pp. 5136–5148, Dec. 2016.
- [35] K. Vasudeva, M. Simsek, D. Lopez-Perez, and I. Guvenc, "Analysis of handover failures in heterogeneous networks with fading," *IEEE Trans. Veh. Technol.*, vol. 66, pp. 6060–6074, Jul. 2017.
- [36] X. Xu, Z. Sun, X. Dai, T. Svensson, and X. Tao, "Modeling and analyzing the cross-tier handover in heterogeneous networks," *IEEE Trans. Wireless Commun.*, vol. 16, pp. 7859–7869, Dec. 2017.
- [37] 3GPP, *TS 36.214 Physical layer; Measurements*, 2018.
- [38] T. Jansen, I. Balan, J. Turk, I. Moerman, and T. Kurner, "Handover parameter optimization in LTE self-organizing networks," in *Proc. IEEE VTC 2010-Fall*, Sep. 2010, pp. 1–5.
- [39] B. Li, H. Zhang, P. Hao, and J. Li, "Sojourn time estimation-based small cell selection in ultra-dense networks," in *2017 IEEE PIMRC*, Oct. 2017, pp. 1–5.
- [40] M. Polese, M. Giordani, M. Mezzavilla, S. Rangan, and M. Zorzi, "Improved handover through dual connectivity in 5g mmwave mobile networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 2069–2084, Sep. 2017.
- [41] M. Mezzavilla, M. Zhang, M. Polese, R. Ford, S. Dutta, S. Rangan, and M. Zorzi, "End-to-end simulation of 5g mmwave networks," *Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2237–2263, thirdquarter 2018.
- [42] D. Matsuo, M. Morita, T. Nobukiyo, and Y. Matsunaga, "Adaptive coverage con-

- trol using cell-specific beamforming based on throughput prediction,” in *2015 IEEE PIMRC*, Aug. 2015, pp. 1360–1364.
- [43] A. Awada, B. Wegmann, I. Viering, and A. Klein, “A son-based algorithm for the optimization of inter-rat handover parameters,” *IEEE Trans. Veh. Technol.*, vol. 62, no. 5, pp. 1906–1923, Jun. 2013.
- [44] H. Cho, S. Shin, G. Lim, C. Lee, and J. Chung, “Lte-r handover point control scheme for high-speed railways,” *IEEE Wireless Commun.*, vol. 24, no. 6, pp. 112–119, Dec. 2017.
- [45] 3GPP, *TS 36.133 Evolved Universal Terrestrial Radio Access (E-UTRA); Requirements for support of radio resource management*, 2018.
- [46] D. A. Brannan, M. F. Espen, and J. J. Gray, *Geometry (2nd Ed.)*. UK: Cambridge University Press, 2012, ch. 5.
- [47] 3GPP, *TR 38.901 5G, Study on channel model for frequencies from 0.5 to 100 GHz*, 2018.
- [48] D. Lopez-Perez, X. Chu, and I. Guvenc, “On the expanded region of picocells in heterogeneous networks,” *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 3, pp. 281–294, jun 2012.
- [49] Z. Becvar and P. Mach, “Adaptive hysteresis margin for handover in femtocell networks,” in *Proc. IEEE ICWMC*, Sep. 2010, pp. 256–261.
- [50] Z. Ren, P. Fertl, Q. Liao, F. Penna, and S. Stanczak, “Street-specific handover optimization for vehicular terminals in future cellular networks,” in *Proc. IEEE VTC 2013-Spring*, Jun. 2013, pp. 1–5.

- [51] S. Sun, T. A. Thomas, T. S. Rappaport, H. Nguyen, I. Z. Kovacs, and I. Rodriguez, "Path loss, shadow fading, and line-of-sight probability models for 5g urban macro-cellular scenarios," in *Proc. 2015 IEEE Globecom Workshops*, Dec. 2015, pp. 1–7.
- [52] 3GPP, *TS 36.521-3 User Equipment (UE) conformance specification; Radio transmission and reception; Part 3: Radio Resource Management (RRM) conformance testing*, 2017.
- [53] A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, and D. Malladi, "A survey on 3gpp heterogeneous networks," *IEEE Wireless Commun.*, vol. 18, no. 3, pp. 10–21, Jun. 2011.
- [54] D. Lopez-Perez, I. Guvenc, and X. Chu, "Theoretical analysis of handover failure and ping-pong rates for heterogeneous networks," in *Proc. IEEE ICC*, Jun. 2012, pp. 6774–6779.
- [55] H. Elshaer, F. Boccardi, M. Dohler, and R. Irmer, "Downlink and uplink decoupling: A disruptive architectural design for 5G networks," in *Proc. IEEE GLOBECOM 2014*, Dec. 2014, pp. 1798–1803.
- [56] M. M. Hasan, S. Kwon, and J. H. Na, "Adaptive mobility load balancing algorithm for LTE small-cell networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2205–2217, Apr. 2018.
- [57] R. Li, Z. Zhao, X. Chen, J. Palicot, and H. Zhang, "TACT: A transfer actor-critic learning framework for energy saving in cellular radio access networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 4, pp. 2000–2011, Apr. 2014.
- [58] K. T. Dinh and S. Kukliński, "Joint implementation of several LTE-SON functions," in *Proc. IEEE GLOBECOM Workshops*, Dec. 2013, pp. 953–957.

- [59] P. Muñoz, R. Barco, J. M. Ruiz-Avilés, I. de la Bandera, and A. Aguilar, “Fuzzy rule-based reinforcement learning for load balancing techniques in enterprise LTE femto-cells,” *IEEE Transactions on Vehicular Technology*, vol. 62, no. 5, pp. 1962–1973, Jun. 2013.
- [60] R. S. Sutton and A. G. Barto, *Introduction to Reinforcement Learning*. Cambridge, MA, USA: MIT Press, 1998.
- [61] A. Galindo-Serrano, L. Giupponi, and G. Auer, “Distributed learning in multiuser OFDMA femtocell networks,” in *Pro. IEEE VTC-Spring 2011*, May 2011, pp. 1–6.
- [62] Q. Zhao, T. Jiang, N. Morozs, D. Grace, and T. Clarke, “Transfer learning: A paradigm for dynamic spectrum and topology management in flexible architectures,” in *Pro. IEEE VTC Fall*, Sep. 2013, pp. 1–5.
- [63] E. Baştuğ, M. Bennis, and M. Debbah, “A transfer learning approach for cache-enabled wireless networks,” in *Proc. IEEE WiOpt*, May 2015, pp. 161–166.
- [64] H. Sanneck, Y. Bouwen, and E. Troch, “Dynamic radio configuration of self-organizing base stations,” in *Pro. IEEE ISWCS*, Sep. 2010, pp. 716–720.
- [65] F. Calabrese, M. Colonna, P. Lovisolo, D. Parata, and C. Ratti, “Real-time urban monitoring using cell phones: A case study in Rome,” *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 1, pp. 141–151, Mar. 2011.
- [66] M. M. Hasan, S. Kwon, and S. Oh, “Frequent-handover mitigation in ultra-dense heterogeneous networks,” *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 1035–1040, Jan. 2019.
- [67] L. Torrey and J. Shavlik, “Transfer learning.” [Online]. Available: <http://pages.cs.wisc.edu/shavlik/abstracts/torrey.handbook09.abstract.html>

