

UCLA

UCLA Electronic Theses and Dissertations

Title

Spectrum Sharing for Massive Internet-of-Things Communications

Permalink

<https://escholarship.org/uc/item/7001j86r>

Author

Hattab, Ghaith

Publication Date

2019

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Spectrum Sharing for Massive Internet-of-Things Communications

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Electrical and Computer Engineering

by

Ghaith Nu'man Mohammed Hattab

2019

© Copyright by
Ghaith Nu'man Mohammed Hattab
2019

ABSTRACT OF THE DISSERTATION

Spectrum Sharing for Massive Internet-of-Things Communications

by

Ghaith Nu'man Mohammed Hattab

Doctor of Philosophy in Electrical and Computer Engineering

University of California, Los Angeles, 2019

Professor Danijela Cabric, Chair

The emergence of the massive Internet-of-things (IoT) market, i.e., applications with large-scale deployment of Internet-enabled devices, is transforming many sectors, including cities, logistics, retail, and agriculture. Nevertheless, a large-scale uptake of such applications is intertwined with connectivity challenges due to the unique traffic characteristics of massive IoT and the high density of IoT devices.

The main objective of this dissertation is to develop spectrum sharing solutions to enable massive IoT connectivity over the unlicensed and licensed bands. To this end, we unravel the interplay between the intra-network sharing and inter-network sharing, i.e., how a high density of low-cost IoT devices should share the spectrum among themselves while still harmoniously coexist with other wireless networks.

We first present a framework to model and analyze ultra-narrowband (UNB) communications, where IoT devices randomly access the unlicensed spectrum using extremely narrowband signals. The framework enables tractable analysis of large-scale networks, identifying fundamental limits and proposing enhancements that help UNB networks support a higher density of low-rate IoT devices. We then present an access scheme that supports applications with higher rates over unlicensed bands with sensing regulations. To this end, we revisit spectrum sensing for massive IoT and propose a sensing-based

system, where base stations with narrowband receivers locally share and process sensing reports to scan a wideband spectrum, identifying resources at a fine spectral resolution and aggressively reusing them over space. Next, for applications with strict quality-of-service guarantees, we consider cellular connectivity, where we develop a shared spectrum access protocol, leveraging drones as IoT data aggregators. The proposed protocol maximizes the energy-efficiency of IoT devices while still protecting existing cellular users from interference. Last, we analyze and optimize the densification of small cells to maximize coverage and rate of users and IoT devices.

To deliver these contributions, we leverage tools from stochastic geometry, communication theory, and optimization, working across the theoretical analysis and practical algorithm development. We validate the effectiveness of the proposed solutions via comprehensive simulations of wide-area networks. We glean several design guidelines and recommendations throughout the dissertation, which can be of relevance to wireless communications engineers, network operators, and standard bodies.

The dissertation of Ghaith Nu'man Mohammed Hattab is approved.

Lara Dolecek

Izhak Rubin

Lieven Vandenberghe

Danijela Cabric, Committee Chair

University of California, Los Angeles

2019

To Areej, Rayyan, and my parents

TABLE OF CONTENTS

List of Figures	xii
List of Tables	xv
Acknowledgments	xvi
Vita	xviii
1 Introduction	1
1.1 Massive IoT Applications and Communications Properties	3
1.2 Current Approaches Over the Unlicensed and Licensed Spectrum	4
1.2.1 Unlicensed-based access	4
1.2.2 Licensed-based access	7
1.3 Challenges and Communications Requirements	9
1.4 Contributions	10
1.4.1 Analysis and optimization of ultra-narrowband communications	10
1.4.2 Revisiting spectrum sensing for massive IoT connectivity	11
1.4.3 Energy-efficient shared spectrum access of massive IoT and UEs	12
1.4.4 Coverage and rate maximization of heterogeneous cellular networks	13
1.5 Dissertation Organization and Related Publications	14
2 Random Access via UNB Communications	15
2.1 Related Work	16
2.2 Chapter Contributions and Organization	17
2.3 System Model	19

2.3.1	Network topology	19
2.3.2	Transmission model	19
2.3.3	Channel model	21
2.3.4	Performance metrics	21
2.4	Performance Analysis of UNB Protocols	22
2.4.1	Performance with nearest BS association	22
2.4.2	Performance with no BS association	25
2.4.3	The role of geographical diversity	26
2.4.4	Random versus pseudorandom frequency hopping	27
2.5	Enhancements and Optimizations of UNB Protocols	29
2.5.1	Low-cost multiband access	29
2.5.2	Optimizing the number of transmissions	34
2.6	Simulation Results	36
2.6.1	Impact of synchronization and hopping patterns	36
2.6.2	Success probability comparison	38
2.6.3	Impact of the number of bands	39
2.6.4	Impact of the number of repetitions	40
2.6.5	Transmission capacity comparison	42
2.7	Conclusions	43
3	Distributed Wideband Sensing-based Access	45
3.1	Related Work	46
3.2	Chapter Contributions and Organization	47
3.3	System Model and Proposed Architecture	49
3.3.1	Network topology	49

3.3.2	Proposed sensing-based architecture	51
3.4	The Sensing Assignment Scheduler	52
3.4.1	A heuristic sensing assignment scheduler	56
3.4.2	Computational complexity and implementation of the heuristic scheduler	57
3.4.3	Numerical validation	58
3.5	Diffusion-based Distributed Sensing	59
3.5.1	Computational complexity and implementation of the distributed sensing algorithm	62
3.5.2	Performance analysis of the deflection coefficient	63
3.5.3	The proposed sensing algorithm as a REM	66
3.6	Distributed Allocation of Spatio-spectral Resources	68
3.6.1	Network utility optimization	69
3.6.2	Fast distributed resource allocation	71
3.6.3	Implementation of the fast allocation algorithm	71
3.7	Simulation Results	72
3.7.1	Performance evaluation of the distributed sensing algorithm	73
3.7.2	Performance evaluation of the distributed resource allocation algorithm	79
3.7.3	Case study: IoT for public parks	82
3.8	Conclusions	84
4	Massive IoT over Drone-assisted Cellular Networks	86
4.1	Related Work	86
4.2	Chapter Contributions and Organization	88

4.3	System Model	89
4.3.1	Network topology	89
4.3.2	Channel models	91
4.3.3	Performance metrics	92
4.4	Proposed Transmission Protocol	94
4.4.1	Characterization of average resources in the proposed and existing protocols	95
4.4.2	Analysis of IoT interference into UEs	98
4.5	IoT Energy-Efficiency Maximization	100
4.5.1	Average IoT EE in the SC-SD case	102
4.5.2	IoT interference constraint	104
4.5.3	SC-SD stochastic optimization framework	105
4.5.4	EE optimization in heterogeneous single-cell multi-drone case . . .	107
4.6	Simulation Results	108
4.6.1	Validation of the stochastic optimization frameworks	109
4.6.2	Performance comparison with existing protocols	113
4.6.3	IoT device lifetime comparison	116
4.7	Conclusions	119
5	Coverage and Rate Maximization over Heterogeneous Cellular Networks . . .	121
5.1	Related Work	122
5.2	Chapter Contributions and Organization	123
5.3	System Model	126
5.3.1	Network topology and channel model	126
5.3.2	Multi-antenna transmission modes	127

5.3.3	Spectrum allocation	127
5.3.4	Biased user association	128
5.4	Coverage Probability Analysis	130
5.4.1	Exact coverage probability	130
5.4.2	Coverage probability approximation	132
5.5	Coverage Maximization Framework	134
5.5.1	Special network settings	135
5.5.2	Suboptimal user association	139
5.5.3	Implementation	140
5.6	Rate Maximization Framework	142
5.6.1	Optimal spectrum allocation	143
5.6.2	Optimal user association	144
5.6.3	Suboptimal multi-objective user association	144
5.6.4	Implementation	146
5.7	Simulation Results	147
5.7.1	Evaluation of the coverage maximization framework	147
5.7.2	Evaluation of the rate maximization framework	151
5.7.3	Performance of the proposed policies in a realistic channel	154
5.8	Conclusions	156
6	Conclusion	158
6.1	Summary of Contributions and Lessons Learned	158
6.2	Future Work	161
6.2.1	Ultra-reliable low-latency communications over wide areas	161
6.2.2	Advanced signal processing of random access signals	162

6.2.3	Spectrum occupancy dissemination for massive IoT	162
6.2.4	Beyond ALOHA-based and sensing-based access	163
A	Performance Analysis of UNB Networks in Chapter 2	165
A.1	Proof of Theorem 1	165
A.2	Proof of Theorem 2	166
B	Analysis of the Sensing Algorithm in Chapter 3	167
B.1	Statistics of the Filtered Energy Measurements	167
B.2	Analysis of the Proposed Sensing Algorithm	168
C	Analysis of Drone-assisted Cellular Networks in Chapter 4	169
C.1	Proof of Theorem 6	169
C.2	Distribution of Signal and Interference Powers at the IoT Device	171
C.3	Convexity Analysis of the EE Maximization Framework	171
D	Analysis of Multi-antenna HetNets in Chapter 5	174
D.1	Coverage Probability Approximation	174
D.2	Concavity of Coverage Maximization	175
D.3	Concavity of Rate Maximization	176
	References	177

LIST OF FIGURES

1.1	There are several emerging IoT markets such as critical IoT, tactile IoT, and massive IoT.	2
2.1	Spectrogram of one Sigfox packet of bandwidth 600Hz. The packet is sent three times at randomly selected frequencies within a multiplexing band of bandwidth 200KHz.	17
2.2	Access can be slotted or unslotted in time and/or frequency.	21
2.3	Two variants of UNB networks are considered: with and without BS association.	22
2.4	Impact of geographical diversity on the network resources needed, i.e., the number of bands and the density of BSs, to achieve a target success probability constraint.	28
2.5	Two deployment approaches: few expensive BSs or many low-cost ones.	30
2.6	Two low-cost multiband access protocols: slotted and unslotted.	31
2.7	Success probability for the Sigfox protocol under different access cases.	38
2.8	Success probability comparison of different UNB protocols.	39
2.9	Success probability with variations of the number of bands, M ($\tau = 0\text{dB}$).	40
2.10	Success probability with variations of N ($\tau = 0\text{dB}$).	41
2.11	Transmission capacity for a given success probability constraint ($\tau = 5\text{dB}$).	42
3.1	An illustration of the network architecture.	53
3.2	Relative gap between the integer program (3.3) and other scheduling algorithms.	59
3.3	Deflection coefficient comparison between the proposed and the non-cooperative sensing algorithms.	65
3.4	Deflection coefficient versus SNR.	67

3.5	A REM for a single indoor WiFi AP using actual RSS measurements and the proposed algorithm.	68
3.6	Illustration of distributed allocation over a small network.	73
3.7	Performance with variations of the energy threshold.	76
3.8	The spatio-spectral footprints. ‘o’ and ‘x’ denote available and busy decisions at their locations, respectively. For narrowband systems, BSs with their assignments are highlighted in a different color.	77
3.9	Impact of communication radius on the performance of cooperative schemes.	79
3.10	Comparison of the different spatio-spectral resource allocation schemes in terms of the sum rate, the number of users served, and UL SINR per device.	81
3.11	Relative INR at incumbents.	82
3.12	An illustration of the large-scale network set-up.	83
3.13	Average number of scheduled devices per scheme.	84
4.1	The drone-assisted cellular architecture and one spatial realization of the network topology.	93
4.2	Comparison of the different TDD transmission protocols.	96
4.3	The UE DL coverage performance with and without IoT devices.	101
4.4	Validation of the theoretical EE.	111
4.5	Validation of the theoretical coverage using the optimal transmit power.	112
4.6	EE performance for different schemes in the SC-MD case (CAT-0 parameters and $\rho_{\text{dB}} = -12\text{dB}$).	113
4.7	UE DL and UL performance for the different protocols.	115
4.8	IoT EE performance under different protocols.	117
4.9	IoT energy-efficiency vs. UE spectral efficiency ($\lambda_{\text{M}} = 10$ and $h_{\text{D}} = 50\text{m}$).	118
4.10	Distribution of IoT device lifetime ($h_{\text{D}} = 50\text{m}$).	120

5.1	Illustration of a two-tier network. Here, $B_1 = 1$ for the macro tier, whereas tier-2 BSs artificially extend the coverage area to offload users from tier-1 BSs.	129
5.2	Validation of the theoretical analysis, where markers denote Monte Carlo simulations.	134
5.3	Impact of small-cell bias and transmit power on coverage ($\tau = 0\text{dB}$).	149
5.4	Coverage performance under different transmission modes ($\tau = 0\text{dB}$).	150
5.5	Impact of network densification on coverage ($\tau = 0\text{dB}$).	152
5.6	Mean rate and utility with variations of B_2 ($B_1 = 1$ and $\tau = 5\text{dB}$).	153
5.7	User mean rate and association probability for different user association policies ($\tau = 5\text{dB}$).	154
5.8	Coverage performance under the 3GPP channel model.	155
5.9	Rate performance under the 3GPP channel model ($U = [8, 4]$).	156

LIST OF TABLES

1.1	Properties and regulations of different unlicensed bands	7
2.1	Examples of UNB technologies	17
2.2	Main parameters	37
3.1	Simulation parameters	74
4.1	Main parameters	110
4.2	Battery lifetime parameters [3GP15a,3GP15b]	119

ACKNOWLEDGMENTS

I set my sight on pursuing a PhD degree almost a decade ago because I have been fascinated by how wireless communication systems have reshaped the way we communicate, work, and live. I believed that the PhD would help me break into the research and development of wireless technologies, not only taking me to their boundaries but also enabling me to push their limits. Little did I know that the PhD years would not only equip me with the skills and knowledge I need to become an independent researcher, but they would also transform the way I approach many aspects of life, technical and personal. After all, I have actually spent almost 50% of my adult life here at UCLA as a PhD student!

My PhD tenure has begun by joining an incredible lab like CORES, for which I'm greatly indebted to its founder and my advisor, Prof. Danijela Cabric. During my graduate career, Danijela has taught me that the research endeavor starts by formulating the right questions and culminates when complex results are fully grasped and elegantly explained to others. Danijela has not only given me the opportunity to explore several research directions, but she has also pushed me forward to pursue them. Looking ahead, I will carry forward what she has taught me, yet I will still miss our group meetings and regular brainstorming sessions!

I'm also grateful to Prof. Lara Dolecek, Prof. Izhak Rubin, and Prof. Lieven Vandenberghe. Before being my thesis committee, they have been great instructors, teaching me the necessary foundations that I have built my dissertation on. I have learned a lot from their courses on matrix analysis, discrete-time stochastic processes, and convex optimization. I'm also very fortunate to have learned from Prof. Ali Sayed. His seminal work on distributed adaptation and learning and his feedback on my developed distributed processing algorithms have been tremendously helpful.

I'm incredibly lucky to be part of CORES lab. I have had great times with my lab mates and collaborators: Mihir Laghate, Shailesh Chaudhari, Han Yan, Chun-Hao Liu,

Samer Hanna, Veljko Boljanovic, Tom Vermeulen, Fereidoun H. Panahi, and Rico Mendrzik. Our daily interactions have really uplifted my spirit. I also have made great friendships beyond the lab: Anas Alanqar, Ahmed Hareedy, Mohammed Karmoose, Hawraa Salami, Mahmoud Rashad, Yahya Ezzeldin, Kun Yuan, Hemant Sagar, and many more. My academic journey has also been enriched with great internships at Nokia Bell Labs and Qualcomm Research. Special thanks to my managers Amitava Ghosh and Farhad Meshkati who have hosted me; I'm also grateful to meet great mentors with immense industry experience in wireless communication systems: Eugene Visotsky, Mark Cudak, Prakash Moorut, Fred Vook, Mostafa Khoshnevisan, Peerapol Tinnakornsriruphap, Rajat Prakash, and Vinay Joseph.



I can now write hundreds of pages about my research and work, but what I cannot inscribe is how immensely lucky I am to have my family around me in my life. I'm truly blessed to have my wife, Areej, and my son, Rayyan, with me throughout this journey; they have brought me jubilation and purpose. Words also cannot describe my gratitude to my parents, Laila and Numan, who have always been there for me and my little family. Without exaggeration, Areej, Rayyan, Laila, and Numan have been as involved as I have been in this PhD journey. Likewise, I'm also grateful for the perpetual support and encouragement from my amazing extended family, Samir and Ghada, my great brothers, cousins, and friends here and back home.

VITA

2008 – 2012 American University of Sharjah, Sharjah, United Arab Emirates

2012 – 2014 Queen's University, Kingston, Ontario, Canada

2014 – 2019 University of California, Los Angeles, United States

Summer 2017 Nokia Bell Labs, Arlington Heights, Illinois, United States

Summer 2018 Qualcomm Wireless R&D, San Diego, California, United States

CHAPTER 1

Introduction

Wireless communication systems have transformed our lives, profoundly impacting our society and economy. Indeed, many wireless devices, e.g., mobile phones, have become a necessity instead of a mere accessory, making humans monthly generate exabytes of data (that is 10^{18} bytes!) [Cis19a]; and while wireless networks are still pushing the limits to provide unprecedented data rates, they are also bringing new disruptive applications, primarily by connecting myriads of physical objects to the Internet, from sensors and machines [AGM15] to home appliances [AZZ17], vehicles [ASS15], plants [VKW17], and even livestock [DDT17].

Connecting these physical *things* has become commonly referred to as the *Internet-of-things* (IoT), and we anticipate to connect billions of devices more in the next few years [Cis19b], potentially creating IoT markets worth trillions of dollars [Mck18]. Fig. 1.1 illustrates the different IoT markets that are set to push the limits of current networks to new frontiers. For example, a WiFi access point (AP) or a cellular base station (BS) can now support tens of users, providing different human-type services, e.g., video streaming, with device data rates of few hundreds of Mbps and latencies of few tens of milliseconds. Yet, with the emergence of IoT applications, we need wireless networks to support the following growing markets:

- **The critical IoT market:** The objective here is to connect factory machines, e.g., automated guided vehicles, at extremely low latency, i.e., 1ms, and ultra-high reliability, i.e., block error rates lower than 10^{-6} . This market is integral to the industry 4.0 vision [CWS18], and it has become one of the use cases identified by the Inter-

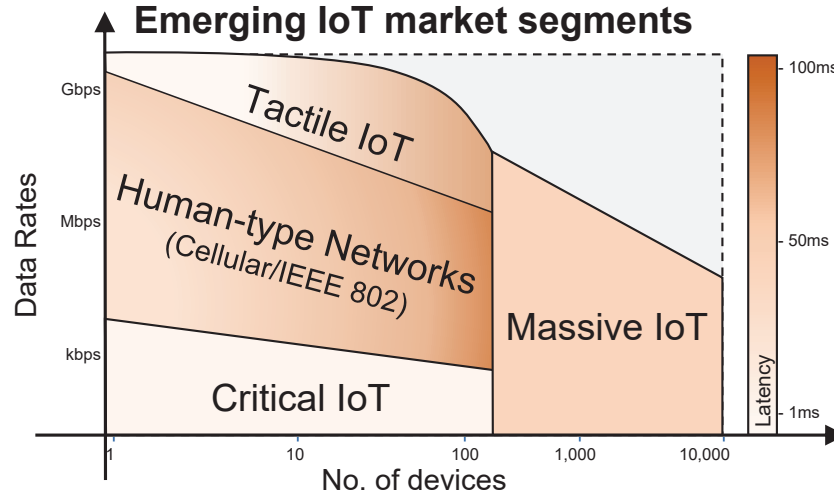


Figure 1.1: There are several emerging IoT markets such as critical IoT, tactile IoT, and massive IoT.

national Mobile Telecommunications (IMT) for next-generation networks [ITU15].

- **The tactile IoT market:** The objective here is to enable seamless interactions between humans and machines via haptic and visual feedback. This market is expected to bring us real-time interactive systems such as virtual and augmented reality [ADA17], and thus the key requirements are high availability, gigabit connectivity, and low latency [SAD16].
- **The massive IoT market:** The objective here is to provide wide-area connectivity to a large number of low-cost IoT devices; for example, this market, also known as massive machine-type communications (mMTC), has become another use case for the IMT-2020 vision, calling future networks to support IoT connection densities of $10^6/\text{km}^2$ [ITU15].

The research in this dissertation, in spirit, is centered around enabling wide-area connectivity for massive IoT applications. Thus, we delve into more details about massive IoT applications, discussing their communications properties and requirements.

1.1 Massive IoT Applications and Communications Properties

Massive IoT refers to applications that require a large-scale deployment of low-cost IoT devices, unlocking a myriad of services across many spheres: cities, logistics, retail, and agriculture. For example, smart cities are now integrating IoT devices with street lights to reduce their power consumption, with traffic signals to improve transportation, and with buildings to monitor their health [MAY17]. In logistics, IoT devices improve fleet operations and asset tracking [Bra15]. In agriculture, IoT sensors can help in water management, pest control, and collecting farm-related physical parameters, e.g., soil moisture [VKW17].

Realizing a large-scale uptake of the aforementioned applications requires wide-area connectivity. Nevertheless, the traffic characteristics of massive IoT set them apart from human-type traffic, requiring wholesale changes to existing wide-area networks. The main traffic properties of massive IoT include:

- **Connection density:** A single BS needs to support thousands of IoT devices, while current BSs can support tens of users.
- **Device types:** A wide landscape of devices including sensors, utility meters, wearables, and surveillance systems. The payload is typically small relative to human-type traffic payloads, e.g., from bytes to few megabytes.
- **Traffic direction:** IoT traffic is predominantly in the uplink, i.e., the IoT device reports sensed information to the network. Downlink traffic is typically limited for control information, acknowledgments, or software updates.
- **Temporal characteristics:** The majority of massive IoT applications are delay-tolerant, where sensors either report data in a periodic manner or when an event is triggered.
- **Key performance indicators:** Many of IoT devices are battery-powered or deployed deep indoors, e.g., utility meters in basements; this centralizes the design of

wireless communication protocols around energy-efficiency and coverage instead of network capacity or spectral efficiency.

The aforementioned properties of massive IoT communications have spurred mobile network operators (MNOs), technology vendors, and standard bodies, to develop tailored wide-area networks in both the unlicensed and licensed bands. Next, we discuss current efforts to enable massive IoT connectivity.

1.2 Current Approaches Over the Unlicensed and Licensed Spectrum

A new class of low-power wide-area (LPWA) networks for IoT applications has emerged, consisting of various proprietary technologies. The two most prevalent ones, which use ALOHA-based access, are LoRa [Sor17], that uses chirp spread spectrum, and Sigfox that relies on frequency hopping (FH) and ultra-narrowband signals [Sig19]. In addition, MNOs are now leveraging the ubiquitous infrastructure of cellular networks to provide wide-area connectivity for IoT services. In fact, two fronts of ongoing cellular standardization can accelerate the adoption of massive IoT over these networks: the first one is the newly introduced user categories tailored for low-cost IoT devices, e.g., LTE-M and NB-IoT [RVX16, WLA17], and the second one is the use of unlicensed access over cellular networks [KJB17, Mul17].

1.2.1 Unlicensed-based access

The unlicensed spectrum, although free, has become an invaluable resource toward enabling massive IoT applications. There are many candidate unlicensed bands for massive IoT, yet each one has specific regulations to follow, which can also vary by the country or region. The major differences among these bands typically include:

- **Duty cycle:** The maximum duration for an active transmission in a one-hour period
- **Transmit power:** The maximum transmit power or the maximum equivalent isotrop-

ically radiated power (EIRP)

- **Polite spectrum access:** Whether sensing, e.g., listen-before-talk (LBT) or dynamic frequency selection (DFS), is required.

We also note that there are other regulations on the spectral mask used by the device, the support of power control, the type of modulation used, and restrictions to indoor deployments. The aforementioned differences have a direct influence on the characteristics of massive IoT networks, and so we discuss the key regulations and properties of three main bands: the sub-1GHz, the 2.4GHz, and the 5GHz bands.

1.2.1.1 The sub-1GHz bands

The sub-1GHz spectrum is primarily used by LPWA networks and WiFi HaLow (IEEE 802.11ah) because of its favorable propagation conditions, i.e., lower path and penetration losses. This helps the network deploy fewer BSs when covering a wide area, compared to networks using higher bands. Nevertheless, the sub-1GHz bands are not universal, and the spectrum available is rather limited. For instance, in the US (or Europe), the band comprises 26MHz (or 7MHz) of the spectrum at 902-928MHz (or 863-870MHz).

Because polite spectrum access is not mandated in sub-1GHz bands in Europe, there are stricter regulations on the transmit power and the duty cycle. For example, the sub-band 868.0-868.6MHz, which is used by LoRa and Sigfox, has a peak radiated power of 14dBm and a one percent duty cycle, i.e., an IoT device can transmit only for a duration of 36s per hour. In the US, the peak power is higher, e.g., 30dBm for FH systems, and there are no restrictions on the duty cycle. However, the dwell time, i.e., the time on air of a single transmission over a particular channel, is enforced for FH systems. Such regulation can be an issue for low-rate IoT networks that use FH. For example, when the bandwidth of the hopping channel is less than 250KHz, then the device needs to use at least 50 hopping frequencies while occupying each one for at most 0.4s within a 20s period.

1.2.1.2 The 2.4GHz bands

Different from the fragmented sub-1GHz bands worldwide, the 2.4GHz band is global, enabling economies of scale. In addition, there are more channels spanning the 2402-2484MHz spectrum, and the peak power is typically 20-30dBm, depending on the region. In addition, there are no regulations on the duty cycle, yet in the US, the dwell time can be limited, whereas polite spectrum access may be further required in Europe. The key issue of this band is the presence of several technologies such as WiFi, Bluetooth, and Zigbee.

1.2.1.3 The 5GHz bands

The wide available spectrum at 5GHz, which is approximately 500MHz, has motivated the development of Third Generation Partnership Project (3GPP) cellular standards with unlicensed access [LMR17]. While the main driver for unlicensed-based cellular networks has been boosting the network capacity, the 5GHz spectrum is now envisioned to also support private IoT networks. For example, the MulteFire network aims to provide IoT-based services by connecting LTE-M and NB-IoT devices over unlicensed carriers, without anchored licensed channels [Mul17]. This spectrum is divided into several bands, and the peak transmit power in each one ranges from 23dBm to 30dBm. In addition, LBT is mandated at 5GHz in Europe and Japan, whereas DFS is required in some bands in the US. For this reason, cellular-based solutions use sensing-based access.

In summary, each band enables different IoT networks, depending on the regulations enforced. While the sub-1GHz spectrum has the lowest propagation losses, some of its regulations restrict the number of packets an IoT device can report per hour, limiting the types of applications that can be used. At 2.4GHz and 5GHz, there are more available channels, and the increase in path loss can be compensated by using a higher transmit power or additional antennas as the wavelength is shorter at these bands relative to sub-1GHz bands. Also, cellular-based solutions can rely on dense deployments of small

Table 1.1: Properties and regulations of different unlicensed bands

	Bands (MHz)	Regions ¹	Duty cycle	Peak power ² (dBm)	polite access	Comments	IoT networks
Sub-1	433-434	Europe	< 10%	10dBm EIRP	No	BW limit may apply	LoRa, Sigfox, Telensa, WavIoT Weightless, IEEE 802.11ah
	779-787	China	< 0.1%	12dBm EIRP	No	–	
	863-870	Europe	0.1 – 10%	14-27dBm EIRP	LBT ³	Divided into subbands ⁴	
	902-928	US	No	24-30dBm	No	Dwell time limit may apply	
2.4	2402-2484 ⁵	Europe	No	20dBm	LBT	–	Ingenu Wi-SUN
		US	No	30dBm	No	Dwell time limit may apply	
5.0	5150-5350 ⁶	Europe	No	23dBm EIRP	LBT/DFS	Indoors only	Cellular variants that support LTE-M and NB-IoT such as MulteFire
		US	No	24-30dBm	DFS	DFS in 5250-5350	
		China	No	23dBm EIRP	DFS	Indoors only	
	5470-5725 ⁷	Europe	No	30dBm EIRP	LBT/DFS	–	
		US	No	24dBm	DFS	–	
		Japan	No	Depends on BW	LBT/DFS	–	
	5725-5875	Europe	No	33dBm EIRP	DFS	Fixed wireless service only	
		US	No	30dBm	No	–	
		China	No	33dBm EIRP	DFS	–	

¹ These are sample regions. In Europe, rules can be found in ETSI documents, whereas they can be found in the Federal Communications Commission (FCC) Title 47 Part 15 for the US.

² Some regulations are defined based on the peak transmit power, whereas others are based on the EIRP, which typically includes antenna gains.

³ The use of LBT is optional and its presence affects the duty cycle limit.

⁴ Different subbands have different duty cycles, peak transmit powers, and additional regulations on the signal bandwidth (BW).

⁵ The 2.4GHz is global, making regulations very similar across countries and regions.

⁶ This can be further divided into 5150-5250MHz and 5250-5350MHz as regulations slightly differ between them in some regions.

⁷ There are other bands for considerations, e.g., 5350-5470MHz and the DSRC band 5850-5925GHz. In China, 5470-5725MHz is under consideration.

cells. Nevertheless, sensing-based access is typically enforced in these bands, and in fact needed to protect IoT devices from the high interference in such crowded bands. Table 1.1 provides a taxonomy of some regulations for these bands in certain regions.

1.2.2 Licensed-based access

3GPP has recognized the potential revenues IoT applications can bring to MNOs; thus, it has been actively enhancing cellular standards to support machine-type applications and achieve the following objectives [3GP15a]:

- **Connection density:** Each cell should be able to support approximately 50,000 devices that generate small payloads.
- **Reduced complexity:** Each device should cost a fraction of current cellular devices, e.g., the target NB-IoT module price is \$5.
- **Extended coverage:** The link budget, or maximum coupling loss, should be +20dB more than what current cellular networks offer.
- **Improved power efficiency:** Each device can live up to 10 years on a battery with a capacity of 5Wh.

To materialize these objectives, 3GPP has introduced several user categories with lower cost and complexity in recent releases, e.g., Rel. 13 [RVX16] and Rel. 14 [HLL17]. The most dominant ones are LTE-M (or CAT-M1) and NB-IoT (or CAT-M2), which are half-duplex single-antenna devices. Compared to more sophisticated user categories, e.g., CAT-4 with full-duplex two-antenna transceivers, the modem complexity of NB-IoT is less than 15% [Nok16], which should motivate technology vendors to build IoT devices at a mass scale or in a disposable manner. To extend the coverage and connect more IoT devices over an LTE channel, LTE-M and NB-IoT use channels of bandwidth 1.4MHz and 200KHz, respectively. Parallel to these hardware complexity reductions, new power classes are introduced to increase the lifetime of these IoT devices. For example, most user categories are required to support up to 23dBm of transmit power, but NB-IoT allows devices to have a 14dBm power limit [HLL17]. Further, new power optimization modes are introduced, e.g., power saving mode (PSM) and extended discontinuous reception (eDRX) [XYW18]. PSM allows the device to shut off its radio frequency (RF) module, entering a new state with lower power consumption compared to the already supported idle state. Similarly, eDRX allows the device to significantly reduce the control channel monitoring periodicity, from 2.56s to 43min [RVX16].

These aforementioned cellular enhancements should bring new revenue streams for MNOs, creating new carrier plans tailored for IoT applications and services.

1.3 Challenges and Communications Requirements

Most challenges in enabling massive IoT connectivity stem from the lack of a dedicated spectrum for IoT networks and the high connection density of low-cost IoT devices.

First, IoT networks have to share the spectrum with other networks over the unlicensed bands or with cellular users over the licensed spectrum. On the one hand, unlicensed-based access needs to address the *intra-network sharing*, i.e., how a large density of low-cost IoT devices should share the spectrum, and the *inter-network sharing*, i.e., the coexistence of IoT networks with other incumbent networks. At the same time, access over the sub-1GHz bands must adhere to restrictions on duty cycles and dwell times, whereas access over higher unlicensed bands requires novel sensing algorithms that can reliably identify access opportunities or *whitespaces*. On the other hand, integrating IoT devices into cellular networks requires harmonious coexistence of these devices with current cellular users, henceforth denoted as UEs.

Second, the high density of IoT devices can quickly congest the network, increasing collisions and interference. For this reason, further enhancements of existing ALOHA-based access are needed to push their current limits and make them scalable. In addition, sensing-based solutions must identify a large number of narrowband channels in a wideband spectrum and aggressively reuse them over space. The former requires sensing at a fine spectral resolution, e.g., sensing a spectrum of 500MHz at 180KHz resolution. The latter requires capturing the footprints' of active incumbents at a fine spatial resolution, e.g., reusing the same channels across BSs or APs that are 100-200m apart. Last, when a large density of IoT devices uses cellular connectivity, their access requests can quickly overload the control channel, delaying or barring existing UEs from access [[AAS17](#), [BCZ15](#)].

1.4 Contributions

This dissertation, fundamentally, aims to develop several spectrum sharing solutions for massive IoT wide-area connectivity. At a high level, the dissertation first examines two approaches over the unlicensed bands: one uses ALOHA random access and the other one is sensing-based access. Then, it explores system-level optimizations for cellular networks, maximizing the energy-efficiency, coverage, and rate of IoT devices while still protecting UEs. The specific contributions of this dissertation are outlined next.

1.4.1 Analysis and optimization of ultra-narrowband communications

We begin with ultra-narrowband (UNB) communications, an emerging paradigm that relies on random, or ALOHA, access. In particular, the IoT device sends packets at any time using extremely narrowband signals, with restrictions primarily on duty cycles and dwell times. The motivations of using random access for massive IoT are twofold. First, it reduces the control overhead, which is critical given that the IoT payload is generally small. Second, it relaxes synchronizing devices with the network, which reduces the complexity, cost, and energy consumption of the IoT device.

The scalability of UNB networks and their coexistence capabilities with other incumbents yet remain unexplored. To this end, in this contribution, we present a framework to analyze and model UNB networks on a large scale. The framework captures the unique properties of UNB communications such as the use of signal repetition, the lack of time-frequency synchronization, and the absence of cell selection. The framework uses stochastic geometry [Hae12], enabling tractable analysis of key performance metrics such as coverage and connection density. Closed-form expressions of several UNB protocols are presented, and the key network parameters that affect their scalability and coexistence capabilities are examined. Equally important, leveraging these analytical expressions, we present low-cost protocol enhancements to further push the limits of UNB networks, enabling a single BS to serve thousands of IoT devices over unlicensed bands.

1.4.2 Revisiting spectrum sensing for massive IoT connectivity

Although UNB communications is a promising paradigm, it is only suitable for very low-rate applications. To provide services to applications that require Mbps data rates, wider channels are required. While the 2.4GHz and 5GHz bands have a lot more channels compared to the sub-1GHz spectrum, a polite spectrum access mechanism is mandated in certain regions, i.e., sensing is needed to operate in these bands.

Existing sensing-based solutions rely on using narrowband spectrum sensors or scanners, and so each BS can only identify available whitespaces within a band, say of bandwidth 20MHz. To identify more spectral opportunities, wideband receivers are needed; the complexity and cost of such receivers scanning hundreds of MHz of spectrum at a spectral resolution of hundreds of KHz can be prohibitive, particularly in dense deployments of sensing BSs. For this reason, the main contribution of this research thrust is to develop a sensing-based system, where we identify available resources over a wideband spectrum at a fine spectral resolution, yet using narrowband receivers. The developed system further aims to increase the connection density of IoT devices by reusing the available spectral resources over space, i.e., at different BSs.

The development of the proposed sensing-based system comprises three algorithms. The first one is a sensing assignment scheduler that intelligently assigns each BS a portion of the spectrum to scan. The second one is a distributed wideband sensing algorithm, where neighboring BSs exchange few sensing reports to diffuse wideband spectrum occupancy information across the network; the algorithm helps each BS attain local information about the incumbents' energy footprints over a wideband spectrum. The last algorithm enables neighboring BSs to coordinate the allocation of spectral resources over space to limit the intra- and inter-network interference. By integrating the algorithms together, we are able to build a system that identifies more spectral resources at lower misdetection compared to existing networks such as MulteFire.

1.4.3 Energy-efficient shared spectrum access of massive IoT and UEs

The ubiquitous infrastructures of cellular networks and their well-established protocol stack have made them a natural choice to provide wide-area coverage for massive IoT applications. Nevertheless, the high connection density of IoT devices can quickly congest the network with their access requests, barring existing UEs from services if they share the spectrum with IoT devices; a straw-man approach to this problem is to split resource blocks (RBs) across UEs and IoT devices. Alternatively, IoT devices can be treated as a different class of lower priority, also referred to as access class barring. These solutions, however, limit the IoT applications that can be supported over cellular networks. For example, orthogonal allocation restricts the spectrum allocated to IoT devices, particularly in dense urban areas where UEs are widespread. Likewise, access class barring, while prevents IoT access requests to flood the network, it does not address the co-channel interference between IoT devices and UEs after access requests are granted.

The main contribution of this research thrust is to develop a spectrum sharing cellular protocol that enables a harmonious coexistence of a large number of IoT devices and UEs. The protocol leverages drones, or unmanned aerial vehicles, as IoT data aggregators. Unlike using small cells or terrestrial data aggregators, drones bring an additional degree of freedom thanks to their mobility; drones also can be deployed on demand, with minimal network planning compared to the deployment of terrestrial gateways.

The proposed shared access protocol aims to balance three intertwined objectives: provide reliable coverage to IoT devices, extend their lifetime, and protect UEs. To accomplish these objectives, we formulate a stochastic framework that optimizes the transmit power of IoT devices to maximize their energy-efficiency while still constraining the IoT interference into UEs. The improvements achieved in the energy-efficiency of IoT communications are directly translated into extended device lifetime. We show, using system-level simulations and the 3GPP evaluation methodology, that the proposed access protocol enables IoT devices to live +15 years on small batteries, with minimal degradation to the spectral efficiency of cellular UEs.

1.4.4 Coverage and rate maximization of heterogeneous cellular networks

The use of drones as data aggregators or BSs has become an emerging trend for future connectivity solutions. However, realizing a large-scale uptake of drone cells still requires overcoming persistent challenges such as drones' flight time and the autonomous control of drone swarm to perform specific tasks. Until then, can we optimize the existing terrestrial infrastructure to improve coverage or rate of IoT devices and cellular UEs?

Cellular networks are becoming denser and more heterogeneous; indeed, they are no longer a network of macro cells, but rather they comprise cells of different coverage areas such as pico and femto cells. Network densification helps in coverage enhancements by reducing the average distance between IoT devices and BSs, and it helps in capacity improvements by aggressively reusing the spectrum across these cells. However, we need to tackle two network bottlenecks, namely user association for load-balanced networking and interference management due to the high channel reuse. For instance, choosing the received signal strength as a user association metric, which typically improves coverage, inevitably overloads macro BSs, leaving many small cells and aggregators underutilized and reducing the long-term network capacity.

The main contribution of this research thrust is to develop low-complexity user association algorithms that maximize the coverage of IoT devices and the rate of high-throughput devices. To this end, analysis of coverage and rate in heterogeneous cellular networks with different multi-antenna transmission modes is first presented. Then, stochastic optimization frameworks to maximize coverage and rate are formulated, and algorithms to optimally solve them are developed. In the proposed algorithms, the device only measures the power of received reference signals from nearby heterogeneous BSs, and then, individually, infers which BS to associate with to maximize the coverage or rate. Although the algorithms are user-centric, i.e., no network coordination or central controller is used, by configuring which BS each device should connect to, we can harness tangible coverage and rate improvements, particularly for cell-edge devices.

1.5 Dissertation Organization and Related Publications

The rest of the dissertation is divided into several chapters as follows

- **Chapter 2:** We present the UNB communications paradigm, the framework developed, and the proposed protocol enhancements. This chapter revises our previous publications in [HC18f, HC19b].
- **Chapter 3:** We present the proposed sensing-based system. This chapter revises the previous publications in [HC18g, HC18b].
- **Chapter 4:** We present a cellular network with drones acting as data aggregators. Then, we develop a transmission protocol and an energy-efficiency maximization framework. This chapter revises our previous publications in [HC19a, HC17, HC18d].
- **Chapter 5:** We present user-centric association algorithms to maximize the coverage and the rate of IoT devices and UEs over cellular networks. This chapter revises our previous publications in [HC18a, HC16b]. Other relevant publications are given in [HC18c, HC18e, HC16a].
- **Chapter 6:** We summarize the research contributions and outline future extensions and research directions.

CHAPTER 2

Random Access via UNB Communications

Massive IoT has spurred the development of a new class of networks, known as low-power wide-area (LPWA) networks [RKS17,CVZ16,XZX15]. LPWA networks rely on the use of the unlicensed spectrum due to its low capital expenditure; they provide long-range connectivity, primarily using bands sub-1GHz due to their favorable propagation conditions, and they use lightweight access protocols to limit the communication overhead and extend the lifetime of end-devices [RKS17]. One popular variant of LPWA networks, which is the focus of this chapter, is the ultra-narrowband (UNB) network.

In UNB networks, data communications is done using extremely narrowband signals to connect a large number of devices, addressing the intra-network sharing problem, and to improve interference robustness by concentrating the signal energy into ultra-narrow channels, addressing the inter-network sharing problem. Indeed, the uplink (UL) signal bandwidth is few hundred Hz. In addition, these networks rely on simple ALOHA-like access protocols, where IoT devices avoid associating and synchronizing with any UNB base station (BS); in essence, IoT devices operate in a broadcast mode, transmitting their packets at any time and frequency. The only restrictions on these devices are primarily related to how many packets per day each can send and the bandwidth of the band that comprises the different channels a device can pick from. Furthermore, to combat the absence of acknowledgments from the network and the presence of incumbent networks, signal repetition is used; in particular, each IoT packet is sent multiple times, one after another, each at a different frequency within the predetermined multiplexing band. The packet is successfully transmitted if any BS decodes any of the packet transmissions, yet no combining of these packets is done nor cooperation with other BSs is used.

2.1 Related Work

A list of few UNB proprietary technologies that currently provide IoT services is given in Table 2.1. Sigfox is one of the most recognized UNB network, and it has been growing globally [Sig17,Sig19]. Fig. 2.1 shows the spectrogram of one Sigfox packet, which is sent three times randomly hopping from one frequency to another. In this chapter, our objective is to model, analyze, and optimize these UNB networks, considering their unique communications properties and examining their scalability, in terms of the connection density of IoT devices, and their robustness to interference from incumbent networks.

Network-level analysis of UNB networks remains unexplored. For instance, early works, led by Sigfox, have considered single cells and ignored the presence of incumbents [GM16,MDG16,MGG16]. The more recent work in [MGG18] has considered the performance of UNB networks when two BSs are used to decode IoT packets, although signal repetition is ignored. These works assume a specific geometry of the network, whereas in this chapter we use stochastic geometry to help derive fundamental limits and performance trends of UNB networks that are not restricted to particular deployment geometries.

In addition, the aforementioned works have quantified the UNB network performance in terms of the probability of collision, i.e., two or more devices picking the same time-frequency resources. This metric is pessimistic and in fact becomes inaccurate when UNB networks share the spectrum with devices that use wider channels. Indeed, a single LoRa device, that transmits signals at a bandwidth of 250KHz, can completely overlap the 200KHz multiplexing band used by Sigfox. Thus, under the aforementioned metric, UNB communications is impossible when the LoRa device is active! In this chapter, we focus on the received signal-to-noise-plus-interference ratio (SINR) of UNB packets, i.e., a packet is successfully decoded if the SINR is above some threshold. Such SINR analysis of IoT communications has been considered in [JDN18,CVS17,AAJ18,HC18a], but these works consider cellular networks, where access is vastly different compared

Table 2.1: Examples of UNB technologies

UNB Technology	Bandwidth (Hz)
Sigfox [Sig19]	600 (US) 100 (Europe)
WavIoT NB-Fi [WAV16]	100
NWave Weightless-N [Wei]	200
Telensa [Tel18]	500

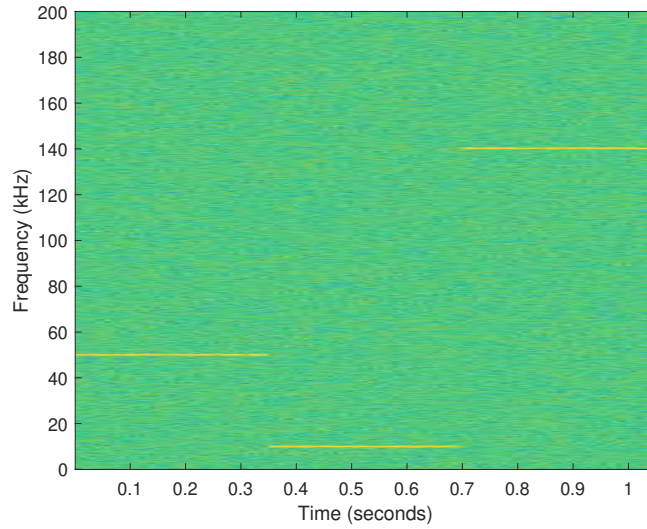


Figure 2.1: Spectrogram of one Sigfox packet of bandwidth 600Hz. The packet is sent three times at randomly selected frequencies within a multiplexing band of bandwidth 200KHz.

to UNB communications. Indeed, with cellular access, devices connect to a specific cell and send each packet once in a coordinated manner.

2.2 Chapter Contributions and Organization

The main contributions of this chapter are twofold. First, we present an analytical framework to model UNB networks in the presence of other interfering networks; the frame-

work is based on stochastic geometry [Hae12], which has become a powerful tool to model cellular networks [ESA17]. Since UNB networks rely on devices transmitting at random times and frequencies, stochastic geometry can help capture the spatial randomness of these networks and obtain theoretical expressions of key performance indicators. To this end, we derive the success probability in closed form, which is defined as the probability of having at least one of the packet's transmissions successfully decoded. The obtained expression is then used to determine the transmission capacity of UNB networks [HAW08, WA12], i.e, the maximum density of IoT devices that can be reliably covered in the network for a given density of UNB BSs. More importantly, the obtained theoretical expressions help answer the following questions: what is the role of geographical diversity achieved by operating in an UL broadcast mode in random access? When is repetition diversity useful? What are the key parameters that affect the scalability and robustness of UNB networks?

Second, the gleaned answers to the aforementioned questions are used to enhance and optimize UNB communication protocols. For example, by identifying a fundamental trade-off between the number of UNB BSs and the hardware complexity of each one, we propose a multiband access protocol that can push the limits of UNB networks at virtually no hardware cost. In addition, we derive the optimal number of repetitions, showing that current Sigfox systems, that use three transmissions per packet, can be suboptimal when the incumbents' density is low.

The rest of the chapter is organized as follows. We present the system model and the key performance metrics in Section 2.3. Performance analysis of UNB protocols and their variants is examined in Section 2.4. UNB protocol enhancements and optimizations are given in Section 2.5. The framework, the gleaned insights, and the protocol recommendations are validated via large-scale Monte Carlo simulations in Section 2.6. Finally, the main conclusions are drawn in Section 2.7.

2.3 System Model

2.3.1 Network topology

We consider a random spatial topology of BSs, IoT devices, and interfering networks. Specifically, for the UNB network, we assume that BSs' locations and IoT devices are generated from homogeneous Poisson Point processes (HPPP) Φ_B and Φ_{IoT} , with densities λ_B and λ_{IoT} , respectively. Likewise, devices of incumbent or interfering networks are generated from another independent HPPP Φ_I with density λ_I . Extension of this framework to include different tiers, or types, of incumbents is straightforward.

2.3.2 Transmission model

The UNB IoT device transmits signals at power P_{IoT} , occupying a bandwidth b . Each signal is sent N times, consecutively over time, yet hopping from one frequency to another. These signals are sent within a predetermined *multiplexing band* of bandwidth B . In this chapter, we generalize the analysis to multiband access, i.e., we assume there are M multiplexing bands, each of bandwidth B . For the temporal generation of IoT traffic, UNB devices randomly send packets over time at the following rate

$$\lambda_T = \frac{K \times T}{T_{\text{tot}}}, \quad (2.1)$$

where K is the number of unique reports, or packets, that can be transmitted over a total duration of T_{tot} , and T is the duration of each transmission. For example, in Sigfox [Sig19], the device can send up to $K = 6$ packets per hour, i.e., $T_{\text{tot}} = 3600\text{s}$. Each Sigfox packet is typically 26 bytes, and thus at $b = 600\text{Hz}$, the transmission duration is $T \approx 0.35\text{s}$. We note that the limits on K and T are due to the sub-1GHz regulations on duty cycles and dwell times (cf. Table 1.1). Finally, incumbent devices transmit signals at power P_I over a bandwidth $B_I \gg b$, where this bandwidth is assumed to be overlapped with the spectrum used by the UNB network. The interfering network could be a WiFi network or another IoT-based network, e.g., LoRa [CVZ16, BRV16].

Since the transmissions are sent over extremely narrowband channels, it is difficult to send a signal at a specific channel and carrier frequency. In fact, the frequency offset in a low-cost oscillator could be higher than the signal bandwidth, making it impractical to assume a perfectly channelized system for UNB networks. For this reason, UNB networks are commonly assumed to have an unslotted frequency access [DGG14]. In particular, let the n -th transmission be sent at carrier frequency f_n ; then, in unslotted frequency access, we have

$$f_n \sim \mathcal{U}\left(f_c - \frac{M \cdot B}{2}, f_c + \frac{M \cdot B}{2}\right), \quad (2.2)$$

where f_c the center frequency of the spectrum used by the UNB network, and $\mathcal{U}(a, b)$ is the uniform distribution on the interval $[a, b]$. While UNB networks are actually unslotted, we generalize the analysis to consider the slotted case in time and/or frequency.

In the slotted case, a device interferes with another if both devices are transmitting over the same slot, which occurs with probability T/T_{tot} and $b/(M \cdot B)$ in time and frequency, respectively. In the unslotted time (or frequency) case, interference between two devices occurs if their signals have any overlap in time (or frequency). For instance, if $\tilde{t}_{n,1}$ and $\tilde{t}_{n,2}$ are the start transmission times of device one and two, then an interference occurs if $|\tilde{t}_{n,1} - \tilde{t}_{n,2}| < T$, and thus two devices interfere with probability $2T/T_{\text{tot}}$. Similarly, if $f_{n,1}$ and $f_{n,2}$ are the center frequencies of the two devices, then an interference occurs if $|f_{n,1} - f_{n,2}| < b/2$, which happens with probability $2b/(M \cdot B)$. In other words, an unslotted system in time (or frequency) doubles the probability of a device interfering with another one in comparison with a slotted system in time (or frequency). For a compact notation, we use $1 \leq \beta_T \leq 2$, where $\beta_T = 1$ and $\beta_T = 2$ denote a slotted and an unslotted system in time, respectively. Note that $\beta_T \in (1, 2)$ can denote different levels of time synchronization or interference overlap tolerance. An identical notation is used for frequency, where we use β_F . Fig. 2.2 illustrates the different access scenarios.

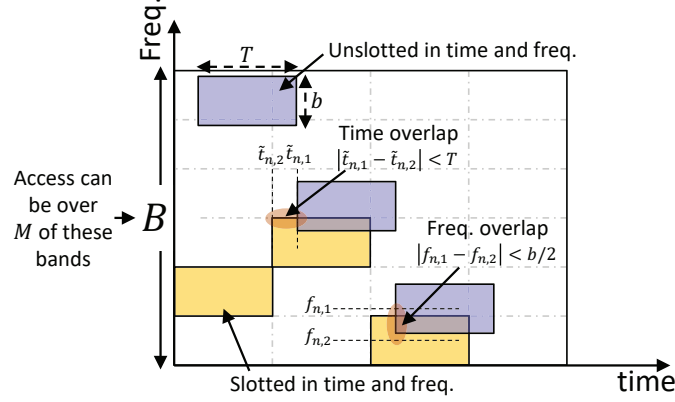


Figure 2.2: Access can be slotted or unslotted in time and/or frequency.

2.3.3 Channel model

For the wireless channel, we consider a power-law path loss model with exponent α , i.e., the path loss between a typical IoT device and a BS at distance y is $y^{-\alpha}$. All channels experience independent Rayleigh fading with unit power, i.e., the channel power gain between the device and a BS is $h \sim \exp(1)$, whereas the channel power gain between an interfering device and a BS is $f \sim \exp(1)$.

2.3.4 Performance metrics

Different UNB protocols are compared in terms of *the transmission success probability*, $\mathbb{P}_s$; it is defined as the probability that at least one of the N messages is decoded successfully, i.e., the signal received SINR is above a certain threshold τ . Further, the success probability can be interpreted as the complementary cumulative distribution function (CDF) of the maximum SINR among the N messages. This expression is also useful to determine *the transmission capacity* of the network, which is defined as the maximum number of devices that can be supported for a given density of BSs and success probability constraint $\gamma \in (0, 1)$. More formally, the success probability can be written as a function of the density of IoT devices, i.e., $\mathbb{P}_s = F(\lambda_{\text{IoT}})$, and thus, the transmission capacity is defined as $\mathbb{C}(\gamma) = \gamma \cdot F^{-1}(\gamma)$ [WA12].

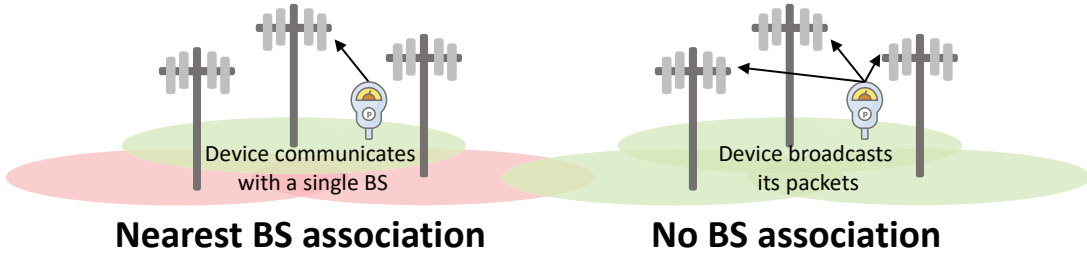


Figure 2.3: Two variants of UNB networks are considered: with and without BS association.

2.4 Performance Analysis of UNB Protocols

In this section, we analyze the performance of two variants of UNB protocols as shown in Fig. 2.3. In the first one, the IoT device appends its packets with a cell ID, i.e., the device communicates with the nearest BS. In the second one, the IoT device operates in a broadcast mode, i.e., there is no BS association. While the latter is adopted by Sigfox, analyzing the former helps highlight the significance of the geographical diversity achieved when the device broadcasts its packets instead of sending them to a single BS.

2.4.1 Performance with nearest BS association

When a device is restricted to connect to the nearest BS, the transmission is considered successful if this particular BS decodes at least one of the N messages. Assuming the nearest BS is j , then we have

$$\mathbb{P}_s^{\text{nearest}} = 1 - \Pr\{\text{No message is decoded at the nearest BS}\}. \quad (2.3)$$

Clearly, a BS cannot decode any of the N messages if the message with the maximum SINR is below the decoding threshold τ . More formally, consider a typical device at a distance x_j from the j -th BS. Then, the SINR of the n -th message at this BS can be

expressed as

$$\text{SINR}_{n,j} = \frac{h_n x_j^{-\alpha}}{\underbrace{\hat{P}_N + \sum_{u \in \Phi_{\text{IoT}}} f_u y_{u,j}^{-\alpha}}_{I_{\text{UNB}}} + \underbrace{\sum_{k \in \Phi_{\text{I}}} \hat{P}_I f_k y_{k,j}^{-\alpha}}_{I_{\text{INC}}}}, \quad (2.4)$$

where $\hat{P}_I = \frac{P_I \cdot b/B_I}{P_{\text{IoT}}}$, $\hat{P}_N = \frac{P_N}{P_{\text{IoT}}}$, P_N is the noise power, $y_{u,j}$ and $y_{k,j}$ are the distances from the j -th BS to an interfering UNB device and an interfering incumbent, respectively, and $\tilde{\Phi}_{\text{IoT}}$ and $\tilde{\Phi}_{\text{I}}$ are the set of interfering UNB devices and incumbents, respectively. Thus, the j -th BS fails to decode the N messages with probability

$$\mathbb{Q}_j = \Pr \left(\max_{n \in \{1, 2, \dots, N\}} \text{SINR}_{n,j} \leq \tau \right), \quad (2.5)$$

which means that the success probability in (2.3) can be rewritten as

$$\mathbb{P}_s^{\text{nearest}} = 1 - \mathbb{E}_x [\mathbb{Q}_j], \quad (2.6)$$

where the expectation is with respect to the location of the nearest BS in Φ_{B} . We can simplify (2.5) as follows

$$\begin{aligned} \mathbb{Q}_j &= \Pr (\text{SINR}_{1,j} \leq \tau, \text{SINR}_{2,j} \leq \tau, \dots, \text{SINR}_{N,j} \leq \tau) \\ &\stackrel{(a)}{=} \mathbb{E}_{\tilde{\Phi}_{\text{IoT}}, \tilde{\Phi}_{\text{I}}, f_u, f_k} \left[\Pr (h_n \leq \tau x_j^\alpha (\hat{P}_N + I_{\text{UNB}} + I_{\text{INC}}))^N \right] \\ &\stackrel{(b)}{=} \left(\mathbb{E}_{\tilde{\Phi}_{\text{IoT}}, \tilde{\Phi}_{\text{I}}, f_u, f_k} \left[(1 - e^{-\tau x_j^\alpha (\hat{P}_N + I_{\text{UNB}} + I_{\text{INC}})}) \right] \right)^N, \end{aligned} \quad (2.7)$$

where (a) follows as each signal is sent over a different channel with independent fading¹ and (b) follows from the CDF of h_n and the fact that each transmission can experience a different and independent set of interfering UNB devices. Note that an assumption here, that we make throughout this chapter, is that each IoT transmission will experience an independent set of interfering incumbents. This assumption simplifies the analysis and is reasonable as we assume incumbent transmitters have shorter transmission durations (recall $B_I \gg b$). For instance, Sigfox transmission duration is $T \approx 0.35\text{s}$ in the US (or 2s

¹Channels can be assumed to be independent even if frequency hopping is done over a narrowband of bandwidth B because the duration of each UNB signal is long enough to justify different channels across messages [Sig17].

in Europe), whereas other incumbents, e.g. LoRa, typically have durations of the order of few tens of milliseconds for similar packet sizes [Sor17].

To obtain a closed form expression of (2.7), we need to determine the spatial process of interfering UNB devices, i.e., $\tilde{\Phi}_{\text{IoT}}$, and interfering incumbent devices, i.e., $\tilde{\Phi}_{\text{I}}$. Since UNB devices randomly pick time-frequency slots, the set of interfering devices $\tilde{\Phi}_{\text{IoT}}$ is essentially a thinned Φ_{IoT} , which is still an HPPP [Hae12]. The density of this thinned process is

$$\tilde{\lambda}_{\text{IoT}} = N \cdot \beta_T \lambda_T \cdot \frac{\beta_F b}{M \cdot B} \cdot \lambda_{\text{IoT}}. \quad (2.8)$$

This can be derived as follows. The original density of IoT devices is λ_{IoT} , and thus the portion of these devices that would have overlapping transmissions in time is $N\beta_T\lambda_T$. However, not all of those devices will have the same frequency as the typical device. Indeed, the probability of a device picking the same frequency as the typical one is $\frac{\beta_F b}{M \cdot B}$, for which (2.8) follows. In a similar manner, the set of interfering incumbents is another HPPP process with density $\tilde{\lambda}_{\text{I}} = \min\{1, \frac{B_{\text{I}}}{M \cdot B}\} \lambda_{\text{I}}$. The next theorem provides the success probability of UNB access with nearest BS association.

Theorem 1. *In an interference-limited network, where $\hat{P}_N \rightarrow 0$, the success probability of UNB access under nearest BS association is given as*

$$\mathbb{P}_s^{\text{nearest}} = 1 - \sum_{k=0}^N \binom{N}{k} (-1)^k \left(1 + k \xi^{-1} \tau^\delta \frac{\tilde{\lambda}_{\text{IoT}} + \hat{P}_{\text{I}}^\delta \tilde{\lambda}_{\text{I}}}{\lambda_{\text{B}}} \right)^{-1}, \quad (2.9)$$

where $\delta = 2/\alpha$ and $\xi = \frac{\sin(\pi\delta)}{\delta\pi}$.

Proof: See Appendix A.1. ■

The transmission capacity, in this case, can be computed numerically when $N > 1$. When $N = 1$, we have the following result.

Corollary 1. *The transmission capacity of UNB access under nearest BS and $N = 1$ is*

$$\mathbb{C}^{\text{nearest}}(\gamma) = \frac{\gamma M \cdot B}{\beta_T \beta_F b \lambda_T} \left(\frac{\xi \tau^{-\delta} \lambda_{\text{B}}}{(\frac{\gamma}{1-\gamma})} - \hat{P}_{\text{I}}^\delta \min\{1, \frac{B_{\text{I}}}{M \cdot B}\} \lambda_{\text{I}} \right). \quad (2.10)$$
■

We have the following key observations. First, the presence of an interfering network reduces the transmission capacity, as the expression in parentheses decreases with λ_I . Second, increasing the number of bands has two gains: it can reduce the density of interfering incumbents, i.e., the value in the parentheses is increased, and it scales the UNB transmission capacity, i.e., the term outside the parentheses is also increased. Note that for $M = 1$, we get the performance of UNB networks with single-band access. Finally, a completely synchronous system quadruples the transmission capacity compared to an asynchronous time-frequency system.

2.4.2 Performance with no BS association

In this section, we consider a UNB network similar to Sigfox, where a device operates in a broadcast mode; hence, the transmission is successful if *at least one BS* decodes at least one of the N messages. In this case, the success probability is

$$\begin{aligned} \mathbb{P}_s &= 1 - \Pr\{\text{No message is successfully decoded at any BS}\} \\ &= 1 - \mathbb{E}_{\Phi_B} \left[\prod_{b \in \Phi_B} \mathbb{Q}_b \right]. \end{aligned} \quad (2.11)$$

Compared to (2.6), the expectation here is with respect to the set of UNB BSs instead of just the distance to the nearest one. The next theorem presents the success probability performance of this protocol.

Theorem 2. *In an interference-limited network, the success probability of UNB access under no BS association is given as*

$$\mathbb{P}_s = 1 - \exp \left(-\xi \tau^{-\delta} H_N \cdot \frac{\lambda_B}{\tilde{\lambda}_{\text{IoT}} + \hat{P}_I^\delta \tilde{\lambda}_I} \right), \quad (2.12)$$

where H_N is the harmonic number, i.e., $H_N = -\sum_{k=1}^N \binom{N}{k} (-1)^k k^{-1}$.

Proof: See Appendix A.2. ■

Using Theorem 2, we can obtain the transmission capacity in closed-form as follows.

Corollary 2. *The transmission capacity of the UNB network with no BS association is*

$$C(\gamma) = \frac{\gamma MB}{\beta_T \beta_F b \lambda_T} \left(\frac{\xi \tau^{-\delta} H_N \lambda_B}{N \ln(\frac{1}{1-\gamma})} - \frac{\hat{P}_I^\delta \min\{1, \frac{B_I}{MB}\} \lambda_I}{N} \right). \quad (2.13)$$

■

Next, we closely examine the performance gains of operating in a broadcast mode instead of connecting to a single BS, as commonly done in traditional wide-area networks.

2.4.3 The role of geographical diversity

In geographical diversity, also known as macro diversity, multiple physically apart BSs are used to receive the same signal. Thus, the signal can experience multiple independently faded channels, substantially improving the transmission reliability. We distinguish this diversity method from spatial diversity, which is used by having multiple antennas at the same BS. The latter is more common in human-centric wireless networks.

In what follows, we assume a single-transmission per packet, i.e., $N = 1$, to factor out the gains achieved by repetition diversity. Let the target success probability be $\mathbb{P}_s \geq \epsilon$ and $B_I \leq M \cdot B$. Then, we have the following corollary.

Corollary 3. *To achieve a success probability of at least ϵ , the following constraint must be satisfied*

$$M \lambda_B \geq c g(\epsilon), \quad (2.14)$$

where $c = \left(\beta_T \lambda_T \beta_F \frac{b}{B} \lambda_{IoT} + \hat{P}_I^\delta \frac{B_I}{B} \lambda_I \right)^{-1} \xi^{-1} \tau^\delta$ and

$$g(\epsilon) = \begin{cases} \frac{\epsilon}{1-\epsilon}, & \text{Nearest BS} \\ \ln\left(\frac{1}{1-\epsilon}\right), & \text{No BS association} \end{cases}. \quad (2.15)$$

Proof: For $N = 1$, we can rewrite (2.9) as follows

$$\mathbb{P}_s^{\text{nearest}} = \left(1 + \xi^{-1} \tau^\delta \frac{\tilde{\lambda}_{IoT} + \hat{P}_I^\delta \tilde{\lambda}_I}{\lambda_B} \right)^{-1}. \quad (2.16)$$

Thus, by setting $\mathbb{P}_s^{\text{nearest}} \geq \epsilon$, we get the desired expression in (2.14). We can also rewrite (2.12) for $\mathbb{P}_s \geq \epsilon$ in a similar manner. ■

We observe from Corollary 3 that as $\epsilon \rightarrow 1$, then the lower bound in (2.14) is significantly higher for nearest BS association, requiring a larger number of bands or UNB BSs. This is also shown in Fig. 2.4a. Put differently, if an IoT network operator wants to deploy λ_B BSs to cover a particular region, then the required density of these BSs, if IoT devices operate in a broadcast mode instead of a point-to-point mode, is given as

$$\lambda_B = \left(\frac{1 - \epsilon}{\epsilon} \right) \ln \left(\frac{1}{1 - \epsilon} \right) \lambda_B^{\text{nearest}}, \quad (2.17)$$

where $\lambda_B^{\text{nearest}}$ is the density of BSs that the operator needs to deploy under nearest BS association. In Fig. 2.4b, we show the ratio $\lambda_B / \lambda_B^{\text{nearest}}$ as $\epsilon \rightarrow 1$. It is evident that geographical diversity reduces the density of BSs required by one to two orders of magnitude at high success probability constraints. Alternatively, from (2.16), we observe that outage, i.e., $1 - \mathbb{P}_s$, decreases linearly with λ_B , whereas from (2.12), it decreases exponentially when no BS association is used.

Likewise, we get similar trends by comparing the two schemes in terms of the transmission capacities in (2.10) and (2.13).

2.4.4 Random versus pseudorandom frequency hopping

The aforementioned analysis considers the IoT device to randomly hop from one frequency to another. In this section, we consider a simpler system, where there is a predefined codebook of hopping patterns at a given time. Hopping patterns in the codebook are orthogonal, i.e., each pattern determines the N channels to be used such that each one has channels orthogonal to all other patterns. The number of codes, over a given time slot, is equal to the number of channels; thus, the probability that two devices, transmitting with some time overlap, pick the same sequence is the same as the probability of two devices picking the same channel in the random hopping scheme.

It can be shown that the densities of interfering IoT devices and incumbents remain

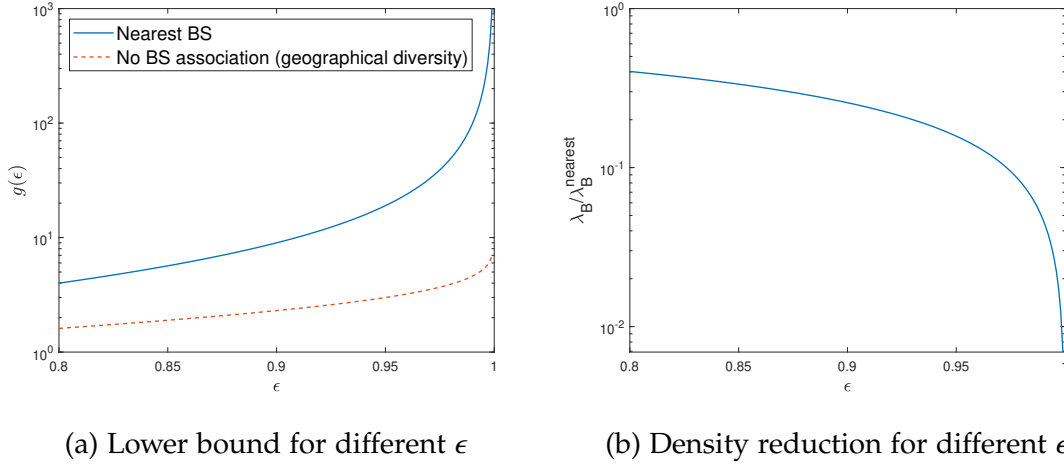


Figure 2.4: Impact of geographical diversity on the network resources needed, i.e., the number of bands and the density of BSs, to achieve a target success probability constraint.

the same even under the pseudorandom (PN) hopping scheme. However, in random hopping, each transmission could interfere with a different set of interfering IoT devices, whereas in PN hopping, if both devices pick the same hopping pattern, they will interfere in each of the N transmissions. Thus, the question we want to answer in this section is the following: for a given density of interfering devices, do we want interfering devices to remain the same or to change across the N transmissions? We answer this question in the following theorem.

Theorem 3. *The success probability performance of the PN hopping scheme is upper bounded by that of the random hopping scheme for both nearest BS and no BS association systems.*

Proof: When the same devices interfere with each other in every transmission, then (2.5) becomes

$$\begin{aligned}
 Q_j^{\text{PN}} &= \mathbb{E}_{\Phi_{\text{IoT}}, f_u} \left[\left(1 - \mathbb{E}_{\Phi_{\text{I}}, f_k} \left[e^{-\tau x_j^\alpha (\hat{P}_N + I_{\text{UNB}} + I_{\text{INC}})} \right] \right)^N \right] \\
 &\stackrel{(a)}{=} \mathbb{E}_{\Phi_{\text{IoT}}, f_u} \left[\sum_{k=0}^N \binom{N}{k} (-1)^k e^{-k(\pi \xi^{-1} (\tau \hat{P}_1)^\delta \tilde{\lambda}_1 x^2 + \tau x_j^\alpha (\hat{P}_N + I_{\text{UNB}}))} \right] \\
 &\stackrel{(b)}{=} \sum_{k=0}^N \binom{N}{k} (-1)^k e^{-\pi \xi^{-1} \tau^\delta (k^\delta \tilde{\lambda}_{\text{IoT}} + k \hat{P}_1^\delta \tilde{\lambda}_1) x^2 - k \tau x_j^\alpha \hat{P}_N},
 \end{aligned} \tag{2.18}$$

where (a) follows by taking the expectation with respect to $\tilde{\Phi}_I$ and f_k and then using the binomial theorem, and (b) follows by taking the expectation with respect to $\tilde{\Phi}_{IoT}$ and f_u . Following the same steps as in Appendix A.1, the success probability in an interference-limited network and with nearest BS association is given as

$$\mathbb{P}_s^{\text{nearest,PN}} = 1 - \sum_{k=0}^N \binom{N}{k} (-1)^k \left(1 + \xi^{-1} \tau^\delta \frac{k^\delta \tilde{\lambda}_{IoT} + k \hat{P}_I^\delta \tilde{\lambda}_I}{\lambda_B} \right)^{-1}. \quad (2.19)$$

Similarly, following the same steps as Appendix A.2, the success probability with no BS association is

$$\mathbb{P}_s^{\text{PN}} = 1 - \exp \left(\xi \tau^{-\delta} \sum_{k=1}^N \binom{N}{k} \frac{(-1)^k \lambda_B}{k^\delta \tilde{\lambda}_{IoT} + k \hat{P}_I^\delta \tilde{\lambda}_I} \right). \quad (2.20)$$

To prove that the performance of PN hopping is upper bounded by that of the random one, we can rewrite (2.18) as $\mathbb{E}[g(z)]$, where $g(z) = z^N$, and hence (2.7) is rewritten as $g(\mathbb{E}[z])$. Using Jensen's inequality, we have $g(\mathbb{E}[z]) \leq \mathbb{E}[g(z)]$, i.e., $Q_j \leq Q_j^{\text{PN}}$, since $g(z)$ is a convex function. Thus the success probability of the random hopping scheme outperforms that of the PN scheme for both nearest and no BS association. ■

To summarize, random hopping inherently achieves an *interference diversity* that can only improve the success probability performance. This diversity is attained because different transmissions experience interference from different sources.

2.5 Enhancements and Optimizations of UNB Protocols

2.5.1 Low-cost multiband access

Corollary 3 has provided us with a lower bound on the UNB network resources needed to meet a success probability target of ϵ . Interestingly, we observe that it is the product $M\lambda_B$ that matters, i.e., we can trade the density of BSs with more bands. For example, the IoT network operator can halve the number of BSs, but then the network has to operate over twice the spectrum bandwidth. The issue, however, is that the BS needs a more complex receiver if wider spectrum is used. This follows because UNB networks are typically asynchronous in frequency, and thus the UNB receiver must sample the

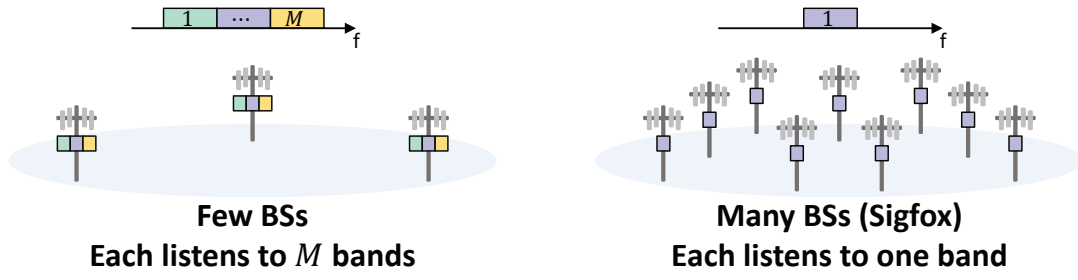


Figure 2.5: Two deployment approaches: few expensive BSs or many low-cost ones.

spectrum at a very high resolution to detect IoT signals. For this reason, existing UNB networks assume that devices send their signals within a single narrowband. For example, in a Sigfox system [Sig17], the BS listens to a spectrum of bandwidth $B = 200\text{KHz}$, where the power spectral density of the band is obtained using the Fast Fourier transform (FFT), with a very small sampling interval [Art17]. For example, assuming that the device sends a signal of bandwidth $b = 100\text{Hz}$ in this multiplexing band, then using a sampling interval of $b/4$ [Art17] requires the FFT size to be at least 2^{14} !

To summarize, if an IoT network operator, such as Sigfox, wants to improve the success probability or increase the transmission capacity, it has two options, as shown in Fig. 2.5. In the first one, the operator can keep the BS deployment density low, but then each BS requires very expensive receivers to scan M bands. In the second one, the operator can use existing low-cost BSs, but it has to deploy them at a higher density.

One question we are interested to answer in this dissertation is: can we improve the network performance if we use M bands, but deploy low-cost BSs, where each one randomly selects one of the M bands? To this end, we present two transmission protocols:

- **Unslotted multiband:** The IoT device sends all of its N transmissions over one randomly selected band.
- **Slotted multiband:** Each transmission of the N packets can be sent to a different band.

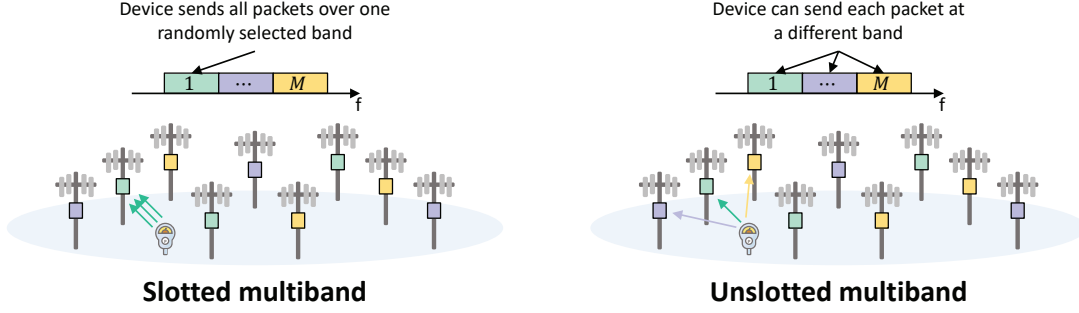


Figure 2.6: Two low-cost multiband access protocols: slotted and unslotted.

We note that the difference between the slotted and unslotted multiband protocols is that in the former, the same set of BSs will listen to all N messages, whereas in the latter, each message could be received at different subsets of BSs. These two low-cost multiband access protocols are illustrated in Fig. 2.6. Next, we analyze both protocols under no BS association and random repetitions. Extensions to the case of nearest BS association and PN repetitions are straightforward.

2.5.1.1 Slotted multiband with no BS association

Let $\mathbb{P}_{s|m}$ denote the success probability given that the typical IoT device picks the m -th band for the transmission of N messages. Then, the success probability is

$$\mathbb{P}_s^{\text{SM}} = \frac{1}{M} \sum_{m=1}^M \mathbb{P}_{s|m}. \quad (2.21)$$

Further, to generalize the analysis, let each BS picks the M -th band with probability p_m . Then, we have the following theorem.

Theorem 4. *In an interference-limited network, the success probability of the slotted multiband protocol with no BS association is given as*

$$\mathbb{P}_s^{\text{SM}} = 1 - \frac{1}{M} \sum_{m=1}^M \exp \left(-\xi \tau^{-\delta} H_N \cdot \frac{p_m \lambda_B}{\tilde{\lambda}_{\text{IoT}} + \tilde{P}_I \tilde{\lambda}_I} \right). \quad (2.22)$$

Proof: We can compute $\mathbb{P}_{s|m}$ as follows. The set of BSs listening to the m -th band is an HPPP process with density $p_m \lambda_B$. Furthermore, the set of interfering IoT devices is the

same as $\tilde{\Phi}_{\text{IoT}}$. To show this, recall that the density of IoT devices with a time overlap is $N\beta_T\lambda_T\lambda_{\text{IoT}}$. Among these devices, only $1/M$ of them, on average, will select the m -th band. Among those that select the m -th band, only $\beta_F b/B$ will select, on average, the same channel as the typical IoT device. Finally, the set of interfering incumbents remains the same, as we assume the incumbent can use any part of the spectrum. To this end, we can use the success probability of the benchmark protocol to compute $\mathbb{P}_{s|m}$, where we replace λ_B in (2.12) with $p_m\lambda_B$. ■

Remark: We note that $\mathbb{P}_s^{\text{SM}}$ is maximized when $p_m = 1/M$. This can be proved using the inequality $\frac{1}{M} \sum_{m=1}^M \exp(p_m z) \geq \exp(\frac{z}{M} \sum_{m=1}^M p_m) = \exp(z/M)$, which holds with equality when $p_m = 1/M$. This is intuitive as devices select a band with equal probability, and thus BSs must follow the same selection procedure to maximize the success performance. An interesting observation is in the absence of an incumbent network, the slotted multiband has the same performance as the single-band protocol when $p_m = 1/M$. This shows that while increasing the number of bands decreases the density of interfering devices, the density of BSs listening to the same band also decreases with the same rate, i.e., there is no gain of using more than one band in this case. However, in the presence of the incumbent network, increasing the number of bands can have different impacts, depending on the bandwidth of the interfering network. We can illustrate this in terms of the transmission capacity, which is given in the following corollary.

Corollary 4. *The transmission capacity of the slotted multiband protocol with $p_m = 1/M$ is*

$$\mathbb{C}^{\text{SM,R}}(\gamma) = \frac{\gamma B}{\beta_T \beta_F b \lambda_T} \left(\frac{\xi \tau^{-\delta} H_N \lambda_B}{N \ln(\frac{1}{1-\gamma})} - \frac{\hat{P}_1^\delta M \min\{1, \frac{B_I}{M \cdot B}\} \lambda_I}{N} \right). \quad (2.23)$$

■

Let $\mathbb{C}^{\text{SB}}(\gamma)$ denote the transmission capacity of the single-band, which can be obtained by substituting $M = 1$ in (2.13). Then, we have the following remarks.

- If $B_I < B$, and hence $B_I < M \cdot B$, then $\mathbb{C}^{\text{SM}}(\gamma) = \mathbb{C}^{\text{SB}}(\gamma)$. This follows because in both protocols, there will be channels with no interfering incumbents. While the

slotted multiband will have more of such channels, the density of BSs listening to a particular band is reduced.

- If $B_I > M \cdot B$, then $\mathbb{C}^{\text{SM}}(\gamma) < \mathbb{C}^{\text{SB}}(\gamma)$. This follows because both protocols will have the same density of interfering incumbents. However, in slotted multiband, fewer BSs listen to the same band compared to the single-band protocol.
- If $B_I > B$ and $B_I < M \cdot B$, then interestingly $\mathbb{C}^{\text{SM}}(\gamma) < \mathbb{C}^{\text{SB}}(\gamma)$. Note that when $B_I > B$, then irrespective of the channels used by the device in the single-band access, there will always be interference from the incumbent network. For the slotted multiband, the density of BSs listening to a particular band is reduced, and as B_I increases, fewer channels will be incumbent-free compared to the case when $B_I < B$.

To summarize, in the slotted multiband protocol, there is a cost due to randomly assigning each BS a single band to listen to, and thus increasing the number of bands is not beneficial, particularly when the bandwidth of the incumbent signal overlaps with all bands.

2.5.1.2 Unslotted multiband with no BS association

In this protocol, each packet can be sent over a different multiplexing band. Assuming that the typical device sends n_m of the N packets over the m -th band, where $\sum_{m=1}^M n_m = N$, then the probability of having at least a single message decoded successfully at any BS is given as

$$\mathbb{P}_{s|\{n_m\}_1^M} = 1 - \prod_{m=1}^M \exp \left(-\xi \tau^{-\delta} H_{n_m} \cdot \frac{p_m \lambda_B}{\tilde{\lambda}_{\text{IoT}} + \hat{P}_I \tilde{\lambda}_I} \right). \quad (2.24)$$

This expressions follows because the m -th band has $p_m \lambda_B$ BSs listening to n_m messages. Let $\mathcal{N} = \{n_1, n_2, \dots, n_M | \sum_{m=1}^M n_m = N, n_m \in \{0, 1, \dots, N\}\}$ be the set of all possible combinations of sending N messages over M bands. Then, the success probability of this protocol is computed by averaging over $\mathcal{N}$, as given in the following theorem.

Theorem 5. *In an interference-limited network, the success probability of the unslotted multi-band protocol with no BS association is given as*

$$\mathbb{P}_s^{\text{UM}} = \frac{1}{M^N} \sum_{\{n_m\} \subset \mathcal{N}} \frac{N!}{n_1! n_2! \cdots n_M!} \mathbb{P}_{s|\{n_m\}_1^M}, \quad (2.25)$$

where the sum is over every possible combination in $\mathcal{N}$.

Proof: Observe that the probability of sending n_m packets out of N over the m th band follows the multinomial distribution, where N is the number of trials and M is the number of possible outcomes for each one. Since each channel is picked with probability $1/M$, then a specific combination $\{n_m\}_{m=1}^M \subset \mathcal{N}$ occurs with probability $\frac{1}{M^N} \frac{N!}{n_1! n_2! \cdots n_M!}$, which completes the proof. ■

The transmission capacity of this protocol can be obtained numerically using the aforementioned expressions.

2.5.2 Optimizing the number of transmissions

In addition to the geographical diversity, UNB networks rely on repetition diversity, where the same packet is sent multiple times over time and/or frequency. For example, Sigfox typically uses three transmissions per packet; however, is this optimal in terms of maximizing the success probability? How many transmissions should each device send? We answer these questions by examining the impact of N on the performance of the UNB network with no BS association.

We can rewrite (2.12) as

$$\mathbb{P}_s = 1 - \exp\left(-\frac{a_1 H_N}{a_2 N + a_3}\right), \quad (2.26)$$

where $a_1 = \xi \tau^{-\delta} \lambda_B$, $a_2 = \beta_T \lambda_T \cdot \frac{\beta_F b}{M \cdot B} \cdot \lambda_{\text{IoT}}$, and $a_3 = \hat{P}_I^\delta \tilde{\lambda}_I$. We further have

$$\begin{aligned} H_N &= - \sum_{k=1}^N \binom{N}{k} (-1)^k k^{-1} \\ &= \sum_{k=1}^N \frac{1}{k}. \end{aligned} \quad (2.27)$$

In essence, H_N represents a processing gain achieved by sending N transmissions; remarkably, this processing gain is identical to that achieved under the selection combining scheme [GZA10]. In other words, repetition diversity is translated into selection combining as the BS does not coherently add the repeated transmissions. Nevertheless, sending the packet multiple times increases power consumption for a given payload (and latency, yet it is typically less critical), and it worsens the intra-network interference due to congesting the network with more transmissions. The following corollary presents the optimal N that maximizes the success probability.

Corollary 5. *The optimal number of transmissions that maximizes $\mathbb{P}_s$ is*

$$N^* = \operatorname{argmin}_{N \in \mathbb{Z}} \left\{ N \left| (1+N)H_N - N > \frac{\hat{P}_I^\delta \tilde{\lambda}_I}{\beta_T \frac{\beta_F b}{M \cdot B} \lambda_T \lambda_{\text{IoT}}} \right. \right\}. \quad (2.28)$$

Further, in the absence of incumbents, then $N^* = 1$, i.e., the IoT device should send a single transmission per packet.

Proof: Let $f(N) = -\frac{a_1 H_N}{a_2 N + a_3}$. If $f(N+1) > f(N)$, then adding an additional transmission decreases the success probability. We can simplify the inequality as follows

$$\begin{aligned} -\frac{a_1 H_{N+1}}{a_2(N+1) + a_3} &> -\frac{a_1 H_N}{a_2 N + a_3} \\ \frac{a_1 H_{N+1}}{a_2(N+1) + a_3} &< \frac{a_1 H_N}{a_2 N + a_3} \\ \frac{H_{N+1}}{H_N} \cdot \frac{a_2 N + a_3}{a_2 N + a_3 + a_2} &< 1. \end{aligned} \quad (2.29)$$

For the above inequality to hold, we must have

$$\frac{H_{N+1}}{H_N} < 1 + \frac{1}{N + a_3/a_2}. \quad (2.30)$$

Since $H_{N+1} = H_N + \frac{1}{N+1}$, we can simplify the above inequality to

$$(1+N)H_N - N > \frac{a_3}{a_2}. \quad (2.31)$$

Further, if $\lambda_I = 0$, then $\tilde{\lambda}_I = 0$, and thus the smallest N that satisfies (2.31) is $N = 1$. ■

It is observed from (2.28) that the optimal number does not depend on the BS density or the decoding threshold. Likewise, if $B_I < M \cdot B$, then (2.28) becomes independent

of M , i.e., the optimal number of transmissions is independent of how many bands are used.

Ideally, the number of transmissions should be reconfigurable, and a lower number of repetitions is recommended in regions where the UNB network is the dominant operator. Not only using lower N , if recommended, reduces power consumption of the device, but it can also widen the scope of services provided by the UNB network. Recall that Sigfox allows the IoT device to report a maximum of six unique packets per hour, as each one is sent three times.² For applications that require constant monitoring with frequent status updates, using Sigfox becomes less appealing; however, if each device uses a single transmission, then the number of unique packets per hour increases to 18.

2.6 Simulation Results

In this section, we validate the theoretical expressions via Monte Carlo simulations, where the performance is averaged out over 10^4 spatial realizations. Unless otherwise stated, we use Sigfox US specifications for the UNB network parameters, which are listed in Table 2.2. The temporal traffic generation assumes each device sends six unique packets per hour, each with duration of $T \approx 0.35$ s. Note that this is the maximum number of packets per hour a Sigfox system currently supports. We further consider LoRa devices as interfering incumbents [BRV16], and assume they have a similar temporal traffic generation. Finally, in all figures, we use markers to denote Monte Carlo simulations and lines to denote the theoretical results.

2.6.1 Impact of synchronization and hopping patterns

We first validate the framework under different transmission access cases, focusing on the existing Sigfox protocol, i.e., single-band access and no BS association (identical

²In Europe, the UNB device has a duty cycle of 36s per hour; since each Sigfox transmission is 2s, the maximum number of transmissions is 18.

Table 2.2: Main parameters

Description	Parameters
IoT signal bandwidth	$b = 600\text{Hz}$
UNB multiplexing band	$B = 200\text{KHz}$
Number of unique packets per hour	$K = 6$ packets
Packet size	26 bytes
Transmission duration	$T = 26 \times 8/b$
Number of transmissions	$N = 3$
Number of bands	$M = 5$
IoT Tx power	$P_{\text{IoT}} = 14\text{dBm}$
IoT density	$\lambda_{\text{IoT}} = 30 \times 10^3/\lambda_B$
Incumbent bandwidth	$B_I = 125\text{KHz}$
Incumbent effective density	$\lambda_I = 10^3 \lambda_T/\lambda_B$
Incumbent Tx power	$P_I = 14\text{dBm}$ (over B_I)
Noise power	$P_N = -146\text{dBm}$ (over b)
Path loss exponent	$\alpha = 3.5$

trends hold for the different protocols, and hence the results are omitted). Fig. 2.7 shows the success probability with variations of the SINR threshold for different hopping patterns and access cases, i.e, different levels of time-frequency synchronization. We observe that the theoretical curves match well with simulations. We further make the following observations. First, there is a significant gain achieved when access is slotted in time and frequency, e.g., the median SINR approximately improves by 10dB compared to an unslotted time-frequency system. Second, a time-slotted system has the same performance as a frequency-slotted one. Since frequency synchronization is difficult to achieve for UNB networks, as signals are very narrowband, more efforts should be centered around implementing a slotted-time system, particularly because UNB signals span a long duration in time. Last, the random repetition scheme shifts the curves by

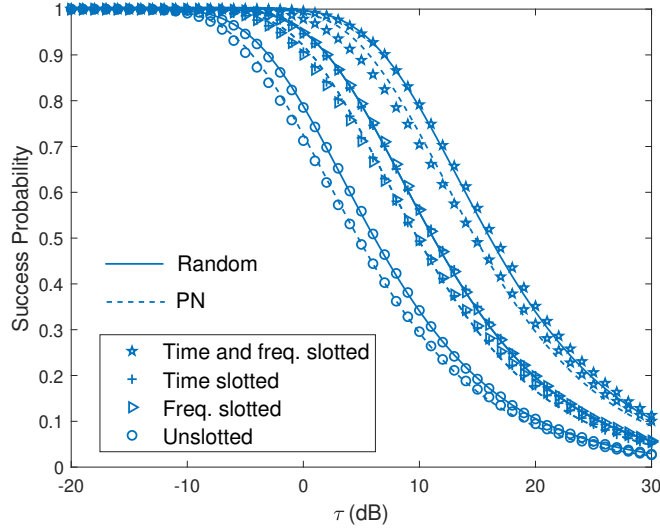


Figure 2.7: Success probability for the Sigfox protocol under different access cases.

approximately 1dB, outperforming the PN scheme, because in the former there is *an interference diversity gain* achieved as each transmission can experience a different set of interferers compared to the PN scheme. In the next results, we only focus on the unslotted time-frequency system, with random hopping patterns.

2.6.2 Success probability comparison

We compare the performance of several UNB protocols in terms of the success probability with variations of the SINR threshold. We consider the following schemes:

- **Benchmark multiband:** Each BS listens to M bands, requiring complex receivers.
- **Unslotted multiband:** This is the proposed protocol, where each BS randomly selects a band and the IoT device sends its packets at different bands.
- **Sigfox:** All BSs listen to a single band, and all IoT devices transmit their packets within that band.
- **Nearest BS:** This is similar to Sigfox, but the IoT device does not operate in a broadcast mode, i.e., it only communicates with the nearest BS.

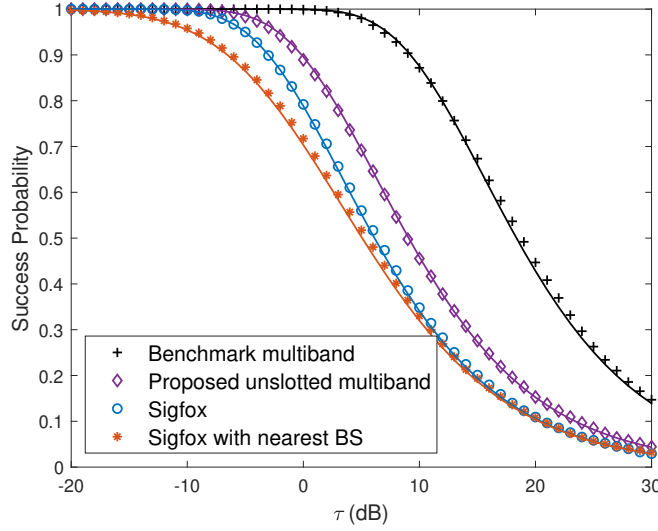


Figure 2.8: Success probability comparison of different UNB protocols.

The success probability performance is shown in Fig. 2.8. We have the following remarks. First, the benchmark multiband scheme significantly improves the success probability compared to Sigfox, e.g., the median SINR improves by approximately 12dB. However, such protocol is impractical to implement due to the high computational complexity of processing a wider band at a very fine resolution. The proposed unslotted multiband protocol provides a practical compromise, where the median SINR improves by 3dB relative to Sigfox, at virtually no hardware cost. Further, an access protocol with no BS association improves the cell-edge performance compared to restricting devices to connect to the nearest BS. For instance, Sigfox with geographical diversity improves the cell-edge SINR, i.e., the 5th percentile of the CDF, by 5dB, and it is further improved by an additional 3dB when the proposed multiband protocol is used.

2.6.3 Impact of the number of bands

We study the success probability performance with variations of the number of multiplexing bands, as shown in Fig. 2.9. We also show the performance of single-band access for reference. It is observed that increasing the number of bands tangibly improves the

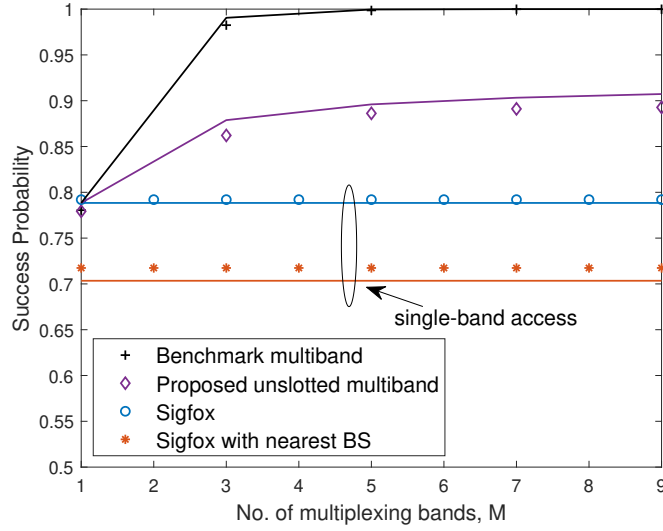


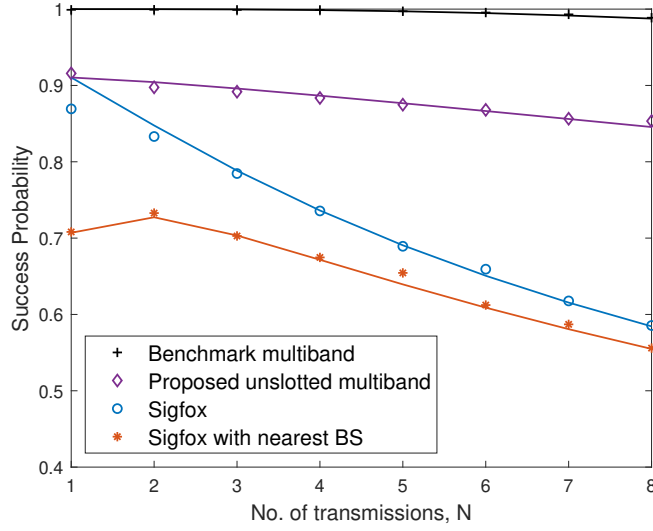
Figure 2.9: Success probability with variations of the number of bands, M ($\tau = 0\text{dB}$).

performance of the benchmark protocol as the density of collisions decreases with M ; in the unslotted multiband protocol, the performance still improves with M , yet the gain saturates, e.g., there is virtually no gain by increasing the spectrum from 1MHz to 1.8MHz. This follows because as M increases, the density of BSs listening to a particular band decreases, reducing the gain of geographical diversity.

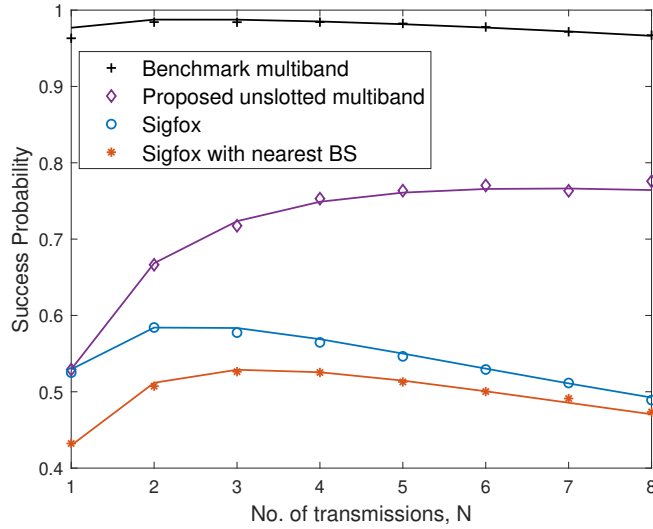
2.6.4 Impact of the number of repetitions

We study the success probability performance for a different number of transmissions. In Fig. 2.10a, we show the success probability of the different access protocols with variations of N for low incumbent density. It is observed that a single transmission per packet is optimal for protocols with geographical diversity; however, the absence of this diversity under nearest BS association pushes the optimal number to two transmissions. Another interesting observation is that multiband access is less sensitive to the intra-network interference compared to single-band access, as allowing the device to have a larger pool of channels to select from reduces collisions.

In Fig. 2.10b, we show the performance when the density of incumbent LoRa devices



(a) Low density: 1,000 LoRa devices/BS



(b) High density: 30,000 LoRa devices/BS

Figure 2.10: Success probability with variations of N ($\tau = 0\text{dB}$).

is the same as UNB devices. We have the following important observations. First, the optimal number under the benchmark multiband access and Sigfox is increased from one to two, which also asserts Corollary 5, where it is shown that the optimal number is independent of M if $B_1 < B \cdot M$. Second, in nearest BS association, the optimal number of transmissions is higher than that of Sigfox to compensate for the absence of geographical

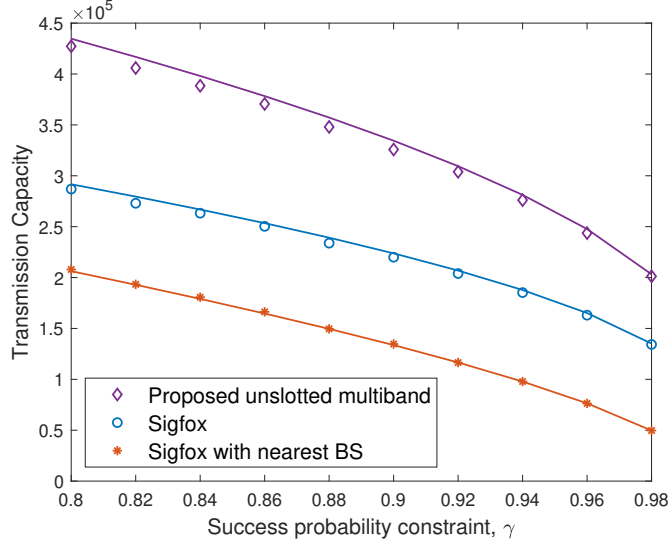


Figure 2.11: Transmission capacity for a given success probability constraint ($\tau = 5\text{dB}$).

diversity. Likewise, the proposed multiband access protocol has also a higher optimal number of transmissions as we achieve a geographical diversity gain by sending different packets at different bands, where each packet is received by a different subset of BSs.

2.6.5 Transmission capacity comparison

We compare the different single-band and proposed multiband protocols in terms of the transmission capacity, i.e., how many IoT devices can be supported by a network for a given success probability constraint. We assume the network is deployed over an area of $25 \times 25\text{km}^2$, and the average number of BSs in that area is 25. It is shown that UNB networks can provide coverage for a very large number of IoT devices. We observe that the proposed multiband protocol achieves the highest transmission capacity. For example, at a success probability constraint of 0.98, each UNB BS can support, on average, 8,000 IoT devices under the proposed protocol compared to only 2,000 IoT devices under Sigfox with nearest BS association.

2.7 Conclusions

In this chapter, an analytical framework has been developed to model and analyze UNB communications, an emerging paradigm that relies on ultra-narrowband signals to tackle the intra-network sharing among IoT devices and the inter-network sharing with incumbent networks. Several variants of UNB protocols are studied and compared in terms of the success probability and transmission capacity, where closed-form expressions are given for these metrics to identify fundamental limits and key performance trends under different network settings.

The analysis has shown that the geographical diversity, the number of signal repetitions, and the number of bands play key roles in the UNB network performance. First, the geographical diversity is essential for random access to support a massive number of IoT devices using only few BSs; otherwise, a high density of BSs is needed to enable massive IoT connectivity. Second, sending many repetitions helps improve robustness to the inter-network interference, yet it increases the intra-network interference. Indeed, it is shown that when the network is dominated by interference from IoT devices, a single transmission maximizes the success probability, whereas higher signal repetitions are needed when the density of incumbents increases or when the geographical diversity is absent. Third, using multiple bands reduces IoT collisions and the interference with incumbents. However, to fully exploit these gains, all BSs are required to listen to all bands. When each BS is restricted to listen to one of the bands, then the diversity achieved by sending several repetitions is not sufficient as the density of UNB BSs listening to a given band is reduced. To mitigate this, each packet transmission should be sent at a different band to exploit a geographical diversity gain, improving the success probability and transmission capacity over single-band protocols.

Other useful design guidelines have been also gleaned from the analysis. For instance, the interference diversity achieved by random hopping can further improve the UNB network performance. In addition, more efforts should be expended on enabling

synchronous access, as it brings tangible performance gains. Because UNB signals are extremely narrowband, they have longer transmission durations, making time synchronization easier to attain.

CHAPTER 3

Distributed Wideband Sensing-based Access

The high density of devices in massive IoT applications has made spectrum sharing a key challenge to fully realize large-scale deployment of these low-cost low-rate devices. In Chapter 2, we have discussed the use of UNB networks, which rely on extremely narrowband signals and random access. UNB networks are a variant of a more general class of low-power wide-area (LPWA) networks [CVZ16].

The vast majority of LPWA networks, in fact, rely on narrowband signals to improve robustness to interference and to connect a large number of devices. However, they have several limitations. First, they are constrained to very low-rate applications, with data rates below 50Kbps [RKS17]. Second, they primarily operate over the sub-1GHz spectrum due to its favorable propagation conditions, yet it is fragmented and has a limited bandwidth, e.g., 26MHz at 915MHz in the US. Third, LPWA networks rely on random access without any polite spectrum access mechanism, e.g., listen-before-talk (LBT), clear channel assessment (CCA), or dynamic frequency selection (DFS). Thus, they cannot be deployed at several unlicensed bands because of the regulatory requirements. For example, Japan and Europe mandate LBT for access at 5GHz, whereas dynamic frequency selection is required in some bands by the Federal Communications Commission (FCC) in the US [KJB17].

The focus of this chapter is to explore an alternative access method for massive IoT over the unlicensed spectrum. In particular, we aim to develop a system that uses spectrum sensing, making it compliant with the regulations at 2.4GHz and 5GHz, where more channels are available.

3.1 Related Work

One prime example of a network that relies on LBT for IoT access is the MulteFire network [Mul17], which is a 3GPP-compliant standard for cellular-like access in the unlicensed spectrum [LMR17].¹ It combines the simplicity of WiFi deployment and the reliability of cellular networks, and it is envisioned to enable a stand-alone unlicensed 5G-NR [3GP17d]. The authors in [LBA16, CLG17] theoretically analyze the coexistence of WiFi and LTE with LBT, showing that both systems benefit from such coexistence. In [RKF18], system-level simulations show that MulteFire outperforms WiFi in coverage and capacity. Alternative to using LTE-like IoT networks, the authors in [RS17] propose to modify the MAC protocol for WiFi itself, where IoT devices are notified by the WiFi access point (AP) about available *WiFi whitespaces* when the latter no longer has packets from/for WiFi devices. The presence of multiple APs is not considered, which could affect the scalability of this protocol. Finally, the aforementioned works, and the MulteFire network, merely support private IoT networks, e.g., for enterprise in-building cases, and thus they do not necessarily scale to support massive IoT unless each base station (BS) or AP is equipped with a wideband spectrum scanner. In this chapter, we develop a sensing assignment scheduler so that each BS learns the occupancy of a wideband spectrum using narrowband scanners.

Other works have proposed the use of compressive sensing techniques to enable low-cost wideband scanners for IoT networks [EI18, ZMQ18]. Nevertheless, cooperation among sensing scanners destroys spatial information about incumbents' energy footprints, i.e., all cooperating scanners arrive at the same sensing decision, and so available channels cannot be reused across space. In this chapter, we propose a truly distributed sensing algorithm, where cooperating scanners can arrive at different decisions, capturing the spatial variations of incumbents' footprints. Finally, the authors in [KLI17]

¹Other cellular-based unlicensed access standards include LTE-U and licensed-assisted access (LAA), yet they require licensed anchor carriers, i.e., unlicensed access is primarily used for carrier aggregation to improve network capacity [KJB17].

propose the use of radio environment maps (REMs) so that different IoT networks share the spectrum with rotating radars. The framework, however, requires radar operators to share radar locations and operations with a REM repository. In our work, we show that the output of the distributed sensing algorithm resembles a REM, which is obtained without any cooperation with incumbents.

3.2 Chapter Contributions and Organization

In this chapter, we propose a sensing-based architecture for unlicensed massive IoT, where BSs, e.g., small cells, are equipped with spectrum scanners to: (i) identify a large number of narrowband channels in a wideband spectrum, as many massive IoT applications have low-rate requirements, and to (ii) aggressively reuse the unlicensed channels over space to accommodate a high density of IoT devices. The former objective requires sensing at a fine *spectral* resolution, and the latter requires capturing the incumbents' footprints at a fine *spatial* resolution.

Our contributions in this chapter are twofold. First, we aim to limit the sensing burden at each BS, as sensing a wideband spectrum at a fine spectral resolution requires complex receivers. To this end, we formulate an integer program, where we optimize sensing assignments across BSs, so each one only senses a subset of the spectrum. The problem is combinatorial with high complexity in dense networks. Thus, we further develop a heuristic low-complexity assignment scheduler and compare its performance to the integer program and its linear relaxation. We note that sensing assignments have been studied before in a different context in [MHW15, LDL15, ZYY15]. For instance, in [MHW15, LDL15], the assignment is done such that each channel is sensed by one device, whereas in this work we require each channel to be sensed by multiple BSs for reliable cooperative sensing. In [ZYY15], the assignment aims to maximize the rate of secondary users, and thus it requires these users to sense all channels before making the assignment. In this chapter, the assignment is done prior to sensing, where we aim to

ensure that each BS, sensing a specific subset of channels, is surrounded by BSs sensing other subsets.

The second contribution is the development of distributed sensing and resource allocation algorithms. In particular, we propose a distributed sensing algorithm, where each BS senses its assigned subset of channels, shares and collects measurements from nearby BSs, and processes the collected data to infer the spectrum occupancy across all channels. Different from the distributed sensing algorithm in [SDM15], each BS may arrive at a different decision as the occupancy of a channel varies over space, achieving high spatial resolution. In addition, we use the combine-then-adapt diffusion algorithm [CRS15] and propose a novel update of the algorithm’s weights to quickly diffuse information about the wideband spectrum at each BS. The proposed sensing algorithm is shown to improve the deflection coefficient [UV08] compared to the non-cooperative energy detection. In fact, the algorithm weights resemble a REM of incumbents’ footprints. The output of the sensing algorithm determines which channels are available at each BS, henceforth denoted as *spatio-spectral blocks*. This output is then used to maximize a network utility function by optimizing which spatio-spectral resources are allocated across BSs, so that high spatial reuse is achieved while still limiting the intra- and inter-network interference. The network optimization is centralized, and thus we develop a fast distributed resource allocation algorithm, where BSs locally coordinate with their neighbors the allocation of the spatio-spectral blocks.

We validate the effectiveness of the proposed architecture via Monte Carlo simulations. Specifically, we compare the distributed sensing algorithm with centralized and non-cooperative sensing. We show that the proposed approach finds more spatio-spectral blocks, with a lower misdetection in comparison with existing methods. We show that the distributed allocation algorithm performs relatively well compared with its centralized implementation, and it provides better coexistence compared to non-cooperative allocation in terms of the number of devices served and the interference-to-noise ratio (INR) at incumbents. In addition, we simulate a large-scale case study, where

a massive IoT network shares a wideband spectrum with hundreds of outdoor WiFi APs. It is shown that the proposed system helps serve significantly more IoT devices compared to the non-cooperative scheme. Finally, a discussion of the computational complexity and the implementation of the proposed algorithms is presented throughout the chapter.

The rest of the chapter is organized as follows. The system model and the proposed sensing-based architecture are presented in Section 3.3. The sensing assignment problem and the proposed low-complexity schedulers are presented in Section 3.4. The distributed sensing algorithm is described in Section 3.5. The resource allocation problem and the proposed fast distributed algorithm are given in Section 3.6. Simulation results are given in Section 3.7 and the conclusions are drawn in Section 3.8.