-
- [68] L. Saker, S. E. Elayoubi, R. Combes, and T. Chahed, "Optimal control of wake up mechanisms of femtocells in heterogeneous networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 664–672, Apr. 2012.
- [69] E. Oh, K. Son, and B. Krishnamachari, "Dynamic base station switching-on/off strategies for green cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 2126–2136, May 2013.
- [70] S. Samarakoon, M. Bennis, W. Saad, and M. Latva-aho, "Dynamic clustering and on/off strategies for wireless small cell networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 2164–2178, Mar. 2016.
- [71] J. Zhu, Y. Song, D. Jiang, and H. Song, "A new deep-q-learning-based transmission scheduling mechanism for the cognitive internet of things," *IEEE Internet Things J.*, vol. 5, no. 4, pp. 2375–2385, Aug. 2018.
- [72] M. T. Nguyen, J. Song, S. Kwon, and K. Kim, "Joint power allocation and beamforming design for dual-connectivity wireless networks," in *2020 IEEE WCNC*, 2020, pp. 1–6.
- [73] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.
- [74] A. Wolf, P. Schulz, M. Dörpinghaus, J. C. S. Santos Filho, and G. Fettweis, "How reliable and capable is multi-connectivity?" *IEEE Trans. Commun.*, vol. 67, no. 2, pp. 1506–1520, Feb. 2019.
- [75] M. Giordani, M. Mezzavilla, S. Rangan, and M. Zorzi, "An efficient uplink multi-

- connectivity scheme for 5G millimeter-wave control plane applications,” *IEEE Trans. Wireless Commun.*, vol. 17, no. 10, pp. 6806–6821, Oct. 2018.
- [76] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H. Mayer, L. Thiele, and V. Jungnickel, “Coordinated multipoint: Concepts, performance, and field trial results,” *IEEE Commun. Mag.*, vol. 49, no. 2, pp. 102–111, Feb. 2011.
- [77] V. Jungnickel, K. Manolakis, W. Zirwas, B. Panzner, V. Braun, M. Lossow, M. Sternad, R. Apelfrojd, and T. Svensson, “The role of small cells, coordinated multipoint, and massive MIMO in 5G,” *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 44–51, May 2014.
- [78] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, “Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels,” *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461–471, Feb. 2004.
- [79] Z. Shen, R. Chen, J. G. Andrews, R. W. Heath, and B. L. Evans, “Low complexity user selection algorithms for multiuser MIMO systems with block diagonalization,” *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3658–3663, Sep. 2006.
- [80] R. Zhang, “Cooperative multi-cell block diagonalization with per-base-station power constraints,” *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1435–1445, Dec. 2010.
- [81] H. Huh, A. M. Tulino, and G. Caire, “Network MIMO with linear zero-forcing beamforming: Large system analysis, impact of channel estimation, and reduced-complexity scheduling,” *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2911–2934, May 2012.
- [82] D. H. N. Nguyen, H. Nguyen-Le, and T. Le-Ngoc, “Block-diagonalization precoding in a multiuser multicell MIMO system: Competition and coordination,” *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 968–981, Feb. 2014.

-
- [83] S. Mosleh, L. Liu, and J. Zhang, "Proportional-fair resource allocation for coordinated multi-point transmission in LTE-advanced," *IEEE Trans. Wireless Commun.*, vol. 15, no. 8, pp. 5355–5367, Aug. 2016.
- [84] M. Ge and D. M. Blough, "High throughput and fair scheduling for multi-ap multiuser MIMO in dense wireless networks," *IEEE/ACM Trans. Netw.*, vol. 26, no. 5, pp. 2414–2427, Oct. 2018.
- [85] S. Sun, T. S. Rappaport, M. Shafi, and H. Tataria, "Analytical framework of hybrid beamforming in multi-cell millimeter-wave systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 11, pp. 7528–7543, Nov. 2018.
- [86] W. Ni and X. Dong, "Hybrid block diagonalization for massive multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 64, no. 1, pp. 201–211, Jan. 2016.
- [87] J. Song, J. Choi, and D. J. Love, "Common codebook millimeter wave beam design: Designing beams for both sounding and communication with uniform planar arrays," *IEEE Trans. Commun.*, vol. 65, no. 4, pp. 1859–1872, Apr. 2017.
- [88] G. R. MacCartney and T. S. Rappaport, "Millimeter-wave base station diversity for 5G coordinated multipoint (CoMP) applications," *IEEE Trans. Wireless Commun.*, vol. 18, no. 7, pp. 3395–3410, July 2019.
- [89] J. Zhang, R. Chen, J. G. Andrews, A. Ghosh, and R. W. Heath, "Networked mimo with clustered linear precoding," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 1910–1921, Apr. 2009.
- [90] W. Cheng, X. Zhang, and H. Zhang, "Statistical-qos driven energy-efficiency optimization over green 5g mobile wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 12, pp. 3092–3107, 2016.