Notation: In this chapter, we use boldface small and capital letters to denote column vectors and matrices, respectively, and use a normal font for scalars. We use $\mathbf{1}_K \in \mathbb{R}^K$ to denote the vector of all ones, $\mathbf{I}$ to denote the identity matrix, and $\mathbf{I}_{(x)}$ to denote the indicator function for the argument x . Finally, we use $\mathbb{E}[\cdot]$ and $\mathbb{V}(\cdot)$ for the expectation and variance operations, respectively.

3.3 System Model and Proposed Architecture

3.3.1 Network topology

We consider an IoT network that consists of a very large number of IoT devices and a dense deployment of K BSs that are connected to the same core network. We denote the set of BSs by $\mathcal{K} = \{1, 2, \dots, K\}$. When cooperation is enabled across BSs, then each k -th BS can exchange data with neighboring BSs, defined by the set $\mathcal{N}_k$. We assume that any BS within distance R from the k -th BS belongs to $\mathcal{N}_k$. We further assume $k \in \mathcal{N}_k$.

IoT devices are randomly deployed over space, and each device connects to the nearest BS. In this chapter, we assume a fully-loaded network, i.e., each BS has IoT devices

connected to. In case a channel can be accessed by a BS, the BS will schedule its associated IoT devices in a round-robin manner, although the framework can consider other scheduling and admission control algorithms.

The IoT network shares an unlicensed wideband spectrum, of bandwidth B , with different incumbent networks, e.g., WiFi, LoRa, etc. Since the majority of massive IoT applications have low-rate requirements, we further assume that the spectrum can be divided into narrowband channels, each of bandwidth $b \ll B$. The set of channels is denoted by $\mathcal{M} = \{1, 2, \dots, M\}$, where $M = \lfloor \frac{B}{b} \rfloor \gg 1$. Incumbent devices can occupy any part of this spectrum, and each one can use many contiguous narrowband channels.

An example of such coexistence of two networks over the unlicensed spectrum is a cellular network sharing the spectrum with WiFi over the unlicensed spectrum at 5GHz, where $B \approx 500\text{MHz}$. In this case, the cellular network deploys K small-cell BSs with LAA or with MulteFire specifications. The cellular network provides IoT-based services using LTE-M or NB-IoT operations [RVX16]. The former requires channels of bandwidth $b = 1.4\text{MHz}$, and thus $M \approx 350$, and the latter requires $b = 180\text{KHz}$, i.e., $M \approx 2775$. WiFi devices can use different channels with bandwidth ranging from 20MHz to 160MHz. We note further that the FCC has opened an inquiry on the use of the 5.9GHz-7.1GHz spectrum for 5G-NR [FCC17], i.e., in this case $B = 1.2\text{GHz}$ and $M \approx 6666$ for NB-IoT operation.

We assume that BSs have spectrum sensors, where energy-based sensing is used at each BS. We note that unlicensed cellular standards, e.g., LAA or MulteFire, all require BSs to perform LBT prior to channel access [LMR17]. In particular, if the energy of a channel exceeds a predetermined threshold τ (e.g., $\tau = -72\text{dBm}$ [Mul17] or $\tau = -62\text{dBm}$ [MCF16]), then the BS initiates a random back-off procedure or defers for an extended CCA period [KJB17]. Let $Y_{k,m}$ be the energy of the m -th channel measured by the k -th BS. Then, we use the following received energy model [LC17]

$$Y_{k,m} = V_{k,m} + \sum_j S_{j,k,m}, \quad (3.1)$$

where $V_{k,m}$ is the noise power and $S_{j,k,m}$ is the received signal power from a j -th incumbent transmitter over the m -th channel. Without loss of generality, we assume noise to be Gaussian, and independent at each channel and BS. For instance, the average noise power over a 20MHz channel is $\mathbb{E}[V_{k,m}] \approx -101\text{dBm}$. For the power received from an incumbent, we primarily use the 3GPP outdoor channel model that includes Rayleigh small-scale fading, shadowing, and different path losses depending on whether the incumbent-BS link is line-of-sight (LOS) or non-LOS [3GP17b, Sec. 7.4]. Such model captures the variations of the received signal over space, i.e., two nearby BSs should receive correlated energy measurements, when fading is averaged out over multiple observations.

3.3.2 Proposed sensing-based architecture

The objective of the proposed architecture is to help a network of dense BSs identify many narrowband channels in a wideband spectrum and reuse the available ones over space. To this end, to achieve a high spatio-spectral resolution map of available resources, the architecture consists of three interconnected components, which are described next.

3.3.2.1 A sensing assignment scheduler

Sensing a wideband spectrum, e.g., $B = 500\text{MHz}$, at a fine spectral resolution, e.g., $b = 180\text{KHz}$, requires high-complexity receivers, which incurs higher costs in dense network deployments. Therefore, we propose a sensing assignment scheduler, where the k -th BS senses $p_k \ll M$ channels to reduce the sensing burden. For instance, for NB-IoT operation, if each BS senses 20MHz out of the 500MHz spectrum, we have $p_k = 111$. To improve the reliability of sensing decisions, it is desired to have each m -th channel sensed by q_m BSs. To summarize, the assignment scheduler aims to solve the following problem: how should sensing tasks be assigned to a dense network of BSs so that each one senses only part of the spectrum, and each channel is sensed by multiple BSs?

3.3.2.2 A distributed wideband sensing algorithm

Once the sensing assignments are completed, each BS locally senses the assigned channels, measuring their received energies. Given that each BS has sensing reports only for their assigned channels, a key problem is how BSs should cooperate and process sensing data across the network so that each one learns the occupancy of the wideband spectrum. To limit the overhead with the core network, the proposed sensing algorithm is implemented in a distributed manner, where data exchange is done only among neighboring BSs. Such algorithm requires an initial phase, where the k -th BS measures the power of a reference signal broadcasted by the j -th neighboring BS, which is denoted by $\hat{P}_{k,j} \forall j \in \mathcal{N}_k$. Such measurement can be used to assess the quality of the sensing reports received from neighboring BSs.

3.3.2.3 A distributed resource allocation algorithm

The output of the sensing algorithm provides BSs with a list of available channels in the wideband spectrum. If each BS then enables data exchange with IoT devices over any of these channels, collisions would occur within the IoT network as multiple nearby BSs could pick the same channel. Thus, we propose a distributed resource allocation algorithm that aims to solve the following problem: given a set of available channels at each BS and its neighborhood, how should these spatio-spectral resources be allocated across BSs, i.e., reused over space, while limiting the intra-network and inter-network interference? An illustration of the network architecture is provided in Fig. 3.1.

3.4 The Sensing Assignment Scheduler

In this section, we present an optimization framework for the sensing assignment problem and propose a low-complexity scheduler that meets the assignment constraints.

Let the set of channels assigned to the k -th BS be denoted as $\mathcal{M}_k \subset \mathcal{M}$. Since this BS

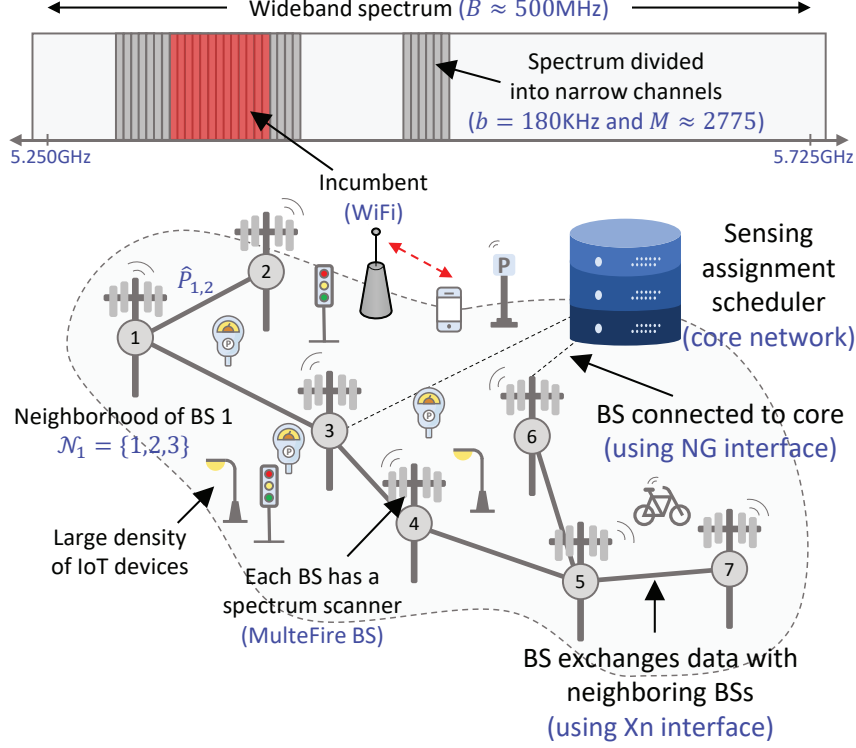


Figure 3.1: An illustration of the network architecture.

will lack information about the channels in $(\mathcal{M}_k)^c = \mathcal{M} \setminus \mathcal{M}_k$, it needs to collect sensing reports from neighboring BSs that sensed channels in $(\mathcal{M}_k)^c$. Nevertheless, collecting reports from distant BSs can be unreliable as the spatial footprints of incumbents vary over space. In other words, the sensing assignment scheduler should ensure that the distance between the k -th BS and any BS sensing channels in $(\mathcal{M}_k)^c$ is minimized.

Let $c_{j,k,m}$ denote the cost of the k -th BS using the j -th BS sensing report of the m -th channel. For instance, the cost can be the quality of the received power of reference signals broadcasted by the j -th BS, or it can be the distance between the two BSs, which is independent of the channel. Further, let $\mathbf{X} \in \mathbb{Z}_2^{K \times M}$ be the assignment matrix, i.e., the (k, m) -th entry $x_{k,m} = 1$ if the k -th BS is assigned to sense the m -th channel, and 0

otherwise. Then, we formulate the following sensing assignment problem²

$$\begin{aligned}
& \underset{\mathbf{X}}{\text{minimize}} && \max_{m \in \{1, 2, \dots, M\}} \sum_{j=1}^K \sum_{k=1}^K c_{j,k,m} x_{k,m} \\
& \text{subject to} && \mathbf{X}^T \mathbf{1}_K = \mathbf{q}, \\
& && \mathbf{X} \mathbf{1}_M = \mathbf{p}, \\
& && \mathbf{X} \in \mathbb{Z}_2^{K \times M},
\end{aligned} \tag{3.2}$$

where $\mathbf{q} = [q_1, q_2, \dots, q_M]^T$ and $\mathbf{p} = [p_1, p_2, \dots, p_K]^T$. In this formulation, the total cost of assigning a given number of BSs to sense an m -th channel is computed, and the objective is to minimize the maximum total costs across all channels, so that the k -th BS can reliably collect sensing reports for any channel in $(\mathcal{M}_k)^c$. The first constraint ensures that each m -th channel is sensed by q_m BSs, and the second constraint ensures that each k -th BS senses only p_k channels. Since channels are narrowband, and incumbents like WiFi devices have wider bandwidth, then adjacent channels can have high frequency correlation, which can be further incorporated in the cost of sensing reports $c_{j,k,m}$ to encourage two neighboring BSs to sense adjacent channels.

The problem in (3.2) is a generalization of the bottleneck assignment problem, which is known to be NP-hard [MT95]. Further, due to the large number of channels, M , it can be impractical to solve, and thus we consider a simpler framework as follows. First, we assume all BSs sense the same number of channels, i.e., $p_k = p \forall k$. In addition, we divide the spectrum into bands, instead of channels, where the number of bands is $L = \lfloor \frac{B}{p \cdot b} \rfloor$. Therefore, each BS will sense a single band, which is a subset of p consecutive channels, and each l -th band will be sensed by $\tilde{q}_l$ BSs. Thus, let $\tilde{\mathbf{X}} \in \mathbb{Z}_2^{K \times L}$ denote the sensing assignment matrix, with $\tilde{x}_{k,l} = 1$ when the k -th BS is assigned the l -th band, and let $\tilde{c}_{j,k,l}$ be the cost of the k -th BS using the j -th BS sensing report of the l -th band. Then, we

²An underlying assumption here is $\mathbf{1}_M^T \mathbf{q} = \mathbf{1}_K^T \mathbf{p}$.

consider the following simpler integer program

$$\begin{aligned}
& \underset{\tilde{\mathbf{X}}}{\text{minimize}} && \max_{l \in \{1, 2, \dots, L\}} \sum_{j=1}^K \sum_{k=1}^K \tilde{c}_{j,k,l} \tilde{x}_{k,l} \\
& \text{subject to} && \tilde{\mathbf{X}}^T \mathbf{1}_K = \tilde{\mathbf{q}}, \\
& && \tilde{\mathbf{X}} \mathbf{1}_L = \mathbf{1}_K, \\
& && \tilde{\mathbf{X}} \in \mathbb{Z}_2^{K \times L},
\end{aligned} \tag{3.3}$$

where the (k, l) -th entry of $\tilde{\mathbf{X}}$ is $\tilde{x}_{k,l}$ and $\tilde{\mathbf{q}} = [\tilde{q}_1, \tilde{q}_2, \dots, \tilde{q}_L]^T$. The optimization problem in (3.3) has a lower complexity than (3.2) as $L \ll M$, and it is practical as each BS will sense a single block of p channels instead of p not-necessarily consecutive narrowband channels. However, it is still a combinatorial problem with high computational complexity when $K \gg 1$.

If we relax the integrality constraint in (3.3), then the problem can be reformulated as a linear program (LP), where its optimal objective value is a lower bound on that of (3.3), yet the optimal variables $\tilde{x}_{k,l}^*$ are not necessarily integral. In addition, a simple rounding to $\tilde{x}_{k,l}^*$ in this case may violate the assignment constraints. Thus, we propose the following rounding algorithm.

Let $\mathcal{I}$ denote the set of optimal variables, minimizing the linear program, that are integral, i.e., $\mathcal{I} = \{(k, l) | \tilde{x}_{k,l}^* \in \{0, 1\}\}$. Similarly, let $\mathcal{F}$ denote the set of optimal variables that are fractional. We define $\bar{q}_l = \sum_{(j,l) \in \mathcal{I}} \tilde{x}_{j,l}^* \forall l$, i.e., the number of BSs fully assigned to sense the l -th band. Note that if $\bar{q}_l = \tilde{q}_l \forall l$, then the optimal solution of the linear program is also optimal for (3.3). Otherwise, if $\bar{q}_l < \tilde{q}_l$ for any l , then we need to find $\tilde{q}_l - \bar{q}_l$ BSs to assign them to those bands. To do so, we first find the BS-band pair with the highest fractional value, i.e., $\tilde{x}_{k,l}^* = \arg\max_{(j,m) \in \mathcal{F}} \tilde{x}_{j,m}^*$. If we have for such pair $\bar{q}_l < \tilde{q}_l$, then we set $\tilde{x}_{k,l}^* = 1$ and $\tilde{x}_{k,m}^* = 0 \forall m \neq l$, and we update $\mathcal{F}$. We repeat this process until $\bar{q}_l = \tilde{q}_l \forall l$. In essence, this algorithm starts with an incomplete assignment, yet with an objective value that is lower than the optimal value in (3.3). Then, the algorithm keeps assigning the remaining BSs to the remaining bands, starting with the BS-band pairs with the highest optimized fractional variables. The scheduler is summarized in Alg. 1.

Algorithm 1 A linear program sensing assignment scheduler

```

1: procedure INPUT( $\mathcal{K}, \{\tilde{c}_{j,k,l}\}, \tilde{\mathbf{q}}$ )
2:   Solve Relax the integrality constraint in (3.3) and solve the resulted linear program
3:   Find  $\mathcal{I}$  and  $\mathcal{F}$ 
4:   Compute  $\bar{q}_l = \sum_{(j,l) \in \mathcal{I}} \tilde{x}_{j,l}^* \forall l$ 
5:   while  $\left(\sum_{l=1}^L \bar{q}_l < K\right)$  do
6:     Find  $\tilde{x}_{k,l}^* = \operatorname{argmax}_{(j,m) \in \mathcal{F}} x_{j,m}^*$ 
7:     if  $(\bar{q}_l < \tilde{q}_l)$  then
8:       Set  $\tilde{x}_{k,l}^* = 1$  and  $\tilde{x}_{k,m}^* = 0 \forall m \neq l$ 
9:       Update  $\mathcal{F} = \mathcal{F} \setminus \{x_{k,l}^*\} \forall l$  and  $\bar{q}_l = \bar{q}_l + 1$ 
10:    end if
11:  end while
12:  Set  $\tilde{x}_{j,m}^* = 0 \forall \tilde{x}_{j,m}^* \in \mathcal{F}$ 
13: end procedure

```

3.4.1 A heuristic sensing assignment scheduler

If we know the BSs' coordinates, we can directly use that information to perform sensing assignments because our objective is to ensure that for any BS k , every band in the spectrum is sensed by at least one BS in vicinity. Indeed, collecting reports from very far BSs should be discouraged as the spatial footprint of an incumbent varies over space. Thus, the set of BSs sensing the l -th band cannot be clustered in a particular area, but rather they should be spread out over the region. This can be accomplished, for instance, by dividing the region into several clusters, where within each one, the entire spectrum is sensed by its members. To this end, we propose the following algorithm.

First, we pick a band, say l , and divide the BSs in the network into $\tilde{q}_l$ clusters $C_{i,l} \subset \mathcal{K} \forall i = 1, 2, \dots, \tilde{q}_l$, based on the coordinates of the BSs. From each cluster $C_{i,l}$, we pick one BS that minimizes the worst cost, i.e., we solve the following problem for each cluster

$$\begin{aligned}
 & \underset{\{\tilde{x}_{k,l} \in C_{i,l}; \tilde{x}_{k,l} \in \{0,1\}\}}{\text{minimize}} && \max_{j \in C_{i,l}} \sum_{k \in C_{i,l}} \tilde{c}_{j,k,l} \tilde{x}_{k,l} \\
 & \text{subject to} && \sum_{k \in C_{i,l}} \tilde{x}_{k,l} = 1
 \end{aligned} \tag{3.4}$$

Algorithm 2 A heuristic sensing assignment scheduler

```
1: procedure INPUT( $\mathcal{K}, \{\tilde{c}_{j,k,l}\}, \tilde{\mathbf{q}}, N$ )
2:   for  $n = 1 \rightarrow N$  do
3:     Set  $\tilde{\mathcal{K}} = \mathcal{K}$ 
4:     Permutate  $l = [1, 2, \dots, L]$ 
5:     for  $u = 1 \rightarrow L$  do
6:       Partition  $\tilde{\mathcal{K}}$  into  $\tilde{q}_{l(u)}$  clusters  $\{C_{i,l(u)}\}_{i=1}^{\tilde{q}_{l(u)}}$ 
7:       for  $i = 1 \rightarrow \tilde{q}_{l(u)}$  do
8:         Solve (3.4) to compute  $\tilde{x}_{k,l(u)}^* \forall k \in C_{i,l(u)}$ 
9:       end for
10:      Update  $\tilde{\mathcal{K}} \rightarrow \tilde{\mathcal{K}} \setminus \{k | \tilde{x}_{k,l(u)}^* = 1 \forall C_{i,l(u)}, i = 1, \dots, \tilde{q}_{l(u)}\}$ 
11:    end for
12:    Store  $\tilde{\mathbf{X}}_n^*$  and its corresponding objective value  $Z_n$ 
13:  end for
14:  Return  $\tilde{\mathbf{X}}_{n^*}^*$ , where  $n^* = \operatorname{argmin}_n Z_n$ 
15: end procedure
```

The optimal solution, in fact, is $\tilde{x}_{k,l}^* = \{1 | k = \operatorname{argmin}_e \sum_{j \in C_{i,l}} \tilde{c}_{j,e,l}\}$. After this iteration, there are $K - \tilde{q}_l$ remaining BSs. Thus, we pick another band, say u , and then cluster the remaining BSs into $\tilde{q}_u$ clusters, and solve (3.4) for each cluster, repeating the process until all bands are completed. Note that the bands that are picked earlier in the procedure will have a lower total cost as there are more BSs to pick from. To combat this, we repeat the whole process multiple times, randomizing the order of picked bands in each time. Then, we pick the one with the lowest maximum total cost.

3.4.2 Computational complexity and implementation of the heuristic scheduler

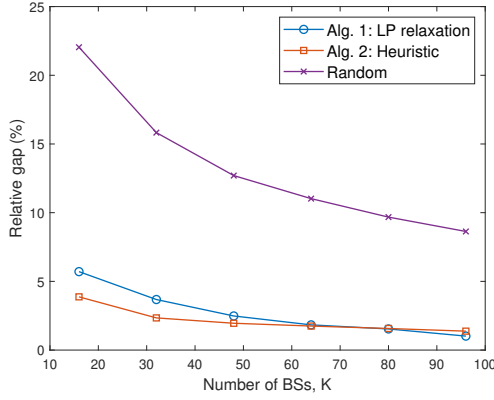
The proposed algorithm is summarized in Alg. 2, and it has overall a polynomial complexity. In particular, the algorithm computations are dominated by clustering the BSs. Using the k -means algorithm, clustering K BSs into q clusters based on their 2D coor-

dinates typically requires computations of order $O(2Kq)$ [BLH16]. Comparing this with solving a linear relaxation of (3.3), the latter requires computational complexity of order $O(K^3L^3)$ using interior-point methods, yet it does not require knowledge of BSs' coordinates.

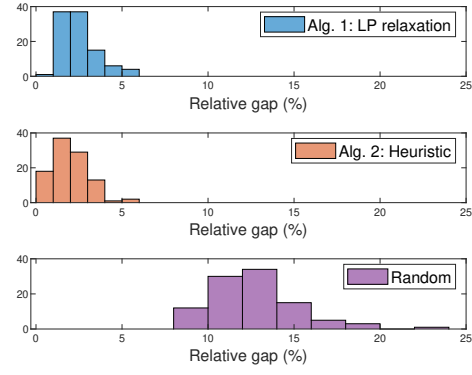
The assignment framework and proposed schedulers are centralized, i.e., they are solved at the core network. Thus, BSs need to forward the cost of collecting reports to the core network, or the BSs' coordinates may need to be known if the cost is the distance between BSs. The core network feeds back the assigned bands (and hence channels) to the BSs. Such two-way communication is done using existing interfaces, e.g., in 5G-NR, the NG interface connects the core network with BSs. We note that sensing assignments do not need to be frequently recomputed, as they primarily need to be updated when the network is changed, e.g., new BSs are deployed or existing ones are removed.

3.4.3 Numerical validation

We compare the solution of the integer program in (3.3), which is solved using MOSEK 8.0 solver with CVX in MATLAB, to the solution of the linear relaxation and Alg. 2. We also show the bottleneck performance of random assignment that satisfies the assignment constraints. Due to the high complexity of solving the integer program, we only compare the solutions using a relatively small number of BSs and bands. Specifically, we run 100 different realizations, where in each one we randomly deploy BSs on an area of $2 \times 2 \text{ km}^2$. We consider the costs to be equal to the distances between BSs. Here, we assume the number of bands to be $L = 4$, and $\tilde{\mathbf{q}} = \tilde{q}\mathbf{1}$, and hence the number of BSs is $K = \tilde{q} \cdot L$. Fig. 3.2a shows the average relative gap between the objective function in (3.3) when evaluated at the optimal solution of the integer program and the solution obtained by the various assignment algorithms for different densities of BSs. It is observed that the proposed algorithms perform very well relative to the integer program, and the gap reduces for a higher density of BSs, as higher density provides more flexibility of sensing assignments. We note that Alg. 2 requires the coordinates of BSs, and such side infor-



(a) Relative gap versus K



(b) Histogram of relative gap ($K = 48$)

Figure 3.2: Relative gap between the integer program (3.3) and other scheduling algorithms.

mation is useful, particularly when the density of BSs is lower. In Fig. 3.2b, we show the histogram of relative gaps across the different realizations when $K = 48$. Unlike random assignment, the proposed algorithms achieve objective values within 5% of the optimal integer program across the different realizations.

3.5 Diffusion-based Distributed Sensing

In this section, we present the distributed wideband sensing algorithm and validate its performance relative to non-cooperative energy detection in terms of the deflection coefficient.

Once sensing tasks are sent to BSs, each one measures the received energy in its assigned channels. For instance, in a MulteFire system, several CCA slots are used, where in each one, an energy measurement is compared to a detection threshold [KJB17]. One major issue that can arise in such system is that raw energy estimates can considerably fluctuate over different CCA windows because of fading, shadowing, or blockage. To mitigate such abrupt changes, an adaptive least-mean-squares (LMS) filter is proposed in [SDM15] so that the energy measurements are softened, and the filter weight is di-

rectly used as a test statistic. In particular, in LMS-based sensing, the BS minimizes the cost function $J(w_m) = \mathbb{E}[(d_{k,m} - w_m Y_{k,m})^2]$ by optimizing w_m , where it is desired to have $d_{k,m} = \mathbb{E}[Y_{k,m}]$ to reduce the variance across different energy measurements. Since the receiver does not have prior information about $\mathbb{E}[Y_{k,m}]$, $d_{k,m}$ is estimated in an online manner. The filter also admits a cooperative implementation, where the objective among cooperating BSs is to minimize $\bar{J}(w_m) = \sum_{k=1}^K \mathbb{E}[(d_{k,m} - w_m Y_{k,m})^2]$ [SDM15].

While it is shown in [SDM15] that the proposed LMS-based approach significantly improves the detection performance in comparison with the energy detector, the approach requires all BSs to have the same w_m to optimize. In other words, a cooperative procedure in this case implies that all BSs will aim to find the optimal estimate w_m^* that minimizes a global cost function, forcing each BS to have the same decision on the occupancy of the m -th channel, irrespective of its location. Since we aim to aggressively reuse channels over space, we need a fine-resolution frequency-space map of the spectrum. To this end, we propose to define a BS-specific cost function, i.e., $J_k(w_{k,m}) = \mathbb{E}[(d_{k,m} - w_{k,m} Y_{k,m})^2]$, where the optimizing variable is BS-dependent. Although each BS has a different $w_{k,m}$ to estimate, nearby BSs are expected to have correlated estimates due to the spatial correlation of energy footprints. Hence, cooperation among neighboring BSs can still improve the spectrum sensing reliability parallel to capturing the spatial variations of the incumbent's footprint. Thus, BSs cooperate to minimize $J_0(\mathbf{w}) = \sum_{k=1}^K J_k(\mathbf{w})$.

We aim to solve this global optimization distributively using diffusion-based distributed algorithms, which are known to be superior to other distributed strategies such as consensus and incremental strategies [CRS15]. In this algorithm, each BS has a vector of weights $\mathbf{w}_k = [w_{k,1}, w_{k,2}, \dots, w_{k,M}]^T$ to compute. Different from existing diffusion-based algorithms [CRS15, WTH17], the k -th BS only optimizes the vector entries corresponding to its assigned subset of channels, instead of optimizing all entries. The other entries will be computed using the measurement reports of the BSs that sense different bands.

The diffusion algorithm is centered around two main stages: the combination stage and the adaptation stage. In the combination stage, each BS shares its estimated $w_{k,m,i-1}$ (the subscript i denotes the iteration number) with its neighbors, where we propose the following combination policy

$$\psi_{k,m,i-1} = \begin{cases} \sum_{j \in \mathcal{N}_k} \alpha_{jk,m,i} w_{j,m,i-1}, & m \in \mathcal{M}_k \\ \sum_{j \in \mathcal{N}_k} \beta_{jk,i} w_{j,m,i-1}, & m \notin \mathcal{M}_k \end{cases}, \quad (3.5)$$

where $\{\alpha_{jk,m,i}, \beta_{jk,i}\}$ are non-negative *combining coefficients* that satisfy $\sum_{j \in \mathcal{N}_k} \alpha_{jk,m,i} = 1$ and $\sum_{j \in \mathcal{N}_k} \beta_{jk,i} = 1$. In the adaptation stage, the weights are updated as follows

$$w_{k,m,i} = \psi_{k,m,i-1} + \mathbf{I}_{(m \in \mathcal{M}_k)} \mu_k Y_{k,m,i} [d_{k,m,i} - Y_{k,m,i} \psi_{k,m,i-1}], \quad (3.6)$$

where μ_k is a constant step-size and $d_{k,m,i}$ is an approximate of $\mathbb{E}[Y_{k,m}]$ [SDM15]. In this work, we use an online first-order filter to compute $\mathbb{E}[Y_{k,m}]$. That is, we have

$$d_{k,m,i} = \zeta d_{k,m,i-1} + (1 - \zeta) Y_{k,m,i}, \quad (3.7)$$

where ζ is a scalar close but less than one. It can be shown that $d_{k,m,i}$ has the same mean as $Y_{k,m}$ but with a lower variance (see Appendix B.1 for the proof).

Parallel to the aforementioned combination and adaptation stages, we perform an online clustering of BSs with similar weights by adapting $\alpha_{jk,m,i}$. If the k -th BS is sensing the m -th channel, it assigns the following combining coefficient to the channel report sent from the neighboring j -th BS [CRS15]

$$\alpha_{jk,m,i} = \frac{(w_{k,m,i-1} + \mu_k \gamma_{k,m,i} - w_{j,m,i-1})^{-2}}{\sum_{j \in \mathcal{N}_k} (w_{k,m,i-1} + \mu_k \gamma_{k,m,i} - w_{j,m,i-1})^{-2}}, \quad (3.8)$$

where $\gamma_{k,m,i} = (d_{k,m,i} - Y_{k,m,i} w_{k,m,i-1}) Y_{k,m,i}$. This weighting mechanism looks at the similarities between the estimated $w_{k,m,i-1}$ and $w_{j,m,i-1}$, where a higher weight is given when these two values are closer to each other. Hence, as the algorithm progresses, cooperating BSs become clustered based on the similarities of their optimal solutions. If the m -th channel is not sensed by the k -th BS, then it will combine reports of this channel from neighboring BSs using the following coefficient

$$\beta_{jk} = \frac{\hat{P}_{k,j}}{\sum_{j \in \mathcal{N}_k \setminus \{k\}} \hat{P}_{k,j}}, \quad (3.9)$$

Algorithm 3 Proposed distributed sensing algorithm implemented by the k -th BS

```
1: procedure INPUT( $\mu_k, N_k, \zeta, \hat{P}_{k,j}$ )
2:   for  $i = 1 \rightarrow N$  do
3:     Measure  $Y_{k,m,i}$  and Compute  $d_{k,m,i} = \zeta d_{k,m,i-1} + (1 - \zeta)Y_{k,m,i}$ 
4:     Compute  $\gamma_{k,m,i} = (d_{k,m,i} - Y_{k,m,i}w_{k,m,i-1})Y_{k,m,i}$ 
5:     Compute weights  $\alpha_{jk,m,i}$  using (3.8) and  $\beta_{jk}$  using (3.9)
6:     Combine  $\psi_{k,m,i-1}$  using (3.5) and Adapt  $w_{k,m,i}$  using (3.6)
7:   end for
8:   Compare  $w_{k,m,N} \geq \lambda_{k,m} \forall k, m$ 
9: end procedure
```

where we have dropped the subscript i as only one reference signal received power, per BS, is used.

After N iterations, each k -th BS will have $w_{k,m,N} \forall m \in \mathcal{M}$, which will be compared with a threshold $\lambda_{k,m}$ to make a decision on that channel. The threshold vector is computed by feeding the algorithm with energy threshold τ , and using the output as a threshold for future samples.

3.5.1 Computational complexity and implementation of the distributed sensing algorithm

The proposed distributed sensing algorithm is given in Alg. 3. Its computational complexity is as follows. For each k -th BS, the combination stage in (3.5) requires $M|N_k|$ operations, i.e., multiplications and additions, whereas the adaptation stage in (3.6) requires $2M|N_k|$ operations. The adaptive combining coefficients require operations of order $2M|N_k|$. Thus, the overall computational complexity of the algorithm, per iteration, is $O(M|N_k|)$. We emphasize that since each BS does not entirely sense the wideband spectrum, the sampling rate can be reduced, mitigating the need for high-rate analog-to-digital converters (ADCs). Alternatively, for the same sampling rate used for the wideband spectrum, the BS acquires more samples when it senses a subset of the spectrum,

enabling finer spectral resolution and more sensing measurements.

The algorithm inevitably requires data exchange, yet it is restricted within the neighborhood of a BS. Thus, the number of data exchange messages transmitted by the BS in each iteration is proportional to $|\mathcal{N}_k|$. In each data exchange, the BS only shares the weights. Data communications between BSs can be done using standard interfaces, e.g., in 5G-NR, the Xn interface is used to connect two BSs using for example wired connections. In addition, control data exchange between the BS and IoT devices can be done using existing methods. For example, In LTE-LAA, a primary licensed carrier is used to carry control information, whereas in MulteFire, discovery reference signals (DRS) are periodically sent [Mul17], where upon decoding, the IoT device can learn which resource blocks have been allocated to it.

3.5.2 Performance analysis of the deflection coefficient

One metric that can be used to quantify the sensing performance of a test statistic T is the deflection coefficient, which is expressed as

$$\delta^2(T) = \frac{(\mathbb{E}_0[T] - \mathbb{E}_1[T])^2}{\mathbb{V}_0(T)}, \quad (3.10)$$

where the subscripts 0 and 1 denote the absence or presence of incumbents, respectively. In essence, sensing is reliable when the distance in the mean between the test statistic under the presence and absence of incumbent is large, or when the variations of the test statistic is low.

It is intractable to theoretically compute $\delta^2(w_{k,m})$, particularly because we use adaptive combining weights and realistic channel models. For this reason, we first use Monte Carlo simulations to show that Alg. 3 improves the deflection coefficient under realistic channels. Then, we consider a simpler model and a special case of the proposed algorithm to theoretically compute the deflection coefficient.

3.5.2.1 Numerical validation of Alg. 3

We consider a small network of $K = 9$ BSs with two different deployments: a grid deployment and a random deployment. We run 1000 realizations, where in each one, we randomly drop a single incumbent that transmits at power 23dBm when it is active. We then compute the deflection coefficients $\delta(Y_{k,m})$ and $\delta(w_{k,m})$ at each BS, which are indexed from the nearest to the furthest from the incumbent. It is assumed $\mu_k = 0.01 \forall k$. In Fig. 3.3a, we show one realization of the incumbent footprint when BSs are deployed on a grid. Here, the channel model is the outdoor 3GPP NR-UMi [3GP17b, Sec. 7.4]. We also show the connections between neighboring BSs. In Fig. 3.3b, we show the average deflection at each BS. It is evident that all BSs, whether they are close or far from the incumbent, have superior deflection coefficients under the proposed sensing algorithm in comparison with the non-cooperative system.

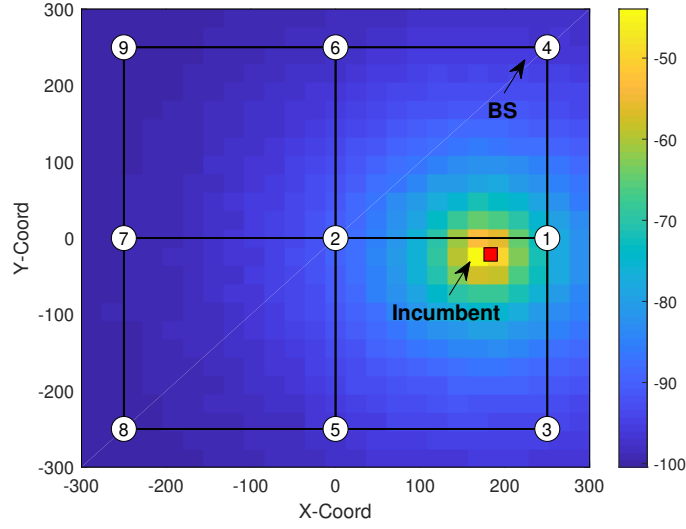
3.5.2.2 Theoretical analysis of the deflection coefficient

Only in what follows, we make the following assumptions for tractable analysis. Specifically, we assume all BSs sense the same channel, have the same objective to minimize, and use non-adaptive combining coefficients α_{jk} which we collect in a matrix $\mathbf{A}$. Under this model, and with slight abuse of notation, we can rewrite (3.5) in a matrix form as $\psi_i = \mathbf{A}^T \mathbf{w}_i$, where $\mathbf{w}_i = [w_{1,i}, w_{2,i}, \dots, w_{K,i}]^T$, i.e., the subscript i here denotes the iteration index, and we drop the subscript m in what follows because all BSs sense the same channel. Similarly, the adaptation stage in (3.6) can be rewritten as

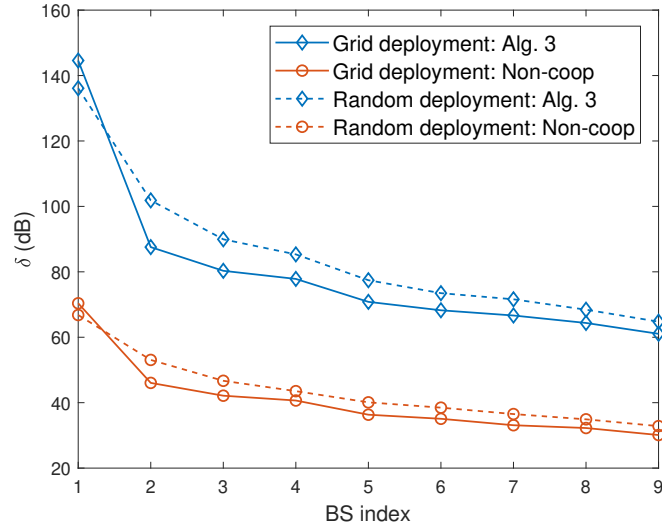
$$\mathbf{w}_i = (\mathbf{I} - \mathbf{M}\mathbf{Y}_i)\mathbf{A}^T \mathbf{w}_{i-1} + \mathbf{M}\phi_i, \quad (3.11)$$

where

$$\begin{aligned} \mathbf{M} &= \text{diag}(\mu_1, \mu_2, \dots, \mu_K), \\ \mathbf{Y}_i &= \text{diag}(Y_{1,i}^2, Y_{2,i}^2, \dots, Y_{K,i}^2), \text{ and} \\ \phi_i &= [Y_{1,i}d_{1,i}, Y_{2,i}d_{2,i}, \dots, Y_{K,i}d_{K,i}]^T. \end{aligned}$$



(a) Incumbent's energy footprint in dBm



(b) Deflection coefficient at different BSs

Figure 3.3: Deflection coefficient comparison between the proposed and the non-cooperative sensing algorithms.

In the steady-state, we have $\mathbf{w}_i \approx \mathbf{w}_{i-1}$, and thus taking the expectation on both sides of (3.11), we get

$$\mathbb{E}_\kappa[\mathbf{w}_i] = \left(\mathbf{I} - (\mathbf{I} - \mathbf{M}\mathbb{E}_\kappa[\mathbf{Y}_i])\mathbf{A}^T \right)^{-1} \mathbf{M}\mathbb{E}_\kappa[\boldsymbol{\phi}_i], \quad (3.12)$$

where $\kappa \in \{0, 1\}$. To evaluate the variance $\mathbb{V}_0(w_{k,i})$, let $\tilde{\mathbf{w}}_i = \mathbf{w}_i - \mathbb{E}_0[\mathbf{w}_i]$. Since the objective is the same across BSs, we have $\mathbb{V}_0(w_{k,i}) \approx \mathbb{E}_0[\|\tilde{\mathbf{w}}_i\|^2]/K$ [Say14]. Let $\mathbf{B} = \mathbb{E}_0[(\mathbf{I} - \mathbf{M}\mathbf{Y}_i)\mathbf{A}^T]$, and let $\mathbf{\Sigma}$ be a nonnegative-definite matrix that we are free to choose. By following the energy-conservation argument [Say14], we can show that

$$\mathbb{E}_0[\tilde{\mathbf{w}}_i^T(\mathbf{\Sigma} - \mathbf{B}\mathbf{\Sigma}\mathbf{B}^T)\tilde{\mathbf{w}}_i] = \mathbb{E}_0[(\boldsymbol{\phi}_i - \mathbf{Y}_i\tilde{\mathbf{w}}_i)^T\mathbf{M}\mathbf{\Sigma}\mathbf{M}(\boldsymbol{\phi}_i - \mathbf{Y}_i\tilde{\mathbf{w}}_i)], \quad (3.13)$$

where $\tilde{\mathbf{w}}_i = \mathbb{E}_0[\mathbf{w}_i]$. By solving the discrete-time Lyapunov equation $\mathbf{\Sigma} - \mathbf{B}\mathbf{\Sigma}\mathbf{B}^T = \mathbf{I}$, we obtain $\mathbf{\Sigma}$. In this case, the left-hand side in (3.13) becomes $\mathbb{E}_0[\|\tilde{\mathbf{w}}_i\|^2]$, and thus the variance can be directly computed.

3.5.2.3 Validation of the theoretical analysis

Assume that the energy sample when the incumbent is absent is the energy of a Gaussian noise sample with variance P_n , and when the incumbent is present, it is the energy of a non-zero mean Gaussian sample with the same variance and with mean $\sqrt{P_s}$. Then, it can be shown that the deflection coefficient of the energy detector is $\delta(Y_k) = \frac{P_s}{\sqrt{2}P_n}$ [SDM15], where P_s/P_n can be interpreted as the signal-to-noise ratio (SNR). We can also obtain simpler expressions for (3.12) and (3.13) under such assumptions (see Appendix B.2 for closed-form expressions). In Fig. 3.4, we show the theoretical and simulated deflection coefficient with variations of the SNR. Here, we consider the grid deployment with $K = 9$ BSs. We assume the same step-size for all BSs (shown in Fig. 3.4), and we use the averaging rule for the combining coefficients [Say14], i.e., $\alpha_{jk} = 1/|\mathcal{N}_k|$ if $j \in \mathcal{N}_k$. It is shown that the theoretical analysis is in agreement with Monte Carlo simulations. Further, Alg. 3 improves δ , and the improvements are higher when smaller step-sizes are used.

3.5.3 The proposed sensing algorithm as a REM

In this section, we show that the weights $w_{k,m}$ can be interpreted as a REM as they capture the variations of received energies over space. REMs can be useful to study

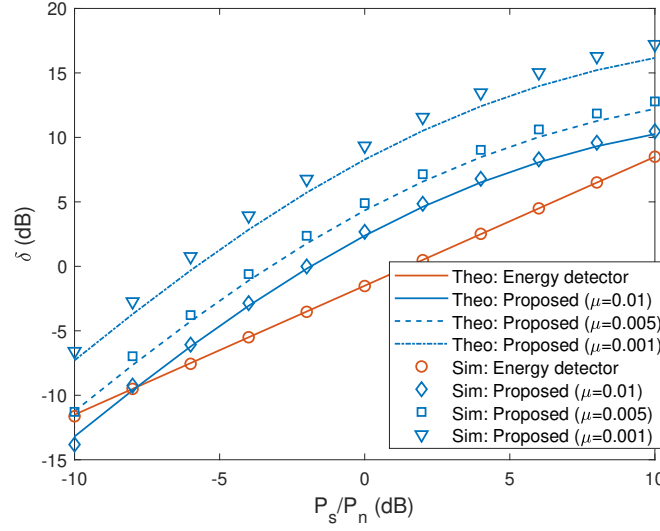


Figure 3.4: Deflection coefficient versus SNR.

the coverage of a network, to optimize spectrum utilization and management, and to improve coexistence of different networks [PZB15,KLI17].

To test the proposed algorithm on real data, we consider an office environment with one WiFi AP, using the set-up and data from [EVS18]. In the set-up, different sensing Raspberry PI boards (RPi3) are placed on the floor on a grid with inter-distance of 20cm. At each sampled location, henceforth denoted by a node, received signal strength (RSS) measurements are collected every three seconds, with a total of 40 RSS measurements per node. Fig. 3.5a shows the REM for this set-up.

Next, we randomly select a subset of nodes, denoted by the set $\mathcal{S}$, and perform Alg. 3 using the RSS measurements at the selected nodes. Each node cooperates with neighboring nodes that are within 2m, and obtain w_k (we drop the subscript m). Then, we use the simple inverse distance weighting spatial interpolation (IDW) method [EVS18] at every j -th node in $(\mathcal{S})^c$, which is expressed as $w_j = \sum_{l \in \mathcal{S}} \beta_{lj} w_l \forall j \in (\mathcal{S})^c$, where here we set β_{lj} to be inversely proportional to the distance squared between the j -th and l -th nodes [EVS18]. In Fig. 3.5b, we show the output of the proposed algorithm when all nodes participate. It is evident that the weights, w_k , capture the same variations as the

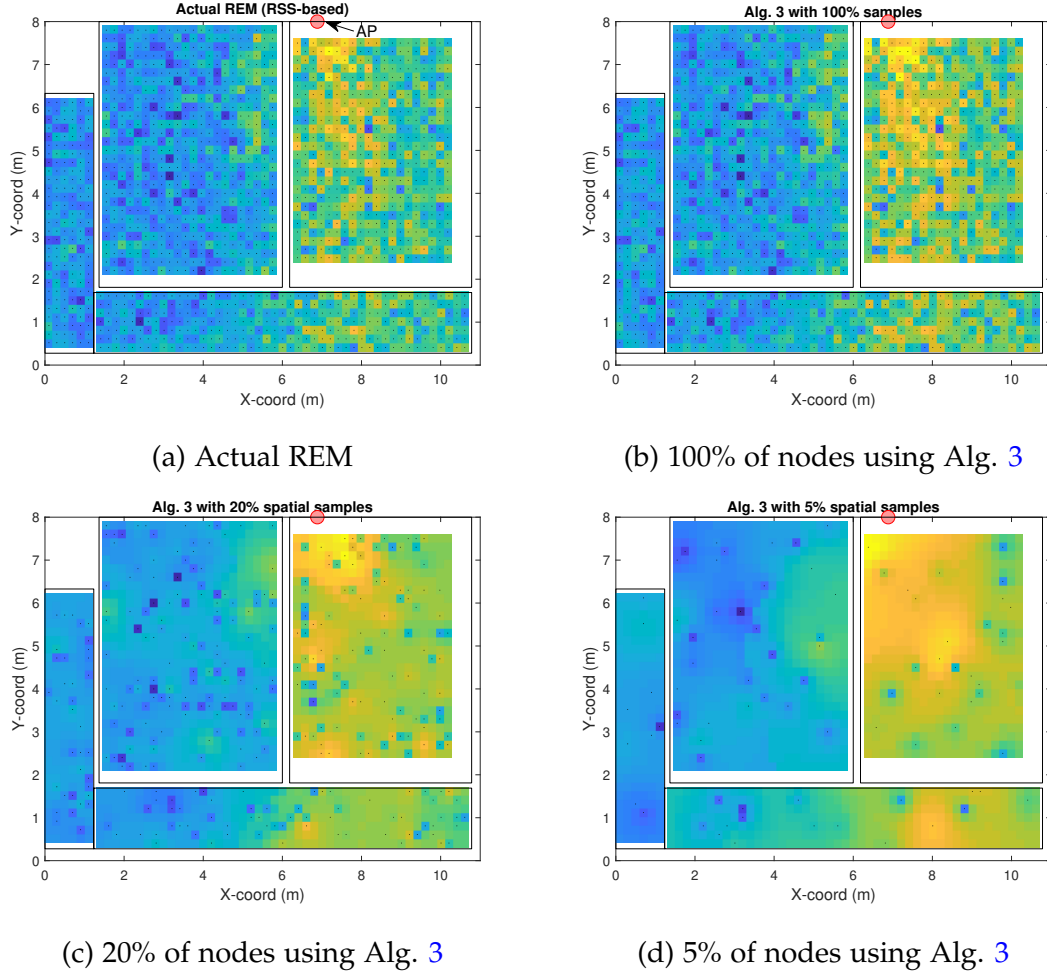


Figure 3.5: A REM for a single indoor WiFi AP using actual RSS measurements and the proposed algorithm.

RSS measurements. In Fig. 3.5c and Fig. 3.5d, we show the weights when 20% and 5% of the nodes participate and the rest use the IDW interpolation, respectively. Even with a lower density of sensing nodes, the weights still provide a good estimate of the REM.

3.6 Distributed Allocation of Spatio-spectral Resources

In this section, we present a network utility optimization where the spatio-spectral blocks, identified by Alg. 3, are allocated across BSs for access. We then present a fast distributed algorithm, where neighboring BSs coordinate resource allocation.

3.6.1 Network utility optimization

If each BS uses all identified spectral opportunities for IoT access, then nearby BSs may interfere with each other. Thus, we propose that neighboring BSs do not use the same channel, where our objective is to allocate the spectral blocks, with the highest spatial reuse under the aforementioned constraint.

We define the reward $r_{k,m}$ if the k -th BS is allowed to use the m -th channel. Since we do not rely on side information or feedback from the associated IoT devices, e.g., no CSI is available beforehand, we aim to merely use the outputs of Alg. 3 to quantify the reward, i.e., $\{w_{k,m,N}, \lambda_{k,m}, d_{k,m,N}\}$. To this end, we propose the following reward value $r_{k,m} = 10 \log_{10}(\frac{\tau}{d_{k,m,N}})$. The motivation behind this reward is twofold. First, the ratio $\frac{\tau}{d_{k,m,N}}$ quantifies how far the filtered energy measurement is from the detection threshold. Indeed, a larger ratio implies that the estimated aggregate incumbents' received power is lower in the m -th channel, and thus using such channel is useful for both the IoT network and the incumbents. Second, the logarithmic function encourages fairness among BSs. Thus, we have the following optimization problem

$$\begin{aligned}
& \underset{\{z_{k,m}\}}{\text{maximize}} && \sum_{m=1}^M \sum_{k=1}^K r_{k,m} z_{k,m} \\
& \text{subject to} && \sum_{j \in N_k} z_{j,m} \leq 1, \quad \forall k, m \\
& && z_{k,m} \leq \mathbb{I}_{(w_{k,m,N} \leq \lambda_{k,m})}, \quad \forall k, m \\
& && z_{k,m} \in \{0, 1\},
\end{aligned} \tag{3.14}$$

where $z_{k,m}$ is the optimizing variable, i.e., $z_{k,m} = 1$ when the k -th BS is allocated the m -th channel for access. The first constraint ensures that within the neighborhood of a BS, a channel is used by at most one BS to limit the intra- and inter-network interference. The second constraint ensures that a BS cannot be assigned a channel that is deemed unavailable by the sensing algorithm.

The integer program in (3.14) can be solved in a distributed manner using the dual decomposition method. In particular, it is observed that the problem can be decomposed into M separable problems, one per channel. For each m -th problem, let $\mathbf{z}_m =$

$[z_{1,m}, z_{2,m}, \dots, z_{K,m}]^T$, and introduce the dual variables $\mathbf{v}_m \geq 0$ to decouple the first constraint in (3.14). Thus, the dual problem for the m -th channel can be shown to be

$$\underset{\mathbf{v}_m \geq 0}{\text{minimize}} \quad \mathbf{1}_K^T \mathbf{v}_m + \sum_{k=1}^K D_k(\mathbf{z}_m, \mathbf{v}_m), \quad (3.15)$$

where $D_k(\mathbf{z}_m, \mathbf{v}_m)$ is the Lagrangian at the k -th BS, which is expressed as

$$D_k(\mathbf{z}_m, \mathbf{v}_m) = \underset{0 \leq z_{k,m} \leq I_{(w_{k,m}, N \leq \lambda_{k,m})}}{\text{minimize}} \quad z_{k,m}(r_{k,m} - \sum_{j \in N_k} v_{j,m}). \quad (3.16)$$

If we fix the dual variables, then the solution to (3.16) is

$$z_{k,m}^* = I_{(w_{k,m}, N \leq \lambda_{k,m})} \times I_{(r_{k,m} \geq \sum_{j \in N_k} v_{j,m})}. \quad (3.17)$$

Similarly, since (3.15) is differentiable, we can solve it using the gradient method, for a given $\mathbf{z}_m$. This process is recursively done until the dual converges. To summarize, the dual decomposition admits the following distributed implementation during the i -th iteration:

- The selection step: Each BS first collects $\{v_{j,m,i}\}$ from its neighboring BSs, and computes (3.17) for each channel, i.e., the BS selects an available m -th channel if there is a non-negative profit defined by $r_{k,m} - \sum_{j \in N_k} v_{j,m,i}$.
- The pricing step: Each BS collects the allocation decisions of neighboring BSs, and then solves the dual problem using the gradient method, i.e., each BS computes

$$v_{k,m,i+1} = \max \left\{ v_{k,m,i} - \tilde{\mu}_{k,i} \left(1 - \sum_{j \in N_k} z_{j,m}^* \right), 0 \right\}, \quad (3.18)$$

where $\tilde{\mu}_{k,i}$ is a predetermined step-size.

The dual variable can be interpreted as a price of selecting a channel. Indeed, if $\sum_{j \in N_k} z_{j,m}^* > 1$ in (3.18) then the price consistently increases in each iteration until only one BS can afford that channel. Although the dual decomposition algorithm is distributed, the performance of the algorithm requires to carefully select or update the step-size, and gradient-based methods can have slow convergence.

3.6.2 Fast distributed resource allocation

In this section, we present a heuristic, yet fast distributed resource allocation algorithm. The key idea is to avoid incremental increases of prices that are needed so only a single BS can afford the current price within its neighborhood. To do so, we define the highest possible reward of selecting an m -th channel in a given neighborhood, i.e., $\bar{r}_{k,m,0} = \max_{j \in \mathcal{N}_k} r_{j,m}$. Then, each BS initially selects the m -th channel if its reward $r_{k,m}$ is at least as large as $\bar{r}_{k,m,0}$. The selection stage is followed by a conflict resolution stage, which is resolved at every BS that has a channel selected within its neighborhood. Such conflict will be resolved by propagating the highest possible rewards among neighboring BSs. In particular, at the i -th message exchange, the k -th BS checks whether a channel is selected by any member in $\mathcal{N}_k$. If $\sum_{j \in \mathcal{N}_k} z_{j,m} \geq 1$, the BS recomputes the possible reward over the channel as follows

$$\bar{r}_{k,m,i+1} = \max_{j \in \mathcal{N}_k} \{z_{j,m} \bar{r}_{j,m,i}\}. \quad (3.19)$$

The BS then gives up the channel in the next iteration if the maximum among the new rewards exceeds its own reward, i.e., $r_{k,m} < \max_{j \in \tilde{\mathcal{N}}_k} \{\bar{r}_{j,m,i}, r_{j,m}\}$, where $\tilde{\mathcal{N}}_k = \{j | j \in \mathcal{N}_k, \sum_{l \in \mathcal{N}_j} z_{l,m} \geq 1\}$ is the set of BSs that participate in the conflict resolution process. A key observation here is that if BS k gives up a channel that it initially selected, then in the next iteration, we must have $k \notin \tilde{\mathcal{N}}_j \forall j \in \mathcal{N}_k$. Thus, the neighbors of BS k can now contend for that channel without worrying about the fact that BS k had a higher reward.

3.6.3 Implementation of the fast allocation algorithm

The proposed fast distributed allocation is summarized in Alg. 4. In terms of implementation, neighboring BSs need to share their original utility once. Then, in every iteration, the BS shares the maximum observed reward and whether it should be included in the conflict resolution process. In our simulations, the algorithm converges within a couple of iterations, and it performs relatively well compared to the centralized solution.

For illustration purposes, we consider a small network with $K = 10$ and a single

Algorithm 4 Proposed fast distributed allocation implemented by the k -th BS for the m -th channel

```

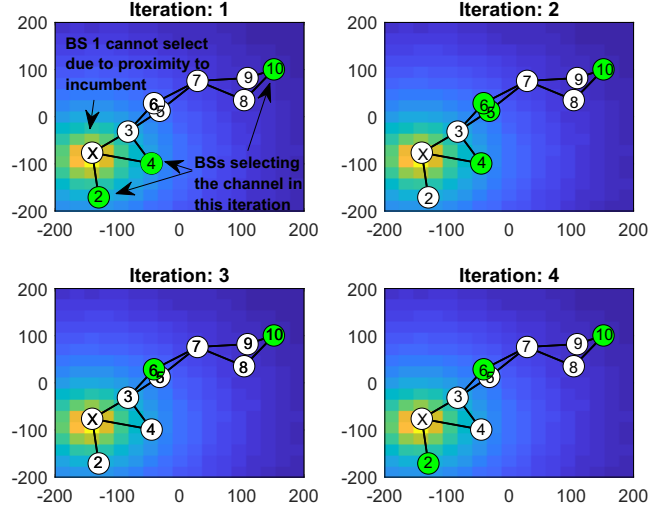
1: procedure INPUT( $r_{k,m}, \mathcal{N}_k$ )
2:   Initialize  $\bar{r}_{k,m,0} = \max_{j \in \mathcal{N}_k} r_{j,m}$  and  $\tilde{\mathcal{N}}_k = \mathcal{N}_k$ 
3:   for  $i = 1 \rightarrow N$  do
4:     Compute  $z_{k,m,i} = \mathbb{I}_{(w_{k,m,N} \leq \lambda_{k,m})} \times \mathbb{I}_{(r_{k,m} \geq \max_{j \in \tilde{\mathcal{N}}_k} \{\bar{r}_{j,m,i}, r_{j,m}\})}$ 
5:     if  $(\sum_{j \in \mathcal{N}_k} z_{j,m,i} \geq 1)$  then
6:       Update  $\bar{r}_{k,m,i+1} = \max_{j \in \mathcal{N}_k} \{z_{j,m,i} \bar{r}_{j,m,i}\}$ 
7:     else
8:       Update  $\bar{r}_{k,m,i+1} = \bar{r}_{k,m,i}$ 
9:     end if
10:    Update  $\tilde{\mathcal{N}}_k = \{j | j \in \mathcal{N}_k, \sum_{l \in \mathcal{N}_j} z_{l,m,i} \geq 1\}$ 
11:  end for
12: end procedure

```

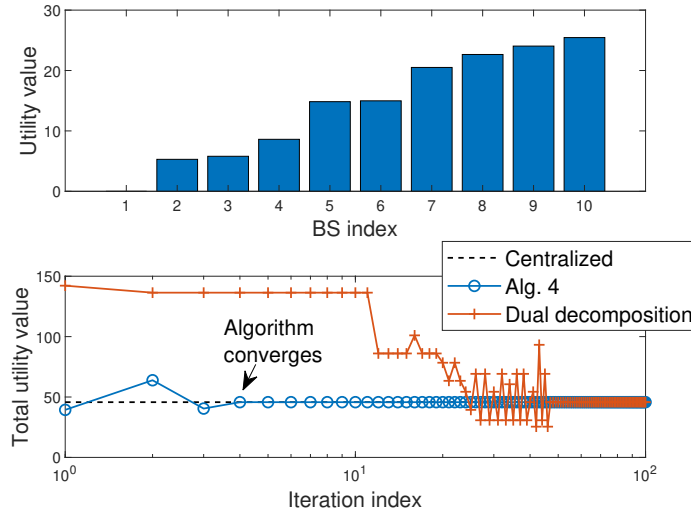
channel with one active WiFi AP. In Fig. 3.6a, we show the decisions made by BSs in each iteration. For example, in the first iteration, only BSs $\{2, 4, 10\}$ select the channel. However, since BS 1 now has two neighbors selecting the channel, a conflict resolution is used in the next iteration. In Fig. 3.6b, we show the utility of each BS, and the total utility over time when Alg. 4 and the dual decomposition are used. The former converges to the solution of the centralized scheme in four iterations, whereas the latter requires 47 iterations.

3.7 Simulation Results

We study the performance of the proposed architecture, comparing it with existing schemes. Specifically, the first two sections focus on a small-scale network, where the distributed sensing and resource allocation algorithms are evaluated, respectively. The last section considers a large-scale case study with massive IoT used for smart public parks. Unless otherwise stated, Table 3.1 lists the values of the parameters used.



(a) Distributed allocation over different iterations



(b) Utility over time

Figure 3.6: Illustration of distributed allocation over a small network.

3.7.1 Performance evaluation of the distributed sensing algorithm

3.7.1.1 Set-up

In this set-up, we randomly deploy 20 incumbent devices, where each one randomly picks one of the four channels, each of bandwidth 20MHz. All incumbents transmit at

Table 3.1: Simulation parameters

Parameter	Small-scale network	Large-scale network
BS deployment	Grid	Random
No. of BSs	$K = 100$	$K = 500$
No. of incumbents	20	2000 WiFi APs
No. of channels	$M = 4$	$M = 350; M = 2775$
Spectrum bandwidth	$B = 80\text{MHz}$	$B = 500\text{MHz}$
Channel bandwidth	$b = 20\text{MHz}$	$b = 1.4\text{MHz}; b = 180\text{KHz}$
Neighborhood radius	$R = 200\text{m}$	$R = 1\text{km}$
Channel model	Outdoor 3GPP NR-UMi [3GP17b]	
Detection threshold	$\tau = -62\text{dBm}$	
Sensing algorithm	$\mu_k = 0.01\forall k$ and $\zeta = 0.95$	

23dBm. For the channel model, we consider the 3GPP NR-UMi model, which is suitable for dense urban areas. We assume that the spectrum is centered around 5.43GHz, and all BSs and WiFi APs are deployed at a height of 10m. We run 100 realizations, where incumbents' locations, their channel usage, and propagation losses are randomized across realizations. We then consider the following schemes:

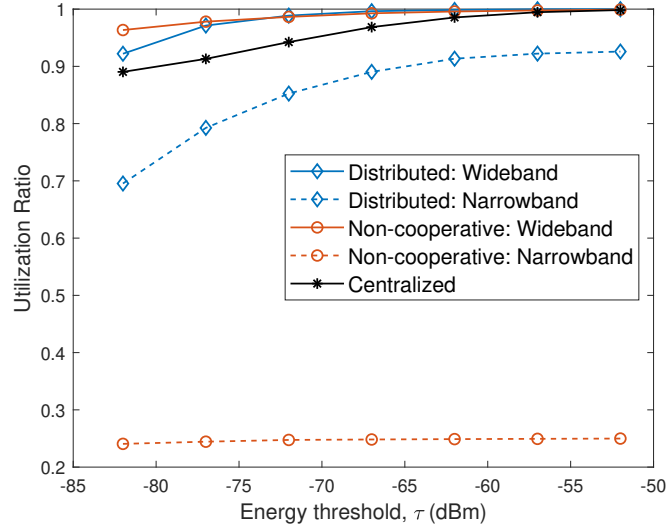
- **Distributed wideband:** We implement Alg. 3, yet assume all BSs sense the entire spectrum.
- **Distributed narrowband:** We implement Alg. 3, where each BS senses a single channel that is assigned via Alg. 2. Similar results are obtained if the scheduler is based on the linear relaxation of (3.3), and thus results are omitted.
- **Centralized:** We divide BSs into 25 clusters, where a cluster head collects energy measurements from the cluster members, and then combines them using equal gain combining, making one decision for the entire cluster. We only consider BSs with wideband sensors.

- **Non-cooperative wideband:** Each BS senses the entire spectrum and makes a local decision about the availability of each channel.
- **Non-cooperative narrowband:** Each BS randomly picks a channel to sense. The BS will not have information about other channels.

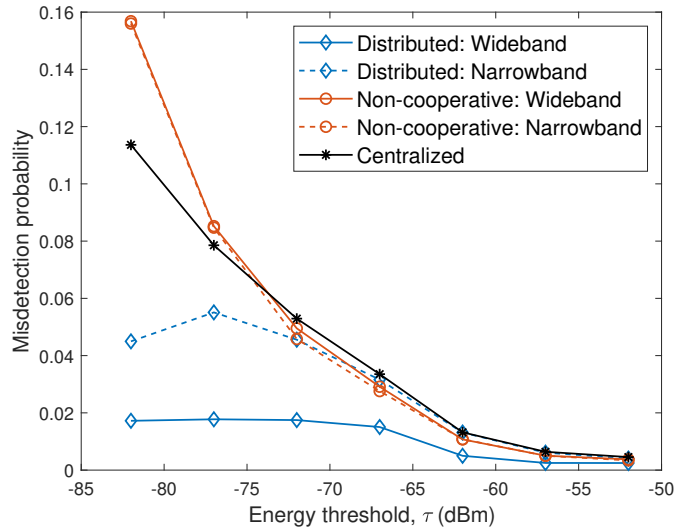
We compare the aforementioned schemes in terms of the *utilization ratio*, defined as the ratio of spatio-spectral blocks that are correctly identified as available by the scheme relative to those found by the genie scheme that knows all correct decisions, and in terms of the *misdetction probability*, which is defined as the probability of incorrectly deciding a spatio-spectral block is available.

3.7.1.2 Impact of detection threshold

We first study the performance under different detection thresholds. Fig. 3.7a shows the utilization ratio with variations of the energy threshold. Clearly, increasing the energy threshold relaxes the coexistence requirement of the IoT network and incumbents, and thus more resources can be reused over space and frequency. Comparing the different schemes, we make the following observations. First, both the proposed and non-cooperative wideband solutions identify the highest number of available resources, as each BS senses the entire spectrum at its location. However, the proposed solution significantly outperforms the non-cooperative one in terms of misdetection, as shown in Fig. 3.7b, since the latter may incorrectly decide a busy channel to be available in the presence of a fading channel and/or shadowing. This is not the case with the proposed diffusion algorithm as cooperation helps enhance the reliability of decisions and reduce the misdetection probability. Second, the distributed narrowband solution significantly outperforms its non-cooperative counterpart although both schemes enforce each BS to sense one channel. This follows because by the end of the diffusion-based sensing procedure, each BS will have occupancy information across all channels, whereas in the non-cooperative one each BS will be limited to the availability of the sensed channel.



(a) Utilization ratio

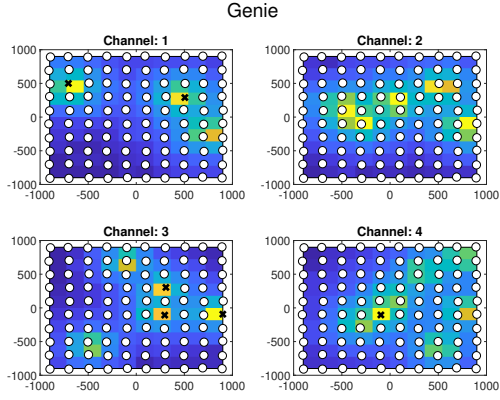


(b) Misdetection probability

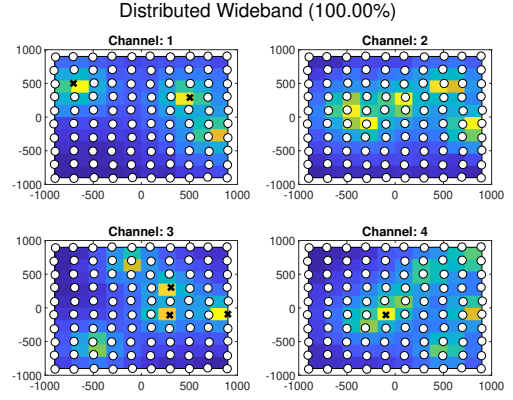
Figure 3.7: Performance with variations of the energy threshold.

Finally, the centralized solution performs well when the detection threshold is high. For lower detection threshold, if a single BS observes high energy, then it can be sufficient to bias all other BSs in the cluster to declare the channel to be busy.

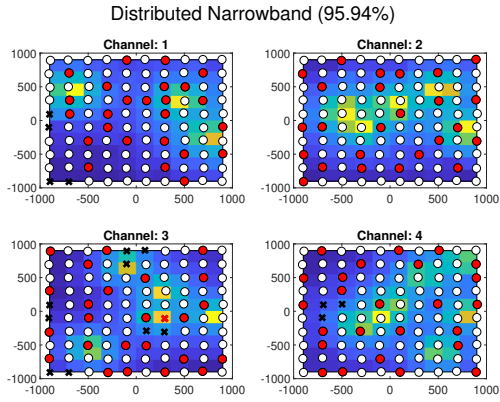
We also show in Fig. 3.8 a snapshot of the spatial energy footprints across all channels, and the corresponding decisions at each BS for every scheme during that realiza-



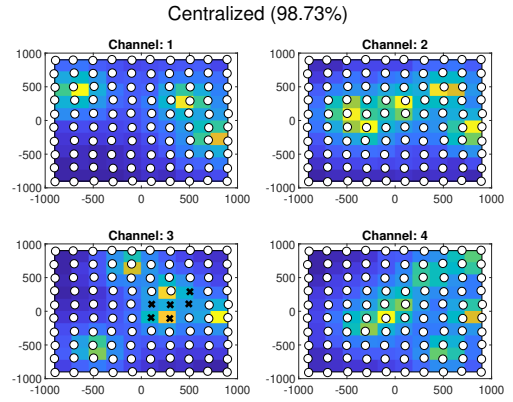
(a) Genie system



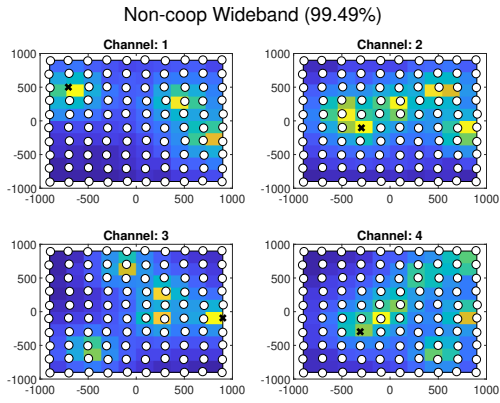
(b) Distributed wideband



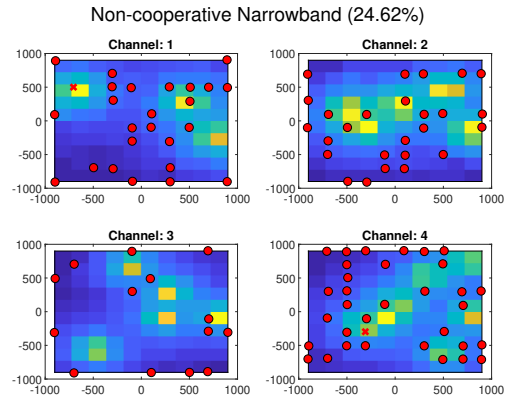
(c) Distributed narrowband



(d) Centralized



(e) Non-cooperative wideband



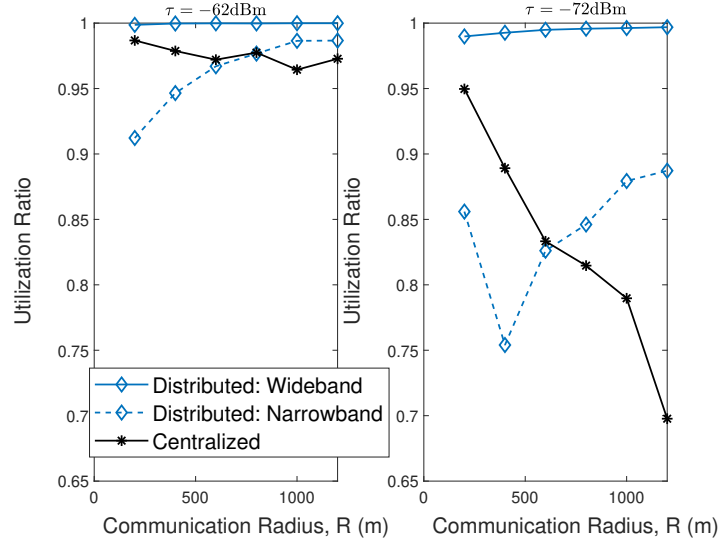
(f) Non-cooperative narrowband

Figure 3.8: The spatio-spectral footprints. ‘o’ and ‘x’ denote available and busy decisions at their locations, respectively. For narrowband systems, BSs with their assignments are highlighted in a different color.

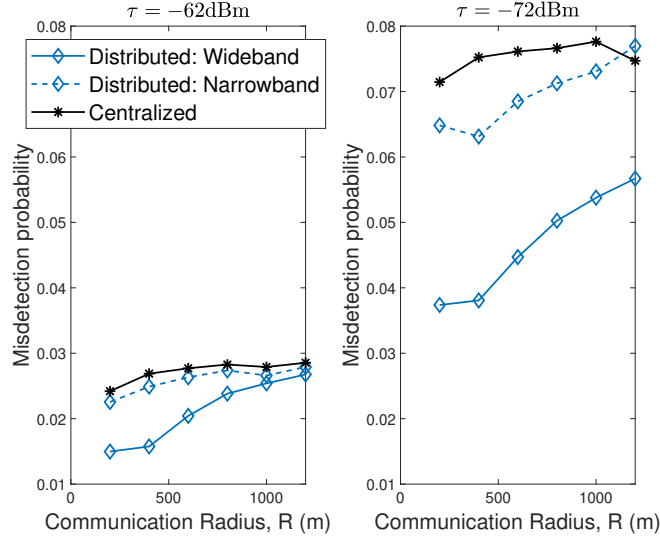
tion, where $\tau = -62\text{dBm}$. For the narrowband schemes, we highlight BSs with assigned channels in a different color. We also show the utilization ratio of the different schemes in that specific snapshot. It is evident that the distributed narrowband algorithm identifies the majority of available resources although each BS is assigned a single band.

3.7.1.3 Impact of communication radius

We study the performance of the cooperative algorithms in terms of the communication radius R , which affects the neighborhood and cluster sizes for the distributed and centralized sensing algorithms, respectively. We note that for a given R , we find the average number of neighbors, and use it to determine how many clusters to use for the centralized scheme. In Fig. 3.9a, we show the utilization ratio for different communication radii. It is observed that increasing R for the distributed sensing algorithm does not destroy the spatial information, but on the contrary, it improves the utilization ratio, particularly for the narrowband system. This follows because each BS collects more sensing reports while still using adaptive weights throughout the sensing process to ensure that reports from nearby BSs are of higher value to those collected from far BSs. Such improvements come at the expense of a slight increase in misdetection as shown in Fig. 3.9b. This follows because a BS that receives a weak incumbent signal could decide the incumbent is inactive when many of its neighboring BSs do not detect the incumbent. In contrast, not only misdetection increases in the centralized scheme, but its utilization ratio also decreases with R . This follows because all members of the cluster arrive to a single global decision, and thus as the cluster size increases, the spatial resolution decreases, with a tangible reduction in the utilization ratio when the detection threshold is very low, e.g., -72dBm .



(a) Utilization ratio



(b) Misdetection probability

Figure 3.9: Impact of communication radius on the performance of cooperative schemes.

3.7.2 Performance evaluation of the distributed resource allocation algorithm

3.7.2.1 Set-up

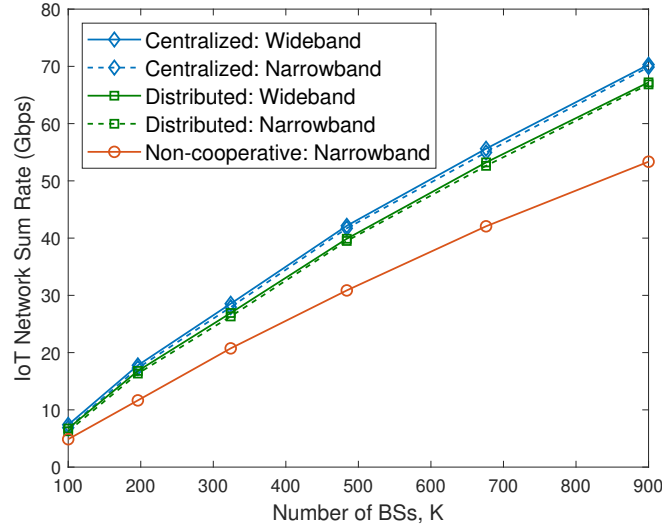
We use the previous set-up, where now we randomly drop a very large number of IoT devices in the network so that each BS has at least four IoT devices to schedule. We

assume the transmit power of IoT devices to be 14dBm. We then consider the following schemes:

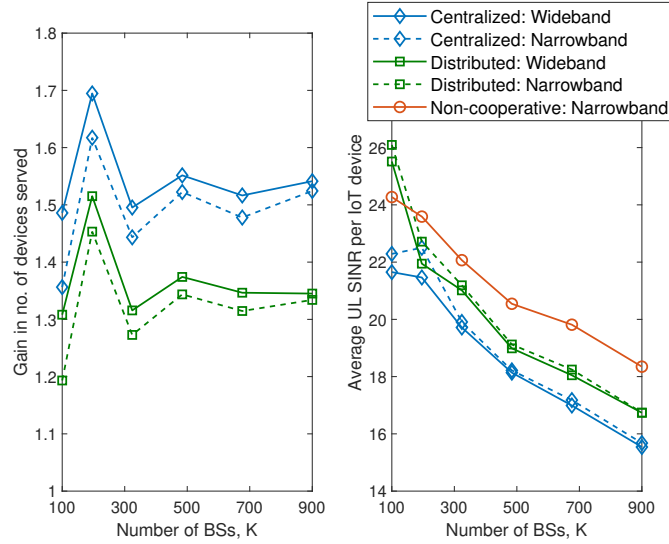
- **Centralized wideband and narrowband:** We use the sensing decisions of the distributed wideband and narrowband sensing algorithms, respectively, and then solve (3.14).
- **Distributed wideband and narrowband:** We use Alg. 4 for resource allocation. The sensing decisions are obtained from the wideband and narrowband sensing algorithms, respectively.
- **Non-cooperative narrowband:** Each BS uses the channel it sensed if it is deemed available.

3.7.2.2 Impact of network densification

We study the performance of the different schemes when we increase the density of BSs, which is accomplished by reducing the inter-site distance between them. In Fig. 3.10a, we show the IoT network sum rate across the 80MHz spectrum. As expected, the centralized framework achieves the highest network capacity because it has global knowledge of the sensing decisions and rewards at each BS. The wideband implementation is slightly better than its narrowband counterpart as the former has very reliable sensing performance (cf. Fig. 3.7). More importantly, the distributed resource allocation performs relatively close to the centralized implementation. Overall, the proposed schemes significantly improve the IoT network sum rate compared to the non-cooperative allocation. This can be explained via Fig. 3.10b. In particular, we show the gain of the proposed schemes over the non-cooperative scheme in terms of the average number of IoT devices served in the network. Focusing on the schemes with narrowband sensing, it is shown that in dense networks, the centralized and distributed narrowband schemes increase the number of devices by 1.5x and 1.3x, respectively, compared to the non-cooperative scheme. Equally important, the mean SINR for the IoT device only slightly degrades un-



(a) Sum rate of the IoT network



(b) Gain in number of devices served and UL SINR per device

Figure 3.10: Comparison of the different spatio-spectral resource allocation schemes in terms of the sum rate, the number of users served, and UL SINR per device.

der the proposed schemes, explaining the overall increase in the IoT network capacity. For example, when the density of BSs is 900, the distributed narrowband scheme adds approximately 300 devices to the network compared to the non-cooperative scheme, yet

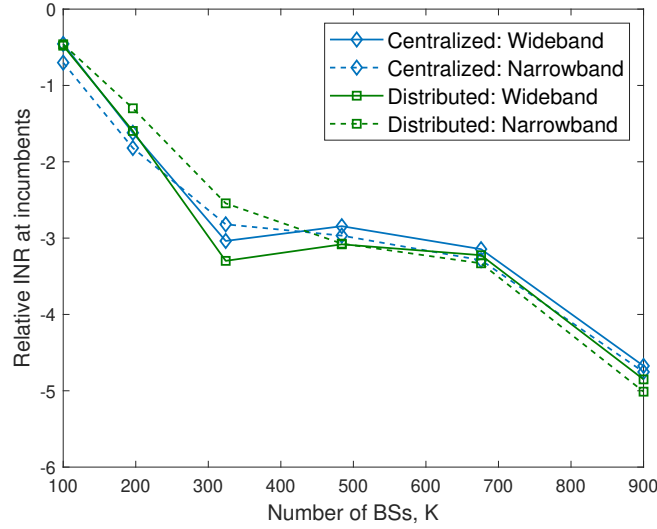


Figure 3.11: Relative INR at incumbents.

the SINR per IoT device under the former is only 1.6dB lower than that under the non-cooperative scheme.

In Fig. 3.11, we show the INR at the incumbents relative to that when non-cooperative allocation is used. It is evident that the incumbents are more protected under the proposed scheme compared to the non-cooperative one, particularly under high network densification. For example, the INR is reduced by approximately 5dB under the distributed narrowband scheme when the BS density is 900 BSs. To summarize, the proposed allocation scheme benefits both the IoT network and the incumbent network.

3.7.3 Case study: IoT for public parks

We test the proposed distributed narrowband system in a more realistic setting. Specifically, we consider a massive IoT application for the public parks in New York City, where sensors and machines can be deployed for water management, tracking traffic activities in the parks, etc. We randomly drop 500 BSs across the city and use the NYC *Open Data* to extract the locations of 2000 outdoor public WiFi APs, where we use their exact coordinates. We treat these APs as incumbents that use the same spectrum. We

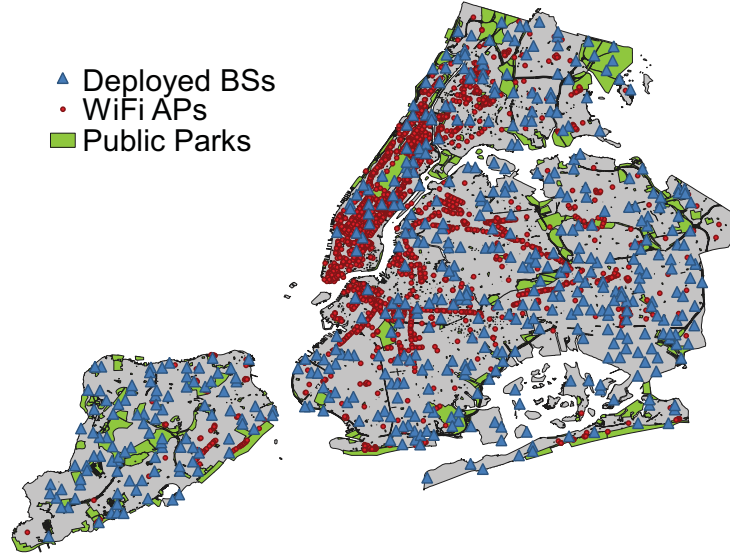


Figure 3.12: An illustration of the large-scale network set-up.

consider two IoT operations: LTE-M and NB-IoT, where the spectrum is channelized into 1.4MHz and 180KHz channels, respectively. Each WiFi AP is assumed to transmit a signal at 30dBm over the signal bandwidth, which is either 20MHz, 40MHz, or 80MHz within the wideband spectrum. In the proposed system, the assignment scheduler, i.e., Alg. 2, assigns each BS a subset of channels with a total bandwidth of 20MHz. Finally, we randomly drop 1×10^5 IoT objects in the public parks and assume all require access to the network. An illustration of the network is given in Fig. 3.12.

Fig. 3.13 shows the average number of devices that are scheduled over channels that are correctly identified as available for each scheme. It is observed that for NB-IoT, almost all devices are scheduled via the proposed and the non-cooperative wideband schemes. Further, the proposed system significantly outperforms the non-cooperative narrowband system, emphasizing that cooperation across neighboring BSs not only enhances the sensing reliability of a channel, but it also helps infer the occupancy of other channels, given a proper sensing assignment. Similar trends hold for LTE-M, yet fewer devices are scheduled in comparison to NB-IoT, as each one requires larger bandwidth.

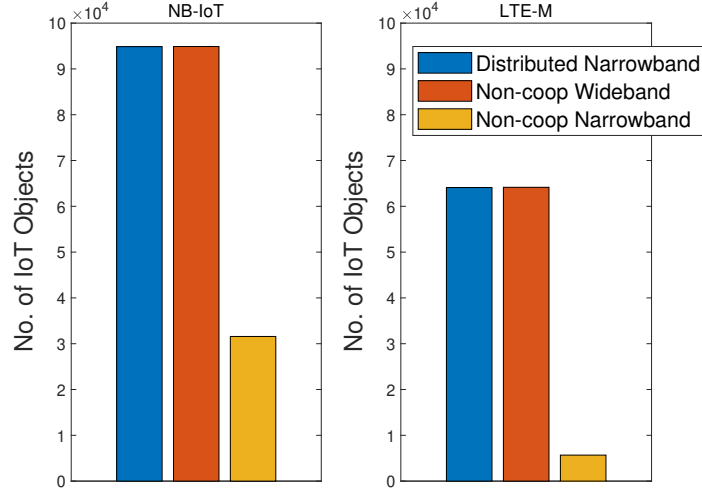


Figure 3.13: Average number of scheduled devices per scheme.

3.8 Conclusions

Providing massive IoT connectivity in the unlicensed spectrum requires exploring a large pool of narrowband channels and aggressively reusing them over space. To obtain a high spatio-spectral map of available resources, sensing a wideband spectrum at a fine resolution is required, while distributed processing of sensing reports helps capture the spatial variations of incumbents' footprints.

In this chapter, we have proposed a sensing assignment scheduler to limit the sensing burden on each BS. In particular, the proposed scheduler ensures that every BS is surrounded by other BSs sensing other parts of the spectrum. We have further developed a distributed sensing algorithm, where BSs locally share and process sensing reports, improving the sensing reliability while maintaining the spatial resolution thanks to the on-line clustering running in parallel to sensing. When each BS obtains the spatio-spectral map at its location, distributed allocation of these resources across BSs is essential to limit the intra- and the inter-network interference. To this end, we have proposed a fast distributed allocation algorithm that ensures a channel is used by a single BS in a given neighborhood. Simulation results have demonstrated the effectiveness of the proposed

architecture in comparison with centralized and non-cooperative schemes. It is shown that distributed sensing helps identify more spatio-spectral blocks at lower misdetection. As a consequence, distributed resource allocation enables connecting more IoT devices while providing higher protection to incumbents.

Acknowledgment

The authors would like to thank Zakarya El-friakh and Ljiljana Simić from the Institute for Networked Systems, RWTH Aachen University, for providing the RSS measurements for the indoor REM.

CHAPTER 4

Massive IoT over Drone-assisted Cellular Networks

Fifth-generation new-radio (5G-NR) is set to unlock new application scenarios at different fronts. One specific use case is the native support of massive machine-type communications (mMTC) [ITU15], which we refer to as massive IoT in this dissertation. Cellular networks have a ubiquitous infrastructure and a well-established protocol stack, making cellular access a natural choice for massive IoT applications.

The support of IoT services over cellular networks has culminated in recent years by the introduction of new user categories tailored for IoT applications, e.g., CAT-0, CAT-M1, and NB-IoT, in addition to enabling extended discontinuous reception (eDRX) to reduce power consumption [3GP15a, RVX16]. Such IoT optimizations, however, are still limited to services that do not require a large deployment of sensors and machines. Indeed, as the density of IoT devices increases, several challenges emerge [SGN16]. First, collisions among IoT devices increase due to the increased number of access requests, making retransmissions more frequent, and thus affecting the energy-efficiency (EE) of IoT devices. Second, interference increases when IoT devices share the spectrum with cellular UEs, degrading the coverage needed for the former and reducing the spectral efficiency (SE) for the latter. For these reasons, our objective is to propose a transmission protocol that enables a harmonious coexistence of massive IoT with existing UEs.

4.1 Related Work

The techniques toward the coexistence of IoT devices and UEs can be broadly classified into two categories: orthogonal-based and sharing-based solutions [RVX16, SGN16].

In the former, resource blocks are split among IoT devices and UEs either over time [LAA14] or frequency [WND17, FFG17] as means to avoid interference and congestion. However, resource partitioning inevitably leads to a spectral efficiency tradeoff between IoT devices and UEs. In contrast, in spectrum sharing, all resource blocks are shared between UEs and IoT devices. To control congestion, several techniques have emerged such as access class barring (ACB) and randomized back-off schemes [JDN18, WW15, MG16]. These methods, nevertheless, do not address the co-channel interference after access requests are granted. Alternative to these approaches, data aggregation has emerged as an effective solution to handle the massive IoT traffic. An experimental study is discussed in [DSG17] showing the benefits of IoT data aggregation on cellular networks. In [KC13, MDA16, TL17, GDZ17], stochastic geometry is used to analyze the coverage performance and/or the energy consumption using single and/or multiple aggregators. These works, however, consider fixed terrestrial aggregators and focus only on the performance of IoT devices, i.e., the coexistence of IoT devices and UEs is not considered. In this chapter, we consider using unmanned aerial vehicles (UAVs) or drones that stop at optimized predetermined locations.

Integrating UAVs with cellular networks has attracted significant attention, e.g., it has become a study item for recent and future 3GPP releases [3GP17c]. Indeed, UAVs are envisioned to become data aggregators (relays) or even base stations (BSs) [BY16], as their mobility brings unparalleled flexibility to cellular networks. For example, UAVs can help realize several IoT applications, e.g., smart cities, by extending the coverage of existing cellular infrastructure, enabling low-power communications with low-cost sensors, and reducing network congestion via offloading some of the traffic generated from massive IoT devices [MGA17]. Implementation and practical considerations for UAV-based IoT platforms are presented in [YJC17, MBT17, YJS18], showing the feasibility of using UAVs for IoT applications. Optimization and analysis of UAV-based aggregation are studied in [ZZL16, BCE18, ZZZ18, MSB17]. For example, the authors in [ZZL16] focus on optimizing the throughput of a single link between a source and a destination,

assisted by a relaying UAV, whereas in this chapter we consider a large-scale cellular network. In [BCE18], the authors focus on minimizing the time it takes the drone to aggregate data samples collected by IoT devices to estimate a field of interest. In [ZZZ18], the authors study optimizing the UAV's trajectory and sensors' wake-up schedules to minimize their energy consumption. In [MSB17], the locations and associations of the drones are optimized to minimize the transmit powers of IoT devices. Compared to the aforementioned works, we focus on optimizing the EE of IoT devices. In addition, we propose a shared spectrum access protocol between UEs and massive IoT, analyzing the impact of IoT interference on UEs.

4.2 Chapter Contributions and Organization

The first contribution in this chapter is the development of a time-division duplexing (TDD) transmission protocol; the protocol enables shared-spectrum access between massive IoT and UEs over the cellular network in the presence of UAVs that act as data aggregators for IoT devices. We use stochastic geometry [Hae12] to characterize the average available resources for UEs and IoT devices and compare the proposed protocol with a sharing-based protocol via ACB as well as resource splitting via frequency partitioning. We also analyze the coverage performance of the UEs in the presence of IoT devices transmitting to their associated drones.

The second contribution is the formulation of a stochastic optimization framework that aims to optimize the transmit power of the IoT device to maximize its average EE subject to a protection criterion to its nearest UEs. For tractable analysis, the framework is given for a single-cell single-drone (SC-SD) scenario, where we derive the average EE in closed form. We further analyze the interference-to-signal ratio (ISR) at the UE in the same cell with IoT devices and use its distribution as a protection constraint. Convexity analysis is presented and the solution of the framework is discussed. Further, we discuss extensions of this framework to the case of the single-cell multi-drone (SC-MD) scenario.

We validate the theoretical expression of the EE and the optimization framework for the SC-SD and SC-MD scenarios via Monte Carlo simulations, showing that the proposed scheme provides significant EE improvements to IoT devices compared to transmitting at the maximum power, which is typically done to extend coverage [3GP15a]. The proposed scheme is further compared to ACB and orthogonal allocation in a large network. Simulations show that the EE is significantly improved for practical drone altitudes, with minimal degradation to the UE's spectral efficiency in the UL and the DL. Such improvements are translated into an increased lifetime of IoT devices, which is validated using the 3GPP evaluation methodology in [3GP15b].

The rest of the chapter is organized as follows. The system model and the key performance metrics are presented in Section 4.3. A comparison of the proposed transmission protocol and existing ones is presented in Section 4.4. The EE SC-SD and SC-MD maximization frameworks are given in Section 4.5. Simulation results and the main conclusions are given in Section 4.6 and Section 4.7, respectively.

4.3 System Model

4.3.1 Network topology

We consider a stochastic network topology, where we model the locations of BSs using the homogeneous Poisson Point process (HPPP) Φ_B with density λ_B [GDZ17]. Each BS is equipped with a multi-antenna array of size M_B and can multiplex U_B users per resource block. During downlink (DL) data communications, the BS transmits at a fixed power of P_B , such that each multiplexed user is equally allocated a power of P_B/U_B . For UEs, we assume that they are generated from an independent HPPP Φ_U with density λ_U , where each one is equipped with a single-antenna system and connects to the nearest BS. During uplink (UL) data communications, the UE transmits at a fixed power of P_U .

Besides providing connectivity to UEs, the cellular network must also provide connectivity to a large number of IoT devices, which are assumed to operate mainly in the

UL [DSG17]. Furthermore, we assume that IoT devices are clustered either inherently, e.g., deploying sensors in hotspots to monitor the same physical phenomenon [GDZ17], or via a clustering algorithm, as done in [MSB17]. To this end, we model the locations of IoT devices using an independent clustered HPPP process. In particular and similar to [GDZ17], we consider the Matern cluster process, where the locations of cluster centers, i.e., parent points, is modeled by the independent HPPP Φ_{Cl} with density λ_{Cl} . In each cluster, IoT devices represent the daughter points of the clustered process, denoted by Φ_{M} , and they are uniformly distributed in a disk of radius R and with density λ_{M} . Thus, the average density of IoT devices in the network is $\lambda_{\text{M}}\lambda_{\text{Cl}}$. Finally, we assume all IoT devices are single-antenna transmitters, and they transmit at a fixed power of P_{M} . Applications where such IoT models are reasonable include the mass deployment of IoT devices across a city, where sensors can be anchored on bridges for infrastructure monitoring, on buildings for utility metering, or in a farm for water management [DSG17]. Such applications are delay tolerant, yet they require reliable coverage and a very long lifetime.

We consider deploying UAVs, e.g., drones, as a middle layer between IoT devices and the cellular infrastructure. The drone acts as a *mobile data aggregator* that is sent by a BS to provide coverage for a cluster of IoT devices. We note that while drones can be used as BSs or relays for UE traffic, as discussed in [BY16], in this work we primarily use them as aggregators for IoT devices that have delay-tolerant traffic. We assume the density of drones, λ_{D} , is equal to the density of clusters, i.e., $\lambda_{\text{D}} = \lambda_{\text{Cl}}$, and each one flies at an altitude of h_{D} . We note that fewer drones can be also used, e.g., a drone can serve multiple clusters by moving from one point to another over time. Further, the drone is equipped with an omni-directional single-antenna cellular transceiver. We discuss the case of a multi-tier drone network, where each tier is defined by a different altitude, in Section 4.5.4.

Since we focus on massive IoT applications with machines and sensors anchored to fixed locations, it is reasonable to assume that the locations of IoT devices in the clus-

ter are registered in a server or a database, and hence they are known to the mobile operator [MSB17, DSG17]. In particular, the drone, which is sent by the BS, moves to predetermined *stop points* for UL data aggregation. In this work and for tractable analysis, the stop point of the l -th drone, $(x_{D,l}, y_{D,l}, h_D)$, is the centroid of its cluster of IoT devices. We note that such location minimizes the distance to the typical IoT device, i.e., it can be shown that $(x_{D,l}, y_{D,l}) = \operatorname{argmin}_{(x,y)} \mathbb{E}_{\Phi_M}[d_{M,l}]$, where $d_{M,l}$ is the 3D distance between a typical IoT device and the drone. Finally, we note that the optimization of the drone's trajectory, e.g., to maximize its lifetime [ZZ17], is outside the scope of this dissertation.

4.3.2 Channel models

For ground-to-ground links, e.g., BS-UE links, we assume a power-law path loss model with a loss of L_0 at a reference distance of 1m and a decaying exponent of α_G . For small-scale fading, we assume a Rayleigh fading channel, with gamma distributed channel power gains, as they model a variety of multi-antenna transmission modes [AEA16]. Specifically, the channel power gains between the UE and the tagged BS and the UE and an interfering BS are, respectively, modeled as $g_B \sim \Gamma(\Delta_B, 1)$ and $f_B \sim \Gamma(\Psi_B, 1)$. For instance, if the BS uses multi-user zero-forcing transmit beamforming, then $\Delta_B = M_B - U_B + 1$ and $\Psi_B = U_B$ [AEA16].

For ground-to-air links, we consider the following popular path loss model for the link between an IoT device and the l -th drone [MSB17, MSB16, AKL14]

$$l_{M \rightarrow D}(d_{M,l}) = \mathbb{P}_{\text{LOS}}(d_{M,l}, h_D) L_0 d_{M,l}^{-\alpha_A} + (1 - \mathbb{P}_{\text{LOS}}(d_{M,l}, h_D)) L_{\text{NLOS}} L_0 d_{M,l}^{-\alpha_A}, \quad (4.1)$$

where $\mathbb{P}_{\text{LOS}}(d_{M,l}, h_D)$ is the line-of-sight (LOS) probability, L_{NLOS} is the excessive path loss due to non-LOS, and α_A is the ground-to-air path loss exponent. Note that, as investigated in [AKL14], the impact of multi-path fading is negligible in such links, and thus it is ignored. The LOS probability is found using the 3GPP UMa-AV channel model,

and it can be expressed as [3GP17c]

$$\mathbb{P}_{\text{LOS}}(d_{\text{M},l}, h_{\text{D}}) = \min \left\{ \frac{\xi_1}{r_{\text{M},l}}, 1 \right\} \left(1 - e^{-r_{\text{M},l}/\xi_2} \right) + e^{-r_{\text{M},l}/\xi_2}, \quad (4.2)$$

where $r_{\text{M},l} = \sqrt{d_{\text{M},l}^2 - h_{\text{D}}^2}$, and ξ_1 and ξ_2 are some constants as given in [3GP17c, Table B-1]. For the link between the BS and the drone, we consider a similar model, which is expressed as

$$l_{\text{B} \rightarrow \text{D}}(d_{\text{B},l}) = L_{\text{STR}} \cdot \left(\mathbb{P}_{\text{LOS}}(d_{\text{B},l}, h_{\text{D}}) L_0 d_{\text{B},l}^{-\alpha_{\text{A}}} + (1 - \mathbb{P}_{\text{LOS}}(d_{\text{B},l}, h_{\text{D}})) L_{\text{NLOS}} L_0 d_{\text{B},l}^{-\alpha_{\text{A}}} \right), \quad (4.3)$$

where the attenuation L_{STR} is due to the fact that the BS's antenna array steering direction either points horizontally or is tilted downwards [3GP17a]. Such attenuation depends on the drone's relative location with the array's boresight. However, we assume it to be constant for tractable analysis, which is reasonable when the drone is flying at heights higher than the BS's antenna height.

The drone-assisted cellular network is shown in Fig. 4.1a, where a drone is sent by a BS to a cluster of IoT devices, stops at $(x_{\text{D},l}, y_{\text{D},l})$ to collect data from IoT devices, and then aggregates the traffic to the BS. An illustration of one realization of the network is also shown in Fig. 4.1b. Finally, we remark that we focus on a single-tier network for easier exposition in the subsequent analysis. However, it is straightforward to extend this work to multi-tier heterogeneous networks as we have done in our prior work in [HC17].

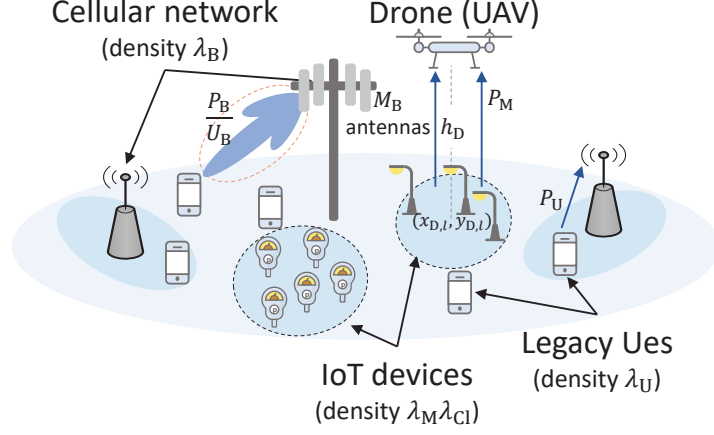
4.3.3 Performance metrics

We consider two metrics: the spectral efficiency (SE) of a typical UE in the UL and DL and the energy efficiency (EE) of a typical IoT device, which are defined next.

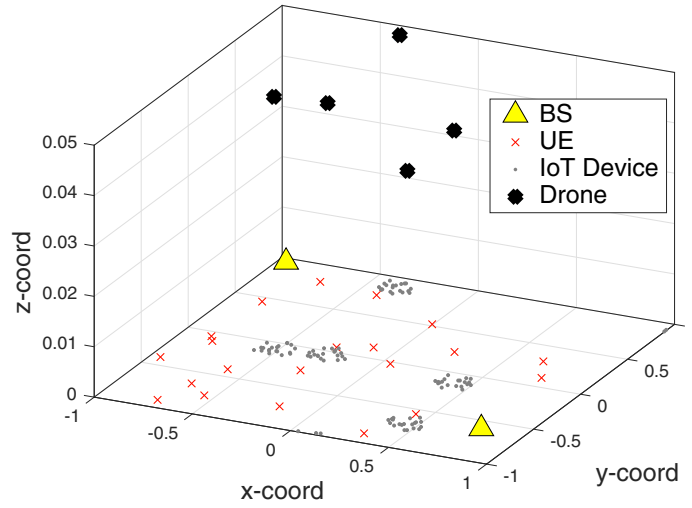
We consider the UE to employ a multi-modulation and coding scheme, and thus the spectral efficiency of a typical UE in the DL/UL is defined as [BL15]

$$C_{\text{U},\xi} = \bar{\beta}_{\text{U},\xi} \sum_{k=1}^{K_{\text{U}}} \log_2(1 + \tau_{\text{U},k}) \mathbf{1}(\tau_{\text{U},k+1} \geq \gamma_{\text{U},\xi} \geq \tau_{\text{U},k}), \quad (4.4)$$

where $\xi \in \{\text{DL}, \text{UL}\}$, $\gamma_{\text{U},\xi}$ is the signal-to-interference-plus-noise ratio (SINR), $\{\tau_{\text{U},k}\}$ are the SINR thresholds, $\mathbf{1}(\cdot)$ is the indicator function, and $\bar{\beta}_{\text{U},\xi}$ is a pre-log term that denotes



(a) An illustrative example of the network



(b) A spatial realization of the network ($\lambda_U = 5\lambda_B$, $\lambda_D = \lambda_B$, and $\lambda_M = 20/\text{cluster}$)

Figure 4.1: The drone-assisted cellular architecture and one spatial realization of the network topology.

the UE long-term available resources in time, frequency, and space. This rate model assumes K_U possible thresholds, and it is a generalization of the single-rate model, i.e., $K_U = 1$, that is commonly used in the literature [LBY15, AEA16]. For tractable analysis, we use the mean load approximation, i.e., we assume that the load $\bar{\beta}_{U,\xi}$ is independent

of $\gamma_{U,\xi}$ [BL15, LBY15]. We relax this assumption when we run Monte Carlo simulations.

Since the majority of IoT devices are battery-powered, the EE of the transmission protocol is critical, which is defined as the ratio of the achieved rate to the total power consumption. More formally, the EE of a typical IoT device is defined as [ZSB16]

$$E_M = \frac{\bar{\beta}_M \sum_{k=1}^{K_M} \log_2(1 + \tau_{M,k}) \mathbf{1}(\tau_{M,k+1} \geq \gamma_M \geq \tau_{M,k})}{P_{CP} + \eta^{-1} P_M}, \quad (4.5)$$

where P_{CP} is a constant that quantifies the circuit power consumption, η is the power amplifier efficiency, and $\bar{\beta}_M$ and γ_M are the IoT device long-term resources and SINR, respectively. We have the following remarks about this metric. First, observe that if the device supports only a single modulation scheme, then, we can set $K_M = 1$. Second, the metric is also relevant to applications where coverage is paramount. For example, the value of $\tau_{M,1}$ determines the minimum coverage needed since a zero rate, and hence zero EE, is achieved if $\gamma_M < \tau_{M,1}$. Last, we only focus on optimizing the link between the IoT device and the drone. The link between the drone and the BS can be optimized separately, e.g., the drone can get closer to the BS to ensure reliable aggregation [MWZ18], the drone can compress data generated from devices performing a similar task [BPZ17], or the drone can itself be a BS [BY16]. Since this link is studied in the aforementioned works, it is outside the scope of this dissertation.

4.4 Proposed Transmission Protocol

In this section, we present the proposed transmission protocol to enable shared spectrum access between massive IoT and UEs over drone-assisted cellular networks. We then analyze the average allocated resources of the UE and the IoT device under the proposed protocol and compare it to those achieved under existing transmission protocols.

We focus on TDD cellular networks, and thus the proposed protocol is divided into two slots: T_1 and T_2 . In the first time slot, the UE operates in the UL, communicating with its tagged BS. Similarly, in this slot, we treat the drone as another UE, which aggregates previously collected data from IoT devices and sends it to its tagged BS. In the

second time slot, the UE operates in the DL, whereas the IoT device operates in the UL, communicating with its associated drone, as shown in Fig. 4.2a. The proposed protocol is motivated as follows.

- **Reducing congestion:** When the UE operates in the UL, it competes for scheduling with drones instead of IoT devices, and thus the channel congestion is significantly reduced compared to the sharing-based TDD shown in Fig. 4.2b.
- **Maximum bandwidth:** The protocol allows the IoT device to share the same time-frequency block with the UE, i.e., no resource splitting is used as done in the orthogonal-based TDD shown in Fig. 4.2c.
- **Limiting the impact of IoT interference:** The shared access paradigm inevitably leads to additional interference from IoT devices into UEs. However, when IoT devices transmit data in the UL, UEs operate in the DL, where their tagged BSs transmit at much higher power than the transmit power of IoT devices as $P_B \gg P_M$.

4.4.1 Characterization of average resources in the proposed and existing protocols

For tractable analysis, we assume that a proportional fair scheduler is used, e.g., round-robin, and thus under the mean load approximation [LBY15], the average load is inversely proportional to the average number of devices connected to the same source.

4.4.1.1 Proposed protocol

Let $\bar{\beta}_{U,UL}^P$ denote the average allocated resources to a typical UE operating in the UL under the proposed protocol. Then, it can be shown that

$$\begin{aligned} \bar{\beta}_{U,UL}^P &\stackrel{(a)}{=} \underbrace{W}_{\text{Freq.}} \times \underbrace{U_B}_{\text{Space}} \times \underbrace{T_1 \left(\frac{1}{N_{B,UL}^P} \right)}_{\text{Time}} \\ &\stackrel{(b)}{=} W \times U_B \times T_1 \left(\frac{\lambda_B}{\lambda_U + \lambda_D} \right), \end{aligned} \tag{4.6}$$

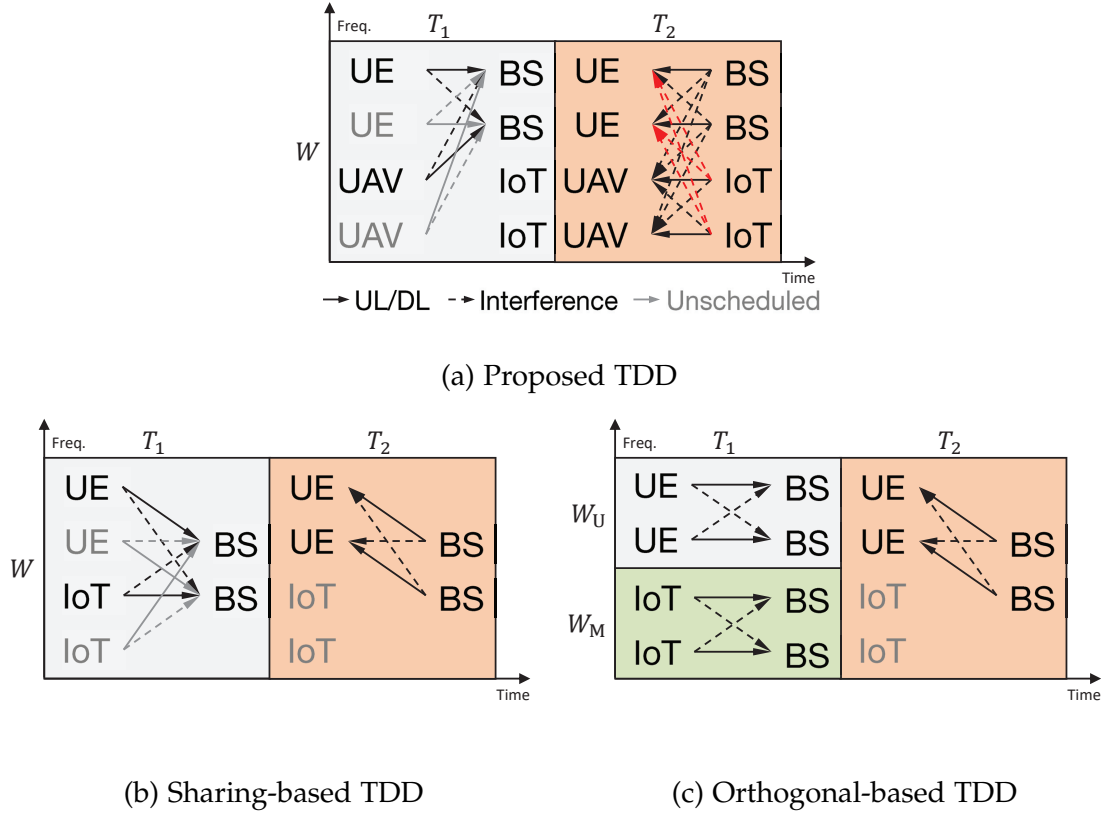


Figure 4.2: Comparison of the different TDD transmission protocols.

where (a) follows since the entire bandwidth W is allocated to the UE, U_B UEs can be spatially multiplexed by the same BS, and the portion of the time allocated to the UE is inversely proportional to the average number of devices connected to the BS, i.e., $N_{B,UL}^P$. Here, (b) follows by showing that the average number of UEs (or drones) connected to the BS under the nearest BS association is λ_U/λ_B (or λ_D/λ_B) [LBY15]. Similarly, the average portion of resources of the UE in the DL is expressed as $\bar{\beta}_{U,DL}^P = W \times U_B \times T_2 \left(\frac{\lambda_B}{\lambda_U} \right)$, which follows since in the second time slot, no IoT devices are connected to the BS under the proposed protocol. For the IoT device, it can be shown that the average portion of resources is given as $\bar{\beta}_M^P = W \times T_2 \lambda_M^{-1}$, which follows since the average number of IoT devices per cluster is λ_M , and the drone multiplexes one IoT device per time-frequency slot. Note that for low-rate applications, the drone can divide the bandwidth W into smaller frequency blocks and serve multiple IoT devices, one per frequency block. This

does not change $\bar{\beta}_M^P$ since the decrease in frequency resources is equally compensated by increased time resources.