- [91] J. Tang, D. K. C. So, E. Alsusa, and K. A. Hamdi, "Resource efficiency: A new paradigm on energy efficiency and spectral efficiency tradeoff," *IEEE Trans. Wireless Commun.*, vol. 13, no. 8, pp. 4656–4669, 2014.
- [92] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser mimo systems," *IEEE Transactions on Communications*, vol. 61, no. 4, pp. 1436–1449, 2013.
- [93] K. N. R. S. V. Prasad, E. Hossain, and V. K. Bhargava, "Energy efficiency in massive mimo-based 5g networks: Opportunities and challenges," *IEEE Wireless Commun.*, vol. 24, no. 3, pp. 86–94, 2017.
- [94] J. Zhang, S. Chen, Y. Lin, J. Zheng, B. Ai, and L. Hanzo, "Cell-free massive mimo: A new next-generation paradigm," *IEEE Access*, vol. 7, pp. 99 878–99 888, 2019.
- [95] L. Bertizzolo, E. Demirors, Z. Guan, and T. Melodia, "Cobeam: Beamforming-based spectrum sharing with zero cross-technology signaling for 5g wireless networks," in *IEEE INFOCOM 2020*, 2020, pp. 1429–1438.
- [96] K. Zu, R. C. de Lamare, and M. Haardt, "Generalized design of low-complexity block diagonalization type precoding algorithms for multiuser mimo systems," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4232–4242, 2013.
- [97] N. N. Moghadam, G. Fodor, M. Bengtsson, and D. J. Love, "On the energy efficiency of mimo hybrid beamforming for millimeter-wave systems with nonlinear power amplifiers," *IEEE Trans. Wireless Commun.*, vol. 17, no. 11, pp. 7208–7221, 2018.
- [98] H. Q. Ngo, L. Tran, T. Q. Duong, M. Matthaiou, and E. G. Larsson, "On the total energy efficiency of cell-free massive mimo," *IEEE Transactions on Green Communications and Networking*, vol. 2, no. 1, pp. 25–39, 2018.

-
- [99] T. M. Hoang, H. Q. Ngo, T. Q. Duong, H. D. Tuan, and A. Marshall, "Cell-free massive mimo networks: Optimal power control against active eavesdropping," *IEEE Trans. Commun.*, vol. 66, no. 10, pp. 4724–4737, 2018.
- [100] M. Alonzo, S. Buzzi, A. Zappone, and C. D'Elia, "Energy-efficient power control in cell-free and user-centric massive mimo at millimeter wave," *IEEE Transactions on Green Communications and Networking*, vol. 3, no. 3, pp. 651–663, 2019.
- [101] E. Björnson and L. Sanguinetti, "Making cell-free massive mimo competitive with mmse processing and centralized implementation," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 77–90, 2020.
- [102] T. Van Chien, E. Björnson, and E. G. Larsson, "Joint power allocation and load balancing optimization for energy-efficient cell-free massive mimo networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6798–6812, 2020.
- [103] C. Shen, T. Chang, K. Wang, Z. Qiu, and C. Chi, "Distributed robust multicell coordinated beamforming with imperfect csi: An admm approach," *IEEE Trans. Signal Process*, vol. 60, no. 6, pp. 2988–3003, 2012.
- [104] L. D. Nguyen, T. Q. Duong, H. Q. Ngo, and K. Tourki, "Energy efficiency in cell-free massive mimo with zero-forcing precoding design," *IEEE Commun. Lett.*, vol. 21, no. 8, pp. 1871–1874, 2017.
- [105] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive mimo versus small cells," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1834–1850, 2017.
- [106] M. M. Zhao, Y. Cai, M. J. Zhao, Y. Xu, and L. Hanzo, "Robust joint hybrid analog-

- digital transceiver design for full-duplex mmwave multicell systems,” *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 4788–4802, 2020.
- [107] S. Kwon and N. B. Shroff, “Energy-efficient sinr-based routing for multihop wireless networks,” *IEEE Trans. Mobile Comput.*, vol. 8, no. 5, pp. 668–681, 2009.
- [108] M. M. Zhao, Y. Cai, M. J. Zhao, Y. Xu, and L. Hanzo, “Robust joint hybrid analog-digital transceiver design for full-duplex mmwave multicell systems,” *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 4788–4802, 2020.
- [109] D. Zhang, Y. Wang, X. Li, and W. Xiang, “Hybridly connected structure for hybrid beamforming in mmwave massive mimo systems,” *IEEE Trans. Commun.*, vol. 66, no. 2, pp. 662–674, 2018.
- [110] H. Kasai, “Fast optimization algorithm on complex oblique manifold for hybrid precoding in millimeter wave mimo systems,” in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2018, pp. 1266–1270.
- [111] J. Liu, X. Liu, and X. Ma, “First-order perturbation analysis of singular vectors in singular value decomposition,” *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3044–3049, 2008.

Appendix A

Proofs in Chapter 2

A.1 Proof of Lemma 1

From (2.19)/(2.20) and (2.19)/(2.21), the following describes conditions for an optimal value of Δ_{01}

$$x_E - v(\text{TTT}_0 + \text{TTT}_1) \cos \theta < x_C \leq x_D - v(\text{TTT}_0 - \tau_R) \cos \theta, \quad (\text{A.1})$$

$$x_F - v(\text{TTT}_0 + \tau_R) \cos \theta < x_C \leq x_F - v(\text{TTT}_0 - \tau_R) \cos \theta. \quad (\text{A.2})$$

Then, the optimal range can-not exist when (A.1) and (A.2) are invalid for all values of Δ_{01} .

Thus, we can derive the following:

$$x_E - v(\text{TTT}_0 + \text{TTT}_1) \cos \theta \geq x_D - v(\text{TTT}_0 - \tau_R) \cos \theta,$$

$$x_F - v(\text{TTT}_0 + \tau_R) \cos \theta \geq x_F - v(\text{TTT}_0 - \tau_R) \cos \theta.$$

After rearrangement, a condition for non-existence of the optimal range can be stated as

$$\min \left\{ \frac{x_E - x_D}{\text{TTT}_1 + \tau_R}, \frac{x_F - x_D}{2\tau_R} \right\} \geq v \cos \theta.$$

Finally, the condition for existence of the optimal range can be expressed as

$$\max \left\{ \frac{x_E - x_D}{\text{TTT}_1 + \tau_R}, \frac{x_F - x_D}{2\tau_R} \right\} \leq v \cos \theta.$$

■

A.2 Proof of Theorem 1

First of all, we denote the dependence of x_C on Δ_{01} and of x_E on Δ_{10} as $f_C(\Delta_{01})$ and $f_E(\Delta_{10})$, respectively. Like wise, $f_D(\gamma_{\min})$ and $f_F(\gamma_{\min})$, respectively, expresses the dependence of x_D and x_F on γ_{\min} . Since handover offset is a discrete value variable, we denote the granularity of Δ_{01} as ε . First, from (2.19), we have

$$\frac{f_C(\Delta_{01}^\dagger) - f_D(\gamma_{\min})}{(\tau_R - \text{TTT}_0) \cos \theta} \leq v. \quad (\text{A.3})$$

We assume that TTT_0 is less than τ_R , so the denominator of (A.3) is positive. For $\Delta'_{01} = \Delta_{01}^\dagger - \varepsilon$ ($\varepsilon > 0$), we have $f_C(\Delta'_{01}) > f_C(\Delta_{01}^\dagger)$, since f_C is a decreasing function of Δ_{01} as explained in Section III-A. As a result, we have

$$v < \frac{f_C(\Delta'_{01}) - f_D(\gamma_{\min})}{(\tau_R - \text{TTT}_0) \cos \theta}.$$

Here, we can choose any v' such that $v' > v$ and

$$\frac{f_C(\Delta'_{01}) - f_D(\gamma_{\min})}{(\tau_R - \text{TTT}_0) \cos \theta} \leq v' < \frac{f_C(\Delta'_{01} - \varepsilon) - f_D(\gamma_{\min})}{(\tau_R - \text{TTT}_0) \cos \theta}.$$

As can be seen, Δ_{01}^\dagger now becomes Δ'_{01} when v increases to v' . Therefore, the optimal range extends to the left when user speed increases.