4.4.1.2 Comparison with sharing-based protocol

In the sharing-based protocol, the IoT device is registered as another UE, as shown in Fig. 4.2b. To control channel congestion, the 3GPP standard proposes one mechanism, namely *access class barring* (ACB), which prioritizes UE traffic over IoT devices [SGN16]. More formally, each BS broadcasts a parameter $0 \leq \kappa \leq 1$ to all IoT devices in vicinity. The IoT device then generates a random number $n \in [0, 1]$ before initiating a channel access request. If $n > \kappa$, then the IoT device does not request access. Clearly, for $\kappa = 1$, the protocol simplifies to a standard sharing protocol that is agnostic to the device type. Let $\bar{\beta}_{U,UL}^S$ denote the average allocated resources to a typical UE operating in the UL under the sharing-based protocol. Then, we have $\bar{\beta}_{U,UL}^S = W \times U_B \times T_1 \left(\frac{\lambda_B}{\lambda_U + \kappa \lambda_M \lambda_{Cl}} \right)$, which follows since the average number of IoT devices per BS is $\lambda_M \lambda_{Cl} / \lambda_B$, yet only a fraction, κ , of them request access. Clearly, for $\bar{\beta}_{U,UL}^S > \bar{\beta}_{U,UL}^P$, we must have $\kappa < \lambda_M^{-1}$, yet this degrades the average resources allocated to the IoT device. Indeed, the average allocated resource of a typical IoT device under the sharing protocol is given as

$$\begin{aligned} \bar{\beta}_M^S &= \mathbb{P}(n > \kappa) \bar{\beta}_{M|n>\kappa}^S + \mathbb{P}(n \leq \kappa) \bar{\beta}_{M|n \leq \kappa}^S \\ &= W \times U_B \times \kappa T_1 \left(\frac{\lambda_B}{\lambda_U + \kappa \lambda_M \lambda_{Cl}} \right) \\ &\stackrel{(a)}{\leq} \bar{\beta}_M^P \left(\frac{U_B \lambda_B}{\lambda_U + \lambda_{Cl}} \right), \end{aligned} \tag{4.7}$$

where (a) follows using $\kappa < \lambda_M^{-1}$ and assuming $T_1 = T_2$. Since the density of UEs is typically higher than the density of BSs, i.e., $\lambda_U \gg U_B \lambda_B$, we get $\bar{\beta}_M^S < \bar{\beta}_M^P$. Finally, for the second time slot, we have $\bar{\beta}_{U,DL}^S = \bar{\beta}_{U,DL}^P$ for the UE.

4.4.1.3 Comparison with orthogonal-based protocol

Alternative to using ACB to resolve channel congestion, 3GPP has also proposed the separation of frequency resources [SGN16], as shown in Fig. 4.2c. Let $W_U \in [0, W]$ be the portion of the spectrum allocated for UEs, then the average portion of allocated resources for a typical UE in the UL under this protocol is $\bar{\beta}_{U,UL}^O = W_U \times U_B \times T_1 \left(\frac{\lambda_B}{\lambda_U} \right)$. If $W_U/W > \lambda_U/(\lambda_U + \lambda_D)$, then we have $\bar{\beta}_{U,UL}^O > \bar{\beta}_{U,UL}^P$. However, this comes at the expense of reducing the resources allocated to the IoT device since we have

$$\begin{aligned} \bar{\beta}_M^O &= W_M \times U_B \times T_1 \left(\frac{\lambda_B}{\lambda_M \lambda_{Cl}} \right) \\ &\stackrel{(a)}{<} W \times U_B \times \frac{T_1}{\lambda_M} \left(\frac{\lambda_B}{\lambda_U + \lambda_{Cl}} \right) \\ &\stackrel{(b)}{<} \bar{\beta}_M^P, \end{aligned} \tag{4.8}$$

where (a) follows from the fact that $W_M = W - W_U$ and thus $\frac{W_U}{W} > \frac{\lambda_U}{\lambda_U + \lambda_{Cl}} \iff \frac{W_M}{W} < \frac{\lambda_{Cl}}{\lambda_U + \lambda_{Cl}}$. In addition, (b) follows assuming $T_1 = T_2$ and the density of UEs is high. Finally and similar to the sharing-based protocol, we have $\bar{\beta}_{U,DL}^O = \bar{\beta}_{U,DL}^P$ for the UE.

To summarize, for the sharing protocol to outperform the proposed one in terms of resource allocation for the UE, we must use aggressive ACB with very small values of κ , which inevitably affects the IoT EE as the average portion of allocated resources is decreased. For the orthogonal allocation to outperform the proposed protocol in terms of the UE performance, nearly all resources should be allocated to the UE, i.e., $W_U \rightarrow W$, since $\lambda_U \gg \lambda_D$, and this also limits the resources allocated to the IoT device.

4.4.2 Analysis of IoT interference into UEs

While the proposed protocol improves the average allocated resources of a typical UE, in comparison with sharing-based and orthogonal-based protocols, the signal-to-interference ratio (SIR) of a typical UE degrades due to the presence of an additional interference term generated from IoT devices (cf. Fig. 4.2a). Specifically, the SIR of a typical UE can be

written as

$$\tilde{\gamma}_{\text{U,DL}}^{\text{P}} = \frac{\frac{P_{\text{B}}}{U_{\text{B}}} g_{\text{B}} L_0 x_{\text{B}}^{-\alpha_{\text{G}}}}{\sum_{y_b \in \Phi'_{\text{B}}} \frac{P_{\text{B}}}{U_{\text{B}}} f_b L_0 y_b^{-\alpha_{\text{G}}} + \sum_{z_m \in \Phi'_{\text{M}}} P_{\text{M}} f_m L_0 z_m^{-\alpha_{\text{G}}}}, \quad (4.9)$$

where Φ'_{B} is the set of interfering BSs, Φ'_{M} is the set of interfering IoT devices, $f_m \sim \Gamma(1, 1)$, x_{B} is the distance to the tagged BS, y_b is the distance to the b -th interfering BS, and z_m is the distance to the m -th interfering IoT device. Next, we derive the coverage probability of the UE under the proposed protocol to highlight the key parameters that affect the UE's coverage.

The coverage probability is defined as $\mathbb{C}_{\text{U,DL}}^{\text{P}}(\tau) \triangleq \mathbb{P}(\tilde{\gamma}_{\text{U,DL}}^{\text{P}} \geq \tau)$, which can be rewritten as

$$\mathbb{C}_{\text{U,DL}}^{\text{P}}(\tau) \stackrel{(a)}{=} 2\pi\lambda_{\text{B}} \int_0^\infty x \mathbb{E}_{g_{\text{B}}} \left[F_{I_{\text{U}}} \left(\frac{g_{\text{B}}}{\tau x^{\alpha_{\text{G}}}} \right) \right] e^{-\pi\lambda_{\text{B}}x^2} dx, \quad (4.10)$$

where (a) follows from the distribution of the distance from the UE to the tagged BS [LBY15] and $F_{I_{\text{U}}}(\cdot)$ is the cumulative distribution function (CDF) of the interference. Using the Gil-Pelaez Inversion theorem [Gil51] to compute the interference CDF, we get the following theorem.

Theorem 6. *The coverage probability of the UE under the proposed protocol is expressed as*

$$\mathbb{C}_{\text{U,DL}}^{\text{P}}(\tau) = \frac{1}{2} - \frac{\Upsilon(\tau, \lambda_{\text{B}}, \lambda_{\text{D}}, \hat{P}_{\text{B,M}})}{\pi}, \quad (4.11)$$

where $\hat{P}_{\text{B,M}} = \text{sinc}^{-1}(\delta_{\text{G}})P_{\text{M}}U_{\text{B}}/P_{\text{B}}$, $\delta_{\text{G}} = 2/\alpha_{\text{G}}$, and

$$\Upsilon(\tau, \lambda_{\text{B}}, \lambda_{\text{D}}, \hat{P}_{\text{B,M}}) = \int_0^\infty \frac{1}{t} \text{Im} \left\{ \frac{(1+jt/\tau)^{-\Delta_{\text{B}}}}{{}_2F_1(\delta_{\text{G}}, \Psi_{\text{B}}; 1-\delta_{\text{G}}; jt) + \frac{\lambda_{\text{D}}}{\lambda_{\text{B}}} \hat{P}_{\text{B,M}}^{\delta_{\text{G}}} (-jt)^{\delta_{\text{G}}}} \right\} dt. \quad (4.12)$$

Proof: See Appendix C.1. ■

The expression is given in a single integral form that can be efficiently evaluated using numerical software. The key insight here is that the degradation of the UE coverage due to the presence of IoT devices depends mainly on two factors: (i) the ratio of the transmit power of the IoT device to the power allocated to the UE, i.e., $P_{\text{M}}/(P_{\text{B}}/U_{\text{B}})$ and (ii) the average number of drones per BS, i.e., $\lambda_{\text{D}}/\lambda_{\text{B}}$. To see this, note that the coverage

probability in the absence of IoT devices is $\mathbb{C}_{\text{U,DL}}^{\text{N}}(\tau) = \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \frac{1}{t} \text{Im} \left\{ \frac{(1+jt/\tau)^{-\Delta_{\text{B}}}}{{}_2F_1(\delta_{\text{G}}, \Psi_{\text{B}}; 1-\delta_{\text{G}}; jt)} \right\} dt$, and thus using stochastic dominance, i.e., $\mathbb{C}_{\text{U,DL}}^{\text{N}}(\tau) \geq \mathbb{C}_{\text{U,DL}}^{\text{P}}(\tau)$, we have

$$\int_0^\infty \frac{\text{Im} \left\{ \frac{(1+jt/\tau)^{-\Delta_{\text{B}}}}{{}_2F_1(\delta_{\text{G}}, \Psi_{\text{B}}; 1-\delta_{\text{G}}; jt)} \right\}}{t} dt \leq \Upsilon(\tau, \lambda_{\text{B}}, \lambda_{\text{D}}, \hat{P}_{\text{B,M}}), \quad (4.13)$$

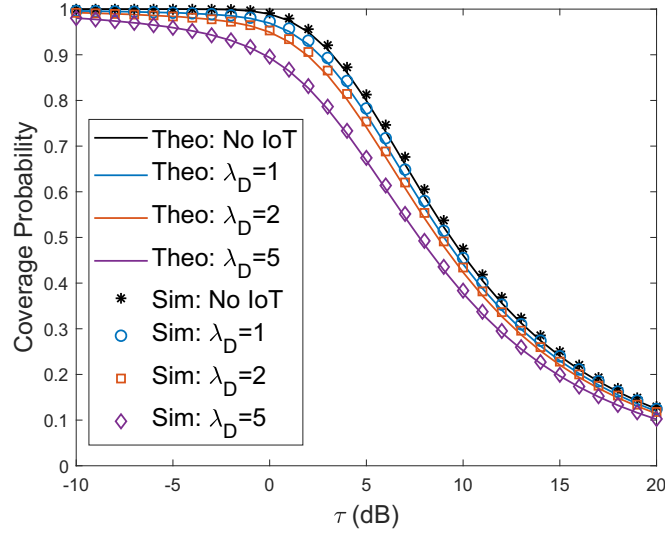
where the gap above decreases as $\lambda_{\text{D}}/\lambda_{\text{B}} \rightarrow 0$ and/or $P_{\text{M}}/(P_{\text{B}}/U_{\text{B}}) \rightarrow 0$.

We validate the theoretical expression with Monte Carlo simulations, using the same parameters in Table 4.1. Fig. 4.3a shows the distribution of coverage in the presence and absence of IoT devices. It is observed that increasing the number of drones decreases the coverage, yet the degradation is not severe, e.g., the median SIR merely degrades by 1.7dB under the proposed protocol with $\lambda_{\text{D}}/\lambda_{\text{B}} = 5$. Recall that here a drone is assigned to a single cluster. However, if a single drone is assigned instead to serve multiple clusters, moving from one stop point to another, then the interference on UEs from IoT transmission decreases, e.g., median SIR degradation is less than 0.36dB when $\lambda_{\text{D}}/\lambda_{\text{B}} = 1$. Nevertheless, the cost of reducing this interference is the increased delay on IoT devices, leading eventually to degradation of the IoT energy efficiency when $\lambda_{\text{M}}\lambda_{\text{Cl}} \gg 1$. To this end, another approach to limit the interference is to reduce the IoT transmit power, as shown in Fig. 4.3b. In the next section, we focus on optimizing the IoT transmit power such that the IoT EE is maximized and the IoT interference is controlled.

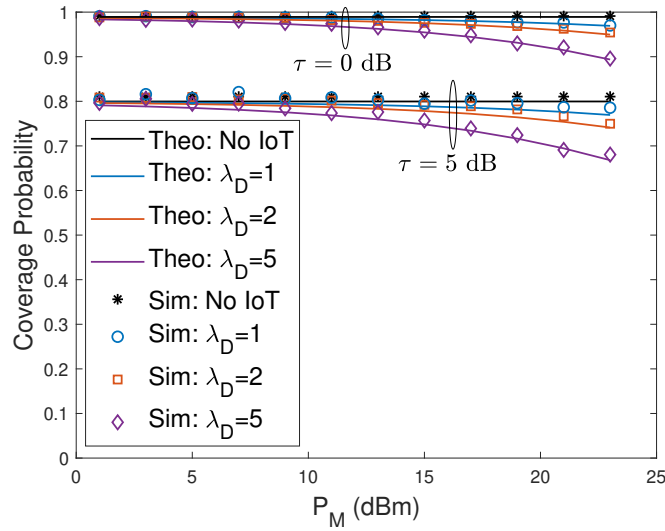
4.5 IoT Energy-Efficiency Maximization

In this section, we present a stochastic optimization framework that maximizes the average EE of an IoT device under the proposed protocol. The framework takes the following form

$$\begin{aligned} & \underset{P_{\text{M}}}{\text{maximize}} && \mathbb{E}[E_{\text{M}}] \\ & \text{subject to} && f(I_{\text{U}}) \leq \epsilon, \\ & && P_{\text{M}} \in \mathcal{P}, \end{aligned} \quad (4.14)$$



(a) Variations of SIR threshold



(b) Variations of P_M

Figure 4.3: The UE DL coverage performance with and without IoT devices.

where the $f(\cdot)$ is an interference constraint to protect UEs and $\mathcal{P}$ is the feasible set of transmit powers. We have the following remarks about the problem in (4.14). First, it is a stochastic optimization framework, as the objective function is the mean of a random variable, and the expectation is taken with respect to spatial realizations. Second, such formulation aims to optimize a single variable P_M , significantly reducing the complexity

of implementation, i.e., the drone merely solves the problem off-line and then broadcasts the optimal transmit power, over a control channel, to all of its IoT devices. Third, the constraint is a function of the interference from IoT devices into UEs. We do not enforce a UE rate constraint because this would incur additional overhead due to the necessary coordination between UEs and UAVs.

There are two challenges to solve (4.14). The first one is that the objective function is intractable due to the ground-to-air channel model from interfering BSs and IoT devices into drones. The second one is that the UE coverage probability in (4.11) is given in an integral form, making it not amenable to use as a UE protection criterion. For these reasons, we focus on optimizing the EE in a single-cell single-drone (SC-SD) setting, i.e., we ignore the interference from BSs and IoT devices outside the cell of the typical IoT device. Such framework is shown to be tractable, which helps glean useful insights. We will further evaluate the solution of the SC-SD framework in the large network case, i.e., many drones and many cells, via Monte Carlo simulations.

4.5.1 Average IoT EE in the SC-SD case

Let $\hat{E}_M$ and $\hat{\gamma}_M$ denote the IoT EE and IoT UL SINR in the SC-SD case, respectively, and let $p = P_M$ for notational simplicity. Then, the average EE under the proposed protocol, i.e., $\bar{E}_M(p) \triangleq \mathbb{E}[\hat{E}_M]$, can be written as

$$\begin{aligned}\bar{E}_M(p) &= \frac{\bar{\beta}_M^P \sum_{k=1}^{K_M} \log_2(1 + \tau_{M,k}) \mathbb{E}[\mathbf{1}(\tau_{M,k+1} \geq \hat{\gamma}_M(p) \geq \tau_{M,k})]}{P_{CP} + \eta^{-1}p} \\ &= \frac{\bar{\beta}_M^P \sum_{k=1}^{K_M} \mu_k \mathbb{P}(\hat{\gamma}_M(p) \geq \tau_{M,k})}{P_{CP} + \eta^{-1}p},\end{aligned}\tag{4.15}$$

where $\mu_1 = \log_2(1 + \tau_{M,1})$ and $\mu_k = \log_2(1 + \tau_{M,k+1}) - \log_2(1 + \tau_{M,k})$ for $k > 1$.

In a single cell, the drone receives signals from IoT devices in its cluster and receives interference from the tagged BS, which operates in the DL. The next proposition presents the distributions of the received desired signal and interference powers at the drone, which will be useful to evaluate $\mathbb{P}(\hat{\gamma}_M(p) \geq \tau_{M,k})$.

Proposition 1. *The distribution of the desired signal power, $S_M = P_M l_{M \rightarrow D}(d_{M,l})$, at a typical drone is expressed as*

$$\mathbb{P}(S_M \leq \tau) = 1 - \frac{(P_M L_M / \tau)^{\delta_A} - h_D^2}{R^2}, \quad (4.16)$$

where $\delta_A = 2/\alpha_A$, $L_M = L_0((1 - L_{\text{NLOS}})\bar{\mathbb{P}}_{\text{LOS},M} + L_{\text{NLOS}})$, $\bar{\mathbb{P}}_{\text{LOS},M}$ is the average of (4.2), which is computed numerically, and $\tau \in [P_M L_M (R^2 + h_D^2)^{-1/\delta_A}, P_M L_M h_D^{-\alpha_A}]$. In addition, the distribution of the interference signal power $I_B = P_B l_{B \rightarrow D}(d_{B,l})$ is given as

$$\mathbb{P}(I_B \leq \tau) = \exp(\pi \lambda_B h_D^2) \cdot \exp\left(-\pi \lambda_B (P_B L_B / \tau)^{\delta_A}\right), \quad (4.17)$$

where $L_B = L_0((1 - L_{\text{NLOS}} L_{\text{STR}})\bar{\mathbb{P}}_{\text{LOS},B} + L_{\text{NLOS}} L_{\text{STR}})$ and $\tau \in [0, P_B L_B h_D^{-\alpha_A}]$.

Proof: See Appendix C.2. ■

Remark: It is observed from (4.16) that as h_D increases, the variations in S_M reduces since $\tau \in [P_M L_M (R^2 + h_D^2)^{-1/\delta_A}, P_M L_M h_D^{-\alpha_A}]$, i.e., the distribution of S_M becomes more concentrated around the median. This follows because the drone location is optimized to reduce the average 2D distance to a typical IoT device, increasing the LOS probability particularly when $h_D \geq R$. Similarly, it is observed from (4.17) that increasing the density of BSs increases the interference as the average distance between the drone and its tagged BS decreases. Furthermore, it can be shown that increasing the drone's height decreases I_B , but not as rapidly as the decrease in S_M .

Using Proposition 1, we derive the coverage probability and the average EE of the IoT device.

Theorem 7. *The coverage probability of an IoT device in the SC-SD case is given as*

$$\mathbb{P}(\hat{\gamma}_M \geq \tau) \approx e^{\pi \lambda_B h_D^2} \cdot \exp\left(-\pi \lambda_B \left(\frac{P_M \tilde{L}_M}{P_B L_B} - P_N\right)^{-\delta_A}\right), \quad (4.18)$$

where $\tilde{L}_M = L_M (R^2 + h_D^2)^{-1/\delta_A}$ and P_N is the noise power. In addition, the average EE of an IoT device is expressed as

$$\bar{E}_M(p) = \frac{\bar{\beta}_M^P e^{\pi \lambda_B h_D^2} \sum_{k=1}^{K_M} \mu_k e^{-\frac{\pi \lambda_B}{(P_B L_B)^{-\delta_A}} \left(\frac{p \tilde{L}_M}{\tau_k} - P_N\right)^{-\delta_A}}}{P_{\text{CP}} + \eta^{-1} p}. \quad (4.19)$$

Proof: The coverage probability can be written as

$$\begin{aligned}\mathbb{P}(\hat{\gamma}_M \geq \tau) &\triangleq \mathbb{P}\left(\frac{S_M}{I_B + P_N} \geq \tau\right) \\ &\approx \mathbb{P}\left(I_B \leq \frac{\tilde{S}_M}{\tau} - P_N\right),\end{aligned}\tag{4.20}$$

where we have used the median received signal power $\tilde{S}_M$ computed from (4.16) to approximate the coverage (recall that S_M has small variations particularly when $h_D \geq R$). Thus, using (4.17), we arrive at (4.18). ■

4.5.2 IoT interference constraint

In this section, we propose to use the distribution of the interference-to-signal ratio (ISR) as a protection criterion. We note that in [WDZ16], the mean of ISR (MISR) has been used as a metric to compare different interference mitigation techniques over cellular networks with stochastic topologies. Yet, using the ISR distribution provides more flexibility compared to the MISR, e.g., the protection criterion can be designed to limit the median, the 95th percentile, etc. The ISR is defined as $\text{ISR}_U = \frac{\mathbb{E}_M[\hat{I}_{U,M}]}{\mathbb{E}_{g_B}[S_B]}$, where $\hat{I}_{U,M}$ is the interference from the typical IoT device to its nearest UE and S_B is the received desired signal power at the UE from its tagged BS. We note, similar to the MISR metric, the expectation is first taken with respect to channel realizations. The distribution is thus given as

$$\begin{aligned}\mathbb{P}(\text{ISR}_U \geq \rho) &\stackrel{(a)}{=} \mathbb{P}\left(\frac{P_M z_M^{-\alpha_G}}{\frac{\Delta_B P_B}{U_B} z_B^{-\alpha_G}} \geq \rho\right) \\ &= \mathbb{P}\left(\frac{z_B}{z_M} \geq \left(\rho \frac{\Delta_B P_B}{U_B P_M}\right)^{1/\alpha_G}\right) \\ &\stackrel{(b)}{=} \mathbb{E}_{z_M} \left[\exp \left(-\pi \lambda_B z_M^2 \left(\rho \frac{\Delta_B P_B}{U_B P_M} \right)^{\delta_G} \right) \right] \\ &\stackrel{(c)}{=} \left[1 + \frac{\lambda_B}{\lambda_U} \left(\rho \frac{\Delta_B P_B}{U_B P_M} \right)^{\delta_G} \right]^{-1},\end{aligned}\tag{4.21}$$

where (a) follows from the mean of desired and interference power channel gains, (b) follows using the complementary CDF of the distance from the UE to its tagged BS, and

(c) follows by taking the expectation with respect to the distance between the IoT device and the nearest UE.

4.5.3 SC-SD stochastic optimization framework

Using (4.19) as an objective function and the ISR expression in (4.21) as an interference constraint, i.e., $f(I_U) \leq \epsilon \Leftrightarrow \mathbb{P}(\text{ISR}_U \geq \rho) \leq \epsilon$, we get the following optimization problem

$$\begin{aligned} & \underset{P_M}{\text{maximize}} && \frac{\sum_{k=1}^{K_M} \mu_k \exp\left(-\frac{\pi \lambda_B}{(P_B L_B)^{-\delta_A}} \left(\frac{P_M L_M}{\tau_k} - P_N\right)^{-\delta_A}\right)}{P_{CP} + \eta^{-1} P_M} \\ & \text{subject to} && P_M \leq \rho \left(\frac{\Delta_B P_B}{U_B}\right) \left(\frac{\lambda_B}{\lambda_U} \cdot \frac{\epsilon}{1-\epsilon}\right)^{1/\delta_G}, \\ & && P_M^{\min} \leq P_M \leq P_M^{\max}, \end{aligned} \quad (4.22)$$

where $P_M^{\min}$ and $P_M^{\max}$ are the minimum and maximum allowable transmit powers, respectively. In what follows, we denote the numerator of the objective function, i.e., the rate, by $r(P_M)$, and the denominator, i.e., power consumption, by $c(P_M)$. The following proposition shows that the problem in (4.22) is quasiconcave, and thus a local maximum is a global one [BV04].

Proposition 2. *The objective function in (4.22) is quasiconcave and unimodal, whereas the constraints are all affine. Hence, the optimization problem is quasiconcave.*

Proof: The proof follows by showing that $r(p)$ is S-shaped in p (see Appendix C.3). ■

Since the problem has a single optimizing variable, a line search is sufficient to solve the problem. However, in what follows we discuss a specific algorithm to solve the problem, as its generalization will be used for the SC-MD case in Section 4.5.4. To solve the optimization problem, we note that if the following condition is satisfied

$$\tau_k \leq \frac{P_M^{\min} \tilde{L}_M}{P_N + P_B L_B \left(\frac{\delta_A \pi \lambda_B}{1 + \delta_A}\right)^{1/\delta_A}} \quad \forall k, \quad (4.23)$$

then the optimization problem can be transformed from a quasiconcave problem into a concave one using the Charnes-Cooper Transform [BV04]. Thus, a standard convex

Algorithm 5 Dinkelbach's algorithm to solve the SC-SD optimization

```

1: procedure ( $\zeta > 0; i = 0; v_i = 0$ )
2:   while  $H(v_i) > \zeta$  do
3:     Compute  $\Gamma = \text{ROOT}(r'(p) - \eta^{-1}v_i)$ 
4:      $p_i^* = \min \left\{ P_M^{\max}, \rho \left( \frac{\Delta_B P_B}{U_B} \right) \left( \frac{\lambda_B}{\lambda_U} \cdot \frac{\epsilon}{1-\epsilon} \right)^{1/\delta_G}, \Gamma \right\}$ 
5:      $H(v_i) = r(p_i^*) - v_i c(p_i^*)$  and  $v_{i+1} = \frac{r(p_i^*)}{c(p_i^*)}$ 
6:      $i = i + 1$ 
7:   end while
8:   return  $P_M^* = p_i^*$ 
9: end procedure

```

optimization solver can be used to solve the problem. When the condition in (4.23) is not satisfied, i.e., for general thresholds, then the problem can be efficiently solved using the Dinkelbach's algorithm [CF91], which converges super-linearly to the global optimal solution. The algorithm that specifically solves (4.22) is summarized in Alg. 5. In essence, the algorithm seeks to find the root of the auxiliary function $H(v) = \max_{p \in \mathcal{P}} \{r(p) - vc(p)\}$, which is achieved at the point that maximizes the optimization problem [CF91]. Finally, a closed-form solution can be derived for the special case $K_M = 1$ and $\alpha_A = 2$. Specifically, the objective function, in this case, is maximized at

$$\begin{aligned}
p_{\text{unconst}}^* &= \frac{\tau_{M,1}(P_N + \pi \lambda_B P_B L_B)}{\tilde{L}_M} \\
&\quad + \frac{\sqrt{\pi \tau_{M,1} \lambda_B P_B L_B (P_N \tau_{M,1} + \tilde{L}_M \eta P_{\text{CP}})}}{\tilde{L}_M},
\end{aligned} \tag{4.24}$$

and thus the optimal transmit power is $P_M^* = \min \{ P_M^{\max}, \rho \left(\frac{\Delta_B P_B}{U_B} \right) \left(\frac{\lambda_B}{\lambda_U} \cdot \frac{\epsilon}{1-\epsilon} \right)^{1/\delta_G}, p_{\text{unconst}}^* \}$. It is observed from (4.24) that the optimal power increases for: (i) higher target threshold $\tau_{M,1}$ to meet the new coverage requirement, (ii) higher BS transmit power P_B or higher BS density λ_B to combat the increased interference from the cellular network, or (iii) higher PA efficiency η to utilize the decrease in power consumption.

4.5.4 EE optimization in heterogeneous single-cell multi-drone case

In this section, we consider the single-cell multi-drone case (SC-MD), where each cell has multiple drones, each serving a cluster of IoT devices. We can further assume each drone belongs to a different tier, i.e., we consider an N -tier drone network such that the l -th tier drone flies at an altitude of $h_{D,l}$ and serves an IoT cluster of radius R_l . The IoT device in a cluster served by the l -th drone transmits at power $P_{M,l}$.

The challenge in the multi-drone is that the objective function can no longer be given in a closed-form expression due to the complicated ground-to-air path loss model between multiple interferers and the drone. Here, the interferers, with respect to a given drone, are the tagged BS and the other IoT devices transmitting to their respective drones in the same cell. To this end, we propose to model the different interference sources in the cell as one Poissonian source with a transmit power equal to the sum of transmit powers of all interferers. Such an approach ensures a tractable formulation and will be validated in the simulations section. Under this modeling assumption, the EE of an IoT device served by the l -th drone tier can be written as

$$\bar{E}_l(\mathbf{P}_M) \triangleq \frac{\bar{\beta}_M^P e^{\pi\lambda_B h_D^2} \sum_{k=1}^{K_M} \mu_k e^{-\frac{\pi\lambda_B \left(\frac{P_{M,l} \bar{L}_{M,l}}{\tau_k} - P_N \right)^{-\delta_A}}}{P_{CP} + \eta^{-1} P_{M,l}}, \quad (4.25)$$

where $\mathbf{P}_M$ is the vector of transmit powers. Next, we discuss two common formulations for the EE optimization in the SC-MD case.

4.5.4.1 Max-min formulation

In this formulation, the objective is to maximize the minimum EE of a typical IoT device. Such formulation is a conservative one that aims to improve the worst case scenario. In

particular, the max-min EE maximization is given as

$$\begin{aligned}
& \underset{\mathbf{P}_M}{\text{maximize}} \quad \min_l \quad \bar{E}_l(\mathbf{P}_M) \\
& \text{subject to} \quad \sum_l P_{M,l} \leq \rho \left(\frac{\Delta_B P_B}{U_B} \right) \left(\frac{\lambda_B}{\lambda_U} \cdot \frac{\epsilon}{1-\epsilon} \right)^{1/\delta_G}, \\
& \quad P_M^{\min} \leq P_{M,l} \leq P_M^{\max} \quad \forall l.
\end{aligned} \tag{4.26}$$

The EE expression in (4.26) is still quasiconcave, and since the objective function is the minimum of quasiconcave functions, it remains quasiconcave. Thus, the max-min framework is quasiconcave, and can be solved using a Dinkelbach's-like algorithm.

4.5.4.2 Sum-EE formulation

In this formulation, the total EE of IoT devices in the cell is used as an objective function, and thus such formulation tends to favor IoT devices with better coverage. More formally, we have

$$\underset{\mathbf{P}_M}{\text{maximize}} \quad \sum_{l=1}^N \bar{E}_l(\mathbf{P}_M), \tag{4.27}$$

subject to the same constraints as in (4.26).

The generalized Dinkelbach's algorithm for both SC-MD formulations is given in Alg. 6, where the implementation of Sum-EE is given in brackets. We note for max-min, the algorithm converges linearly to the global optimal solution [CF91]. However, for the total EE formulation, the objective function is not necessarily quasiconcave, as the sum of quasiconcave functions does not preserve quasiconcavity. To this end, solving this problem globally, with limited complexity, remains an open problem, i.e., although Alg. 6 has become popular to solve sum of ratios, it does not necessarily arrive at the global optimal solution, if it exists [SS03].

4.6 Simulation Results

In this section, we first validate the theoretical analysis and the stochastic optimization frameworks, focusing on SC-SD and SC-MD scenarios. Then, we consider larger net-

Algorithm 6 Generalized Dinkelbach's algorithm to solve the SC-MD optimization

```
1: procedure ( $\zeta > 0; i = 0; v_i = 0$ )
2:   while  $H(v) > \zeta$  do
3:      $\mathbf{p}_{M,i}^\star = \underset{\mathbf{p} \in \hat{\mathcal{P}}}{\operatorname{argmax}} \left\{ \min_l \{r_l(\mathbf{p}) - v_i c_l(\mathbf{p})\} \right\}$             $\left[ \mathbf{p}_{M,i}^\star = \underset{\mathbf{p} \in \hat{\mathcal{P}}}{\operatorname{argmax}} \left\{ \sum_l r_l(\mathbf{p}) - v_{i,l} c_l(\mathbf{p}) \right\} \right]$ 
4:      $H(v_i) = \min_l \{r_l(\mathbf{p}_{M,i}^\star) - v_i c_l(\mathbf{p}_{M,i}^\star)\}$             $\left[ H(\{v_{i,l}\}) = \sum_l r_l(\mathbf{p}_{M,i}^\star) - v_{i,l} c_l(\mathbf{p}_{M,i}^\star) \right]$ 
5:      $v_{i+1} = \min_l \frac{r_l(\mathbf{p}_{M,i}^\star)}{c_l(\mathbf{p}_{M,i}^\star)}$             $\left[ v_{i+1,l} = \frac{r_l(\mathbf{p}_{M,i}^\star)}{c_l(\mathbf{p}_{M,i}^\star)} \right]$ 
6:      $i = i + 1$ 
7:   end while
8:   return  $\mathbf{P}_M^\star = \mathbf{p}_{M,i}^\star$ 
9: end procedure
```

works, where we compare the proposed protocol with sharing-based and orthogonal-based protocols.

Unless otherwise stated, we use the simulation parameters given in Table 4.1. In each spatial realization of the network, we generate BSs, UEs, and IoT devices according to their distributions. Each drone then moves to their predetermined stop points. We then generate the channel power gains according to their distributions and compute the large-scale fading for all links to compute their SINR, where a thermal noise power, with spectral density -174dBm/Hz , is considered at all receivers. Then the SE of UEs is computed via (4.4) and the EE IoT devices is computed via (4.5) for the different transmission protocols. We note that we use the actual load and LOS probability, instead of their mean, to compute the performance metrics.

4.6.1 Validation of the stochastic optimization frameworks

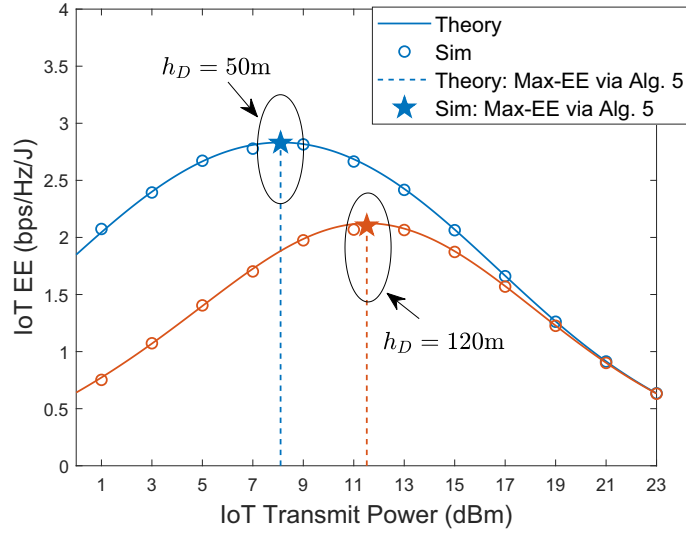
In this section, we focus on a single cell to validate the theoretical analysis. We further compare the EE performance of the SC-SD and SC-MD optimization frameworks with a scheme where the IoT device transmits at maximum power, i.e., $P_M = 23\text{dBm}$ [3GP15a].

Table 4.1: Main parameters

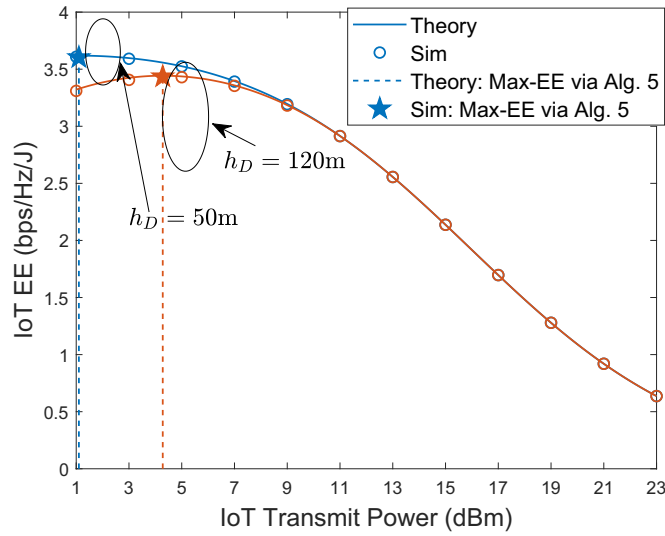
Description	Parameters
Path loss	$\alpha_G = 3.5; \alpha_A = 2.2$ [3GP17c]; $L_0 = -38\text{dB}$; $L_{\text{NLOS}} = -20\text{dB}$ [MSB16]; $L_{\text{STR}} = -30\text{dB}$ [3GP17c]
BS parameters	$M_B = 32; U_B = 4$; transmission mode: multi-user zero-forcing
UE parameters	$P_U = 23\text{dBm}$; $\{\tau_{U,k}\}$ from -5 to 30dB
IoT device parameters	$P_M^{\max} = 23\text{dBm}$; $P_M^{\min} = 1\text{dBm}$; $\{\tau_{M,k}\}$ vary from -5 to 10dB ; $R = 50\text{m}$
Power consumption	$P_{\text{CP}} = 90\text{mW}$, $\eta = 0.44$ [3GP15b]

4.6.1.1 Impact of IoT transmit power on the EE

We consider two IoT categories: CAT-0 and NB-IoT. In the former, the IoT device shares the entire band with the UE, i.e., $W = 20\text{MHz}$ and $P_B = 46\text{dBm}$. In the latter, the IoT only shares a single resource block, i.e., $W = 180\text{KHz}$ and $P_B = 32\text{dBm}$. For interference protection, we assume $\epsilon = 0.5$ and $\rho_{\text{dB}} = -6\text{dB}$, i.e., the median ISR should not exceed -6dB . The performance of EE under different transmit powers is shown in Fig. 4.4a and Fig. 4.4b for CAT-0 and NB-IoT, respectively. It is shown that the theoretical expression in (4.19) matches well with Monte Carlo simulations. Further, the EE is significantly improved using the proposed SC-SD framework compared to the max-power scheme, e.g., the EE in CAT-0 of the proposed SC-SD framework is 4.5x and 3.3x that of max-power for $h_D = 50\text{m}$ and $h_D = 120\text{m}$, respectively. This follows because the drone's location is optimized to minimize the 2D distance to the IoT device, requiring low transmit power for reliable coverage. Third, increasing the drone's height has two effects: an increase in the optimal transmit power and a decrease in the EE. These follow because as h_D increases, the received signal power decreases more rapidly than interference, degrading the coverage. Thus, the IoT device must transmit at higher power to combat the degradation, decreasing its EE. Finally, the NB-IoT operation is more efficient than CAT-0 since



(a) EE: CAT-0



(b) EE: NB-IoT

Figure 4.4: Validation of the theoretical EE.

in the former the IoT device shares a smaller number of carriers with the UE, reducing the interference. In Fig. 4.5, we validate the coverage approximation in (4.18), which is shown to be in good agreement with Monte Carlo simulations. It is also evident here that using drones at lower altitudes is critical to provide high coverage.

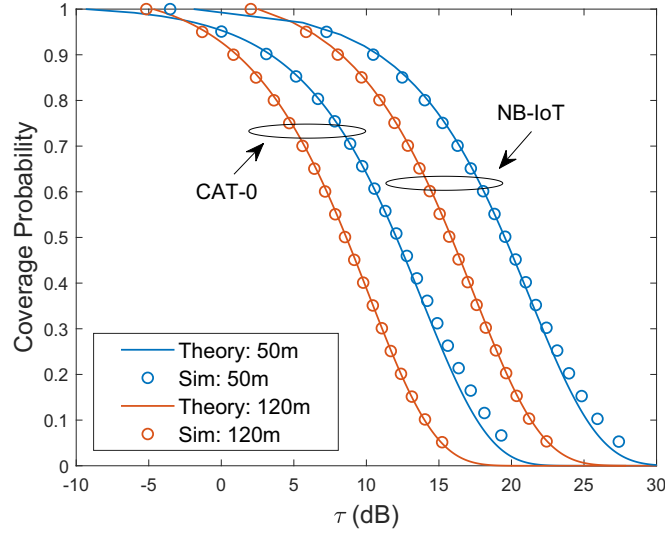
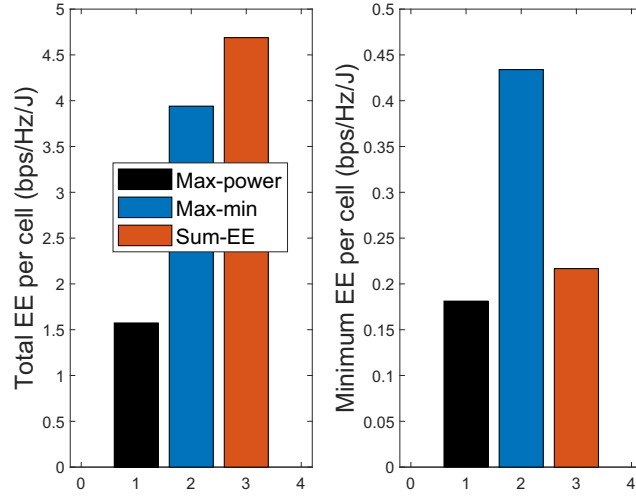


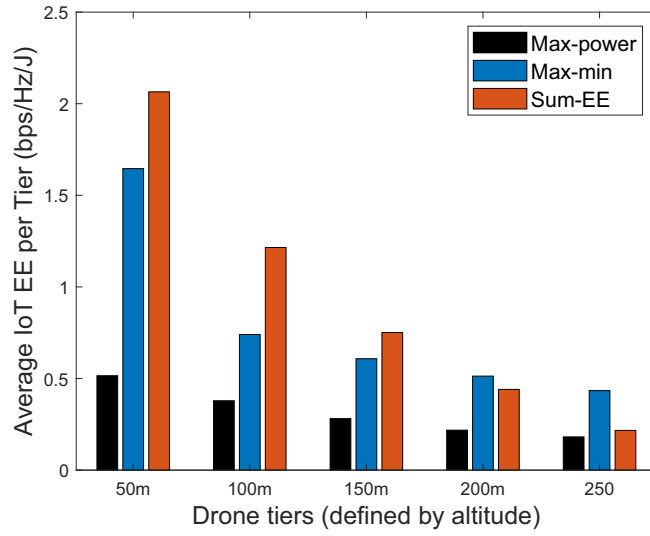
Figure 4.5: Validation of the theoretical coverage using the optimal transmit power.

4.6.1.2 Validation of the SC-MD case

Next, we study the EE performance in the presence of multiple drones in the same cell. Drones fly at one of the following altitudes: $\mathbf{h}_D = [50, 100, 150, 200, 250]$ m. In Fig. 4.6a, we show the total and the minimum EE of scheduled IoT devices per cell. We compare the max-power scheme to *Max-min* and *Sum-EE* schemes, where power allocation is done using Alg. 6. We note that the EE is computed via Monte Carlo simulations. As expected, the *Sum-EE* formulation achieves the highest total EE in a cell, yet the *Max-min* formulation provides tangible improvements to the worst case scenario, e.g., the minimum EE is 2.4x and 2.0x the minimum EE under *Max-power* and *Sum-EE*, respectively. We further show the EE of IoT devices that belong to the different drones in Fig. 4.6b. It is observed that the *Max-min* solution aims to improve the EE of devices connected to the drone with the highest altitude as this tier achieves the lowest EE. In contrast, the *Sum-EE* formulation favors the drones with lower altitudes as they provide higher EE for IoT devices.



(a) Sum and Minimum EE



(b) Average EE per drone tier (defined by altitude)

Figure 4.6: EE performance for different schemes in the SC-MD case (CAT-0 parameters and $\rho_{\text{dB}} = -12\text{dB}$).

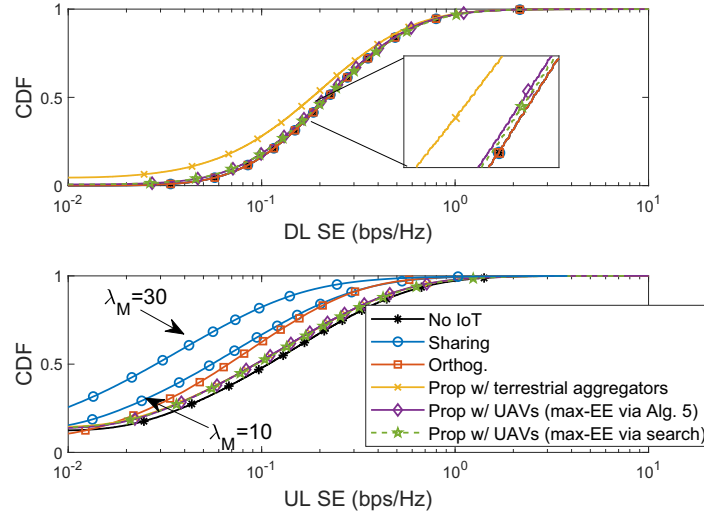
4.6.2 Performance comparison with existing protocols

In this section, we compare the performance of the proposed protocol with existing ones in large networks, and hence the results are obtained via Monte Carlo simulations. We

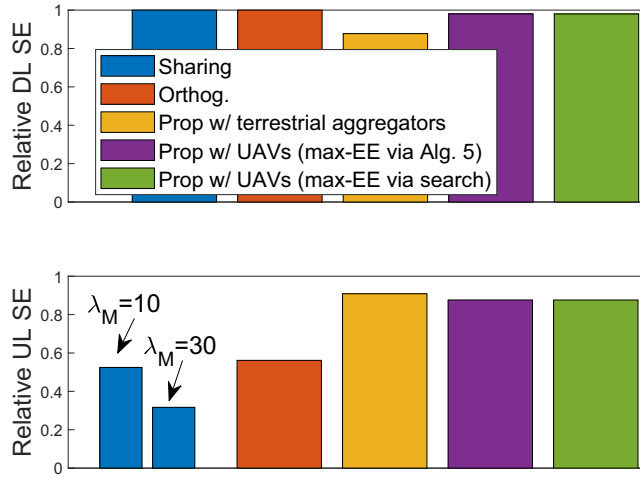
also show the performance of UEs in the absence of IoT devices as a benchmark. For the proposed protocol, we consider two schemes: *Max-EE* using the SC-SD framework (Alg. 5), and *Max-EE* using exhaustive search, where we use extensive network simulations to search for the optimal power. To highlight the benefits of using UAVs, we also consider the proposed protocol with max-power and terrestrial aggregators, e.g., small cells are deployed at the centroid of the clusters instead of UAVs. We assume $\lambda_U = 50\lambda_B$ and $\lambda_{Cl} = \lambda_D = 5\lambda_B$. Unless otherwise stated, we consider $\delta = 1$, $W_U = 0.5W$, and $h_D = 50\text{m}$. We show the results assuming CAT-0 IoT devices (similar trends are observed for NB-IoT and hence omitted).

4.6.2.1 UE performance comparison

We first study the performance of the UE in the DL and the UL under different protocols. We note that both the orthogonal-based and the proposed protocols decouple the UE UL performance from the IoT density as in the former IoT devices use different frequency blocks, and in the latter IoT devices connect to drones instead of BSs. This is not the case for the sharing-based protocol, where we study the UE's performance for $\lambda_M = 10$ and $\lambda_M = 30$. Fig. 4.7a shows the distribution of the SE performance across all UEs and Fig. 4.7b shows the mean SE relative to the benchmark, i.e., the ratio of the mean SE under the given protocol to the mean SE in the absence of IoT devices. Since our model considers IoT devices to only operate in the UL, the DL SE of sharing-based and orthogonal-based protocols is the same as the benchmark. For the proposed protocol, it is shown that the DL degradation is minimal when UAVs are used and the transmit power is optimized to limit the IoT interference. More importantly, the proposed protocols outperform existing ones in the UL, e.g., the relative mean UL SE of the proposed protocols is 1.5x and 2.7x the orthogonal-based and sharing-based ($\lambda_M = 30$) protocols, respectively. This follows due to the resource splitting in the orthogonal-based protocol and the high congestion in the sharing-based one.



(a) Distribution of the SE



(b) Mean SE relative to the benchmark

Figure 4.7: UE DL and UL performance for the different protocols.

4.6.2.2 IoT performance comparison

We then study the EE of the IoT device under the different protocols. In Fig. 4.8a, the EE is shown for different densities of IoT devices. As expected, as the number of IoT devices increases, the EE decreases under all protocols. Yet, the proposed protocol achieves the

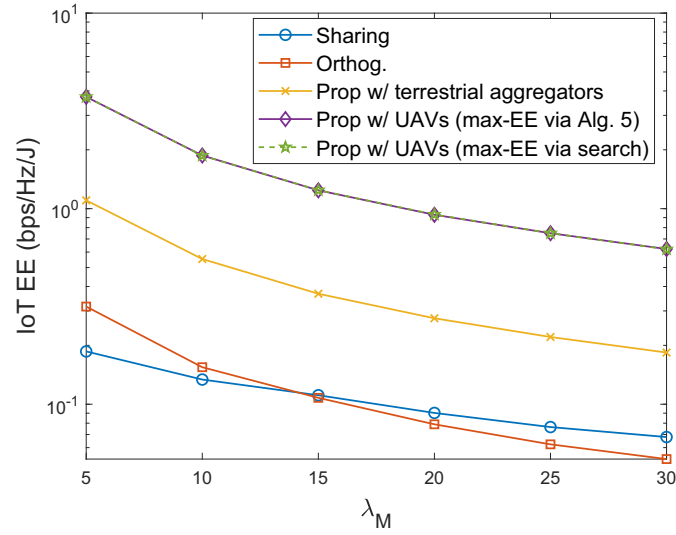
highest EE and scales better with λ_M compared to existing ones. It is also observed that it is beneficial to use UAVs over terrestrial aggregators as the former provides higher LOS, i.e., the EE improves by 3x when aerial aggregators are used instead of terrestrial ones. In Fig. 4.8b, we study the EE for different drone's altitudes. We show the performance under the existing protocols and the proposed one with terrestrial aggregators, which do not depend on h_D , for reference. It is observed that the proposed protocol is beneficial for lower altitudes. At very high altitudes, the received signal power is lower, and the LOS probability with interfering BSs, from different cells, increases. We note that in many regions, e.g., North America, Europe, China, etc., the maximum legal altitude for drones is roughly 120m (400ft), and thus the proposed solution is superior in practical scenarios.

4.6.2.3 Impact of ACB and resource splitting

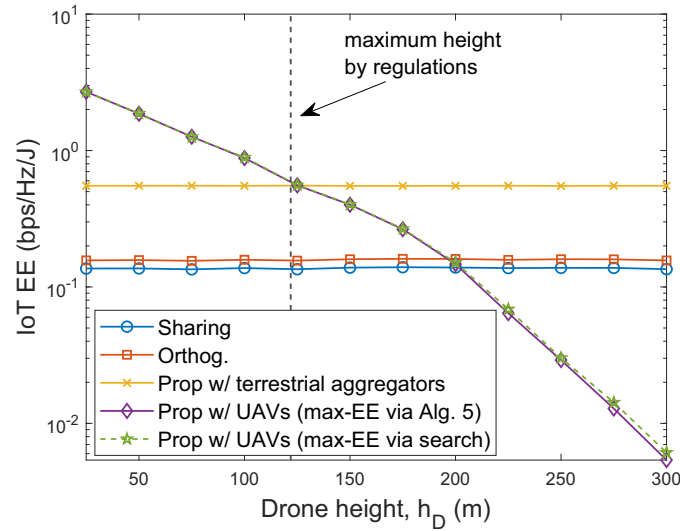
Fig. 4.9 shows the IoT EE versus the UE SE in the UL under different ACB thresholds κ (Fig. 4.9a) and different frequency allocation ratios W_U (Fig. 4.9b). The performance of the proposed protocol, which does not depend on these parameters, is also shown for reference. It can be seen that existing protocols have operating points with higher UE SE performance in comparison with proposed ones. However, high ACB threshold and frequency partitioning ratio are needed, and thus the IoT device will be limited with time and frequency resources, respectively. Finally, it is shown the SC-SD framework leads to a nearly identical performance to exhaustive search, yet the latter requires extensive simulations to solve.

4.6.3 IoT device lifetime comparison

In this section, we show that the EE improvements are translated into extended battery lifetime. To this end, we follow the 3GPP evaluation methodology to compute the lifetime of IoT devices for the different schemes [3GP15b, HC18d]. In particular, the IoT device sends N_{rep} reports per day to the network. For each report, the device operates



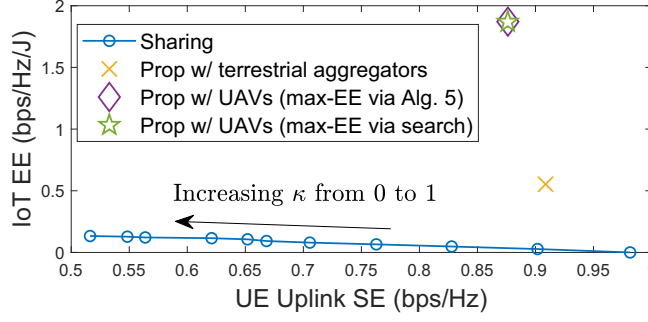
(a) Variations of IoT density ($h_D = 50\text{m}$)



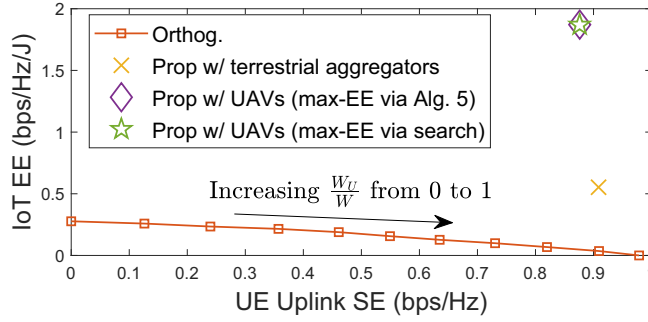
(b) Variations of drone's height ($\lambda_M = 10$)

Figure 4.8: IoT EE performance under different protocols.

in the following stages: standby, idle, transmission, and reception. Let P_S , P_I , and P_{RX} denote the power consumption of the standby, idle, and reception stages. Similarly, let T_S , T_I , and T_{RX} be their durations. For a fair comparison, we assume the energies consumed for these three stages are the same across the schemes, i.e., the schemes only differ in the energy consumed during the transmission stage as our focus is on the UL.