Secondly, we can rearrange (2.20) with $\Delta_{01} = \Delta_{01}^*$ as follows

$$\frac{f_F(\gamma_{\min}) - f_C(\Delta_{01}^*)}{(\text{TTT}_0 + \tau_R) \cos \theta} < v.$$

For $\Delta''_{01} = \Delta^*_{01} + \epsilon$ ($\epsilon > 0$), we have $f_C(\Delta''_{01}) < f_C(\Delta^*_{01})$. As a result, we have

$$v \leq \frac{f_F(\gamma_{\min}) - f_C(\Delta''_{01})}{(\text{TTT}_0 + \tau_R) \cos \theta}.$$

Here, we can choose any v'' such that $v'' > v$ and

$$\frac{f_F(\gamma_{\min}) - f_C(\Delta''_{01})}{(\text{TTT}_0 + \tau_R) \cos \theta} < v'' \leq \frac{f_F(\gamma_{\min}) - f_C(\Delta''_{01} + \epsilon)}{(\text{TTT}_0 + \tau_R) \cos \theta}.$$

We can see that Δ'_{01} becomes Δ^*_{01} when $v = v''$. Likewise, we can obtain a similar result with the ping-pong problem in (2.21). Therefore, the optimal range extends to the right when user speed increases. ■

A.3 Proof of Theorem 2

First, we reorganize (2.19) as follows $\text{TTT}_0 \leq \tau_R - \left(f_C(\Delta^{\dagger}_{01}) - f_D(\gamma_{\min}) \right) / v \cos \theta$.

For $\Delta'_{01} = \Delta^{\dagger}_{01} + \epsilon$ ($\epsilon > 0$), we have $f_D(\Delta'_{01}) < f_D(\Delta^{\dagger}_{01})$ since f_D is a decreasing function of Δ_{01} as explained in Section III-A. As a result, we have:

$$\tau_R - \frac{f_C(\Delta^{\dagger}_{01}) - f_D(\gamma_{\min})}{v \cos \theta} < \tau_R - \frac{f_C(\Delta'_{01}) - f_D(\gamma_{\min})}{v \cos \theta}.$$

Thus, we can choose $\text{TTT}'_0 > \text{TTT}_0$ such that

$$\tau_R - \frac{f_C(\Delta^{\dagger}_{01} - \epsilon) - f_D(\gamma_{\min})}{v \cos \theta} < \text{TTT}'_0 \leq \tau_R - \frac{f_C(\Delta'_{01}) - f_D(\gamma_{\min})}{v \cos \theta}.$$

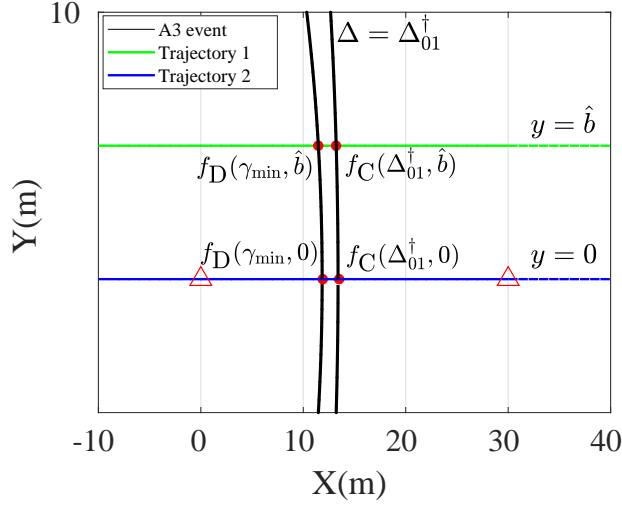
We can see that Δ'_{01} becomes Δ^{\dagger}_{01} when TTT_0 increases to TTT'_0 . In other words, Δ^{\dagger}_{01} increases when TTT_0 rises.

Secondly, (2.20) can be rearranged as

$$\text{TTT}_0 > \frac{f_F(\gamma_{\min}) - f_C(\Delta^*_{01})}{v \cos \theta} - \tau_R.$$

For $\Delta''_{01} = \Delta^*_{01} + \epsilon$ ($\epsilon > 0$), we have

$$\text{TTT}_0 \leq \frac{f_F(\gamma_{\min}) - f_C(\Delta''_{01})}{v \cos \theta} - \tau_R.$$

Figure A.1: Trajectories with $\theta = 0$

We are able to select $\text{TTT}_0'' > \text{TTT}_0$ such that

$$\frac{f_F(\gamma_{\min}) - f_E(\Delta_{01}'')}{v \cos \theta} - \tau_R < \text{TTT}_0'' \leq \frac{f_F(\gamma_{\min}) - f_E(\Delta_{01}'' + \varepsilon)}{v \cos \theta} - \tau_R.$$

So, Δ_{01}'' becomes Δ_{01}^* when TTT_0 increases to TTT_0'' . Likewise, we can obtain a similar result with the ping-pong problem in (2.21). ■