(a) Variations of ACB threshold



(b) Variations of frequency partitioning ratio

Figure 4.9: IoT energy-efficiency vs. UE spectral efficiency ($\lambda_M = 10$ and $h_D = 50\text{m}$).

The transmission duration is

$$T_{\text{TX}}^{(\nu)} = \frac{B/\beta_M^{(\nu)}}{\sum_{k=1}^{K_M} \log_2(1 + \tau_{M,k}) \mathbf{1}(\tau_{M,k+1} \geq \gamma_M^{(\nu)} \geq \tau_{M,k})}, \quad (4.28)$$

where B is the total size of the data transmitted per report, which includes the connection request, the data packet, and the acknowledgment [3GP15b, Table 4]. The superscript ν denotes the scheme used. Similarly, $P_{\text{TX}}^{(\nu)} = P_{\text{CP}} + \eta^{-1} P_M^{(\nu)}$. Thus, the total energy consumed per day is given as [3GP15b]

$$E_{\text{IoT}}^{(\nu)} = N_{\text{rep}} \left(T_{\text{TX}}^{(\nu)} P_{\text{TX}}^{(\nu)} + T_{\text{RX}} P_{\text{RX}} + T_{\text{I}} P_{\text{I}} \right) + T_{\text{S}} P_{\text{S}}. \quad (4.29)$$

Let the IoT battery's capacity be C_{IoT} Wh, then the device lifetime in years is given as $Y^{(\nu)} = \frac{C_{\text{IoT}}}{E_{\text{IoT}}^{(\nu)}} \times \frac{3600}{365}$. Note that increasing the IoT transmit power is expected to improve the SINR, increasing the rate and decreasing $T_{\text{TX}}^{(\nu)}$. Yet, this comes at the expense of increased power consumption during the transmission stage, i.e., higher $P_{\text{TX}}^{(\nu)}$.

Table 4.2: Battery lifetime parameters [3GP15a, 3GP15b]

Description	Parameters
IoT device	NB-IoT with $W = 180\text{KHz}$, $C_{\text{IoT}} = 5\text{Wh}$ $B = 229\text{bytes}$ [3GP15b, Table 4], and $N_{\text{rep}} = 12$
Powers	$P_{\text{RX}} = 90\text{mW}$, $P_{\text{I}} = 3\text{mW}$, and $P_{\text{S}} = 0.015\text{mW}$ [3GP15b, Table 1]
Durations	$T_{\text{RX}} = 565\text{ms}$, $T_{\text{I}} = 22451\text{ms}$, and $T_{\text{S}} = 86400\text{s}$ [3GP15b, Table 6]

Using the parameters given in Table 4.2 [3GP15b], Fig. 4.10 shows the CDF of the IoT device lifetime for each scheme. We have two observations. First, the improvements in EE under the proposed scheme are translated into tangible enhancements to the IoT device lifetime, e.g., the median lifetime is increased by more than four and six years compared to sharing and orthogonal-based schemes, respectively. Second, due to the aggregators' proximity to IoT devices, the variance in lifetime across devices is lower under the proposed schemes. When UAVs are used instead of terrestrial aggregators, the variance is further reduced thanks to the better LOS conditions, where it is shown that almost all devices have a lifetime of +16 years.

4.7 Conclusions

In this chapter, we have proposed a TDD protocol for a shared spectrum access between massive IoT and cellular UEs with UAVs acting as data aggregators. Using stochastic geometry, it is shown that the protocol improves the average allocated resources of IoT devices and UEs compared to resource splitting and ACB, yet UEs experience additional interference from IoT devices. Thus, we have optimized the transmit power of the IoT device to maximize its energy-efficiency while constraining the interference on UEs. The analysis of the stochastic optimization framework is validated via simulations for

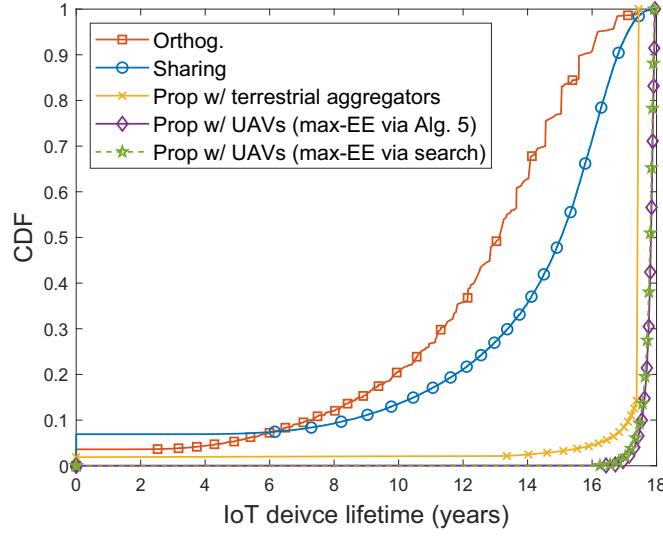


Figure 4.10: Distribution of IoT device lifetime ($h_D = 50\text{m}$).

the single-cell case. For larger networks, Monte Carlo simulations have shown that the proposed system significantly improves the EE, and hence the IoT device lifetime, with minimal degradation on the UE spectral-efficiency.

The key insights gleaned from this chapter are as follows. First, it is beneficial for drones to fly at lower altitudes as higher altitudes increase the LOS with interferers, forcing the IoT device to increase its transmit power to combat coverage degradation, which eventually affects its EE. In case of drones, belonging to the same BS, fly at different altitudes, then maximizing the minimum EE and the total EE amount to prioritizing devices connected to drones at the highest and the lowest altitudes, respectively. Second, the extended coverage mode, where IoT devices transmit at maximum power, is not necessarily energy efficient as the gain in coverage does not outweigh the loss in power consumption.

CHAPTER 5

Coverage and Rate Maximization over Heterogeneous Cellular Networks

The use of UAVs, or drones, as aerial cells and data aggregators can revolutionize wireless communication systems, adding new dimensions to the services that can be provided by technology vendors and MNOs. Nevertheless, there remain practical challenges that transcend wireless connectivity with aerial devices, including the limited flight time and the control of swarm drones. Until these challenges are addressed, the use of small cells or terrestrial aggregators remains an economical choice that is widely followed by cellular networks.

The dense deployment of small cells has reshaped cellular networks; they are no longer a set of macro cells, but rather they comprise different types of BSs, of varying capabilities and coverage areas. The heterogeneous network structure, also known as (HetNet), has been perceived as one of the major steps to meet two key requirements for fifth-generation (5G) networks: (i) the support of sensors and machines that comprise a wide spectrum of IoT applications [NLV15, SRI15, 3GP15a], and (ii) the 1000x capacity enhancement [ABC14]. Specifically, machine-type traffic requires reliable coverage, e.g., sensory data for utility metering, whereas human-type traffic typically requires high-throughput links, e.g., video streaming. In this case, network densification helps in coverage enhancements by reducing the average distance between sensors and BSs, whereas it helps in capacity improvements by aggressively reusing the spectrum.

To attain the envisioned throughput and coverage enhancements, there are two key network bottlenecks that must be overcome, namely user association for load-balanced

networking and interference management due to the high channel reuse. For instance, choosing the received signal strength as a user association metric, which typically improves coverage, inevitably overloads macro BSs, leaving many small cells underutilized and reducing the long-term network throughput. Moreover, even if user association is optimized for load balancing, the co-channel deployment of BSs of vastly different transmit powers and coverage areas can introduce intolerable interference [LGR11]. For instance, UEs that connect to low-power small-cell BSs may experience high interference from active macro BSs in the downlink, limiting the link reliability. For these two reasons, extensive research efforts pertaining to HetNets have centered around the load balancing and the inter-tier interference problems. In this chapter, we aim to develop user-centric association and spectrum partitioning schemes for multi-antenna HetNets to maximize coverage and rate.

5.1 Related Work

The user association problem, i.e., finding the best links between every user and every BS in the network to optimize a specific objective, can be formulated as a combinatorial problem [YRC13, BBP16, HYM17]. However, the complexity of solving this problem in ultra-dense networks has motivated other approaches such as developing distributed algorithms, trading complexity with additional UE-BS message exchange [HYM17], or using load-aware access methods, where BSs frequently broadcast traffic load information [TSH14, STH16]. Alternatively, BS selection can be user-centric, where users merely measure the received signal power of BSs' reference signals and make decisions independent of other users [LWC16]. The majority of user-centric algorithms are based on *cell range expansion*, where UEs weigh the received power measurements with predetermined offsets during BS selection [ASY14]. Using stochastic geometry, a tool that has become central to model and analyze cellular networks [ESA17], biasing has been shown to improve rate, coverage, and delay outage in single-antenna HetNets [JSX12, SDA13, ZQG17] and multi-antenna HetNets [HC18e, GDV14, HWE15, HC16a].

All aforementioned works consider spectrum sharing, i.e., all BSs use the same spectrum, and hence biasing provides rate gains by balancing the load across the network. Nevertheless, the coverage performance can be poor due to the high interference resulted from co-channel deployment. Hence, *frequency partitioning* has become a popular approach to tackle the inter-tier interference, where the macro tier is allocated a set of carriers that are orthogonal to those allocated for small-cell BSs [LGR11, GYY16]. In this case, the key question becomes how much of the spectrum each tier should get. Such question has been addressed jointly with user association in [DT16], yet no optimization framework is given. Optimization frameworks of joint user association and spectrum allocation have been presented in [SA14, BL15, LBY15] to maximize rate-based functions. However, these works are limited to single-antenna HetNets although BSs have become universally endowed with multi-antenna transceivers. Hence, the joint optimization framework must consider the multi-antenna transmission mode used since it characterizes both the *array gain* resulted from the directionality of multi-antenna systems, and the *multiplexing gain* resulted from serving multiple UEs in a single time-frequency block. In addition, the aforementioned prior work has focused on optimizing rate-based performance metrics, ignoring coverage. These metrics are reasonable when spectrum sharing is used since coverage can be maximized by connecting to the BS that provides the maximum received power, i.e., using *max-power* association [SBA14]. However, in orthogonal allocation, this may not be the case. For instance, although connecting to the macro tier provides better signal reception, it entails significantly higher interference in comparison with connecting to a small-cell BS. This calls for revisiting the optimality of max-power association for coverage maximization, which is critical for several IoT applications [DSG17].

5.2 Chapter Contributions and Organization

In this Chapter, we consider a downlink cellular network with multi-antenna BSs. Using a stochastic geometry approach, we aim to optimize two different key performance

indicators: coverage and rate. The former is critical for machine-type traffic, where high coverage is desired for low signal-to-interference-ratio (SIR) target thresholds, and the latter is important for human-type traffic.

Since both performance indicators depend on the *coverage probability*¹, the first contribution of this chapter is to derive the exact analytical coverage performance of a typical user in multi-antenna HetNets. Different from [LZA16], we consider a frequency-partitioned network instead of spectrum sharing as means to mitigate the inter-tier interference. Although the expression is given in a simple form, it still requires numerical evaluation, making it difficult to explicitly understand the impact of different network settings on the coverage performance and renders it impractical for optimization. Hence, we use the Gil-Pelaez Inversion theorem to approximate the coverage probability in closed form. The derived coverage approximation generalizes the analysis of single-tier networks in [RG14] to multi-tier networks with different multi-antenna transmission modes.

The second contribution of this chapter is the maximization of coverage by optimizing the tier association probabilities. Under the proposed coverage probability approximation, the coverage maximization problem is shown to be concave. Furthermore, under special network settings, we derive the optimal user association in closed form. The structure of the optimal solution motivates us to propose a suboptimal user association strategy, for any network setting, that can be implemented by weighting the received powers by a bias factor that is given in closed form, or by reconfiguring the transmit power of small cells. Two interesting observations are gleaned from the coverage maximization framework. First, user association should exploit the multi-antenna transmission scheme since it is shown that the user should connect to the tier with the highest ratio of the array gain to the multiplexing gain. Practically, this means that sensors and machines should be served on time-frequency blocks that do not spatially

¹In this chapter, the rate depends on the coverage probability since a zero-rate is achieved when the UE is in outage.

multiplex other users in space. Second, under the proposed user association strategy, the *SIR invariance* property holds, i.e., the area spectral efficiency linearly increases with the density of BSs, without degrading coverage. This is a very desired property, that has been shown to only hold for single-antenna HetNets [JSX12] and not for multi-antenna HetNets [DKA13, AEA16].

The third contribution of this chapter is the maximization of a proportional-fair rate-based utility function [BBP16], that encourages fairness, by jointly optimizing user association and the spectrum allocated to each tier. The framework is shown to be jointly concave under the coverage probability approximation. In addition, we show that the optimal spectrum allocation factor is equal to the portion of users connected to the tier, where the latter can be optimized, via user association, using a standard convex optimization solver. Interestingly, the proportional-fair rate-based utility maximization framework, henceforth denoted as *rate maximization*, can be interpreted as a scalarized multi-objective optimization [BJD14], with two objectives: maximizing coverage and maximizing the average number of UEs served per resource block. Such formulation motivates a suboptimal approach, where the association probabilities are obtained in closed form. In particular, we show that the user association, in essence, tries to match the users' distribution across tiers to the *supply* of each tier, which is given in terms of a rate-based function.

Simulation results are presented to validate the theoretical expressions with Monte Carlo simulations. They also show that the proposed coverage-maximizing and rate-maximizing strategies provide significant gains in coverage and rate, respectively, when compared to the standard max-power and small-cell range expansion schemes. Such coverage and rate improvements still hold when we test the proposed frameworks in a realistic 3GPP channel.

The rest of the chapter is organized as follows. The system model is described in Section 5.3. The exact theoretical coverage expression and its closed-form approximation are given in Section 5.4. The coverage and rate maximization frameworks are presented

in Section 5.5 and Section 5.6, respectively, along with their suboptimal formulation and implementation. Simulation results are given in Section 5.7, and the main concluding remarks are discussed in Section 5.8.

5.3 System Model

5.3.1 Network topology and channel model

We assume a cellular network that consists of K tiers, where BSs' locations of each tier are sampled from a homogeneous Poisson point process (HPPP) Φ_k with BS density of λ_k per unit area. Each tier is completely characterized by the set $\{\Phi_k, \lambda_k, P_k, M_k, U_k\}$, where P_k is the total transmit power of the k -th tier BS, M_k is the number of antennas deployed by that BS, and U_k is the number of users that can be simultaneously served over a single time-frequency slot. We assume equal power allocation, i.e., each BS allocates P_k/U_k per user.

We further assume that UEs are sampled from another independent HPPP Φ_u with density λ_u , and all of them have single-antenna receivers. Using the properties of the HPPP, we focus on the performance of a user located at the origin, henceforth denoted as the typical user.

The typical UE is served by a BS located at $x_k \in \mathbb{R}^2$, also denoted as the tagged BS. The received signal power during data communications at the typical UE is expressed as

$$P_{R,k} = \frac{P_k}{U_k} g_{x_k} l(\|x_k\|), \quad (5.1)$$

where g_{x_k} is the effective channel power gain seen at the UE, i.e., the channel power after linear precoding, and $l(\|x_k\|) = L_0(\|x_k\|/d_0)^{-\alpha}$ is the path loss where d_0 is a reference distance, L_0 is the path loss at that distance, and $\alpha > 2$ is the path-loss exponent.

5.3.2 Multi-antenna transmission modes

The multi-antenna transmission mode determines the channel power gain, g_{x_k} , between the user and the tagged BS as well as the channel power gain, f_{y_k} , between the user and an interfering BS located at y_k . As discussed in [GDV14, HYA14, AEA16], when beamforming is implemented at the BS, the channel power gains can be modeled as $g_{x_k} \sim \Gamma(\Delta_k, 1)$ and $f_{y_j} \sim \Gamma(\Psi_j, 1)$, where $\Gamma(k, \theta)$ is the Gamma random variable with shape k and scale θ . For instance, when *single-user beamforming* (SU-BF) is used, the BS serves a single user, i.e., the multiplexing gain is one, yet this provides the highest array gain as all antennas are used, i.e., $\Delta_k = M_k$ [AEA16]. The interference in this case comes from other BSs, each serving a single user, i.e., $\Psi_j = 1$. In contrast, in *multi-user zero-forcing* (MU-ZF) beamforming, the BS spatially multiplexes $U_k > 1$ users using the zero-forcing precoder, i.e., the multiplexing gain is U_k . Serving more users comes at the expense of reduced array gain, which can be shown to be $\Delta_k = M_k - U_k + 1$ [AEA16]. Note that since each interfering BS from the j -th tier serves U_j users, we have $\Psi_j = U_j$ [AEA16]. Finally, in the presence of imperfect or quantized channel state information (CSI), it can be shown, as discussed in [DKA13, KA12], that the intra-cell interference is not completely canceled, which leads to slight degradation of the user throughput during data communications. This performance degradation occurs irrespective of the user association policy used, and hence it is ignored in the subsequent analysis.

5.3.3 Spectrum allocation

We consider orthogonal spectrum allocation as means for interference management. Specifically, BSs of different tiers do not use the same carriers, mitigating the inter-tier interference. Hence, we aim to optimize how much spectrum each tier should be allocated to maximize a rate utility function. In particular, the k -th tier is allocated $\eta_k W$ (Hz) of spectrum, where η_k denotes the spectrum allocation factor that needs to be optimized, and W is the total spectrum.

For tractability of the analysis, we make the following two widely adopted assumptions. First, we assume that each k -th tier BS has at least U_k users connected to.² Second, the BS will use a proportionally fair scheduling algorithm [YRC13], so that the long-term resources available at the typical user are proportional to $1/N_k$ [GDV14, LBY15], where N_k is the average number of users per BS in the k -th tier. Note that it is shown in [YRC13] that equal time sharing of spectral resources maximizes a proportionally fair utility.

5.3.4 Biased user association

During the cell search and selection procedure, the BS broadcasts reference signals at maximum power using a single antenna, i.e., broadcasting is omni-directional [GZA10]. Thus, beamforming is only used after user association. In this case, using cell range expansion, the user connects to the nearest BS in the k -th tier that satisfies

$$k = \arg \max_{j \in \{1, 2, \dots, K\}} B_j \underbrace{P_j l(\|x_j\|)}_{P_{\text{RSRP},j}}, \quad (5.2)$$

where B_j is a bias factor that weighs the averaged received signal power $P_{\text{RSRP},j}$ to artificially extend/shrink the coverage area of the BS. We note that the small-scale fading is neglected in (5.2) as the user computes the average power of the reference signal to avoid the ping-pong effect [DT16]. An illustration of this strategy is shown in Fig. 5.1.

Using (5.2), it can be shown that the association probability, namely the probability

²The assumption of fully-loaded networks is common in the literature [GDV14, LBY15, LZA16, AEA16], and it is warranted when $\lambda_u \gg \lambda_k \forall k$. Alternatively, one can assume that the k -th tier BS is active with probability p , which can be computed as discussed in [DT16]. In this case, we only need to replace λ_k with $\tilde{\lambda}_k = p\lambda_k$ [AEA16]. We note that in running simulations, we drop the fully-loaded assumption.

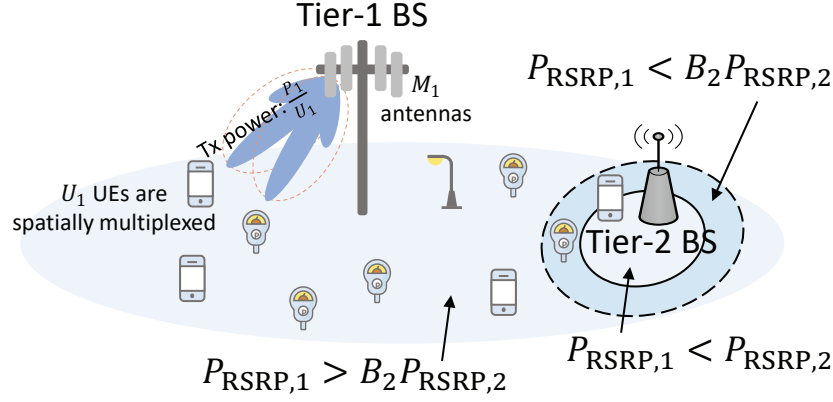


Figure 5.1: Illustration of a two-tier network. Here, $B_1 = 1$ for the macro tier, whereas tier-2 BSs artificially extend the coverage area to offload users from tier-1 BSs.

that the typical user connects to a k -th tier BS, is expressed as [GDV14]

$$\begin{aligned}
 A_k &= \mathbb{P} \left(B_k P_k l(\|x_k\|) \geq \arg\max_{j \neq k} B_j P_j l(\|x_j\|) \right) \\
 &= \prod_{j=1, j \neq k}^K \mathbb{E}_{x_k} \left[\mathbb{P} \left(\|x_j\| \geq \left(\frac{P_j B_j}{P_k B_k} \right)^{1/\alpha} \|x_k\| \right) \right] \\
 &\stackrel{(a)}{=} 2\pi\lambda_k \int_0^\infty x e^{-\pi \sum_{j=1, j \neq k}^K \left(\frac{P_j B_j}{P_k B_k} \right)^{2/\alpha} \lambda_j x^2 - \pi\lambda_k x^2} dx \\
 &= \frac{\lambda_k (B_k P_k)^{2/\alpha}}{\sum_{j=1}^K \lambda_j (B_j P_j)^{2/\alpha}},
 \end{aligned} \tag{5.3}$$

where (a) follows from using the null probability of the HPPP [JSX12] and then taking the expectation with respect to the distance between the typical user and the nearest k -th tier BS, i.e., $\|x_k\| \triangleq x$ [JSX12]. Clearly, A_k is directly proportional to B_k .

While biasing has been standardized for cellular HetNets, the 3GPP standard does not provide a mechanism to set the bias factors or offsets in (5.2). Thus, a key question is whether these factors, or virtual coverage areas, can be optimized to maximize a desired objective function for multi-antenna HetNets. Specifically, we aim to optimize the bias factors to maximize coverage and rate. This will be accomplished by optimizing the association probabilities in (5.3), as it can be shown that, up to a scaling factor, there is a one-to-one map between A_k and B_k .

5.4 Coverage Probability Analysis

In this section, we provide the coverage probability of frequency-partitioned multi-antenna HetNet. Specifically, we derive the exact coverage performance, and then present a closed-form approximation that enables us to learn key design insights regarding user association.

5.4.1 Exact coverage probability

The coverage probability quantifies the reliability of the transmission link. More formally, let γ_k denote the SIR when connecting to the k -th tier, i.e.,

$$\gamma_k = \frac{\frac{P_k}{U_k} g_{x_k} l(\|x_k\|)}{\sum_{y_k \in \Phi'_k} \frac{P_k}{U_k} f_{y_k} l(\|y_k\|)}, \quad (5.4)$$

where Φ'_k is the set of interfering BSs. Thus, the coverage probability is defined as $\mathbb{P}_{\text{cov},k}(\tau) = \mathbb{P}(\gamma_k \geq \tau)$, where τ is the SIR threshold. The most common approach to compute this expression is as follows. First, the expression is rewritten as the complementary cumulative distribution function (CDF) of the channel power gain g_{x_k} , i.e., $\mathbb{P}_{\text{cov},k}(\tau) = \mathbb{P}(g_{x_k} \geq \tau \|x_k\|^\alpha \sum_{y_k \in \Phi'_k} f_{y_k} \|y_k\|^{-\alpha})$. Second, by evaluating the complementary CDF of $g_{x_k} \sim \Gamma(\Delta_k, 1)$, the coverage becomes expressed as a function of the Laplace transform of the interference. Once this is evaluated, we take the expectation with respect to the random distance between the typical user and the tagged BS to get the final expression, which is given in the following theorem.

Theorem 8. *When frequency-partitioning is used in multi-antenna HetNets, the coverage probability of a typical user that connects to the k -th tier BS is expressed as*

$$\mathbb{P}_{\text{cov},k}(\tau) = \frac{\|(\mathbf{Q}_k + c_k \mathbf{I}_{\Delta_k})^{-1}\|_1}{A_k}, \quad (5.5)$$

where $c_k = \frac{1}{A_k} - 1$, $\mathbf{I}_{\Delta_k}$ is the identity matrix of size $\Delta_k \times \Delta_k$, $\|\cdot\|_1$ is the l_1 -induced matrix norm,

and $\mathbf{Q}_k$ is a lower triangular Toeplitz matrix given as

$$\mathbf{Q}_k = \begin{bmatrix} q_0 & & & & \\ q_1 & q_0 & & & \\ q_2 & q_1 & q_0 & & \\ \vdots & \ddots & \ddots & \ddots & \\ q_{\Delta_k-1} & \cdots & q_2 & q_1 & q_0 \end{bmatrix}, \quad (5.6)$$

where $q_i = \tau^i \frac{\Gamma(U_k+i)}{\Gamma(U_k)\Gamma(i+1)} \frac{2/\alpha}{(2/\alpha)-i} {}_2F_1\left(U_k+i, i-\frac{2}{\alpha}; i-\frac{2}{\alpha}+1; -\tau\right)$ and ${}_2F_1(a, b; c; z)$ is the Gauss hypergeometric function [JZ14].

Proof: Using the complementary CDF of g_{x_k} , we get [GDV14]

$$\mathbb{P}_{\text{cov},k}(\tau) = \mathbb{E}_s \left[\sum_{n=0}^{\Delta_k-1} \frac{s^n}{n!} \mathbb{E}_{I_k} [I_k^n e^{-sI_k}] \right] \stackrel{(a)}{=} \mathbb{E}_s \left[\sum_{n=0}^{\Delta_k-1} \frac{(-s)^n}{n!} \mathcal{L}_{I_k}^{(n)}(s) \right], \quad (5.7)$$

where $s = \tau \|x_k\|^\alpha$, $\mathcal{L}_{I_k}^{(n)}(s)$ is the n -th derivative of the Laplace transform of I_k , and (a) follows using the Laplace transform property $\mathbb{E}_{I_k}[I_k^n e^{-sI_k}] = (-1)^n \mathcal{L}_{I_k}^{(n)}(s)$. We then compute (5.7) using the same methodology as in [LZA16], and hence we only provide a sketch of the proof. The main difference compared to [LZA16] is that there is no inter-tier interference due to the orthogonal allocation of spectrum across tiers, and thus we can write $\mathcal{L}_{I_k}^{(n)}(s)$ in a recursive form as follows

$$\mathcal{L}_{I_k}^{(n)}(s) = \sum_{i=0}^{n-1} \binom{n-1}{i} \left[\pi \int_{x_j^2}^{\infty} (-1)^{n-i} (U_j)_{n-i} \frac{\left(v^{-\alpha/2}\right)^{n-i}}{(1 + sv^{-\alpha/2})^{n-i}} dv \right] \mathcal{L}_{I_k}^{(i)}(s), \quad (5.8)$$

where $(z)_{(n)} = \frac{\Gamma(z+n)}{\Gamma(z)}$ is the Pochhammer symbol. Denoting $v_n = \frac{(-s)^n}{n!} \mathcal{L}_{I_k}^{(n)}(s)$, we can then solve for v_n using principles of power series and their inverses similar to [LZA16]. ■

While Theorem 8 provides the coverage performance in a simple form, it is difficult to explicitly understand how different transmission modes and network settings affect the coverage. Hence, we aim next to approximate the coverage probability.

5.4.2 Coverage probability approximation

Instead of evaluating the Laplace transform and its derivative in (5.7), we seek an alternative approach using the Gil-Pelaez Inversion theorem [RG14]. Thus, the key steps are as follows. We first write the coverage performance as a function of the interference CDF. This CDF is then evaluated using the Gil-Pelaez inversion theorem, which states that for a continuous random variable Z , we can express its CDF as [Gil51]

$$F_Z(z) = \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \frac{1}{u} \text{Im} \{ \varphi_Z(u) e^{-juz} \} du, \quad (5.9)$$

where $\varphi_Z(u) = \mathbb{E}_Z[\exp(juZ)]$ is the characteristic function (CF) and $\text{Im}\{\cdot\}$ is the imaginary part of the input argument. After computing the interference CDF, we take the expectation with respect to the distance between the typical user and its tagged BS to get the coverage expression.

With slight abuse of notation, let the distance between the typical user and its tagged BS be denoted as $\|x_k\| \triangleq x$, which has the PDF $p_X(x) = (2\pi\lambda_k/A_k)x \exp(-\pi\lambda_k x^2/A_k)$ [JSX12]. Then, we can write the coverage probability of the k -th tier as

$$\begin{aligned} \mathbb{P}_{\text{cov},k}(\tau) &= \int_x \mathbb{P}(\gamma_k \geq \tau | X = x) p_X(x) dx \\ &\stackrel{(a)}{=} \frac{2\pi\lambda_k}{A_k} \int_0^\infty x \mathbb{E}_{g_{x_k}} \left[F_{I_k} \left(\frac{g_{x_k}}{\tau x^\alpha} \right) \right] e^{-\pi\lambda_k x^2/A_k} dx, \end{aligned} \quad (5.10)$$

where (a) follows by averaging over the distribution of the channel power gain, $\mathbb{E}_{g_{x_k}}[\cdot]$ is the expectation with respect to g_{x_k} , and $F_{I_k}(\cdot)$ is the CDF of the interference. Using (5.9), the interference CDF is expressed as

$$F_{I_k} \left(\frac{g_{x_k}}{\tau x^\alpha} \right) = \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \frac{1}{t} \text{Im} \{ \varphi_{I_k}(x^\alpha t) e^{-jt g_{x_k}/\tau} \} dt, \quad (5.11)$$

where we have used the substitution $t = ux^{-\alpha}$. The CF of interference can be shown to be [AEA16]

$$\begin{aligned} \varphi_{I_k}(tx^\alpha) &= \mathbb{E}_{I_k} \left[\exp \left(jtx^\alpha \sum_{y_k \in \Phi'_k} f_{y_k} \|y_k\|^{-\alpha} \right) \right] \\ &\stackrel{(a)}{=} \exp \left(-2\pi\lambda_k \int_x^\infty y \mathbb{E}_{f_{y_k}} [1 - e^{jtx^\alpha f_{y_k} y^{-\alpha}}] dy \right) \\ &\stackrel{(b)}{=} \exp \left(\pi\lambda_k x^2 \left(1 - \mathbb{E}_{f_{y_k}} [\Psi(f_{y_k}, \alpha, t)] \right) \right), \end{aligned} \quad (5.12)$$

where (a) follows from the probability generating functional of the HPPP [ESA17], (b) follows by carrying out the integral, $\Psi(f_{y_k}, \alpha, t) = {}_1F_1(-2/\alpha; 1 - 2/\alpha; jt f_{y_k})$, and ${}_1F_1(a; b; c)$ is the confluent hypergeometric function [JZ14]. Plugging (5.12) in (5.11), and then evaluating (5.10), we arrive at the following closed-form approximation of the coverage probability.

Theorem 9. *The coverage probability of the k -th tier in a frequency-partitioned multi-antenna HetNet can be approximated as*

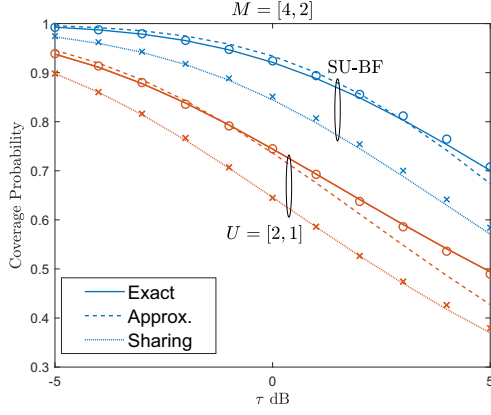
$$\mathbb{P}_{\text{cov},k}(\tau) \approx 1 - \left(\frac{A_k U_k}{A_k U_k + \beta_\tau} \right)^{\Delta_k}, \quad (5.13)$$

where $\beta_\tau = (\alpha - 2)/(2\tau)$.

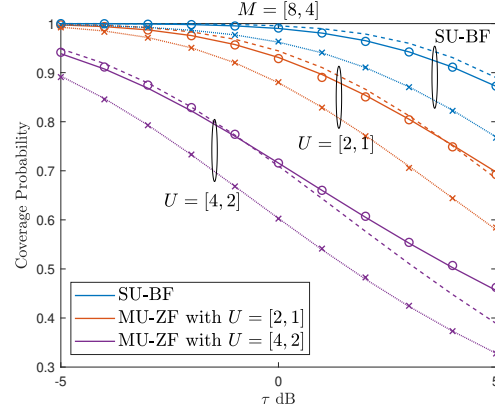
Proof: See Appendix D.1. This expression follows by using a first-order approximation of $\mathbb{E}_{f_{y_k}}[\Psi(f_{y_k}, \alpha, t)]$, and then evaluating the integral in (5.11). ■

It is shown that coverage depends on the user association and the multi-antenna transmission mode. Specifically, the k -th tier outage probability, i.e., $\mathbb{P}_{\text{out},k}(\tau) = 1 - \mathbb{P}_{\text{cov},k}(\tau)$, decreases exponentially with the array gain, and it is directly proportional with the multiplexing gain.

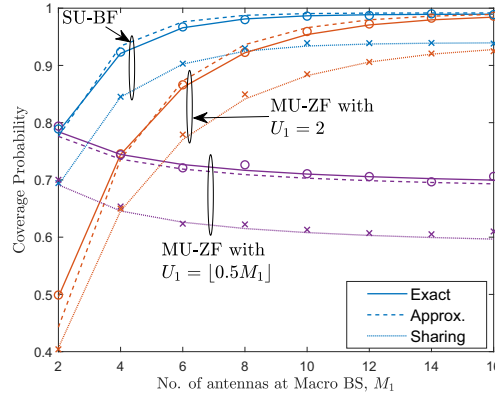
Fig. 5.2 validates the exact and approximated theoretical expressions with Monte Carlo simulations. We also show the performance of spectrum sharing for reference, where we use the analytical expression given in [LZA16, eq. (12)]. Here, a two-tier network, with max-power user association, is considered where $\lambda = [1, 5]$ per unit area, $P = [46, 23]$ dBm, $\alpha = 3.5$, and $L_0 \approx -38$ dB, which corresponds to the path loss at 1m when operating at 2GHz [JSX12]. It is shown that increasing M_k for a fixed U_k improves the coverage performance, as the array gain increases with M_k , whereas the inter-cell interference remains the same since the multiplexing gain is fixed. In contrast, for MU-ZF with $U_1 = \lfloor 0.5M_1 \rfloor$, the coverage degrades as M_1 increases because more users are spatially multiplexed, increasing the inter-cell interference. In addition, orthogonal allocation, as expected, provides better coverage than spectrum sharing, particularly when multiplexing is used, making the former suitable for machine-type applications



(a) SIR Threshold ($M = [4, 2]$)



(b) SIR Threshold ($M = [8, 4]$)



(c) No. of antennas ($M_2 = 2$, $U_2 = 1$, and $\tau = 0\text{dB}$)

Figure 5.2: Validation of the theoretical analysis, where markers denote Monte Carlo simulations.

with high coverage requirements. Finally, it is evident that the theoretical approximation is accurate, and hence we use it in the subsequent maximization frameworks.

5.5 Coverage Maximization Framework

In this section, we aim to develop user association algorithms that maximize the coverage of a frequency-partitioned HetNet. Specifically, we have the following optimization

problem

$$\begin{aligned} & \underset{\{A_k\}}{\text{maximize}} && \mathbb{P}_{\text{cov}}(\tau) = \sum_{k=1}^K A_k \mathbb{P}_{\text{cov},k}(\tau, A_k) \\ & \text{subject to} && \sum_{k=1}^K A_k = 1, \quad A_k \geq 0, \quad k = 1, 2, \dots, K, \end{aligned} \quad (5.14)$$

where we explicitly write the k -th tier coverage as a function of A_k . We have the following proposition.

Proposition 3. *The coverage maximization framework is a convex optimization framework.*

Proof: The optimization problem in (5.14) has a concave objective function (see Appendix D.2) and affine constraints, and hence it is concave. ■

Using a standard convex optimization solver, we can obtain the optimal association probabilities, i.e., $A_k^\star$, and map them to their respective bias factors using (5.3).

5.5.1 Special network settings

Under special network settings, we can find $A_k^\star$ in closed form, which provides useful design guidelines for user association. To this end, consider the following equivalent optimization problem, where $\mathbb{P}_{\text{out}}(\tau) = 1 - \mathbb{P}_{\text{cov}}(\tau)$,

$$\underset{\{A_k\}}{\text{minimize}} \quad \mathbb{P}_{\text{out}}(\tau) = \sum_{k=1}^K A_k \left(\frac{U_k A_k}{U_k A_k + \beta_\tau} \right)^{\Delta_k}, \quad (5.15)$$

subject to the same constraints as in (5.14). Let $L(A_k, z_k, \nu)$ be the Lagrangian of (5.15) with the dual variables z_k and ν . Then, from the Karush-Kuhn-Tucker (KKT) condition $\frac{\partial}{\partial A_k} L(A_k^\star, z_k^\star, \nu^\star) = 0$, we have

$$z_k^\star = \left(\frac{U_k A_k^\star}{U_k A_k^\star + \beta_\tau} \right)^{\Delta_k} \left(1 + \frac{\Delta_k \beta_\tau}{U_k A_k^\star + \beta_\tau} \right) + \nu^\star. \quad (5.16)$$

Using the complementary slackness condition, i.e., $z_k^\star A_k^\star = 0$, and then multiplying both sides by $A_k^\star$, we get

$$\nu^\star = - \sum_{k=1}^K A_k^\star \left(\frac{U_k A_k^\star}{U_k A_k^\star + \beta_\tau} \right)^{\Delta_k} \left(1 + \frac{\Delta_k \beta_\tau}{U_k A_k^\star + \beta_\tau} \right), \quad (5.17)$$

where we have summed over k and used the primal feasibility condition $\sum_{k=1}^K A_k^\star = 1$. Define the set $\mathcal{K} = \{k : A_k^\star > 0\} \subseteq \{1, 2, \dots, K\}$, then we have for all $j \in \mathcal{K}$

$$\left(\frac{U_j A_j^\star}{U_j A_j^\star + \beta_\tau} \right)^{\Delta_j} \left(1 + \frac{\Delta_j \beta_\tau}{U_j A_j^\star + \beta_\tau} \right) = \sum_{k=1}^K A_k^\star \left(\frac{U_k A_k^\star}{U_k A_k^\star + \beta_\tau} \right)^{\Delta_k} \left(1 + \frac{\Delta_k \beta_\tau}{U_k A_k^\star + \beta_\tau} \right) \forall j \in \mathcal{K}. \quad (5.18)$$

We can simplify this further as will be discussed in the following special cases.

5.5.1.1 Identical transmission modes

Consider a cellular network with identical transmission modes, i.e., $M_k = M$, $U_k = U$, and $\Delta_k = \Delta$. Then, we have the following result.

Corollary 6. *The user association policy that maximizes the coverage probability of a typical UE in a HetNet with identical transmission modes is equal-tier association, i.e., $A_k^\star = 1/K$.*

Proof: We have under this network setting

$$\left(\frac{U A_k^\star}{U A_k^\star + \beta_\tau} \right)^\Delta \cdot \left(1 + \frac{\Delta \beta_\tau}{U A_k^\star + \beta_\tau} \right) = \underbrace{-\nu^\star}_{\text{constant}}, \quad (5.19)$$

which means that $A_k \forall k \in \mathcal{K}$ are equal, and thus using the primal feasibility condition $\sum_{k=1}^K A_k^\star = 1$, we get $A_k = 1/|\mathcal{K}|$, where $|\cdot|$ is the cardinality of a set. We now prove that $\mathcal{K} = \{1, 2, \dots, K\}$, i.e., $|\mathcal{K}| = K$. Indeed, the minimum outage can be shown to be

$$\begin{aligned} \mathbb{P}_{\text{out}}^\star(\tau) &= \left(\frac{U}{U + |\mathcal{K}| \beta_\tau} \right)^\Delta \\ &\geq \left(\frac{U}{U + K \beta_\tau} \right)^\Delta, \end{aligned} \quad (5.20)$$

where the equality holds for $|\mathcal{K}| = K$. ■

We note that maximizing coverage in spectrum sharing requires $|\mathcal{K}| = 1$ as shown in [LZA16], whereas Corollary 6 shows that for frequency-partitioned HetNet, all tiers should be used to maximize coverage. This is intuitive since each tier occupies its own spectrum, and hence switching a tier off does not affect the interference seen at a UE using a different set of carriers.

Corollary 7. *Under optimal user association, increasing the number of tiers improves the coverage performance.*

Proof: The maximum coverage can be shown to

$$\mathbb{P}_{\text{cov}}^{\star}(\tau) = 1 - \left(\frac{U}{U + K\beta_{\tau}} \right)^{\Delta}, \quad (5.21)$$

which increases as K increases. ■

Corollary 7 shows how increasing the number of tiers improves coverage, particularly when few UEs are multiplexed. This can be attributed to the fact that the UE must select one of the K nearest BSs, with each one belonging to a different tier. Hence, as this number increases, the optimal association exploits this *tier diversity* to improve coverage. We emphasize that adding a new tier to the network is not equivalent to adding more BSs to an existing tier since frequency partitioning is tier-based, i.e., the new tier will use an orthogonal set of carriers.

5.5.1.2 FSDMA HetNets

Consider a cellular network with full spatial-division-multiple-access (FSDMA), where the number of multiplexed users is the same as the number of antennas, i.e., $U_k = M_k > 1$ [GDV14, KA12], then the optimal user association is given in the following result.

Corollary 8. *The user association policy that maximizes the coverage probability of a typical user in FSDMA HetNet is*

$$A_k^{\star} = \frac{1/M_k}{\sum_{k=1}^K 1/M_k}, \quad (5.22)$$

and the coverage performance is upper bounded by that achieved in single-antenna HetNets.

Proof: Since $U_k = M_k$ and $\Delta_k = 1$ in FSDMA, the left hand-side of (5.18) can be rewritten as

$$\frac{(M_j A_j^{\star})(M_j A_j^{\star} + 2\beta_{\tau})}{(M_j A_j^{\star} + \beta_{\tau})^2} = \underbrace{-\nu^{\star}}_{\text{constant}}, \quad \forall j \in \mathcal{K}. \quad (5.23)$$

Thus, a necessary and sufficient condition for this to hold is to have $M_k A_k^\star = M_j A_j^\star \forall j, k \in \mathcal{K}$, i.e., $A_k^\star = \frac{M_j}{M_k} A_j^\star$. By using the primal feasibility condition $\sum_{k=1}^K A_k^\star = 1$, we get $A_k^\star = M_k^{-1} / (\sum_{k=1}^K M_k^{-1})$. Note that $|\mathcal{K}| = K$ since

$$\begin{aligned} \mathbb{P}_{\text{cov}}^\star(\tau) &= \frac{\sum_{k=1}^{|\mathcal{K}|} M_k^{-1}}{\beta_\tau^{-1} + \sum_{k=1}^{|\mathcal{K}|} M_k^{-1}} \\ &\stackrel{(a)}{\leq} \frac{\sum_{k=1}^K M_k^{-1}}{\beta_\tau^{-1} + \sum_{k=1}^K M_k^{-1}} \\ &\stackrel{(b)}{\leq} \frac{K}{\beta_\tau^{-1} + K}, \end{aligned} \tag{5.24}$$

where (a) holds with equality when all tiers have non-zero association probabilities, and (b) is achieved with equality when $M_k = 1 \forall k$. ■

This shows that from a coverage perspective, IoT devices should be biased to tiers that multiplex a smaller number of UEs since the aggregate interference over the time-frequency slot is lower. We also note that Corollary 7 still holds for FSDMA networks. Note that FSDMA compensates the poor coverage with the increased pre-log term, i.e., long-term available resources, due to the high multiplexing gain.

5.5.1.3 SU-BF HetNets

Consider a cellular network with SU-BF, i.e., $U_k = 1$ and $\Delta_k = M_k$, then optimality can be achieved by solving $\forall k$

$$\left(\frac{A_k^\star}{A_k^\star + \beta_\tau} \right)^{M_k} \cdot \left(1 + \frac{M_k \beta_\tau}{A_k^\star + \beta_\tau} \right) = \underbrace{-v^\star}_{\text{constant}}, \tag{5.25}$$

which we cannot solve analytically. Instead, we approximate the objective function in (5.15) as

$$A_k \left(\frac{A_k}{A_k + \beta_\tau} \right)^{M_k} \approx A_k \left(\frac{A_k / M_k}{A_k / M_k + \beta_\tau} \right), \tag{5.26}$$

which follows from $(1+x)^r \approx 1+rx$ for small x . Using a similar KKT analysis, the left hand-side of (5.18) under this approximation is given as

$$\frac{(M_j^{-1}\hat{A}_j^\star)(M_j^{-1}\hat{A}_j^\star + 2\beta_\tau)}{(M_j^{-1}\hat{A}_j^\star + \beta_\tau)^2} = \underbrace{-\nu^\star}_{\text{constant}}, \quad \forall j \in \mathcal{K}. \quad (5.27)$$

A sufficient and necessary condition for this to hold is $\hat{A}_k^\star = \frac{M_k}{M_j} \hat{A}_j^\star \forall j, k \in \mathcal{K}$, and thus by using the feasibility condition $\sum_{k=1}^K \hat{A}_k^\star = 1$, we get $\hat{A}_k^\star = M_k / (\sum_{k=1}^K M_k)$. Note that $|\mathcal{K}| = K$ since

$$\begin{aligned} \mathbb{P}_{\text{cov}}^\star(\tau) &\approx \frac{\sum_{k=1}^{|\mathcal{K}|} M_k}{\beta_\tau^{-1} + \sum_{k=1}^{|\mathcal{K}|} M_k} \\ &\stackrel{(a)}{\leq} \frac{\sum_{k=1}^K M_k}{\beta_\tau^{-1} + \sum_{k=1}^K M_k}, \end{aligned} \quad (5.28)$$

where (a) holds with equality when all tiers have non-zero association probabilities. To summarize, in SU-BF HetNets, IoT devices should favor tiers that provide higher array gain for coverage enhancements.

5.5.2 Suboptimal user association

We observe that when the multi-antenna transmission modes across tiers only provide multiplexing gains, e.g., FSDMA HetNets, then $A_k^\star \propto \frac{1}{U_k} \triangleq \frac{1}{M_k}$, whereas when the modes only provide array gains, e.g., SU-BF HetNets, then $\hat{A}_k^\star \propto \Delta_k \triangleq M_k$. Finally, when all tiers have the same transmission mode, then any tier can be equally picked, i.e., $A_k^\star = 1/K$. Motivated by these observations, we propose the following suboptimal association scheme

$$\hat{A}_k = \frac{\Delta_k / U_k}{\sum_{k=1}^K \Delta_k / U_k}, \quad (5.29)$$

which is a generalization to all aforementioned special networks. The intuition behind this policy is as follows. The ratio of the array gain Δ_k to the multiplexing gain U_k is a measure of the average SIR. Indeed, it can be shown that, when conditioned on the locations of BSs of the associated tier ϕ_k , we have $\mathbb{E}[\gamma_k | \phi_k] \approx \frac{\Delta_k}{U_k} \left(\frac{I(\|x\|)}{\sum_{y_k \in \phi'_k} I(\|y_k\|)} \right)$, where $\phi'_k = \phi_k \setminus \{x_k\}$. If two tiers have the same ratio, then both will provide similar coverage,

and hence their respective association probabilities should be similar. This observation is in line with [AEA16], where it is shown via simulations that the performance of transmission modes differ according to their relative ratios Δ_k/U_k . We further note that this policy is optimal in HetNets with identical transmission modes and FSDMA HetNets. We end this section with the following interesting observation.

Corollary 9. *Under the association strategy that achieves the probabilities $\hat{A}_k$ in (5.29), the coverage performance is independent of the BSs' densities.*

Proof: Plugging (5.29) in $\mathbb{P}_{\text{cov}}(\tau)$, we get a function that only depends on Δ_k , U_k , and β_τ . ■

This property is commonly known in the literature as the *SIR invariance* property [JSX12]. It implies that we can add or remove BSs from the same tier, without worrying about the coverage performance. Intuitively, reducing the density of the tier's BSs increases the distance to the tagged BS, i.e., degrades signal power, but it also decreases the number of interfering BSs, i.e., the interference is reduced. Similarly, increasing the density of BSs increases the density of interferers, but the received signal power is also improved. To this end, it is more beneficial to increase the density of BSs since this translates into higher area spectral efficiency, i.e., $\log_2(1 + \tau) \sum_{k=1}^K \lambda_k U_k \mathbb{P}_{\text{cov},k}(\tau)$, which is in this case linear in λ_k .

5.5.3 Implementation

We can arrive at the desired association probabilities in (5.29) by either *artificially* changing the coverage areas via B_k or *physically* changing the coverage areas via P_k .

5.5.3.1 Cell-range expansion design

The 3GPP cellular standard specifies the procedure to share the bias values between BSs and UEs. For our proposed method, the user weighs the received power measurements

during cell association by the following bias factors

$$B_k^\star = \frac{1}{P_k} \left(\frac{\Delta_k/U_k}{\lambda_k} \right)^{\alpha/2}, \quad (5.30)$$

which can be shown by making (5.29) equal to (5.3). By convention, (5.30) can be further normalized such that $B_1 = 1$. We note that obtaining the information needed to compute (5.30) can be done by embedding them in reference signals, or it can be known beforehand since they are typically specified for each UE category [NLV15].

5.5.3.2 Transmit power design

Alternative to optimizing the bias values, we can directly optimize the power levels and set $B_k = 1 \forall k$. Specifically, the transmit powers must be set to

$$P_k^\star = P_c \left(\frac{\Delta_k/U_k}{\lambda_k} \right)^{\alpha/2}, \quad (5.31)$$

where P_c is some constant, which can be set such that the transmit power of the macro cell is 46dBm [3GP17a]. This design shows that to maximize coverage, the BS must transmit reference signals at power that is proportional to the average distance between the typical user and the tagged BS (it can be shown that $\mathbb{E}[\|x_k\|] = \sqrt{A_k/4\lambda_k}$). In addition, the transmit power should be proportional to the transmission mode that will be used for data communications, i.e., the disparity in coverage areas during cell discovery and data communications should be reduced. This can be achieved either by: (i) broadcasting reference signals using omni-directional antennas, to which scaling the transmit power is necessary, or (ii) broadcasting reference signals over specific directions, i.e., beamforming is not only used during communications, but also during cell association. The concept of directional user association has become a defining feature of cellular networks operating at millimeter wave frequencies [ANR17, HVC18a, HVC18b], and thus the above discussion suggests that it can be also beneficial sub-6GHz for coverage maximization, yet a procedure for directional cell association becomes needed.

5.6 Rate Maximization Framework

While maximizing the coverage is critical in HetNets, particularly for IoT applications that require deep coverage [Ind17], the framework in (5.14) does not necessarily maximize the rate [SDA13], which is more critical for human-type traffic. This follows because the rate does not only depend on the SIR but also on the load seen at the tagged BS [HC18e]. To this end, we use the following widely adopted throughput metric of a user connected to the k -th tier [GDV14,LBY15,BL15]

$$R_{\tau|k} \approx \frac{\eta_k}{N_k/U_k} W \log_2(1 + \tau) \mathbb{P}(\gamma_k \geq \tau), \quad (5.32)$$

where recall that η_k is the portion of spectrum allocated to the k -th tier and the approximation follows because $N_k = \lambda_u A_k / \lambda_k$ is the *average* number of users per BS in the k -th tier [JSX12]. This expression entails that UEs in outage, i.e., $\gamma_k < \tau$, achieve zero rate³. Let $\hat{R}_{\tau|k} = \frac{\lambda_k U_k \eta_k}{A_k} \mathbb{P}(\gamma_k \geq \tau)$, i.e., the constant $W \log_2(1 + \tau) / \lambda_u$ is ignored, then we consider the following proportional-fair utility function [LBY15,BBP16]

$$U_\tau = \sum_{k=1}^K A_k \log(\hat{R}_{\tau|k}). \quad (5.33)$$

Using the logarithm encourages fairness by giving higher rewards to low-rate UEs compared with high-rate UEs. Thus, the rate utility maximization framework is given as [LBY15,BBP16]

$$\begin{aligned} \underset{\{A_k\}, \{\eta_k\}}{\text{maximize}} \quad & \sum_{k=1}^K A_k \log\left(\frac{\lambda_k U_k \eta_k}{A_k}\right) + A_k \log\left(1 - \left(\frac{A_k U_k}{A_k U_k + \beta_\tau}\right)^{\Delta_k}\right), \\ \text{subject to} \quad & \sum_{k=1}^K A_k = 1, \quad A_k \geq 0, \quad k = 1, 2, \dots, K, \\ & \sum_{k=1}^K \eta_k = 1, \quad \eta_k \geq 0, \quad k = 1, 2, \dots, K. \end{aligned} \quad (5.34)$$

This framework is a generalization to the one given for single-antenna HetNets in [LBY15], and it is a convex optimization problem, as shown in the following proposition.

³In the case of multiple modulation and coding schemes, we can use N threshold values $\tau_1, \dots, \tau_N$ with respective rates $\log_2(1 + \tau_1), \dots, \log_2(1 + \tau_N)$. It can be shown that convexity of the rate maximization framework still holds, but this affects the tractability of the problem. Hence, we consider a single threshold as commonly done in the literature [LBY15,GDV14,LZA16].

Proposition 4. *The rate maximization framework is a convex optimization problem.*

Proof: The objective function is jointly concave with respect to A_k and η_k (see Appendix D.3), and the constraints are all affine. ■

Since each constraint either depends on A_k or η_k , i.e., the constraints are independent, with respect to the optimization variables, and since the objective function is jointly concave in these variables, then we can first maximize over η_k and sequentially maximize over A_k [BV04].

5.6.1 Optimal spectrum allocation

The following proposition shows the optimal spectrum allocation factor that maximizes (5.34).