A.4 Proof of Theorem 3

In (2.19), (2.20), and (2.21), the term $v \cos \theta$ is a projection of velocity v on the x -axis. Therefore, $\theta = 0$ gives the maximum impact of v on the optimal range, which is explained by Theorem 1. With $\theta = 0$, UE trajectories are parallel to the x -axis. We denote x_C and x_D as $f_C(\Delta_{01}, b)$ and $f_D(\Delta_{01}, b)$, respectively, to express the dependence of x_C and x_D on Δ_{01} and b . We rearrange the condition for too-late handover avoidance in (2.19) as $f_C(\Delta_{01}^\dagger, b) - f_D(\gamma_{\min}, b) \leq v(\tau_R - \text{TTT}_0)$. There exists \hat{b} such that $f_C(\Delta_{01}^\dagger, \hat{b}) - f_D(\gamma_{\min}, \hat{b}) = v(\tau_R - \text{TTT}_0)$. As we can see in Fig. A.1, we have

$$f_C(\Delta_{01}^\dagger, b) - f_D(\gamma_{\min}, b) < f_C(\Delta_{01}^\dagger, \hat{b}) - f_D(\gamma_{\min}, \hat{b}) = v(\tau_R - \tau_{TT0}),$$

for $0 \leq b < \hat{b}$ (we consider only $b \geq 0$ due to geometrical symmetry). For $\Delta'_{01} = \Delta_{01}^\dagger - \epsilon$ ($\epsilon > 0$), we have $f_C(\Delta'_{01}, b) > f_C(\Delta_{01}^\dagger, b)$, because $f_C(\Delta_{01}, b)$ is a decreasing function of Δ_{01} . As a result, $f_C(\Delta'_{01}, b) - f_D(\gamma_{\min}, b) > f_C(\Delta_{01}^\dagger, b) - f_D(\gamma_{\min}, b)$. Also, there exists Δ'_{01} such that

$$f_C(\Delta'_{01}, b) - f_D(\gamma_{\min}, b) = v(\tau_R - \tau_{TT0}).$$

Thus, the reduction of b decreases Δ_{01}^\dagger and is lower-bounded by 0. Similarly, we prove that $f(\Delta_{01}^*, b)$ is a decreasing function of b . Therefore, the reduction of b can broaden the optimal range $[\Delta_{01}^\dagger, \Delta_{01}^*]$. As a result, the optimum value of b is 0. This is the line going directly through Cell 0 and Cell 1. ■

This page intentionally left blank.

Appendix B

Proofs in Chapter 5

B.1 Proof of Lemma 1

Without lack of generality, we consider a matrix $\mathbf{A} \in C^{N \times M}$ with rank N and its SVD

$$\mathbf{A} = \mathbf{U}^1 \boldsymbol{\Sigma}^1 (\mathbf{V}^1)^H + \mathbf{U}^0 \boldsymbol{\Sigma}^0 (\mathbf{V}^0)^H,$$

where \mathbf{U}^1 , $\boldsymbol{\Sigma}^1$, \mathbf{V}^1 corresponds to M non-zero singular values; and \mathbf{U}^0 , $\boldsymbol{\Sigma}^0$, and \mathbf{V}^0 corresponds to $M - N$ zero singular values.

An error matrix $\Delta \mathbf{A}$ of \mathbf{A} perturbed the SVD such that

$$\mathbf{A} + \Delta \mathbf{A} = (\mathbf{U} + \Delta \mathbf{U}) (\boldsymbol{\Sigma} + \Delta \boldsymbol{\Sigma}) (\mathbf{V}^H + \Delta \mathbf{V}^H).$$

We estimate $\Delta \mathbf{V}$ as follows. From properties of eigenvalues and eigenvectors, we have $\mathbf{A} v_i = u_i \sigma_i$ and $\mathbf{A}^H u_i = \sigma_i v_i$, and thus obtain

$$\mathbf{A}^H \mathbf{A} v_i = \sigma_i^2 v_i. \tag{B.1}$$

Similar to [111], by differentiating both sides of B.1, we have

$$\Delta \mathbf{A}^H \mathbf{A} v_i + \mathbf{A}^H \Delta \mathbf{A} v_i + \mathbf{A}^H \mathbf{A} \Delta v_i \approx \sigma_i^2 \Delta v_i + 2\sigma_i \Delta \sigma_i v_i \tag{B.2}$$

Since Δv_i is not the eigenvector of \mathbf{A} , we can express Δv_i as a linear combination of the right singular vectors of \mathbf{A} as

$$\Delta v_i = \sum_{k=1, k \neq i}^N g_{ik} v_k, \quad (\text{B.3})$$

where g_{ik} is coefficient. Hence, we rewrite (B.2) as

$$\Delta \mathbf{A}^H \mathbf{A} v_i + \mathbf{A}^H \Delta \mathbf{A} v_i + \mathbf{A}^H \mathbf{A} \sum_{k=1, k \neq i}^N g_{ik} v_k \approx \sigma_i^2 \sum_{k=1, k \neq i}^N g_{ik} v_k + 2\sigma_i \Delta \sigma_i v_i. \quad (\text{B.4})$$

Multiply both sides of (B.4) with v_j^H ($j \neq i$) and note that $v_j^H v_i = 0$, we obtain the following

$$v_j^H (\Delta \mathbf{A}^H \mathbf{A} v_i + \mathbf{A}^H \Delta \mathbf{A} v_i) + g_{ij} \sigma_j^2 \approx g_{ij} \sigma_i^2.$$

And g_{ij} is obtained by

$$g_{ij} \approx \frac{v_j^H (\Delta \mathbf{A}^H \mathbf{A} v_i + \mathbf{A}^H \Delta \mathbf{A} v_i)}{\sigma_i^2 - \sigma_j^2}.$$

So, Δv_i in (B.3) is computed as

$$\Delta v_i \approx \sum_{j=1, j \neq i}^N v_i \frac{v_j^H (\Delta \mathbf{A}^H \mathbf{A} v_i + \mathbf{A}^H \Delta \mathbf{A} v_i)}{\sigma_i^2 - \sigma_j^2}. \quad (\text{B.5})$$

We rewrite (B.5) as

$$\begin{aligned} \Delta v_i &\approx \sum_{j=1, j \neq i}^N \frac{v_j v_j^H (\Delta \mathbf{A}^H \mathbf{A} + \mathbf{A}^H \Delta \mathbf{A}) v_i}{\sigma_i^2 - \sigma_j^2} + \sum_{j=N+1}^M \frac{v_j v_j^H (\Delta \mathbf{A}^H \mathbf{A} + \mathbf{A}^H \Delta \mathbf{A}) v_i}{\sigma_i^2} \\ &= \mathbf{V}^1 \boldsymbol{\Psi}_i \mathbf{V}^{1H} (\Delta \mathbf{A}^H \mathbf{A} + \mathbf{A}^H \Delta \mathbf{A}) v_i + \mathbf{V}^0 \mathbf{V}^{0H} (\Delta \mathbf{A}^H \mathbf{A} + \mathbf{A}^H \Delta \mathbf{A}) v_i \sigma_i^{-2}, \end{aligned} \quad (\text{B.6})$$

where $\boldsymbol{\Psi}_i$ is a matrix with

$$\boldsymbol{\Psi}_i = \text{diag} \left(\frac{1}{\sigma_i^2 - \sigma_1^2}, \dots, 0, \frac{1}{\sigma_i^2 - \sigma_{i+1}^2}, \dots, \frac{1}{\sigma_i^2 - \sigma_N^2} \right).$$

Since $\mathbf{A} v_i = u_i \sigma_i$, $\mathbf{V}^{1H} \mathbf{A}^H = \boldsymbol{\Sigma}^1 \mathbf{U}^{1H}$, and $\mathbf{V}^{0H} \mathbf{A}^H = 0$, equation (B.6) can be rewritten as

$$\begin{aligned} \Delta v_i &\approx \mathbf{V}^1 \boldsymbol{\Psi}_i \left(\mathbf{V}^{1H} \Delta \mathbf{A}^H u_i \sigma_i + \boldsymbol{\Sigma}^1 \mathbf{U}^{1H} \Delta \mathbf{A} v_i \right) \\ &\quad + \mathbf{V}^0 \mathbf{V}^{0H} \Delta \mathbf{A}^H u_i \sigma_i^{-1}. \end{aligned}$$

Thus, we have the following

$$\begin{aligned}\Delta \mathbf{V}^1 &\approx \mathbf{V}^1 \Psi_i \left(\mathbf{V}^{1H} \Delta \mathbf{A}^H \mathbf{U}^1 \Sigma^1 + \Sigma^1 \mathbf{U}^{1H} \Delta \mathbf{A} \mathbf{V}^1 \right) \\ &\quad + \mathbf{V}^0 \mathbf{V}^{0H} \Delta \mathbf{A}^H \mathbf{U}^1 \Sigma^{-1}\end{aligned}\tag{B.7}$$

By replacing \mathbf{A} , $\Delta \mathbf{A}$, \mathbf{V}^1 , \mathbf{V}^0 , and σ_i^2 with $\tilde{\mathbf{H}}_u \bar{\mathbf{V}}_u^0$, $\Delta \tilde{\mathbf{H}}_u \bar{\mathbf{V}}_u^0$, $\hat{\mathbf{V}}_u^1$, $\hat{\mathbf{V}}_u^0$, and $(\hat{\Sigma}_u^1)_{ii}$ respectively, in (B.7), we prove Lemma 1. ■

B.2 Proof of Lemma 2

Multiply both sides of (B.2) with v_i^H , we obtain the following

$$\begin{aligned}v_i^H (\Delta \mathbf{A}^H \mathbf{A} v_i + \mathbf{A}^H \Delta \mathbf{A} v_i + \mathbf{A}^H \mathbf{A} \Delta v_i) \\ \approx v_i^H (\sigma_i^2 \Delta v_i + 2\sigma_i \Delta \sigma_i v_i)\end{aligned}\tag{B.8}$$

After rearrangement and applying the fact that $\mathbf{A} v_i = u_i \sigma_i$ and $v_i^H \mathbf{A}^H = u_i^H \sigma_i$, we rewrite (B.8) as

$$v_i^H \Delta \mathbf{A}^H u_i \sigma_i + u_i^H \sigma_i \Delta \mathbf{A} v_i + v_i^H \sigma_i^2 \Delta v_i \approx v_i^H \sigma_i^2 \Delta v_i + 2\sigma_i \Delta \sigma_i.$$

Thus, we have

$$\Delta \sigma_i = \frac{1}{2} (v_i^H \Delta \mathbf{A}^H u_i + u_i^H \Delta \mathbf{A} v_i)\tag{B.9}$$

and

$$\Delta \Sigma = \text{diag}(\Delta \sigma_1, \dots, \Delta \sigma_N)\tag{B.10}$$

By replacing \mathbf{A} , $\Delta \mathbf{A}$, \mathbf{V}^1 , \mathbf{V}^0 , and σ_i^2 with $\tilde{\mathbf{H}}_u \bar{\mathbf{V}}_u^0$, $\Delta \tilde{\mathbf{H}}_u \bar{\mathbf{V}}_u^0$, $\hat{\mathbf{V}}_u^1$, $\hat{\mathbf{V}}_u^0$, and $(\hat{\Sigma}_u^1)_{ii}$ respectively, in (B.10), we prove Lemma 2. ■

B.3 Proof of Theorem 1

With $\Delta \mathbf{V}^1$ and $\Delta \widehat{\boldsymbol{\Sigma}}_u^1$, we rewrite (5.13) as

$$\begin{aligned} y' &= \left(\widehat{\mathbf{U}}_u^1 + \Delta \widehat{\mathbf{U}}_u^1 \right)^H \left(\widehat{\mathbf{U}}_u^1 + \Delta \widehat{\mathbf{U}}_u^1 \right) \left(\widehat{\boldsymbol{\Sigma}}_u^1 + \Delta \widehat{\boldsymbol{\Sigma}}_u^1 \right) \\ &\quad \times \left(\widehat{\mathbf{V}}_u^{1H} + \Delta \widehat{\mathbf{V}}_u^{1H} \right) \widehat{\mathbf{V}}_u^1 \mathbf{Q}_u^{\frac{1}{2}} s_u + \left(\widehat{\mathbf{U}}_u^1 + \Delta \widehat{\mathbf{U}}_u^1 \right)^H \\ &\quad \times \left(\widetilde{\mathbf{H}}_u + \Delta \widetilde{\mathbf{H}}_u \right) \sum_{v \in \mathcal{U}, v \neq u} \bar{\mathbf{V}}_v^0 \widehat{\mathbf{V}}_v^1 \mathbf{Q}_v^{\frac{1}{2}} s_v. \end{aligned}$$