Proposition 5. *The optimal spectrum allocation factor of the k -th tier is equal to the association probability of that tier, i.e.,*

$$\eta_k^\star = A_k. \quad (5.35)$$

Proof: This follows by using the KKT condition, where the optimal allocation factor minimizes the Lagrangian of (5.34), i.e.,

$$\frac{\partial}{\partial \eta_k} \tilde{L}(A_k, \eta_k^\star, \tilde{z}_k^\star, \tilde{v}^\star) = \frac{A_k}{\eta_k^\star} + \tilde{v}^\star - \tilde{z}_k^\star = 0, \quad (5.36)$$

where $\tilde{v}$ is the dual variable associated with the equality constraint $\sum_{k=1}^K \eta_k = 1$ and $\tilde{z}_k$ is the dual variable associated with the inequality constraint $\eta_k \geq 0$. Using the primal feasibility condition $\sum_{k=1}^K \eta_k^\star = 1$ and the complementary slackness condition $\tilde{z}_k^\star \eta_k^\star = 0$, we get $\tilde{v}^\star = -1$, and thus $\eta_k^\star = A_k$. ■

This is an intuitive spectrum allocation approach since it requires the operator to divide the spectrum across tiers according to the portion of users that connect to each one, i.e., $\lambda_u A_k$. Such allocation can be designed at the core network, where the BSs periodically update the load information. Alternatively, the allocation can be designed

at each macro cell, where small cells in vicinity periodically send the load information to the macro cell using the X2 interface [GZA10].

5.6.2 Optimal user association

Using the optimal spectrum allocation factor in the rate maximization framework, we get

$$\begin{aligned} & \underset{\{A_k\}}{\text{maximize}} && \sum_{k=1}^K A_k (\log(\lambda_k U_k) + \log(1 - (\frac{A_k U_k}{A_k U_k + \beta_\tau})^{\Delta_k})), \\ & \text{subject to} && \sum_{k=1}^K A_k = 1, \quad A_k \geq 0, \quad k = 1, 2, \dots, K. \end{aligned} \quad (5.37)$$

This is concave in A_k , and thus the global solution can be evaluated numerically using a standard convex optimization solver. Comparing this problem to the coverage maximization one in (5.14), then ignoring the $\log(\cdot)$, it is observed the rate maximization still requires maximizing the coverage probability as in (5.14). The key difference between these frameworks is that there is another objective in (5.37), which is to maximize the average number of users served in a resource block, i.e., $\lambda_k U_k$. In other words, the rate maximization problem is a scalarized multi-objective problem [BJD14] that seeks to maximize coverage and the number of users served, with each objective being equally weighted. This interpretation motivates us to propose an alternative suboptimal framework as discussed next.

5.6.3 Suboptimal multi-objective user association

We propose the following suboptimal approach to the rate maximization problem. The first objective is to minimize $\log(1/(\lambda_k U_k))$, and the second objective is to minimize the logarithm of the outage probability, i.e., $\log(\mathbb{P}(\gamma_k \leq \tau))$. That is, we have the following optimization problem

$$\begin{aligned} & \underset{\{A_k\}, \{y_k\}}{\text{minimize}} && \sum_{k=1}^K A_k \log(1/(\lambda_k U_k)) + A_k \Delta_k \log\left(\frac{A_k}{y_k}\right), \\ & \text{subject to} && y_k = A_k + \frac{\beta_\tau}{U_k}, \quad k = 1, 2, \dots, K, \\ & && \sum_{k=1}^K A_k = 1, \quad A_k \geq 0, \quad k = 1, 2, \dots, K. \end{aligned} \quad (5.38)$$

We note that $A_k \log(A_k/y_k)$ is commonly known as the *relative entropy* of (A_k, y_k) , and it is convex [BV04]. Thus, this problem is jointly convex in (A_k, y_k) .

Let us introduce the dual variables $\hat{\mu}_k$, $\hat{\nu}$, and $\hat{z}_k$ for the first, second, and third constraints, respectively. Then, the Lagrangian is expressed as

$$\begin{aligned} \hat{L}(A_k, y_k, \hat{\mu}_k, \hat{z}_k, \hat{\nu}) \\ = -\hat{\nu} + \sum_{k=1}^K -\beta_\tau \frac{\hat{\mu}_k}{U_k} + y_k \hat{\mu}_k - A_k \Delta_k \log(y_k) + A_k \left(\log \left(\frac{A_k}{\lambda_k U_k} \right) - \hat{\mu}_k - \hat{z}_k + \hat{\nu} \right). \end{aligned} \quad (5.39)$$

By computing $\frac{\delta \hat{L}}{\delta y_k} = 0$, we get $y_k^\bullet = \frac{A_k \Delta_k}{\mu_k}$. Thus, replacing $y_k^\bullet$ in (5.39), the dual function becomes

$$\hat{D} = \underset{\{A_k\}}{\text{minimize}} \left\{ -\hat{\nu} - \beta_\tau \sum_{k=1}^K \frac{\hat{\mu}_k}{U_k} + \sum_{k=1}^K A_k (\Delta_k \log(\hat{\mu}_k) - r_k - \hat{\mu}_k - \hat{z}_k + \hat{\nu}) \right\}, \quad (5.40)$$

where $r_k = \log(\lambda_k U_k) + \Delta_k(\log(\Delta_k) - 1)$. Clearly, $\hat{D} = -\infty$ unless $\hat{\nu} = \hat{z}_k + \hat{\mu}_k - \Delta_k \log(\hat{\mu}_k) + r_k$, for which we have $\hat{D} = -\hat{\nu} - \beta_\tau \sum_{k=1}^K \frac{\hat{\mu}_k}{U_k}$. Since the dual variable $\hat{z}_k \geq 0$, this is equivalent to $\hat{\nu} \geq \hat{\mu}_k - \Delta_k \log(\hat{\mu}_k) + r_k$. Thus, the dual optimization problem becomes

$$\begin{aligned} \underset{\{\hat{\mu}_k\}, \hat{\nu}}{\text{minimize}} \quad & \hat{\nu} + \beta_\tau \sum_{k=1}^K \frac{\hat{\mu}_k}{U_k}, \\ \text{subject to} \quad & \hat{\nu} \geq \hat{\mu}_k - \Delta_k \log(\hat{\mu}_k) + r_k. \end{aligned} \quad (5.41)$$

This problem has a linear objective function. Using the KKT conditions $y_k^\bullet = A_k^\bullet + \frac{\beta_\tau}{U_k}$ and $y_k^\bullet = \frac{A_k^\bullet \Delta_k}{\mu_k^\bullet}$, we get $A_k^\bullet = \frac{\hat{\mu}_k^\bullet \beta_\tau}{U_k(\Delta_k - \hat{\mu}_k^\bullet)}$, where $\hat{\mu}_k^\bullet$ is the variable that minimizes (5.41). Thus, using the primal feasibility $\sum_{k=1}^K A_k^\bullet = 1$, the association probability that minimizes (5.38) becomes

$$A_k^\bullet = \frac{\hat{\mu}_k^\bullet [U_k(\Delta_k - \hat{\mu}_k^\bullet)]^{-1}}{\sum_{k=1}^K \hat{\mu}_k^\bullet [U_k(\Delta_k - \hat{\mu}_k^\bullet)]^{-1}}. \quad (5.42)$$

The interpretation of this association probability is discussed next. Using the KKT conditions, it can be shown that to maintain the primal feasibility condition, we must have $\hat{\mu}_k^\bullet \leq \left(\frac{U_k}{\beta_\tau + U_k} \right) \Delta_k$, and we must also have $\hat{\mu}_k^\bullet > 0$, otherwise $\hat{\nu} = \infty$. The latter constraint shows that $A_k^\bullet > 0 \forall k$. Also, we can show that $\hat{\nu}^\bullet = \sum_{k=1}^K A_k^\bullet (\hat{\mu}_k^\bullet - \Delta_k \log(\hat{\mu}_k^\bullet) + r_k)$, and since the inequality constraint is satisfied with equality at the optimal dual value

and $\hat{\mu}_k^\bullet > 0$, the following condition must hold for all tiers

$$r_k - (\Delta_k \log(\hat{\mu}_k^\bullet) - \mu_k^\bullet) = \underbrace{\nu^\bullet}_{\text{constant}}. \quad (5.43)$$

Let $g_k(\mu) = \Delta_k \log(\mu) - \mu$, which is an increasing function with respect to $\mu \in (0, \frac{U_k \Delta_k}{\beta_\tau + U_k})$.

Then, we can infer the following.

- If all tiers use the same transmission mode, i.e., $g_k(\mu) = g_j(\mu)$, then higher $\hat{\mu}_k$ is assigned to tiers with higher densities, which implies higher association probability is assigned to denser tiers since, on average, they serve more users. Indeed, if $\lambda_k > \lambda_j$, we must have $g_k(\hat{\mu}_k) > g_j(\hat{\mu}_j)$ to maintain the condition in (5.43), i.e., $\hat{\mu}_k > \hat{\mu}_j$.
- If all tiers have identical density and multiplexing gain, then for $\Delta_k > \Delta_j$, we must have $\hat{\mu}_k > \hat{\mu}_j$, i.e., higher association probabilities are assigned to tiers with higher array gain when the number of users served across tiers is the same.⁴

From the above observations, and by rewriting (5.43) as $g_k(\hat{\mu}_k^\bullet) = r_k - \nu^\bullet$, we can think of the right-hand side as the total supply of the k -th tier. The k -th tier supply is primarily determined by r_k , which captures the average number of users that can be served by the tier, and the array gain of a single link. Thus, $\hat{\mu}_k^\bullet$ is set to drive the demand, and hence users' distribution across tiers, to an equilibrium point, i.e., $\nu^\bullet$. For instance, if $r_k > r_j$, then $\hat{\mu}_k^\bullet > \hat{\mu}_j^\bullet \iff A_k^\bullet > A_j^\bullet$.

5.6.4 Implementation

The proposed association strategy in (5.42) can be implemented using cell range expansion. To this end, the supply of each tier, r_k , must be computed in advance, possibly offline at the core network, and then broadcasted via reference signals, along with the

⁴Although in this case $\exists \hat{\mu}_k, \hat{\mu}_j$ such that $\hat{\mu}_k < \hat{\mu}_j$ and $g_k(\hat{\mu}_k) > g_j(\hat{\mu}_j)$, we can show that these points do not satisfy (5.43). This can be proven by studying two cases: $\hat{\mu}_k, \hat{\mu}_j < 1$ and $\hat{\mu}_k, \hat{\mu}_j > 1$ and then showing that for each one we must have $\hat{\mu}_k > \hat{\mu}_j$ so that (5.43) holds.

transmission mode to be used. The user then solves (5.41) using a standard convex optimization solver to compute the optimal demand $\hat{\mu}_k^\bullet$ of each tier. Finally, the UE weighs the received signal power by

$$\hat{B}_k = \frac{1}{P_k} \left(\frac{\hat{\mu}_k^\bullet}{\lambda_k U_k (\Delta_k - \hat{\mu}_k^\bullet)} \right)^{\alpha/2}. \quad (5.44)$$

5.7 Simulation Results

Unless otherwise stated, we consider a two-tier network with a macro tier being overlaid with a small-cell tier, where $\lambda = [1, 5]$ per unit area, $\lambda_u = 100$ per unit area, $P = [46, 23]$ dBm, $\alpha = 3.5$, $L_0 \approx -38$ dB, and $W = 20$ MHz. Unless otherwise stated, we assume $M = [8, 4]$ and $U = [4, 1]$. The locations of BSs and UEs are generated according to their HPPPs over a circular disk of radius 10 km. We average the coverage and rate performance of the typical user over 10^4 spatial and channel realizations. The fully-loaded network assumption is dropped, and hence we compute the actual load on each BS to calculate the rate in (5.32), i.e., we replace N_k in (5.32) with $\hat{N}_{b_k}$, where $\hat{N}_{b_k}$ is the actual number of users during that spatial realization. Similarly, the channel power gain from an interfering k -th tier BS $\hat{b}_k$ is distributed as $\Gamma(\min\{U_k, \hat{N}_{b_k}\}, 1)$.

5.7.1 Evaluation of the coverage maximization framework

In this section, we compare the coverage performance of the following schemes: (i) the optimal bias computed from (5.14), the suboptimal bias computed from (5.30), max-power where the user connects to the BS with maximum received signal power, i.e., $B_k = 1 \forall k$, and the small-cell range expansion (SC-RE), where a 10 dB bias value is used for small-cells, i.e., $B_2 = 10$. In what follows, lines denote the theoretical curves, whereas markers denote Monte Carlo simulations.

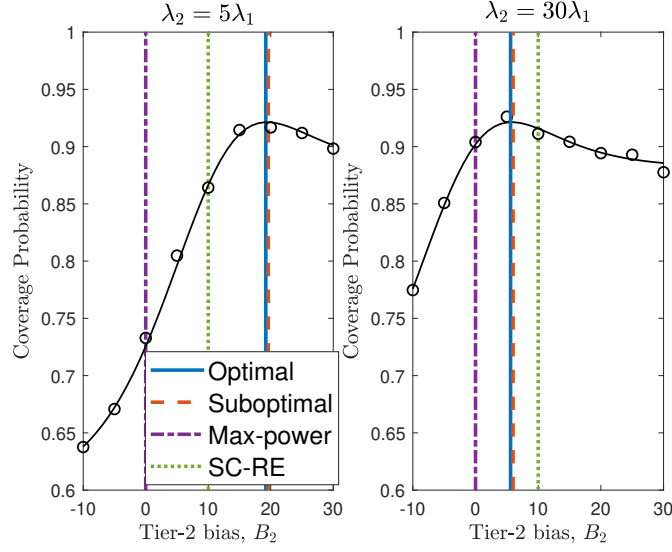
5.7.1.1 Impact of small-cell bias and transmit power

Fig. 5.3a shows the coverage probability for different small-cell bias values, where by convention, we assume $B_1 = 1$ (0dB). It is observed that the theoretical curve matches well with simulations, and that the maximization framework, which uses the coverage approximation, indeed maximizes coverage. Similarly, the suboptimal solution provides nearly identical solution to the optimal one. In addition, different bias values should be used, depending on the density of small cells. Indeed, increasing the density from $\lambda_2 = 5\lambda_1$ to $\lambda_2 = 30\lambda_1$ decreases the optimal bias from 19dB to 5.5dB, as small cells become closer to the user, improving the received signal power, and hence mitigating the need for higher bias values.

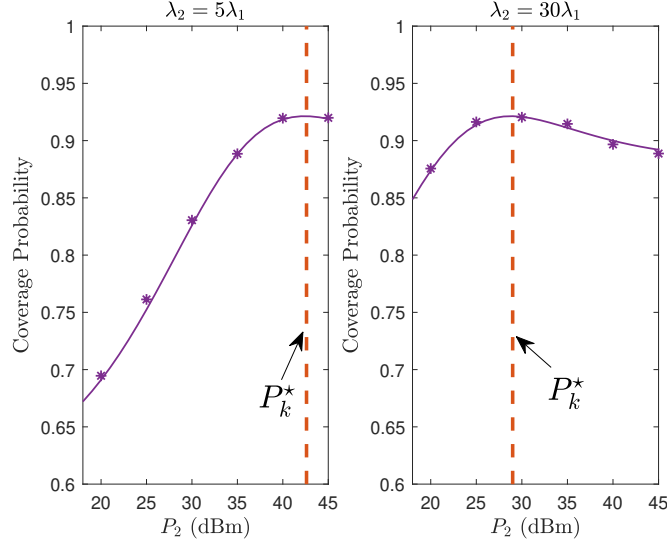
Fig. 5.3b shows the coverage performance of max-power association as the transmit power of small-cell BSs is varied. We also show the proposed transmit power computed from (5.31), where P_c is set such that $P_1 = 46\text{dBm}$. It is evident that the transmit power plays a key role in coverage when cell range expansion is not used, and that there exists an optimal value for which coverage is maximized in frequency-partitioned HetNet. This follows because for low P_2 , the performance is dominated by the received signal from the tagged BS, which is weak, whereas for high P_2 , the performance is dominated by the interference from other BSs, which is high. Both effects have a detrimental impact, and there exists a recommended value in between at which coverage can be maximized. Finally, increasing λ_2 reduces the required transmit power since the average communication range between the UE and the BS is reduced.

5.7.1.2 Impact of transmission mode

Fig. 5.4a and Fig. 5.4b illustrate the coverage with variations of M_1 and U_1 , respectively. As expected, for a fixed number of multiplexed users, increasing the number of antennas increases the array gain, and hence improves coverage. However, when $U_1 = \lfloor 0.5M_1 \rfloor$, the inter-cell interference increases, affecting the coverage as shown in Fig. 5.4a. It is also



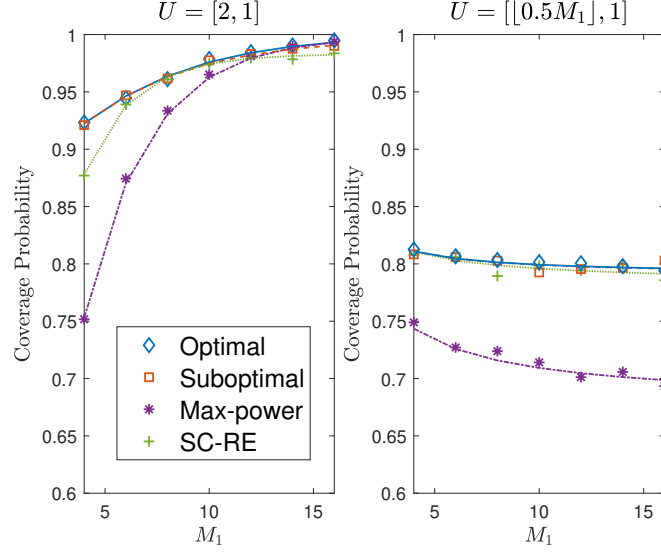
(a) Variations of B_2 ($B_1 = 1$ and $\tau = 0\text{dB}$)



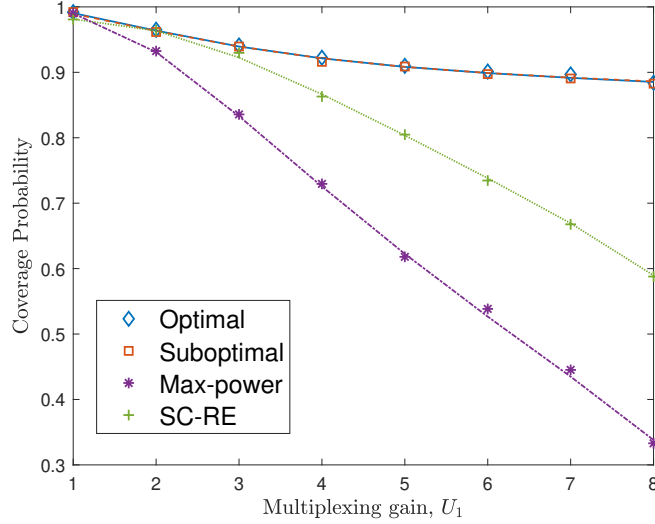
(b) Variations of P_2 ($\tau = 0\text{dB}$)

Figure 5.3: Impact of small-cell bias and transmit power on coverage ($\tau = 0\text{dB}$).

shown that SC-RE performs well for the given transmission modes used in Fig. 5.4a. In contrast, SC-RE and max-power perform poorly when the multiplexing gain is increased, for a fixed M_1 , as shown in Fig. 5.4b. This follows because the macro tier multiplexes more users as U_1 increases, increasing the inter-cell interference. Thus, a higher B_2 is needed to move the users to small cells, where inter-cell interference is lower.



(a) Variations of M_1



(b) Variations of U_1 ($M_1 = 8$)

Figure 5.4: Coverage performance under different transmission modes ($\tau = 0\text{dB}$).

5.7.1.3 Impact of network densification

Fig. 5.5a shows the coverage as the number of tiers increases from a single macro tier to a four-tier HetNet with $M = [8, 4, 4, 4]$, $U = [4, 1, 1, 1]$, and $P = [46, 23, 20, 17]\text{dBm}$. For SC-RE, $B = [1, 10, 10, 10]$. As verified in Corollary 7, coverage improves with K as the UE

has a larger pool of BSs to select from. In addition, increasing heterogeneity reveals that max-power association can provide be a poor strategy even from a coverage perspective.

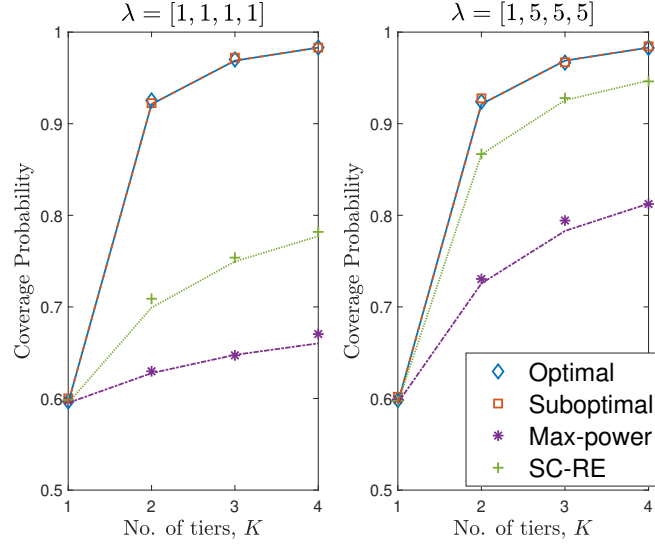
Fig. 5.5b shows the coverage performance as the density of small cells increases in a two-tier network. As proven in Corollary 9, the proposed coverage-maximizing user association strategy maintains the SIR invariance property, which generally does not hold for arbitrary user association in multi-antenna HetNets [AEA16]. For instance, it is observed that for max-power or SC-RE association, coverage can either improve or degrade as the density increases, depending on the transmission mode. Specifically, in SU-BF, the coverage area of the macro tier is actually better during data communication than that during cell discovery, yet this array gain is not utilized since max-power and SC-RE offloads more UEs to small cells as λ_2 increases. In contrast, in MU-ZF, the macro tier's coverage area is smaller during data communications due to equal-power allocation, and hence offloading users to small cells is more beneficial for high λ_2 .

5.7.2 Evaluation of the rate maximization framework

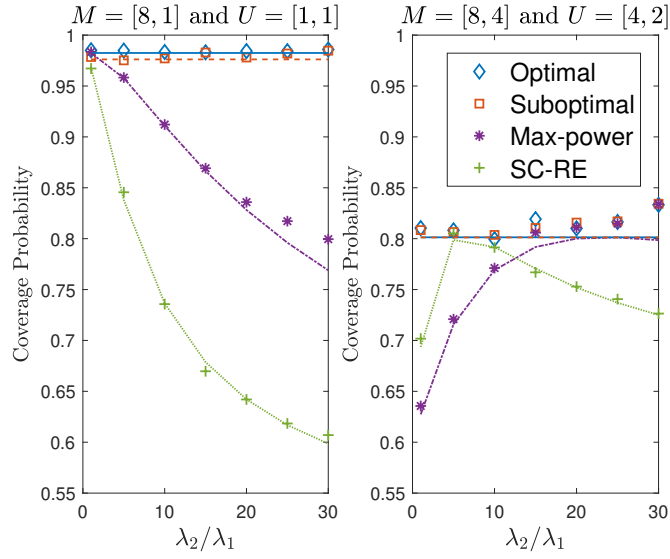
In this section, we compare the rate performance of the following schemes: (i) the optimal bias computed from (5.37), the suboptimal bias computed from (5.44), i.e., the solution of (5.38), max-power, and SC-RE. For all schemes, we use the optimal resource allocation, given in (5.35).

5.7.2.1 Impact of small-cell bias

Fig. 5.6 shows the average rate of a typical UE with variations of the small-cell bias. We also show the theoretical utility value (5.33) and the bias values computed for the different schemes. It is observed that the optimal policy indeed maximizes the rate utility, improving the rate of the typical user. Also, the proposed schemes tend to assign high small-cell bias values because $\lambda_2 = 5\lambda_1$, and hence the average number of users served by the small-cell tier is higher than that of the macro tier, i.e., $U_2\lambda_2 > U_1\lambda_1$.



(a) Variations of K



(b) Variations of λ_2

Figure 5.5: Impact of network densification on coverage ($\tau = 0\text{dB}$).

5.7.2.2 Impact of small cell densification

Fig. 5.7a illustrates the average rate of the typical UE as the density of small cells increases. It is evident that network densification provides substantial rate gains for the UE, and the gains can be further increased when the proposed user association strategies

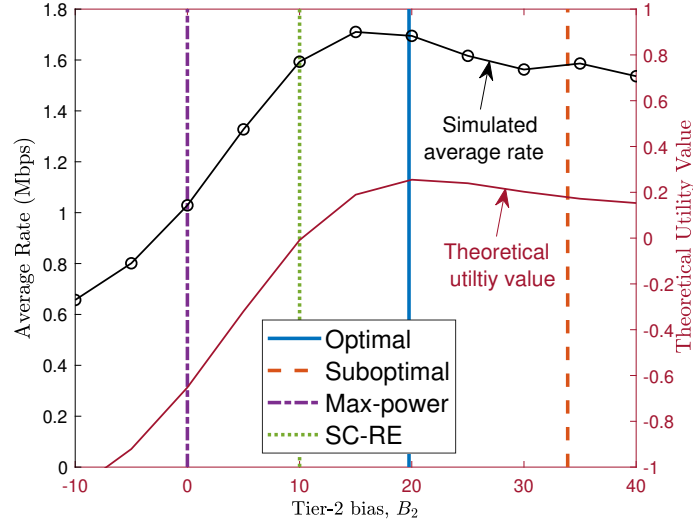
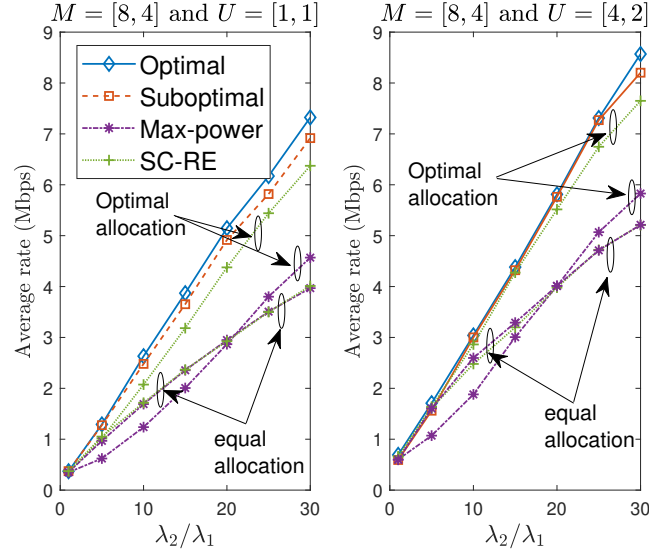
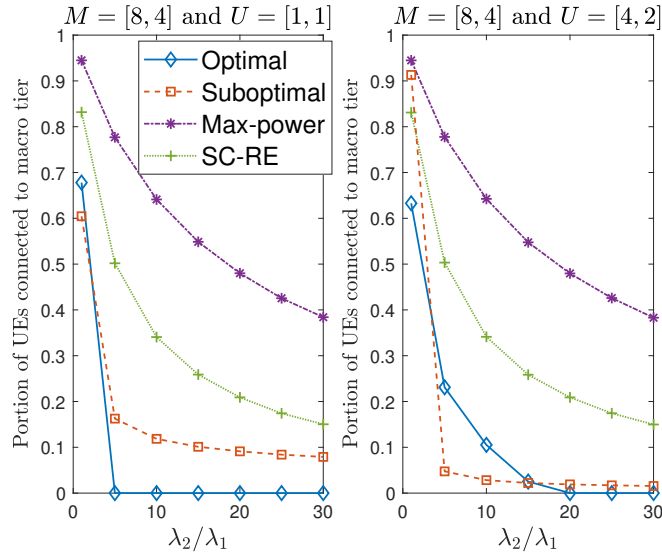


Figure 5.6: Mean rate and utility with variations of B_2 ($B_1 = 1$ and $\tau = 5\text{dB}$).

are used. Specifically, as the density of small-cell BSs increases, load-balancing becomes more critical. Thus, the optimal user association exploits the high density of small cells to connect users to different BSs in order to improve the rate. It is also clear that the sub-optimal user association has a similar performance to the optimal one. We also show the performance of existing methods when equal resource allocation is used, i.e., $\eta_k = 1/K$. As can be seen, the rate degrades when resource allocation is not optimized, particularly when the density of small cells is high. We also show in Fig. 5.7b the portion of users connected to the macro tier as the density of small cells increases. Interestingly, rate maximization entails aggressive offloading from the macro tier to the small-cell tier, when the latter has a higher density. This notably suggests that almost the entire spectrum should be allocated to small cells for rate maximization. This is due to the fact that the small cell tier serves on average $\lambda_2 U_2 \gg \lambda_1 U_1$ in ultra-dense networks, and hence a UE connecting to a small-cell BS has a higher chance to be scheduled more frequently. In other words, spectrum sharing becomes the superior paradigm in terms of the mean rate in ultra-dense networks.



(a) Average rate for different small cell densities



(b) Portion of users connected to macro tier

Figure 5.7: User mean rate and association probability for different user association policies ($\tau = 5\text{dB}$).

5.7.3 Performance of the proposed policies in a realistic channel

In this section, we present simulations that validate the proposed policies using the 3D-UMi 3GPP channel model [3GP17a, Table 7.2-1], which is suitable for dense urban areas.

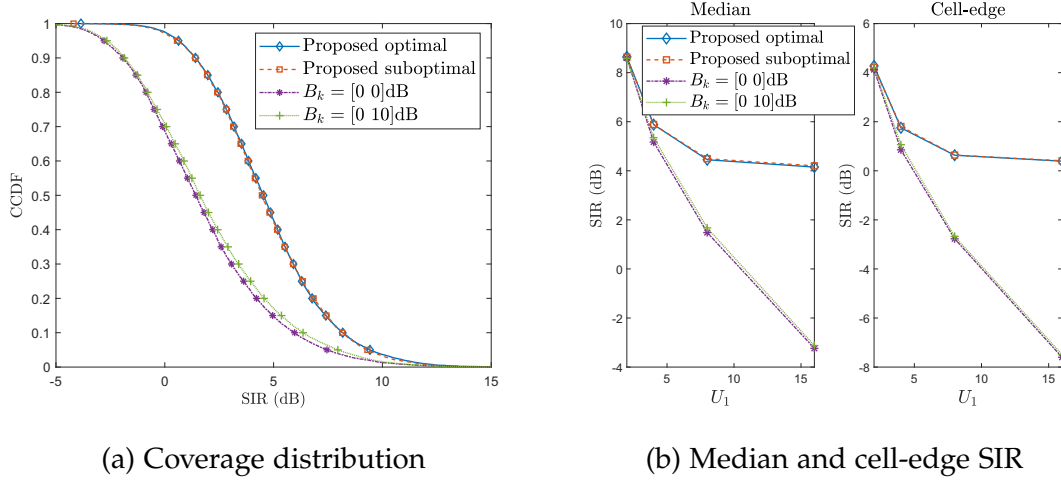


Figure 5.8: Coverage performance under the 3GPP channel model.

The model considers a multi-slope path loss with different attenuation, depending on whether the link is line-of-sight (LOS) or non-LOS, where the probability of LOS is given in [3GP17a, Table 7.2-2]. Log-normal shadowing is also added, with a standard deviation of 3dB and 4dB in LOS and NLOS, respectively [3GP17a]. We assume the antenna height of the BS and the UE are 10m and 1.5, respectively [3GP17a], and the operating channel is centered at 2GHz. We further assume $M = [32, 16]$ and, unless otherwise stated, the density of small cells is 5x the density of macro cells.

Fig. 5.8a shows the SIR coverage when $U = [8, 1]$ and Fig. 5.8b shows the median and cell-edge, i.e., 5th percentile, SIR when the number of multiplexed users in the macro cell is varied. It is evident that the proposed coverage-based association policies still bring tangible gains, particularly when the macro tier multiplexes more users. Fig. 5.9b shows the CDF of the rate coverage to capture the performance of all users in the network, and in Fig. 5.9b we show the median and cell-edge rates with variations of small cell density. It is observed that the proposed rate-based association significantly improves the cell-edge rate, particularly in ultra-dense networks, e.g., the cell-edge rate approximately quadruples using the proposed policies when $\lambda_2 = 30\lambda_1$, compared to max-power association.

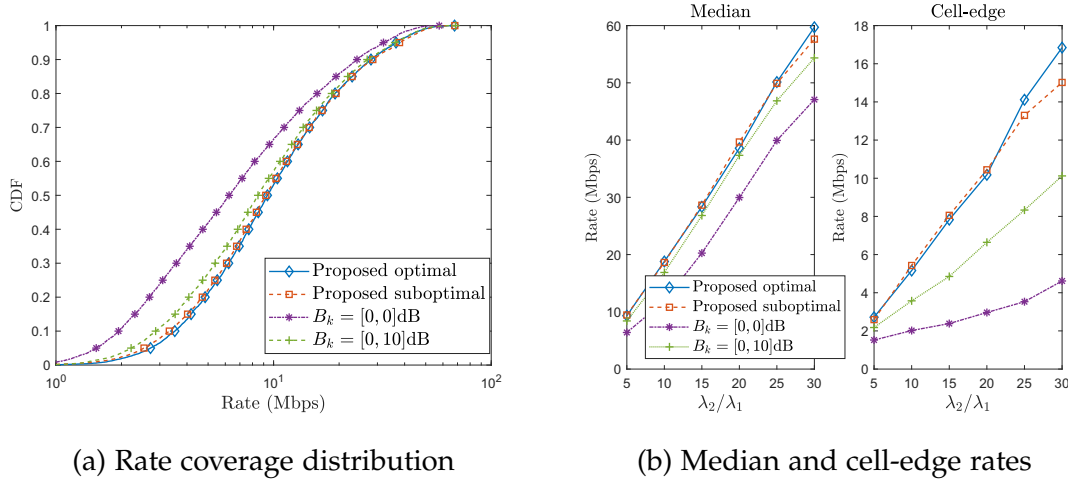


Figure 5.9: Rate performance under the 3GPP channel model ($U = [8, 4]$).

5.8 Conclusions

The heterogeneity of cellular networks is not only tied to the different types of BSs, but also to the various types of UEs, e.g., machine-type and human-type devices. Network densification, hence, must be designed to provide coverage enhancements for the former and enable high throughput links for the latter. To this end, we have formulated two frameworks to maximize the coverage and a proportional-fair rate utility, by optimizing user association and resource allocation.

Both frameworks depend on the coverage probability, which we first derive in a simple form and then approximate it to obtain a closed-form expression that is amenable for statistical optimization. Specifically, the approximation is shown to be concave with the association probability, making it suitable for optimization problems with coverage-dependent objective functions and optimizing variables that depend on user association. There can be a variety of objective functions that can be used with this expression, and in this work we have demonstrated two frameworks.

The coverage maximization framework has led to several key insights. First, it is shown that the user must connect to the BS with the highest ratio of the array gain to the multiplexing gain, revealing that max-power association performs poorly when a BS

multiplexes several UEs. Practically, this suggests that in order to maximize the coverage for IoT devices, they should not be multiplexed by other devices in the same time-frequency block. Second, the proposed policy maintains the SIR invariance property, where coverage is independent of the density of BSs in a given tier. Thus, to improve the coverage, the spectrum should be further orthogonalized, i.e., increase the number of tiers instead of increasing the number of BSs per tier.

In the rate maximization framework, we have shown that the portion of the spectrum allocated to one tier should be proportional to how many UEs are expected to be connected to that tier. Interestingly, in a network with high small cell densification, it is shown that almost the entire spectrum should be allocated to the small cell tier, motivating co-channel deployment for rate maximization. In addition, user association breaks down into matching the user distribution across tiers to the supply of each tier, which is defined in terms of a rate-based function.

CHAPTER 6

Conclusion

In conclusion, the dissertation presents several spectrum sharing solutions that enable wide-area connectivity for massive IoT applications. The emphasis in this dissertation has been put on the intra-network and inter-network sharing challenges due to the lack of a dedicated spectrum for IoT applications and the high density of IoT devices, along with their unique traffic characteristics. The dissertation leverages tools from stochastic geometry, communication theory, and optimization to develop practical algorithms and glean system-level design guidelines for massive IoT communications.

6.1 Summary of Contributions and Lessons Learned

In Chapter 2, we have developed the first stochastic geometry framework that models UNB networks, analyzing their scalability and robustness to interference from incumbent networks. Several variants of UNB protocols have been examined, and further protocol enhancements and optimizations have been proposed. The key lessons that we have learned from the analysis are:

- **Geographical diversity is paramount:** Networks that rely on coordinated access naturally enforce the user to synchronize with a single BS. Interestingly, in random access, where time-frequency synchronization is absent, the lack of cell selection is not only beneficial, but it is in fact critical thanks to the attained geographical diversity. Even under a sparse density of BSs, UNB signals have long-range propagation, reaching far away BSs and harnessing the geographical diversity gains.

- **We can trade spectrum with BSs:** UNB networks operate over the unlicensed bands, which are free to use. However, the sub-1GHz spectrum is limited, and thus network densification with UNB BSs can replace the need to use a wider spectrum. If BSs are using narrowband receivers, we need the IoT device to transmit its packets at different multiplexing bands; otherwise, a high density of BSs becomes necessary.

In Chapter 3, we have revisited the requirements of spectrum sensing to enable a wider landscape of massive IoT applications. To this end, we have demonstrated the feasibility of using narrowband receivers to learn the occupancy of a wideband spectrum at a fine spectral resolution; with careful distributed processing of sensing reports, we also have been able to obtain radio maps to identify which channels can be reused over space. The key lessons that we have learned from the development of the sensing-based system are:

- **We can do a lot with narrowband receivers:** A straw-man approach to explore more whitespaces is the use of wideband receivers. Such receivers are still expensive, particularly if spectrum scanning is done at a fine spectral resolution. We have demonstrated in Chapter 3 that even if narrowband receivers are used, we can still learn the occupancy of the wideband spectrum at the desired spectral resolution. This is accomplished through an intelligent sensing assignment scheduler and a distributed dissemination of occupancy information about the different bands of the spectrum.
- **Cooperation beyond improving sensing reliability:** The literature is rich with cooperative sensing algorithms, where multiple sensing nodes collaborate to infer the spectrum occupancy of a particular band. In essence, cooperation brings diversity gains, tangibly improving the sensing reliability. However, we have demonstrated that BSs can still cooperate even if they do not sense the same bands. Such cooperation helps diffuse occupancy information across the network, capturing the spatial variations of incumbents' footprints.

In Chapter 4, we have proposed a shared spectrum access protocol that enables a harmonious coexistence of massive IoT with cellular UEs. The protocol leverages drones as data aggregators, maximizes the energy-efficiency of IoT devices, and protects UEs from harmful interference. We have gleaned the following design guidelines:

- **Line-of-sight is not always beneficial:** The use of drones has traditionally been motivated because of their mobility and the fact that ground users can be in good line-of-sight (LOS) with the drone, reducing propagation losses. Further, as the drone flies at higher altitudes, the path loss exponent typically decreases and the LOS probability increases as there are fewer obstructions [3GP17a]. However, in a large network, this also means that the drone will have a higher LOS with more interfering devices. Indeed, we have shown that the energy-efficiency of communications deteriorates when drones fly at higher altitudes due to the increased LOS with interfering ground BSs and UEs.
- **Extended coverage mode without max-power transmission is needed:** 3GPP has defined an extended coverage mode where the IoT device can transmit at maximum power [3GP15a]. This helps provide deep coverage for utility sensors deployed inside buildings and basements. However, transmitting at maximum power is shown to significantly reduce the device lifetime; thus, novel coding and decoding schemes are still needed to improve coverage without the need to increase the transmit power of the IoT device.

In Chapter 5, we have analyzed the coverage performance in HetNets, taking into account the multi-antenna transmission mode used. We have then developed user association algorithms to maximize the coverage of IoT devices and the rate of high-rate cellular devices. We have gleaned the following insights:

- **Coverage depends on the ratio of the array gain to the multiplexing gain:** BSs are universally endowed with multi-antenna systems and are now envisioned to have massive MIMO arrays. One could expect if the IoT device connects to a BS

with a large-scale antenna system, the IoT device can experience better coverage due to the attained array gain. However, if the BS multiplexes other users on the same time-frequency resource block, the coverage can tangibly deteriorate. For this reason, the measured power of received reference signals should be weighted by a coefficient that depends on the perceived ratio of the array gain to the multiplexing gain for a given transmission mode.

- **Spectrum sharing maximizes the average capacity of dense networks:** When we have densified the network with small cells of different tiers, the optimal spectrum allocation policy is to allow small cells to reuse all RBs rather than breaking the spectrum into orthogonal channels across the different tiers of the network.

6.2 Future Work

The different generations of wireless communication systems have been naturally defined by disruptive technological features that set them apart from previous generations. The next revolution is centered around connecting machines and IoT devices, with seamless interactions with humans. This dissertation has focused on a particular class of IoT applications, namely providing a wide-area connectivity for massive IoT. To take the contributions of this dissertation forward, there remain exciting thrusts that require further exploration.

6.2.1 Ultra-reliable low-latency communications over wide areas

Ultra-reliable low-latency communications (URLLC), also known as critical IoT, is a central use case in 5G-NR. However, standardization efforts have focused on industrial applications, where factory machines wirelessly connect to the Internet. Thus, state-of-the-art 3GPP-compliant solutions are suitable for a moderate density of IoT devices, creating a need for novel techniques that extend URLLC over wider areas. For example, a large density of connected vehicles, in urban cities, can tremendously benefit from *low-*

latency wide-area (LLWA) connectivity. While our proposed solutions provide the desired coverage and connection density, they are tailored for delay-tolerant applications. It is interesting to revisit the proposed systems as well as existing LPWAs networks, adding the latency dimension to the protocol design and converting them into LLWA networks.

6.2.2 Advanced signal processing of random access signals

Random access networks have successfully penetrated the massive IoT market. In this dissertation, we have focused on the analysis and development of UNB networks, a variant of LPWA networks, which rely on UNB signals, repetition diversity, and geographical diversity. Nevertheless, coherent combining of signal transmissions is not considered nor cooperation among BSs. It is expected that advanced processing of these transmissions, that are repeated over frequency with replicas also at different BSs, can further extend the coverage of these networks. The challenge here is that these UNB signals are of unknown frequency, and thus novel distributed signal processing algorithms at BSs are needed to decode a high density of UNB signals.

6.2.3 Spectrum occupancy dissemination for massive IoT

In sensing-based access, we have used BSs to perform sensing and identify available whitespaces. Once these resources are identified, the BS has to broadcast the spectrum occupancy information to IoT devices. This typically assumes the availability of a control channel, which can become unavailable due to the unpredictable activity of incumbent devices using the unlicensed spectrum. Thus, it is critical to revisit the spectrum occupancy dissemination procedure; for example, one idea is to have floating control channels, where their density and locations in the spectrum adapt to the activity of IoT devices and incumbents. Equally important, IoT devices should be able to quickly find these control channels, using their low-cost battery-constrained hardware.

6.2.4 Beyond ALOHA-based and sensing-based access

ALOHA-based and sensing-based access remain the dominating spectrum sharing methods for massive IoT over the unlicensed spectrum. Yet, there are alternative options that are worth exploring such as database-assisted access and spectrum access systems.

To begin with, the TV white space (TVWS) has been recently utilized to provide broadband connectivity, particularly in rural areas. For example, there is approximately 180MHz of spectrum (512-698MHz) that can be used by IoT networks given that they do not interfere with incumbents such as TV stations. To this end, a geolocation database is used to identify which channels can be used. One example of using TVWS is Microsoft's FarmBeats project [Mic16], where IoT sensors collect farm-related physical parameters, e.g., soil moisture, and these measurements are forwarded to the cloud using the TVWS. In addition, regulatory and standard bodies are exploring different parts of the spectrum for future unlicensed-based access. A primary example is the FCC's notice of proposed rulemaking regarding expanding unlicensed-based access to the 5.9-7.1GHz bands [FCC17]. For example, the subbands in 6.425-6.525GHz and 6.875-7.125GHz are proposed for low-power indoor operations, whereas the subbands in 5.925-6.425GHz and 6.525-6.875GHz can be used for outdoor applications, yet an *automated frequency control* (AFC) system is required to protect incumbents such as fixed-satellite service. While the FCC is currently seeking more comments on several aspects of the AFC system, it is currently envisioned to be a simple database, for which BSs can request a list of available frequencies at their locations. For this reason, future research can focus on the development of AFC systems that consider the traffic properties of wide-area IoT networks.

In addition to licensed and unlicensed access, a third emerging paradigm is the licensed-shared access (LSA) or the spectrum access system (SAS), which comprises different tiers of users with varying access privileges. For example, the FCC has freed up 150MHz in the 3.5GHz CBRS band and introduced three tiers of access: incumbent access, priority access license (PAL), and general authorized access (GAA). PAL users can

acquire spectrum via an auction, whereas GAA users do not need a license. However, access priorities are given to incumbents followed by PAL users, e.g., if a PAL user is active over a channel used by the GAA user, the latter must vacate the channel immediately. Acquiring a PAL adds additional costs on IoT network operators, and thus GAA may be desirable to pursue instead, yet a dynamic spectrum access mechanism would be needed to ensure that incumbents and PAL users are protected from any interference. The interplay between these different tiers and the policy aspect of spectrum sharing are two interesting research thrusts that require more attention from the academic community.

APPENDIX A

Performance Analysis of UNB Networks in Chapter 2

A.1 Proof of Theorem 1

We can simplify the expression in (2.7) as follows

$$\begin{aligned}
 Q_j &= \left(\mathbb{E}_{\tilde{\Phi}_{\text{IoT}}, \tilde{\Phi}_{\text{I}}, f_u, f_k} \left[\left(1 - e^{-\tau x_j^\alpha (\hat{P}_N + I_{\text{UNB}} + I_{\text{INC}})} \right) \right] \right)^N \\
 &= \left(1 - \mathbb{E}_{\tilde{\Phi}_{\text{IoT}}, \tilde{\Phi}_{\text{I}}, f_u, f_k} \left[e^{-\tau x_j^\alpha (\hat{P}_N + I_{\text{UNB}} + I_{\text{INC}})} \right] \right)^N \\
 &= \left(1 - e^{-\tau x_j^\alpha \hat{P}_N} \mathbb{E}_{\tilde{\Phi}_{\text{IoT}}, f_u} [e^{-\tau x_j^\alpha I_{\text{UNB}}}] \mathbb{E}_{\tilde{\Phi}_{\text{I}}, f_k} [e^{-\tau x_j^\alpha I_{\text{INC}}}] \right)^N.
 \end{aligned} \tag{A.1}$$

Let $s = \tau x_j^\alpha \hat{P}_I^\delta$, then we have

$$\begin{aligned}
 \mathbb{E}_{\tilde{\Phi}_{\text{I}}, f_k} [e^{-\tau x_j^\alpha I_{\text{INC}}}] &= \mathbb{E}_{\tilde{\Phi}_{\text{I}}, f_k} \left[\exp \left(-s \sum_{k \in \tilde{\Phi}_{\text{I}}} f_k y_{k,j}^{-\alpha} \right) \right] \\
 &\stackrel{(a)}{=} \exp \left(-2\pi \tilde{\lambda}_{\text{I}} \int_0^\infty \left(y - \frac{y}{1+sy^{-\alpha}} \right) dy \right) \\
 &= \exp \left(-\pi \xi^{-1} (\tau \hat{P}_{\text{I}})^\delta \tilde{\lambda}_{\text{I}} x^2 \right),
 \end{aligned} \tag{A.2}$$

where (a) follows using the probability generating functional (PGFL) of the interfering incumbent HPPP and the characteristic function (CF) of $f \sim \exp(1)$. We note that the integral limit starts at zero because the interfering incumbent can be arbitrarily close to the BS. Likewise, let $\tilde{s} = \tau x_j^\alpha$, then we have

$$\begin{aligned}
 \mathbb{E}_{\tilde{\Phi}_{\text{IoT}}, f_u} [e^{-\tau x_j^\alpha I_{\text{UNB}}}] &= \mathbb{E}_{\tilde{\Phi}_{\text{IoT}}, f_u} \left[\exp \left(-\tilde{s} \sum_{u \in \tilde{\Phi}_{\text{IoT}}} f_u y_{u,j}^{-\alpha} \right) \right] \\
 &= \exp \left(-\pi \xi^{-1} \tau^\delta \tilde{\lambda}_{\text{IoT}} x^2 \right).
 \end{aligned} \tag{A.3}$$

Plugging the above expressions into (2.7) and using the binomial theorem, we get

$$\begin{aligned}
 Q_j &= \left(1 - e^{-\pi \xi^{-1} \tau^\delta (\tilde{\lambda}_{\text{IoT}} + \hat{P}_{\text{I}}^\delta \tilde{\lambda}_{\text{I}}) x^2 - \tau x_j^\alpha \hat{P}_N} \right)^N \\
 &= \sum_{k=0}^N \binom{N}{k} (-1)^k e^{-k \pi \xi^{-1} \tau^\delta (\tilde{\lambda}_{\text{IoT}} + \hat{P}_{\text{I}}^\delta \tilde{\lambda}_{\text{I}}) x^2 - k \tau x_j^\alpha \hat{P}_N},
 \end{aligned} \tag{A.4}$$

and hence

$$\begin{aligned}
\mathbb{P}_s^{\text{nearest}} &= 1 - \mathbb{E}_x [\mathbb{Q}_j] \\
&\stackrel{(a)}{=} 1 - 2\pi\lambda_B \sum_{k=0}^N \left\{ \binom{N}{k} (-1)^k \int_0^\infty x e^{-\pi x^2 (\lambda_B + k\xi^{-1}\tau^\delta (\tilde{\lambda}_{\text{IoT}} + \hat{P}_I^\delta \tilde{\lambda}_I))} dx \right\} \\
&= 1 - \sum_{k=0}^N \binom{N}{k} (-1)^k \left(1 + \frac{k\xi^{-1}\tau^\delta (\tilde{\lambda}_{\text{IoT}} + \hat{P}_I^\delta \tilde{\lambda}_I)}{\lambda_B} \right)^{-1},
\end{aligned} \tag{A.5}$$

where (a) follows from the distribution of the distance from the IoT device to its nearest BS, i.e., $f(x) = 2\pi\lambda_B x \exp(-\pi\lambda_B x^2)$ [Hae12], and the assumption that the network is interference limited, i.e., $\hat{P}_N \rightarrow 0$.

A.2 Proof of Theorem 2

We can simplify (2.11) by compute the following

$$\mathbb{E}_{\Phi_B} [\prod_{b \in \Phi_B} \mathbb{Q}_b] = \exp \left(2\pi\lambda_B \sum_{k=1}^N \left\{ \binom{N}{k} (-1)^k \int_0^\infty x e^{-\pi\xi^{-1}k\tau^\delta (\tilde{\lambda}_{\text{IoT}} + \hat{P}_I^\delta \tilde{\lambda}_I)x^2 - k\tau x^\alpha \hat{P}_N} dx \right\} \right). \tag{A.6}$$

Pluggin (A.6) in (2.11), we get the exact success performance. We can further simplify the expression assuming an interference limited network, i.e., $\hat{P}_N \rightarrow 0$. In this case, we have

$$\int_0^\infty x e^{-\pi\xi^{-1}\tau^\delta (k^\delta \tilde{\lambda}_{\text{IoT}} + k\hat{P}_I^\delta \tilde{\lambda}_I)x^2} dx = \frac{\xi\tau^{-\delta}/(2\pi)}{k^\delta \tilde{\lambda}_{\text{IoT}} + k\hat{P}_I^\delta \tilde{\lambda}_I}, \tag{A.7}$$

and hence

$$\mathbb{E}_{\Phi_B} [\prod_{b \in \Phi_B} \mathbb{Q}_b] = \exp \left(\xi\tau^{-\delta} \frac{\lambda_B}{\tilde{\lambda}_{\text{IoT}} + \hat{P}_I^\delta \tilde{\lambda}_I} \sum_{k=1}^N \binom{N}{k} \frac{(-1)^k}{k} \right). \tag{A.8}$$

Plugging (A.7) in (A.6) and then using (2.11), we arrive at (2.12).

APPENDIX B

Analysis of the Sensing Algorithm in Chapter 3

B.1 Statistics of the Filtered Energy Measurements

We show that $d_{k,m}$ has the same mean as $Y_{k,m}$, yet with lower variance. Indeed, we can rewrite (3.7) recursively as

$$d_{k,m,i} = \zeta^i d_{k,m,0} + (1 - \zeta) \sum_{l=0}^{i-1} \zeta^l Y_{k,m,i-l}. \quad (\text{B.1})$$

Let $d_{k,m,0} = Y_{k,m,0}$ and assume that $\{Y_{k,m,l}\}_{l=0}^i$ are independent and identically distributed random variables, then we have

$$\begin{aligned} \mathbb{E}[d_{k,m,i}] &= \zeta^i \mathbb{E}[Y_{k,m,0}] + (1 - \zeta) \mathbb{E}[Y_{k,m,i-1}] \sum_{l=0}^{i-1} \zeta^l \\ &\stackrel{(i \rightarrow \infty)}{=} \mathbb{E}[Y_{k,m}], \end{aligned} \quad (\text{B.2})$$

where in the last equality we consider the steady-state behavior. Similarly, the variance in steady-state can be computed as

$$\begin{aligned} \mathbb{V}(d_{k,m,i}) &\stackrel{(i \rightarrow \infty)}{=} (1 - \zeta)^2 \mathbb{V} \left(\sum_{l=0}^{i-1} \zeta^l Y_{k,m,i-l} \right) \\ &= (1 - \zeta)^2 \mathbb{V}(Y_{k,m}) \sum_{l=0}^{i-1} \zeta^{2l} \\ &= \frac{(1 - \zeta)^2}{1 - \zeta^2} \mathbb{V}(Y_{k,m}), \end{aligned} \quad (\text{B.3})$$

which is much smaller than $\mathbb{V}(Y_{k,m})$.

B.2 Analysis of the Proposed Sensing Algorithm

We compute $\mathbb{E}_0[\mathbf{w}_i]$ when $Y_{k,i}$ is the energy of a Gaussian noise with variance P_n . To this end, it can be shown $\mathbb{E}_0[\mathbf{Y}_i] = 3P_n^2\mathbf{I}$. Similarly, the k -th entry in $\mathbb{E}_0[\boldsymbol{\phi}_i]$ is $\mathbb{E}_0[Y_{k,i}d_{k,i}] = (3 - 2\zeta)P_n^2$. To compute $\mathbb{E}_1[\mathbf{w}_i]$ when $Y_{k,i}$ is the energy of a Gaussian sample with variance P_n and mean $\sqrt{P_s}$, we have $\mathbb{E}_1[\mathbf{Y}_i] = (P_s^2 + 6P_sP_n + 3P_n^2)\mathbf{I}$ and $\mathbb{E}_1[Y_{k,i}d_{k,i}] = (1 - \zeta)(P_s^2 + 6P_sP_n + 3P_n^2) + \zeta(P_s + P_n)^2$.