Via approximation, we have

$$\begin{aligned} y' &\approx \left(\widehat{\boldsymbol{\Sigma}}_u^1 + \Delta \widehat{\boldsymbol{\Sigma}}_u^1 + \widehat{\boldsymbol{\Sigma}}_u^1 \Delta \widehat{\mathbf{V}}_u^{1H} \widehat{\mathbf{V}}_u^1 \right) \mathbf{Q}_u^{\frac{1}{2}} s_u + \left(\widehat{\mathbf{U}}_u^1 + \Delta \widehat{\mathbf{U}}_u^1 \right)^H \Delta \widetilde{\mathbf{H}}_u \sum_{v \in \mathcal{U}, v \neq u} \bar{\mathbf{V}}_v^0 \widehat{\mathbf{V}}_v^1 \mathbf{Q}_v^{\frac{1}{2}} s_v \\ &\approx \left(\widehat{\boldsymbol{\Sigma}}_u^1 + \Delta \widehat{\boldsymbol{\Sigma}}_u^1 + \widehat{\boldsymbol{\Sigma}}_u^1 \Delta \widehat{\mathbf{V}}_u^{1H} \widehat{\mathbf{V}}_u^1 \right) \mathbf{Q}_u^{\frac{1}{2}} s_u + \widehat{\mathbf{U}}_u^{1H} \Delta \widetilde{\mathbf{H}}_u \sum_{v \in \mathcal{U}, v \neq u} \bar{\mathbf{V}}_v^0 \widehat{\mathbf{V}}_v^1 \mathbf{Q}_v^{\frac{1}{2}} s_v \\ &= \underbrace{\left(\widehat{\boldsymbol{\Sigma}}_u^1 + \Delta \widehat{\boldsymbol{\Sigma}}_u^1 \right) \mathbf{Q}_u^{\frac{1}{2}} s_u}_{y_{\text{desired}}} + \underbrace{\widehat{\boldsymbol{\Sigma}}_u^1 \Delta \widehat{\mathbf{V}}_u^{1H} \widehat{\mathbf{V}}_u^1 \mathbf{Q}_u^{\frac{1}{2}} s_u}_{y_{\text{self}}} + \underbrace{\widehat{\mathbf{U}}_u^{1H} \Delta \widetilde{\mathbf{H}}_u \sum_{v \in \mathcal{U}, v \neq u} \bar{\mathbf{V}}_v^0 \widehat{\mathbf{V}}_v^1 \mathbf{Q}_v^{\frac{1}{2}} s_v}_{y_{\text{intf}}}, \end{aligned}$$

where $\|y_{\text{desired}}\|_2^2$, $\|y_{\text{self}}\|_2^2$, and $\|y_{\text{intf}}\|_2^2$ denote the power of the desired signal, self-interference signal, and interference signal from other UEs.

The expected SINR of UE u under imperfect CSI is given as

$$\mathbb{E} [\Gamma_u] \approx \frac{\mathbb{E} \left[\|y_{\text{desired}}\|_2^2 \right]}{\mathbb{E} \left[\|y_{\text{self}}\|_2^2 \right] + \mathbb{E} \left[\|y_{\text{intf}}\|_2^2 \right] + \sigma_n^2}. \quad (\text{B.11})$$

First, We compute $\mathbb{E} \left[\|y_{\text{desired}}\|_2^2 \right]$ as follows

$$\mathbb{E} \left[\|y_{\text{desired}}\|_2^2 \right] = \mathbb{E} \left[\left\| \left(\widehat{\boldsymbol{\Sigma}}_u^1 + \Delta \widehat{\boldsymbol{\Sigma}}_u^1 \right) \mathbf{Q}_u^{\frac{1}{2}} s_u \right\|_2^2 \right] \approx \mathbf{Q}_u \left(\widehat{\boldsymbol{\Sigma}}_u^{1^2} + \mathbb{E} \left[\Delta \widehat{\boldsymbol{\Sigma}}_u^{1^2} \right] \right).$$

For single stream, we compute $\mathbb{E} \left[\left(\Delta \widehat{\boldsymbol{\Sigma}}_u^1 \right)_{11}^2 \right]$ as follows

$$\begin{aligned} \mathbb{E} \left[\left(\Delta \widehat{\boldsymbol{\Sigma}}_u^1 \right)_{11}^2 \right] &= \frac{1}{4} \mathbb{E} \left[\left\| \left(\widehat{\mathbf{V}}_u^1 \right)_{:1}^H \left(\Delta \widetilde{\mathbf{H}}_u \overline{\mathbf{V}}_u^0 \right)^H \left(\widehat{\mathbf{U}}_u^1 \right)_{:1} + \left(\widehat{\mathbf{U}}_u^1 \right)_{:1}^H \Delta \widetilde{\mathbf{H}}_u \overline{\mathbf{V}}_u^0 \left(\widehat{\mathbf{V}}_u^1 \right)_{:1} \right\|_2^2 \right] \\ &\approx \frac{2}{4} \mathbb{E} \left[\left\| \left(\widehat{\mathbf{U}}_u^1 \right)_{:1}^H \Delta \widetilde{\mathbf{H}}_u \overline{\mathbf{V}}_u^0 \left(\widehat{\mathbf{V}}_u^1 \right)_{:1} \right\|_2^2 \right] \\ &= \frac{1}{2} \mathbb{E} \left[\text{vec} \left\{ \left(\widehat{\mathbf{U}}_u^1 \right)_{:1}^H \Delta \widetilde{\mathbf{H}}_u \overline{\mathbf{V}}_u^0 \left(\widehat{\mathbf{V}}_u^1 \right)_{:1} \right\} \text{vec} \left\{ \left(\widehat{\mathbf{U}}_u^1 \right)_{:1}^H \Delta \widetilde{\mathbf{H}}_u \overline{\mathbf{V}}_u^0 \left(\widehat{\mathbf{V}}_u^1 \right)_{:1} \right\}^H \right] \end{aligned}$$

By applying the property of Kronecker and vectorization $\text{vec} \{ \mathbf{ABC} \} = \text{vec} \{ \mathbf{C}^T \otimes \mathbf{A} \} \text{vec} \{ \mathbf{B} \}$, we have

$$\begin{aligned} \mathbb{E} \left[\left(\Delta \widehat{\boldsymbol{\Sigma}}_u^1 \right)_{11}^2 \right] &\approx \mathbb{E} \left[\text{vec} \left(\left(\overline{\mathbf{V}}_u^0 \left(\widehat{\mathbf{V}}_u^1 \right)_{:1} \right)^T \otimes \left(\widehat{\mathbf{U}}_u^1 \right)_{:1}^H \right) \right. \\ &\quad \left. \text{vec} \left(\Delta \widetilde{\mathbf{H}}_u \right) \text{vec} \left(\Delta \widetilde{\mathbf{H}}_u \right)^H \text{vec} \left(\left(\overline{\mathbf{V}}_u^0 \left(\widehat{\mathbf{V}}_u^1 \right)_{:1} \right)^* \otimes \left(\widehat{\mathbf{U}}_u^1 \right)_{:1} \right) \right] \\ &= \mathbb{E} \left[\text{vec} \left(\left(\overline{\mathbf{V}}_u^0 \left(\widehat{\mathbf{V}}_u^1 \right)_{:1} \right)^T \otimes \left(\widehat{\mathbf{U}}_u^1 \right)_{:1}^H \right) \sigma_e^2 \mathbf{I}_{M_{\text{rf}} \times M_{\text{rf}}} \right. \\ &\quad \left. \text{vec} \left(\left(\overline{\mathbf{V}}_u^0 \left(\widehat{\mathbf{V}}_u^1 \right)_{:1} \right)^* \otimes \left(\widehat{\mathbf{U}}_u^1 \right)_{:1} \right) \right] \\ &= \frac{1}{2} \sigma_e^2. \end{aligned}$$

Hence, we obtain

$$\mathbb{E} \left[\|y_{\text{desired}}\|_2^2 \right] \approx Q_u \left(\left(\widehat{\boldsymbol{\Sigma}}_u^1 \right)_{11}^2 + \frac{1}{2} \sigma_e^2 \right) \quad (\text{B.12})$$

Next, we compute $\mathbb{E} \left[\|y_{\text{self}}\|_2^2 \right]$ as follows

$$\begin{aligned} &\mathbb{E} \left[\|y_{\text{self}}\|_2^2 \right] \\ &= \mathbb{E} \left[\left\| \widehat{\boldsymbol{\Sigma}}_u^1 \Delta \widehat{\mathbf{V}}_u^1 \widehat{\mathbf{V}}_u^1 \mathbf{Q}_u^{\frac{1}{2}} s_u \right\|_2^2 \right] = \mathbb{E} \left[\left\| \widehat{\boldsymbol{\Sigma}}_u^1 \left(\widehat{\mathbf{V}}_u^1 \mathbf{T}_u + \widehat{\mathbf{V}}_u^0 \overline{\mathbf{V}}_u^0 \Delta \widetilde{\mathbf{H}}_u \widehat{\mathbf{U}}_u^1 \widehat{\boldsymbol{\Sigma}}_u^{1^{-1}} \right)^H \widehat{\mathbf{V}}_u^1 \mathbf{Q}_u^{\frac{1}{2}} s_u \right\|_2^2 \right] \\ &= \mathbb{E} \left[\left\| \widehat{\boldsymbol{\Sigma}}_u^1 \mathbf{T}_u^H \widehat{\mathbf{V}}_u^1 \widehat{\mathbf{V}}_u^1 \mathbf{Q}_u^{\frac{1}{2}} s_u + \left(\overline{\mathbf{V}}_u^0 \Delta \widetilde{\mathbf{H}}_u \widehat{\mathbf{U}}_u^1 \widehat{\boldsymbol{\Sigma}}_u^{1^{-1}} \right)^H \widehat{\mathbf{V}}_u^{0H} \widehat{\mathbf{V}}_u^1 \mathbf{Q}_u^{\frac{1}{2}} s_u \right\|_2^2 \right]. \end{aligned}$$

Since $\widehat{\mathbf{V}}_u^{0H} \widehat{\mathbf{V}}_u^1 = 0$, we have

$$\mathbb{E} \left[\|y_{\text{self}}\|_2^2 \right] = \mathbb{E} \left[\left\| \widehat{\boldsymbol{\Sigma}}_u^1 \mathbf{T}_u^H \widehat{\mathbf{V}}_u^1 \widehat{\mathbf{V}}_u^1 \mathbf{Q}_u^{\frac{1}{2}} s_u \right\|_2^2 \right].$$

Since we consider single stream transmission, from (5.17), $\Psi_u = 0$, thus, from (5.16), $\mathbf{T}_u^H = 0$. Therefore

$$\mathbb{E} \left[\|y_{\text{self}}\|_2^2 \right] = 0 \quad (\text{B.13})$$

Last, we compute $\mathbb{E} \left[\|y_{\text{intf}}\|_2^2 \right]$ as follows

$$\begin{aligned} \mathbb{E} \left[\|y_{\text{intf}}\|_2^2 \right] &= \mathbb{E} \left[\left\| \sum_{v \in \mathcal{U}, v \neq u} \left(\hat{\mathbf{U}}_u^1 \right)_{:1} \Delta \tilde{\mathbf{H}}_u \bar{\mathbf{V}}_v^0 \left(\hat{\mathbf{V}}_v^1 \right)_{:1} Q_v^{\frac{1}{2}} s_v \right\|_2^2 \right] \\ &\approx \sum_{v \in \mathcal{U}, v \neq u} \mathbb{E} \left[\left\| \left(\hat{\mathbf{U}}_u^1 \right)_{:1} \Delta \tilde{\mathbf{H}}_u \bar{\mathbf{V}}_v^0 \left(\hat{\mathbf{V}}_v^1 \right)_{:1} Q_v^{\frac{1}{2}} s_v \right\|_2^2 \right] \\ &= \sum_{v \in \mathcal{U}, v \neq u} Q_v \mathbb{E} \left[\text{vec} \left(\left(\hat{\mathbf{U}}_u^1 \right)_{:1} \Delta \tilde{\mathbf{H}}_u \bar{\mathbf{V}}_v^0 \left(\hat{\mathbf{V}}_v^1 \right)_{:1} \right) \text{vec} \left(\left(\hat{\mathbf{U}}_u^1 \right)_{:1} \Delta \tilde{\mathbf{H}}_u \bar{\mathbf{V}}_v^0 \left(\hat{\mathbf{V}}_v^1 \right)_{:1} \right)^H \right] \\ &= \sum_{v \in \mathcal{U}, v \neq u} \mathbb{E} \left[\text{vec} \left(\left(\bar{\mathbf{V}}_v^0 \left(\hat{\mathbf{V}}_v^1 \right)_{:1} \right)^T \otimes \left(\hat{\mathbf{U}}_u^1 \right)_{:1}^H \right) \text{vec} \left(\Delta \tilde{\mathbf{H}}_u \right) \right. \\ &\quad \left. \text{vec} \left(\Delta \tilde{\mathbf{H}}_u \right)^H \text{vec} \left(\left(\bar{\mathbf{V}}_v^0 \left(\hat{\mathbf{V}}_v^1 \right)_{:1} \right)^* \otimes \left(\hat{\mathbf{U}}_u^1 \right)_{:1} \right) \right] \\ &= \sigma_e^2 \sum_{v \in \mathcal{U}, v \neq u} Q_v \end{aligned} \quad (\text{B.14})$$

By putting $\mathbb{E} \left[\|y_{\text{desired}}\|_2^2 \right]$, $\mathbb{E} \left[\|y_{\text{self}}\|_2^2 \right]$, and $\mathbb{E} \left[\|y_{\text{intf}}\|_2^2 \right]$ from (B.12), (B.13), and (B.14), respectively, into (B.11), we proved the Theorem 1. ■