To compute the variance, we can expand $\tilde{\mathbf{w}}_i^T \boldsymbol{\Sigma} \tilde{\mathbf{w}}_i$, and then take the expectation on both sides to get (3.13). Note that in this case, $\mathbf{B} = (\mathbf{I} - 3P_n^2\mathbf{M})\mathbf{A}^T$. Also, expanding the right-hand side of (3.13), we get

$$\mathbb{E}_0[(\boldsymbol{\phi}_i - \mathbf{Y}_i \bar{\mathbf{w}}_i)^T \mathbf{M} \boldsymbol{\Sigma} \mathbf{M} (\boldsymbol{\phi}_i - \mathbf{Y}_i \bar{\mathbf{w}}_i)] = \text{Tr}(\mathbf{M} \boldsymbol{\Sigma} \mathbf{M} \bar{\mathbf{Z}}) + \bar{\mathbf{w}}_i^T \mathbf{M} \mathbb{E}_0[\mathbf{Y}_i \boldsymbol{\Sigma} \mathbf{Y}_i] \mathbf{M} \bar{\mathbf{w}}_i - 2\bar{\mathbf{w}}_i^T \bar{\mathbf{c}}, \quad (\text{B.4})$$

where $\text{Tr}(\cdot)$ is the trace operation, $\bar{\mathbf{Z}} = \mathbb{E}_0[\boldsymbol{\phi}_i \boldsymbol{\phi}_i^T]$, and $\bar{\mathbf{c}} = \mathbb{E}_0[\mathbf{Y}_i \mathbf{M} \boldsymbol{\Sigma} \mathbf{M} \boldsymbol{\phi}_i]$. It can be shown that the k -th diagonal element of $\bar{\mathbf{Z}}$ is $\mathbb{E}_0[\phi_{k,i}^2] = (1 - \zeta)^2 P_n^4 (105 + \frac{30}{1-\zeta} + \frac{9\zeta^2}{1-\zeta^2} + \frac{6\zeta^3}{(1+\zeta)(1-\zeta^2)})$, and the off-diagonal (l, j) -th element is $\mathbb{E}_0[\phi_{l,i} \phi_{j,i}] = (3 - 2\zeta)^2 P_n^4$. Similarly, the k -th diagonal element of $\mathbb{E}_0[\mathbf{Y}_i \boldsymbol{\Sigma} \mathbf{Y}_i]$ is $\mathbb{E}_0[Y_{k,i}^4] \sigma_{k,k} = 105P_n^4 \sigma_{k,k}$, where $\sigma_{k,k}$ is the k -th diagonal element of $\boldsymbol{\Sigma}$. Also, the off-diagonal (l, j) -th element is $\mathbb{E}_0[Y_{l,i}^2 Y_{j,i}^2] \sigma_{l,j} = 9P_n^4 \sigma_{l,j}$. Finally, to compute $\bar{\mathbf{c}}$, let $\boldsymbol{\theta}_k = [\mu_1 \mu_k \sigma_{1,k}, \mu_2 \mu_k \sigma_{2,k}, \dots, \mu_K \mu_k \sigma_{K,k}]$, i.e., the k -th column vector of $\mathbf{M} \boldsymbol{\Sigma} \mathbf{M}$. Then, the k -th element of $\bar{\mathbf{c}}$ is

$$\begin{aligned} \mathbb{E}_0[\boldsymbol{\phi}_i^T \boldsymbol{\theta}_k Y_{k,i}^2] &= \mathbb{E}_0 \left[Y_{k,i}^2 \sum_{j=1}^K Y_{j,i} d_{j,i} \mu_j \mu_k \sigma_{j,k} \right] \\ &= \mu_k^2 \sigma_{k,k} \mathbb{E}_0[Y_{k,i}^3 d_{k,i}] + \mu_k \mathbb{E}_0[Y_{k,i}^2] \mathbb{E}_0[Y_{j,i} d_{j,i}] \sum_{j=1, j \neq k}^K \mu_j \sigma_{j,k} \\ &= (105 - 90\zeta) \mu_k^2 \sigma_{k,k} P_n^4 + 3(3 - 2\zeta) P_n^4 \mu_k \sum_{j=1, j \neq k}^K \mu_j \sigma_{j,k}. \end{aligned} \quad (\text{B.5})$$

Computing $\boldsymbol{\Sigma}$ from the discrete-time Lyapunov equation $\boldsymbol{\Sigma} - \mathbf{B} \boldsymbol{\Sigma} \mathbf{B}^T = \mathbf{I}$, and plugging the aforementioned expressions in (3.10), we get the final theoretical expression for the deflection coefficient, which is shown to be in good agreement with Monte Carlo simulations in Fig. 3.4.

APPENDIX C

Analysis of Drone-assisted Cellular Networks in Chapter 4

C.1 Proof of Theorem 6

Recall that the aggregate interference at the UE is given as

$$I_U = \underbrace{\sum_{y_b \in \Phi'_B} f_b y_b^{-\alpha_G}}_{I_{U,B}} + \underbrace{\sum_{z_m \in \Phi'_M} \frac{P_M}{P_B/U_B} f_m z_m^{-\alpha_G}}_{I_{U,M}}, \quad (\text{C.1})$$

where we have normalized the interference by $L_0 P_B/U_B$. Using the Gil-Pelaez Inversion theorem [Gil51] to evaluate the interference CDF, we get

$$F_{I_U} \left(\frac{g_B}{\tau x^{\alpha_G}} \right) = \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \text{Im} \left\{ \varphi_{I_U}(x^{\alpha_G} t) e^{-j t g_B / \tau} \right\} dt, \quad (\text{C.2})$$

where $\varphi_{I_U}(x^{\alpha_G} t) = \mathbb{E}_{I_U} [\exp(j t x^{\alpha_G} I_U)]$ is the characteristic function (CF), which is given as

$$\begin{aligned} \varphi_{I_U}(x^{\alpha_G} t) &\stackrel{(a)}{=} \mathbb{E}_{I_{U,B}} \left[\exp \left(j t x^{\alpha_G} \sum_{y_b \in \Phi'_B} f_b y_b^{-\alpha_G} \right) \right] \\ &\quad \times \mathbb{E}_{I_{U,M}} \left[\exp \left(j t x^{\alpha_G} \sum_{z_m \in \Phi'_M} \frac{P_M}{P_B/U_B} f_m z_m^{-\alpha_G} \right) \right], \end{aligned} \quad (\text{C.3})$$

where (a) follows since interfering BSs are independent from the interfering IoT devices. Using the probability-generating functional of the HPPP process [Hae12], it can be shown that $\varphi_{I_{U,B}}(x^{\alpha_G} t) = \exp(\pi \lambda_B x^2 (1 - \mathbb{E}_{f_b} [\Omega(f_b, \delta_G, t)]))$, where $\Omega(f_b, \delta_G, t) = {}_1F_1(-\delta_G; 1 - \delta_G; j t f_b)$, and ${}_1F_1(a; b; c)$ is the confluent hypergeometric function [JZ14]. Sim-

ilarly, we have

$$\begin{aligned}
\varphi_{I_{U,M}}(x^{\alpha_G}t) &\stackrel{(a)}{=} \exp \left(-2\pi\lambda_D \int_0^\infty y \left(1 - \mathbb{E}_{f_m} \left[e^{jt \frac{U_B P_M f_m x^{\alpha_G}}{P_B y^{\alpha_G}}} \right] \right) dy \right) \\
&\stackrel{(b)}{=} \exp \left(-\delta_G \pi \lambda_D x^2 \int_0^\infty l^{-(1+\delta_G)} \left(1 - \frac{1}{1 - jtl \frac{P_M}{P_B/U_B}} \right) dl \right) \\
&= \exp \left(-\pi \lambda_D x^2 \hat{P}_{B,M}^{\delta_G}(-jt)^{\delta_G} \right),
\end{aligned} \tag{C.4}$$

where (a) follows using the fact that the set of interfering IoT devices can be modeled as a set of Poisson interferers with density of λ_D as one IoT device is scheduled per drone per time-frequency slot, and (b) follows using the CF of f_m and then using the substitution $l = x^{\alpha_G} y^{-\alpha_G}$. Here, the integral lower limit is zero as the set of interfering IoT devices is independent of the UE's location, i.e., no protection zone is present unlike the case of interfering BSs which cannot be closer than the tagged BS. To summarize, we have

$$\varphi_{I_U}(x^{\alpha_G}t) = \exp \left(\pi \lambda_B x^2 \left(1 - \mathbb{E}_f[\Omega(f_b, \delta_G, t)] - (\lambda_D/\lambda_B) \hat{P}_{B,M}^{\delta_G}(-jt)^{\delta_G} \right) \right). \tag{C.5}$$

Thus, the coverage becomes

$$\mathbf{C}_{U,DL}^P(\tau) = \frac{1}{2} - 2\lambda_B \int_0^\infty \frac{1}{t} \text{Im} \{ \varphi_g(-t/\tau) \Xi(t) \} dt, \tag{C.6}$$

where $\varphi_{g_B}(-t/\tau) = (1 + jt/\tau)^{-\Delta_B}$ is the CF of g_B and

$$\begin{aligned}
\Xi(t) &= \int_0^\infty x e^{-\pi \lambda_B x^2 \left(\mathbb{E}_f[\Omega(f_b, \delta_G, t)] + (\lambda_D/\lambda_B) \hat{P}_{B,M}^{\delta_G}(-jt)^{\delta_G} \right)} dx \\
&= \frac{(2\pi\lambda_B)^{-1}}{\mathbb{E}_f[\Omega(f_b, \delta_G, t)] + (\lambda_D/\lambda_B) \hat{P}_{B,M}^{\delta_G}(-jt)^{\delta_G}}.
\end{aligned} \tag{C.7}$$

Plugging (C.7) in (C.6) and using $\mathbb{E}_f[\Omega(f_b, \delta_G, t)] = {}_2F_1(\Psi_G, U_B; 1 - \delta_G; jt)$, we arrive at (4.11).

C.2 Distribution of Signal and Interference Powers at the IoT Device

Let the IoT device be at a distance x_M from the drone. Then, the received signal power at the drone is given as

$$\begin{aligned}
 \mathbb{P}(S_M \leq \tau) &= \mathbb{P}(P_M L_0 x_M^{-\alpha_A} L_0((1 - L_{\text{NLOS}})\mathbb{P}_{\text{LOS}}(x_M, h_D) + L_{\text{NLOS}}) \leq \tau) \\
 &\stackrel{(a)}{\approx} \mathbb{P}(P_M L_M x_M^{-\alpha_A} \leq \tau) \\
 &= 1 - F_{x_M} \left(\left(\frac{P_M L_M}{\tau} \right)^{1/\alpha_A} \right) \\
 &\stackrel{(b)}{=} \begin{cases} 1 - \frac{(P_M L_M / \tau)^{\delta_A} - h_D^2}{R^2}, & \frac{P_M L_M}{(R^2 + h_D^2)^{1/\delta_A}} \leq \tau \leq \frac{P_M L_M}{h_D^{1/\delta_A}}, \\ 0, & \text{otherwise} \end{cases}
 \end{aligned} \tag{C.8}$$

where (a) follows by using the mean LOS probability instead of the actual one, and (b) follows using the CDF of the distance between the typical IoT device and the drone $F_{x_M}(\cdot)$, which can be found using the fact that the IoT device is randomly distributed over an area of radius R centered around the 2D coordinates of the drone. Similarly, let r_B denote the distance between the tagged BS and the drone, then

$$\begin{aligned}
 \mathbb{P}(I_B \leq \tau) &\approx \mathbb{P}(P_B L_B r_B^{-\alpha_A} \leq \tau) \\
 &= 1 - F_{r_B} \left(\left(\frac{P_B L_B}{\tau} \right)^{1/\alpha_A} \right).
 \end{aligned} \tag{C.9}$$

Recall that the distance between a point in $\mathbb{R}^2$ and the nearest BS is distributed as $f_{y_B}(y) = 2\pi\lambda_B y \exp(-2\pi\lambda_B y^2)$ [LBY15]. Thus, the distribution of $r_B = \sqrt{y_B^2 + h_D^2}$ can be shown to be $f_{r_B}(r) = \frac{2r f_{y_B}(\sqrt{r^2 - h_D^2})}{\sqrt{r^2 - h_D^2}}$. Thus, we can compute $F_{r_B}(r) = \int_{h_D}^{\infty} f_{r_B}(r) dr$ to arrive at (4.17).

C.3 Convexity Analysis of the EE Maximization Framework

we first derive useful properties for $r(p)$. In particular, $r(p)$ is a non-negative sum of coverage probabilities, each is non-decreasing with the transmit power, and thus $r(p)$ is non-decreasing with p . Thus, the t -sublevel sets, i.e., $\mathcal{R}_{t,\text{sub}} = \{p | r(p) \leq t\}$, and the t -superlevel sets, i.e., $\mathcal{R}_{t,\text{sup}} = \{p | r(p) \geq t\}$, are convex, and hence $r(p)$ is quasilinear [BV04].

To show that $r(p)$ is S-shaped, we take the 2nd derivative with respect to p to get

$$\begin{aligned} \frac{d^2 r(p)}{dp^2} &= \delta_A \pi \lambda_B \tilde{L}_M^2 (P_B L_B)^{\delta_A} \sum_{k=1}^{K_M} \frac{\mu_k e^{-\pi \lambda_B \left(\frac{p \tilde{L}_M - P_N}{\tau_{M,k} P_B L_B} \right)^{-\delta_A}}}{\tau_{M,k}^2 \left(\frac{p \tilde{L}_M}{\tau_{M,k}} - P_N \right)^{2(1+\delta_A)}} \\ &\quad \times \left\{ \delta_A \pi \lambda_B (P_B L_B)^{\delta_A} - (1 + \delta_A) \left(\frac{p \tilde{L}_M}{\tau_{M,k}} - P_N \right)^{\delta_A} \right\}. \end{aligned} \quad (C.10)$$

Note that $\left(\frac{p \tilde{L}_M - P_N}{\tau_{M,k} P_B L_B} \right) \geq 0$ for any feasible $\tau_{M,k}$ and the derivative is non-increasing with p . Furthermore, there exists a point $p^\bullet$ such that $\frac{d^2 r(p)}{dp^2} \geq 0$ for $p \leq p^\bullet$ and $\frac{d^2 r(p)}{dp^2} \leq 0$ for $p \geq p^\bullet$, and hence the function is S-shaped. To prove that $r(p)$ is concave when the condition in (4.23) holds, it is sufficient to show that the coverage $\mathbb{P}(\hat{\gamma}_M \geq \tau_k)$ is concave in p as $r(p)$ is a non-negative sum of the coverage probabilities. This can be shown by taking the second derivative of the coverage and showing that it is non-positive if and only if

$$\delta_A \pi \lambda_B (P_B L_B)^{\delta_A} - (1 + \delta_A) \left(\frac{p \tilde{L}_M - P_N}{\tau_k P_B L_B} \right)^{\delta_A} \leq 0, \quad (C.11)$$

which is satisfied when the condition in (4.23) holds.

To prove that the objective function in (4.22) is quasiconcave, then it is sufficient to prove that the superlevel sets $\mathcal{E}_{t,\text{sup}} = \{p \mid \frac{r(p)}{c(p)} \geq t\}$ are convex. Note since the objective is non-negative, then we only need to consider the case for $t \geq 0$ and prove that $\{p \mid r(p) - c(p)t \geq 0\}$ is convex. We consider two cases. First, for $p \geq p^\bullet$, $r(p)$ is concave, and thus for $p_1, p_2 \in \mathcal{E}_{t,\text{sup}}$ and $\beta \in [0, 1]$ we have

$$\begin{aligned} &r(\beta p_1 + (1 - \beta)p_2) - c(\beta p_1 + (1 - \beta)p_2)t \\ &\stackrel{(a)}{\geq} \beta r(p_1) + (1 - \beta)r(p_2) - c(\beta p_1 + (1 - \beta)p_2)t \\ &\stackrel{(b)}{=} \beta (r(p_1) - c(p_1)t) + (1 - \beta) (r(p_2) - c(p_2)t) \\ &\geq 0, \end{aligned} \quad (C.12)$$

where (a) follows because $r(p)$ is concave and (b) follows because $c(p)$ is affine, and thus, $\mathcal{E}_{t,\text{sup}}$ is convex. Second, for $p \leq p^\bullet$, $r(p)$ is convex. Taking the first derivative of $\bar{E}(p)$, we

get

$$\frac{d\bar{E}(p)}{dp} = \frac{\eta^{-1}(pr'(p) - r(p)) + P_{\text{CP}}r'(p)}{c^2(p)}. \quad (\text{C.13})$$

The second sum term $r'(p) \geq 0$ as $r(p)$ is a non-decreasing function in p . In addition, the first sum term $pr'(p) - r(p) \geq 0$ because $r(p)$ is convex, i.e., for any convex differentiable function $g(x)$ with $g(0) = 0$, we have $g(y) \geq g(x) + (y - x)g'(x)$ [BV04], and by setting $y = 0$ we get $xg'(x) \geq g(x)$. Thus, the derivative is non-negative, i.e., it is a non-decreasing function in $p \leq p^\bullet$, and hence the superlevel sets are convex, which completes the proof that the objective function is quasiconcave. To prove that the objective function is unimodal, note that $\bar{E}(0) = \bar{E}(\infty) = 0$, and since $\bar{E}(p)$ is quasiconcave, then it has to be unimodal.

Finally, since the objective function is quasiconcave and the constraints are all affine, then the EE maximization framework is quasiconcave.

APPENDIX D

Analysis of Multi-antenna HetNets in Chapter 5

D.1 Coverage Probability Approximation

Using (5.11) and (5.12) in (5.10), we get

$$\mathbb{P}_{\text{cov},k}(\tau) = \frac{1}{2} - \frac{2\lambda_k}{A_k} \int_0^\infty \frac{1}{t} \text{Im}\{\varphi_{g_{x_k}}(-t/\tau)\Omega(t)\}dt, \quad (\text{D.1})$$

where

$$\begin{aligned} \Omega(t) &= \int_0^\infty x e^{\pi\lambda_k x^2 (1 - \mathbb{E}_{f_{y_k}}[\Psi(f_{y_k}, \alpha, t)] - 1/A_k)} dx \\ &= \frac{(2\pi\lambda_k)^{-1}}{\mathbb{E}_{f_{y_k}}[\Psi(f_{y_k}, \alpha, t)] + 1/A_k - 1}, \end{aligned} \quad (\text{D.2})$$

and the CF of g_{x_k} is $\varphi_{g_{x_k}}(-t/\tau) = (1 + jt/\tau)^{-\Delta_k}$, which follows using the fact that $g_{x_k} \sim \Gamma(\Delta_k, 1)$. We can further rewrite $\mathbb{E}_{f_{y_k}}[\Psi(f_{y_k}, \alpha, t)]$ as follows

$$\begin{aligned} \mathbb{E}_{f_{y_k}}[\Psi(f_{y_k}, \alpha, t)] &\stackrel{(a)}{=} \sum_{n=0}^\infty \frac{(-\alpha/2)_{(n)}}{(1 - \alpha/2)_{(n)} n!} (jt)^n \mathbb{E}_{f_{y_k}}[f_{y_k}^n] \\ &\stackrel{(b)}{=} {}_2F_1(-2/\alpha, U_k; 1 - 2/\alpha; jt), \end{aligned} \quad (\text{D.3})$$

where (a) follows from the series representation of the confluent hypergeometric function [JZ14] and (b) follows from evaluating the n -th order moment of $f_{y_k} \sim \Gamma(U_k, 1)$ [SA05] and then using the series representation of the Gauss hypergeometric function [JZ14]. Using the first-order approximation of ${}_2F_1(\cdot)$, we get ${}_2F_1(-2/\alpha, U_k; 1 - 2/\alpha; jt) \approx 1 - \frac{2jtU_k}{\alpha-2}$.

Thus, plugging this in (D.1), we arrive at

$$\mathbb{P}_{\text{cov},k}(\tau) \approx \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \frac{1}{t} \text{Im} \left\{ \frac{(1 + jt/\tau)^{-\Delta_k}}{1 - 2jtA_kU_k/(\alpha-2)} \right\} dt, \quad (\text{D.4})$$

where the approximation here follows from using the first-order expression of ${}_2F_1(\cdot)$. The accuracy of this approximation is higher when the numerator of the integrand in (D.4) behaves more like a delta function [WG03, RG14], which occurs for high Δ_k or low τ . Such regimes are of practical interest as in massive MIMO systems, we have $\Delta_k \gg 1$ [HYA14], and for IoT-based applications with deep coverage requirements, τ is typically low [Ind17]. Let $\sigma = \tau^{-1}$ and $\zeta_k = 2A_k U_k / (\alpha - 2)$. Then, using $\text{Im}(z) = (z - z^*)/(2j)$, the integrand in (D.4) becomes

$$\frac{1}{2j} \int_0^\infty \frac{1}{t} \left(\frac{(1 + j\sigma t)^{-\Delta_k}}{1 - j\zeta_k t} - \frac{(1 - j\sigma t)^{-\Delta_k}}{1 + j\zeta_k t} \right) dt = \left(\frac{\zeta_k}{\sigma + \zeta_k} \right)^{\Delta_k} \pi - \frac{\pi}{2}, \quad (\text{D.5})$$

and thus (D.4) becomes

$$\begin{aligned} \mathbb{P}_{\text{cov},k}(\tau) &\approx \frac{1}{2} - \frac{1}{\pi} \left[\left(\frac{\zeta_k}{\sigma + \zeta_k} \right)^{\Delta_k} \pi - \frac{\pi}{2} \right] \\ &= 1 - \left(\frac{A_k U_k}{A_k U_k + \beta_\tau} \right)^{\Delta_k}, \end{aligned} \quad (\text{D.6})$$

which completes the proof.

D.2 Concavity of Coverage Maximization

It is sufficient to show that $f(A_k) = A_k \mathbb{P}_{\text{cov},k}(\tau)$ is concave, since the sum of concave functions is concave. To prove that $f(A_k)$ is concave, we show its second derivative is nonpositive on its domain. Indeed, it can be shown that

$$\frac{\partial}{\partial A_k} f(A_k) = 1 - \left(\frac{U_k A_k}{U_k A_k + \beta_\tau} \right)^{\Delta_k} \left(1 + \frac{\beta_\tau \Delta_k}{A_k U_k + \beta_\tau} \right), \quad (\text{D.7})$$

and thus

$$\frac{\partial^2}{\partial A_k^2} f(A_k) = -\Delta_k(1 + \Delta_k) \left(\frac{\beta_\tau}{U_k A_k + \beta_\tau} \right)^2 \left(\frac{U_k A_k}{U_k A_k + \beta_\tau} \right)^{\Delta_k}. \quad (\text{D.8})$$

Since $\Delta_k, U_k, \beta_\tau \geq 0$, we have $\frac{\partial^2}{\partial A_k^2} f(A_k) \leq 0$.

D.3 Concavity of Rate Maximization

To prove that the objective function in (5.34) is concave, it is sufficient to prove that $f_1(A_k, \eta_k) = A_k \log(\eta_k/A_k)$ is jointly concave in (A_k, η_k) and $f_2(A_k) = A_k \log \left(1 - \left(\frac{A_k U_k}{A_k U_k + \beta_\tau} \right)^{\Delta_k} \right)$ is concave. Note that $f_1(A_k, \eta_k)$ is the perspective function of $h(z) = \log(z)$, i.e., $f_1(t, z) = t \cdot h(z/t)$ which is jointly concave in (t, z) when $h(z)$ is concave [BV04]. Thus, $f_1(A_k, \eta_k)$ is jointly concave in (A_k, η_k) . To prove the concavity of $f_2(A_k)$, we show that the second derivative is nonpositive on its domain. Let $\xi_k = [A_k U_k / (A_k U_k + \beta_\tau)]^{\Delta_k}$, then it can be shown that

$$\frac{d^2}{dA_k^2} f_2(A_k) = \frac{\beta_\tau^2 \Delta_k \xi_k (\xi_k - 1 - \Delta_k)}{A_k (\beta_\tau + A_k U_k)^2 (\xi_k - 1)^2}. \quad (\text{D.9})$$

The denominator is always positive. In addition, since $\xi_k < 1$, we have $\xi_k - 1 - \Delta_k < 0$. Thus, the second derivative of $f_2(A_k)$ is nonpositive on its domain, and hence it is concave. Since the objective function is a nonnegative sum of $f_1(A_k, \eta_k)$ and $f_2(A_k)$, then it is concave.

REFERENCES

- [3GP15a] 3GPP. “Cellular System Support for Ultra Low Complexity and Low Throughput Internet of Things, Release 13.” TS 45.820, November 2015.
- [3GP15b] 3GPP. “NB-IoT: Battery lifetime evaluation.” R1-156006, October 2015.
- [3GP17a] 3GPP. “Study on 3D channel model for LTE.” TR 36.889, May 2017.
- [3GP17b] 3GPP. “Study on channel model for frequencies from 0.5 to 100 GHz.” TS 38.901, December 2017.
- [3GP17c] 3GPP. “Study on Enhanced LTE Support for Aerial Vehicles.” TR 36.777, December 2017.
- [3GP17d] 3GPP. “Study on NR-based Access to Unlicensed Spectrum.” TS 38.889, 3GPP, February 2017.
- [AAJ18] O. L. Alcaraz López, H. Alves, P. H. Juliano Nardelli, and M. Latva-aho. “Aggregation and Resource Scheduling in Machine-Type Communication Networks: A Stochastic Geometry Approach.” *IEEE Trans. Wireless Commun.*, **17**(7):4750–4765, July 2018.
- [AAS17] T. P. C. d. Andrade, C. A. Astudillo, L. R. Sekijima, and N. L. S. d. Fonseca. “The Random Access Procedure in Long Term Evolution Networks for the Internet of Things.” *IEEE Commun. Mag.*, **55**(3):124–131, March 2017.
- [ABC14] 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.*, **32**(6):1065–1082, June 2014.
- [ADA17] A. Aijaz, M. Dohler, A. H. Aghvami, V. Friderikos, and M. Frodigh. “Realizing the Tactile Internet: Haptic Communications over Next Generation 5G Cellular Networks.” *IEEE Wireless Commun.*, **24**(2):82–89, April 2017.
- [AEA16] L. H. Afify, H. ElSawy, T. Y. Al-Naffouri, and M. S. Alouini. “A Unified Stochastic Geometry Model for MIMO Cellular Networks With Retransmissions.” *IEEE Trans. Wireless Commun.*, **15**(12):8595–8609, December 2016.
- [AGM15] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash. “Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications.” *Commun. Surveys Tuts.*, **17**(4):2347–2376, 2015.
- [AKL14] A. Al-Hourani, S. Kandeepan, and S. Lardner. “Optimal LAP Altitude for Maximum Coverage.” *IEEE Wireless Commun. Lett.*, **3**(6):569–572, December 2014.

- [ANR17] A. Alkhateeb, Y. H. Nam, M. S. Rahman, J. Zhang, and R. W. Heath. "Initial Beam Association in Millimeter Wave Cellular Systems: Analysis and Design Insights." *IEEE Trans. Wireless Commun.*, **16**(5):2807–2821, May 2017.
- [Art17] C. Artigue. "Method for searching for a useful signal in a multiplexing band.", September 2017. US Patent 9,768,897.
- [ASS15] K. M. Alam, M. Saini, and A. E. Saddik. "Toward Social Internet of Vehicles: Concept, Architecture, and Applications." *IEEE Access*, **3**:343–357, 2015.
- [ASY14] J. G. Andrews, S. Singh, Q. Ye, X. Lin, and H. S. Dhillon. "An overview of load balancing in HetNets: old myths and open problems." *IEEE Wireless Commun.*, **21**(2):18–25, April 2014.
- [AZZ17] Mussab Alaa, AA Zaidan, BB Zaidan, Mohammed Talal, and Miss Laiha Mat Kiah. "A review of smart home applications based on Internet of Things." *Journal of Network and Computer Applications*, **97**:48–65, 2017.
- [BBP16] D. Bethanabhotla, O. Y. Bursalioglu, H. C. Papadopoulos, and G. Caire. "Optimal User-Cell Association for Massive MIMO Wireless Networks." *IEEE Trans. Wireless Commun.*, **15**(3):1835–1850, March 2016.
- [BCE18] Osama M. Bushnaq, Abdulkadir Celik, Hesham ElSawy, Mohamed-Slim Alouini, and Tareq Y. Al-Naffouri. "Aeronautical Data Aggregation and Field Estimation in IoT Networks: Hovering & Traveling Time Dilemma of UAVs." *arXiv preprint arXiv:1810.08035*, October 2018.
- [BCZ15] Andrea Biral, Marco Centenaro, Andrea Zanella, Lorenzo Vangelista, and Michele Zorzi. "The challenges of M2M massive access in wireless cellular networks." *Digital Communications and Networks*, **1**(1):1–19, 2015.
- [BJD14] E. Bjornson, E. A. Jorswieck, M. Debbah, and B. Ottersten. "Multiobjective Signal Processing Optimization: The way to balance conflicting metrics in 5G systems." *IEEE Signal. Proc. Mag.*, **31**(6):14–23, November 2014.
- [BL15] W. Bao and B. Liang. "Rate Maximization Through Structured Spectrum Allocation and User Association in Heterogeneous Cellular Networks." *IEEE Trans. Commun.*, **63**(11):4510–4524, November 2015.
- [BLH16] Olivier Bachem, Mario Lucic, S Hamed Hassani, and Andreas Krause. "Approximate K-Means++ in Sublinear Time." In *AAAI*, February 2016.
- [BPZ17] Alessandro Biason, Chiara Pielli, Andrea Zanella, and Michele Zorzi. "On the Energy/Distortion Tradeoff in the IoT." *arXiv preprint arXiv:1702.03695*, February 2017.
- [Bra15] Gregory Braun. "The Internet of Things and the Modern Supply Chain." Technical report, C3 Solutions, August 2015.

- [BRV16] Martin C. Bor, Utz Roedig, Thiemo Voigt, and Juan M. Alonso. “Do LoRa Low-Power Wide-Area Networks Scale?” In *Proceedings of the 19th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, pp. 59–67, November 2016.
- [BV04] S.P. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.
- [BY16] I. Bor-Yaliniz and H. Yanikomeroglu. “The New Frontier in RAN Heterogeneity: Multi-Tier Drone-Cells.” *IEEE Commun. Mag.*, **54**(11):48–55, November 2016.
- [CF91] Jean-Pierre Crouzeix and Jacques A Ferland. “Algorithms for generalized fractional programming.” *Mathematical Programming*, **52**(1):191–207, 1991.
- [Cis19a] Cisco. *Cisco Visual Networking Index: Forecast and Trends, 2017–2022*. Cisco Systems, Inc., February 2019.
- [Cis19b] Cisco. *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2017–2022*. Cisco Systems, Inc., February 2019.
- [CLG17] C. Cano, D. J. Leith, A. Garcia-Saavedra, and P. Serrano. “Fair Coexistence of Scheduled and Random Access Wireless Networks: Unlicensed LTE/WiFi.” *IEEE/ACM Trans. Netw.*, **25**(6):3267–3281, December 2017.
- [CRS15] J. Chen, C. Richard, and A. H. Sayed. “Diffusion LMS Over Multitask Networks.” *IEEE Trans. Signal Process.*, **63**(11):2733–2748, June 2015.
- [CVS17] M. Centenaro, L. Vangelista, S. Saur, A. Weber, and V. Braun. “Comparison of Collision-Free and Contention-Based Radio Access Protocols for the Internet of Things.” *IEEE Trans. Commun.*, **65**(9):3832–3846, September 2017.
- [CVZ16] M. Centenaro, L. Vangelista, A. Zanella, and M. Zorzi. “Long-range communications in unlicensed bands: the rising stars in the IoT and smart city scenarios.” *IEEE Wireless Commun.*, **23**(5):60–67, October 2016.
- [CWS18] B. Chen, J. Wan, L. Shu, P. Li, M. Mukherjee, and B. Yin. “Smart Factory of Industry 4.0: Key Technologies, Application Case, and Challenges.” *IEEE Access*, **6**:6505–6519, 2018.
- [DDT17] Ousmane Dieng, Babacar Diop, Ousmane Thiare, and Congduc Pham. “A Study on IoT Solutions for Preventing Cattle Rustling in African Context.” In *Proceedings of the Second International Conference on Internet of Things, Data and Cloud Computing, ICC ’17*, pp. 153:1–153:11, New York, NY, USA, 2017. ACM.
- [DGG14] Minh-Tien Do, Claire Goursaud, and Jean-Marie Gorce. “On the benefits of random FDMA schemes in ultra narrow band networks.” In *International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, pp. 672–677, May 2014.

- [DKA13] H. S. Dhillon, M. Kountouris, and J. G. Andrews. "Downlink MIMO HetNets: Modeling, Ordering Results and Performance Analysis." *IEEE Trans. Wireless Commun.*, **12**(10):5208–5222, October 2013.
- [DSG17] Z. Dawy, W. Saad, A. Ghosh, J. G. Andrews, and E. Yaacoub. "Toward Massive Machine Type Cellular Communications." *IEEE Wireless Commun.*, **24**(1):120–128, February 2017.
- [DT16] Y. Dhungana and C. Tellambura. "Multichannel Analysis of Cell Range Expansion and Resource Partitioning in Two-Tier Heterogeneous Cellular Networks." *IEEE Trans. Wireless Commun.*, **15**(3):2394–2406, March 2016.
- [EI18] W. Ejaz and M. Ibnkahla. "Multiband Spectrum Sensing and Resource Allocation for IoT in Cognitive 5G Networks." *IEEE Internet Things J.*, **5**(1):150–163, February 2018.
- [ESA17] H. ElSawy, A. Sultan-Salem, M. S. Alouini, and M. Z. Win. "Modeling and Analysis of Cellular Networks Using Stochastic Geometry: A Tutorial." *Commun. Surveys Tuts.*, **19**(1):167–203, 2017.
- [EVS18] Zakarya El-friakh, Andra M. Voicu, Shaham Shabani, Ljiljana Simić, and Petri Mähönen. "Crowdsourced Indoor Wi-Fi REMs: Does the Spatial Interpolation Method Matter?" In *IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*, Oct. 2018.
- [FCC17] FCC. "Expanding Flexible Use in Mid-Band Spectrum Between 3.7 and 24 GHz.", August 2017.
- [FFG17] Z. Feng, Z. Feng, and T. A. Gulliver. "Biologically Inspired Two-Stage Resource Management for Machine-Type Communications in Cellular Networks." *IEEE Trans. Wireless Commun.*, **16**(9):5897–5910, September 2017.
- [GDV14] A. K. Gupta, H. S. Dhillon, S. Vishwanath, and J. G. Andrews. "Downlink Multi-Antenna Heterogeneous Cellular Network With Load Balancing." *IEEE Trans. Commun.*, **62**(11):4052–4067, November 2014.
- [GDZ17] J. Guo, S. Durrani, X. Zhou, and H. Yanikomeroglu. "Massive Machine Type Communication With Data Aggregation and Resource Scheduling." *IEEE Trans. Commun.*, **65**(9):4012–4026, September 2017.
- [Gil51] J. Gil-Pelaez. "Note on the Inversion Theorem." *Biometrika*, **38**(3-4), December 1951.
- [GM16] C. Goursaud and Y. Mo. "Random unslotted time-frequency ALOHA: Theory and application to IoT UNB networks." In *Proc. 23rd Int. Conf. Telecommunications (ICT)*, pp. 1–5, May 2016.

- [GYY16] X. Ge, J. Ye, Y. Yang, and Q. Li. "User Mobility Evaluation for 5G Small Cell Networks Based on Individual Mobility Model." *IEEE J. Sel. Areas Commun.*, **34**(3):528–541, March 2016.
- [GZA10] Arunabha Ghosh, Jun Zhang, Jeffrey G. Andrews, and Rias Muhamed. *Fundamentals of LTE*. Pearson Education (US), 2010.
- [Hae12] Martin Haenggi. *Stochastic Geometry for Wireless Networks*. Cambridge University Press, 2012.
- [HAW08] Andrew M Hunter, Jeffrey G Andrews, and Steven Weber. "Transmission capacity of ad hoc networks with spatial diversity." *IEEE Trans. Wireless Commun.*, **7**(12):5058–5071, December 2008.
- [HC16a] G. Hattab and D. Cabric. "Inter-Tier Interference Mitigation in Multi-Antenna HetNets: A Resource Blanking Approach." In *Proc. IEEE Global Communications Conf. (GLOBECOM)*, pp. 1–6, December 2016.
- [HC16b] G. Hattab and D. Cabric. "Joint Resource Allocation and User Association in Multi-Antenna Heterogeneous Networks." In *Proc. IEEE Global Communications Conf. (GLOBECOM)*, pp. 1–7, December 2016.
- [HC17] G. Hattab and D. Cabric. "Energy-efficient massive cellular IoT shared spectrum access via mobile data aggregators." In *IEEE 13th Int. Conf. Wireless and Mobile Computing(WiMob)*, pp. 1–6, October 2017.
- [HC18a] G. Hattab and D. Cabric. "Coverage and Rate Maximization via User Association in Multi-Antenna HetNets." *IEEE Trans. Wireless Commun.*, **17**(11):7441–7455, November 2018.
- [HC18b] G. Hattab and D. Cabric. "Distributed Wideband Spatio-Spectral Sensing for Unlicensed Massive IoT Communications." In *Proc. IEEE Global Communications Conf. (GLOBECOM)*, pp. 1–7, December 2018.
- [HC18c] G. Hattab and D. Cabric. "Long-Term Rate-Based User-Centric Association for Downlink Multi-Antenna HetNets." In *Proc. IEEE Int. Conf. Communications Workshops (ICC Workshops)*, pp. 1–6, May 2018.
- [HC18d] G. Hattab and D. Cabric. "Performance Analysis of Uplink Cellular IoT Using Different Deployments of Data Aggregators." In *Proc. IEEE Global Communications Conf. (GLOBECOM)*, pp. 1–6, December 2018.
- [HC18e] G. Hattab and D. Cabric. "Rate-Based Cell Range Expansion for Downlink Massive MIMO Heterogeneous Networks." *IEEE Wireless Commun. Lett.*, **7**(3):296–299, June 2018.

- [HC18f] G. Hattab and D. Cabric. "Spectrum Sharing Protocols based on Ultra-Narrowband Communications for Unlicensed Massive IoT." In *Proc. IEEE Int. Symp. Dynamic Spectrum Access Networks (DySPAN)*, pp. 1–10, October 2018.
- [HC18g] Ghaith Hattab and Danijela Cabric. "Distributed Wideband Sensing-based Architecture for Unlicensed Massive IoT Communications." *submitted to IEEE Trans. on Cogn. Commun. Netw.*, *arXiv preprint arXiv:1812.06238*, December 2018.
- [HC19a] Ghaith Hattab and Danijela Cabric. "Energy-Efficient Massive IoT Shared Spectrum Access over UAV-enabled Cellular Networks." *CoRR*, **abs/1808.08006**, February 2019.
- [HC19b] Ghaith Hattab and Danijela Cabric. "Unlicensed Spectrum Sharing for Massive Internet-of-Things Communications." *submitted to IEEE Wireless Commun. Mag.*, *arXiv preprint arXiv:1903.01504*, March 2019.
- [HLL17] A. Hoglund, X. Lin, O. Liberg, A. Behravan, E. A. Yavuz, M. Van Der Zee, Y. Sui, T. Tirronen, A. Ratilainen, and D. Eriksson. "Overview of 3GPP Release 14 Enhanced NB-IoT." *IEEE Network*, **31**(6):16–22, November 2017.
- [HVC18a] G. Hattab, E. Visotsky, M. Cudak, and A. Ghosh. "Interference Mitigation via Beam Range Biasing for 5G mmWave Coexistence with Incumbents." In *Proc. IEEE 5G World Forum (5GWF)*, pp. 210–214, July 2018.
- [HVC18b] G. Hattab, E. Visotsky, M. Cudak, and A. Ghosh. "Toward the Coexistence of 5G MmWave Networks with Incumbent Systems beyond 70 GHz." *IEEE Wireless Commun.*, **25**(4):18–24, August 2018.
- [HWE15] A. He, L. Wang, M. El Kashlan, Y. Chen, and K. K. Wong. "Spectrum and Energy Efficiency in Massive MIMO Enabled HetNets: A Stochastic Geometry Approach." *IEEE Commun. Lett.*, **19**(12):2294–2297, December 2015.
- [HYA14] K. Hosseini, W. Yu, and R. S. Adve. "Large-Scale MIMO Versus Network MIMO for Multicell Interference Mitigation." *IEEE J. Sel. Top. Signal Process.*, **8**(5):930–941, October 2014.
- [HYM17] Q. Han, B. Yang, G. Miao, C. Chen, X. Wang, and X. Guan. "Backhaul-Aware User Association and Resource Allocation for Energy-Constrained HetNets." *IEEE Trans. Veh. Technol.*, **66**(1):580–593, January 2017.
- [Ind17] Industry White Paper. "Coverage Analysis of LTE-M Category-M1." Technical report, 2017.
- [ITU15] ITU-R. "IMT Vision – Framework and overall objectives of the future development of IMT for 2020 and beyond." M. 2083-0, ITU-R, September 2015.

- [JDN18] N. Jiang, Y. Deng, A. Nallanathan, X. Kang, and T. Q. S. Quek. "Analyzing Random Access Collisions in Massive IoT Networks." *IEEE Trans. Wireless Commun.*, **17**(10):6853–6870, October 2018.
- [JSX12] H. S. Jo, Y. J. Sang, P. Xia, and J. G. Andrews. "Heterogeneous Cellular Networks with Flexible Cell Association: A Comprehensive Downlink SINR Analysis." *IEEE Trans. Wireless Commun.*, **11**(10):3484–3495, October 2012.
- [JZ14] A. Jeffrey and D. Zwillinger. *Table of Integrals, Series, and Products*. Elsevier LTD, Oxford, 2014.
- [KA12] M. Kountouris and J. G. Andrews. "Downlink SDMA with Limited Feedback in Interference-Limited Wireless Networks." *IEEE Trans. Wireless Commun.*, **11**(8):2730–2741, August 2012.
- [KC13] T. Kwon and J. M. Cioffi. "Random Deployment of Data Collectors for Serving Randomly-Located Sensors." *IEEE Trans. Wireless Commun.*, **12**(6):2556–2565, June 2013.
- [KJB17] H. Kwon, J. Jeon, A. Bhorkar, Q. Ye, H. Harada, Y. Jiang, L. Liu, S. Nagata, B. L. Ng, T. Novlan, J. Oh, and W. Yi. "Licensed-Assisted Access to Unlicensed Spectrum in LTE Release 13." *IEEE Commun. Mag.*, **55**(2):201–207, February 2017.
- [KLI17] Z. Khan, J. J. Lehtomaki, S. I. Iellamo, R. Vuohtoniemi, E. Hossain, and Z. Han. "IoT Connectivity in Radar Bands: A Shared Access Model Based on Spectrum Measurements." *IEEE Commun. Mag.*, **55**(2):88–96, February 2017.
- [LAA14] A. Laya, L. Alonso, and J. Alonso-Zarate. "Is the Random Access Channel of LTE and LTE-A Suitable for M2M Communications? A Survey of Alternatives." *Commun. Surveys Tuts.*, **16**(1):4–16, 2014.
- [LBA16] Y. Li, F. Baccelli, J. G. Andrews, T. D. Novlan, and J. C. Zhang. "Modeling and Analyzing the Coexistence of Wi-Fi and LTE in Unlicensed Spectrum." *IEEE Trans. Wireless Commun.*, **15**(9):6310–6326, September 2016.
- [LBY15] Y. Lin, W. Bao, W. Yu, and B. Liang. "Optimizing User Association and Spectrum Allocation in HetNets: A Utility Perspective." *IEEE J. Sel. Areas Commun.*, **33**(6):1025–1039, June 2015.
- [LC17] M. Laghate and D. Cabric. "Cooperatively Learning Footprints of Multiple Incumbent Transmitters by Using Cognitive Radio Networks." *IEEE Trans. on Cogn. Commun. Netw.*, **3**(3):282–297, September 2017.
- [LDL15] J. Lai, E. Dutkiewicz, R. P. Liu, and R. Vesilo. "Opportunistic Spectrum Access with Two Channel Sensing in Cognitive Radio Networks." *IEEE Trans. Mobile Comput.*, **14**(1):126–138, January 2015.

- [LGR11] D. Lopez-Perez, I. Guvenc, G. de la Roche, M. Kountouris, T. Q. S. Quek, and J. Zhang. "Enhanced intercell interference coordination challenges in heterogeneous networks." *IEEE Wireless Commun.*, **18**(3):22–30, June 2011.
- [LMR17] M. Labib, V. Marojevic, J. H. Reed, and A. I. Zaghloul. "Extending LTE into the Unlicensed Spectrum: Technical Analysis of the Proposed Variants." *IEEE Communications Standards Magazine*, **1**(4):31–39, December 2017.
- [LWC16] D. Liu, L. Wang, Y. Chen, M. El Kashlan, K. K. Wong, R. Schober, and L. Hanzo. "User Association in 5G Networks: A Survey and an Outlook." *Commun. Surveys Tuts.*, **18**(2):1018–1044, 2016.
- [LZA16] C. Li, J. Zhang, J. G. Andrews, and K. B. Letaief. "Success Probability and Area Spectral Efficiency in Multiuser MIMO HetNets." *IEEE Trans. Commun.*, **64**(4):1544–1556, April 2016.
- [MAY17] Y. Mehmood, F. Ahmad, I. Yaqoob, A. Adnane, M. Imran, and S. Guizani. "Internet-of-Things-Based Smart Cities: Recent Advances and Challenges." *IEEE Commun. Mag.*, **55**(9):16–24, September 2017.
- [MBT17] N. H. Motlagh, M. Bagaa, and T. Taleb. "UAV-Based IoT Platform: A Crowd Surveillance Use Case." *IEEE Commun. Mag.*, **55**(2):128–134, February 2017.
- [MCF16] A. Mukherjee, J. F. Cheng, S. Falahati, H. Koorapaty, D. H. Kang, R. Karaki, L. Falconetti, and D. Larsson. "Licensed-Assisted Access LTE: coexistence with IEEE 802.11 and the evolution toward 5G." *IEEE Commun. Mag.*, **54**(6):50–57, June 2016.
- [Mck18] McKinsey Global Institute. "The Internet of Things: How to capture the value of IoT." Technical report, McKinsey&Company, May 2018.
- [MDA16] D. Malak, H. S. Dhillon, and J. G. Andrews. "Optimizing Data Aggregation for Uplink Machine-to-Machine Communication Networks." *IEEE Trans. Commun.*, **64**(3):1274–1290, March 2016.
- [MDG16] Y. Mo, M. T. Do, C. Goursaud, and J. M. Gorce. "Optimization of the predefined number of replications in a Ultra Narrow Band based IoT network." In *Proc. Wireless Days (WD)*, pp. 1–6, March 2016.
- [MG16] F. Morvari and A. Ghasemi. "Two-Stage Resource Allocation for Random Access M2M Communications in LTE Network." *IEEE Commun. Lett.*, **20**(5):982–985, May 2016.
- [MGA17] H. Menouar, I. Guvenc, K. Akkaya, A. S. Uluagac, A. Kadri, and A. Tuncer. "UAV-Enabled Intelligent Transportation Systems for the Smart City: Applications and Challenges." *IEEE Commun. Mag.*, **55**(3):22–28, March 2017.

- [MGG16] Y. Mo, C. Goursaud, and J. M. Gorce. "Theoretical analysis of UNB-based IoT networks with path loss and random spectrum access." In *IEEE Int. Symp. Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pp. 1–6, September 2016.
- [MGG18] Y. Mo, C. Goursaud, and J. Gorce. "Multiple Base Stations Diversity for UNB Systems: Theoretical Analysis and Performances." In *Proc. Computers and Communications (ISNCC) 2018 Int. Symp. Networks*, pp. 1–6, June 2018.
- [MHW15] R. Mochaourab, B. Holfeld, and T. Wirth. "Distributed Channel Assignment in Cognitive Radio Networks: Stable Matching and Walrasian Equilibrium." *IEEE Trans. Wireless Commun.*, **14**(7):3924–3936, July 2015.
- [Mic16] Microsoft Research. "FarmBeats: IoT for agriculture." Technical report, April 2016.
- [MSB16] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah. "Unmanned Aerial Vehicle With Underlaid Device-to-Device Communications: Performance and Tradeoffs." *IEEE Trans. Wireless Commun.*, **15**(6):3949–3963, June 2016.
- [MSB17] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah. "Mobile Unmanned Aerial Vehicles (UAVs) for Energy-Efficient Internet of Things Communications." *IEEE Trans. Wireless Commun.*, **16**(11):7574–7589, November 2017.
- [MT95] Silvano Martello and Paolo Toth. "The bottleneck generalized assignment problem." *Eur. J. Oper. Res.*, **83**(3):621–638, June 1995.
- [Mul17] MulteFire Alliance. "MulteFire Release 1.0 Technical Paper." Technical report, January 2017.
- [MWZ18] Weidong Mei, Qingqing Wu, and Rui Zhang. "Cellular-Connected UAV: Uplink Association, Power Control and Interference Coordination." *arXiv preprint arXiv:1807.08218*, July 2018.
- [NLV15] G. Naddafzadeh-Shirazi, L. Lampe, G. Vos, and S. Bennett. "Coverage enhancement techniques for machine-to-machine communications over LTE." *IEEE Commun. Mag.*, **53**(7):192–200, July 2015.
- [Nok16] Nokia. "LTE Evolution for IoT Connectivity." Technical report, July 2016.
- [PZB15] Jordi Pérez Romero, Andreas Zalonis, Lila Boukhatem, Adrian Kliks, Aikaterini Koutlia, Nikos Dimitriou, and Reben Kurda. "On the use of radio environment maps for interference management in heterogeneous networks." *IEEE Commun. Mag.*, **53**(8):184–191, August 2015.
- [RG14] M. Di Renzo and P. Guan. "Stochastic Geometry Modeling of Coverage and Rate of Cellular Networks Using the Gil-Pelaez Inversion Theorem." *IEEE Commun. Lett.*, **18**(9):1575–1578, September 2014.

- [RKF18] C. Rosa, M. Kuusela, F. Frederiksen, and K. I. Pedersen. "Standalone LTE in Unlicensed Spectrum: Radio Challenges, Solutions, and Performance of MulteFire." *IEEE Commun. Mag.*, **56**(10):170–177, October 2018.
- [RKS17] U. Raza, P. Kulkarni, and M. Sooriyabandara. "Low Power Wide Area Networks: An Overview." *Commun. Surveys Tuts.*, **19**(2):855–873, Second quarter 2017.
- [RS17] A. Rajandekar and B. Sikdar. "O-MAC: Opportunistic MAC Protocol for M2M Communication in WiFi White Spaces." *IEEE Commun. Lett.*, **21**(11):2440–2443, November 2017.
- [RVX16] A. Rico-Alvarino, M. Vajapeyam, H. Xu, X. Wang, Y. Blankenship, J. Bergman, T. Tirronen, and E. Yavuz. "An overview of 3GPP enhancements on machine to machine communications." *IEEE Commun. Mag.*, **54**(6):14–21, June 2016.
- [SA05] Marvin K Simon and Mohamed-Slim Alouini. *Digital communication over fading channels*, volume 95. John Wiley & Sons, 2005.
- [SA14] S. Sadr and R. S. Adve. "Tier Association Probability and Spectrum Partitioning for Maximum Rate Coverage in Multi-Tier Heterogeneous Networks." *IEEE Commun. Lett.*, **18**(10):1791–1794, October 2014.
- [SAD16] M. Simsek, A. Aijaz, M. Dohler, J. Sachs, and G. Fettweis. "5G-Enabled Tactile Internet." *IEEE J. Sel. Areas Commun.*, **34**(3):460–473, March 2016.
- [Say14] Ali H Sayed. "Adaptation, learning, and optimization over networks." *Foundations and Trends® in Machine Learning*, **7**(4-5):311–801, 2014.
- [SBA14] S. Singh, F. Baccelli, and J. G. Andrews. "On Association Cells in Random Heterogeneous Networks." *IEEE Wireless Commun. Lett.*, **3**(1):70–73, February 2014.
- [SDA13] S. Singh, H. S. Dhillon, and J. G. Andrews. "Offloading in Heterogeneous Networks: Modeling, Analysis, and Design Insights." *IEEE Trans. Wireless Commun.*, **12**(5):2484–2497, May 2013.
- [SDM15] I. Sobron, P. S. R. Diniz, W. A. Martins, and M. Velez. "Energy Detection Technique for Adaptive Spectrum Sensing." *IEEE Trans. Commun.*, **63**(3):617–627, March 2015.
- [SGN16] E. Soltanmohammadi, K. Ghavami, and M. Naraghi-Pour. "A Survey of Traffic Issues in Machine-to-Machine Communications Over LTE." *IEEE Internet Things J.*, **3**(6):865–884, December 2016.
- [Sig17] Sigfox. "Sigfox Technical Overview." Technical report, May 2017.
- [Sig19] Sigfox. "Sigfox Radio Specifications v1.3." Technical report, February 2019.

- [Sor17] N. Sornin. “LoRaWAN 1.1 Specification.” Technical report, LoRA Alliance, October 2017.
- [SRI15] H. Shariatmadari, R. Ratasuk, S. Iraj, A. Laya, T. Taleb, R. Jäntti, and A. Ghosh. “Machine-type communications: current status and future perspectives toward 5G systems.” *IEEE Commun. Mag.*, **53**(9):10–17, September 2015.
- [SS03] Siegfried Schaible and Jianming Shi. “Fractional programming: the sum-of-ratios case.” *Optimization Methods and Software*, **18**(2):219–229, 2003.
- [STH16] U. Siddique, H. Tabassum, E. Hossain, and D. I. Kim. “Channel-Access-Aware User Association With Interference Coordination in Two-Tier Downlink Cellular Networks.” *IEEE Trans. Veh. Technol.*, **65**(7):5579–5594, July 2016.
- [Tel18] Telensa. “Traffic Adaptive Street Lighting: What is it and how does it work?” Technical report, Telensa Ltd, 2018.
- [TL17] U. Tefek and T. J. Lim. “Relaying and Radio Resource Partitioning for Machine-Type Communications in Cellular Networks.” *IEEE Trans. Wireless Commun.*, **16**(2):1344–1356, February 2017.
- [TSH14] H. Tabassum, U. Siddique, E. Hossain, and M. J. Hossain. “Downlink Performance of Cellular Systems With Base Station Sleeping, User Association, and Scheduling.” *IEEE Trans. Wireless Commun.*, **13**(10):5752–5767, October 2014.
- [UV08] J. Unnikrishnan and V. V. Veeravalli. “Cooperative Sensing for Primary Detection in Cognitive Radio.” *IEEE J. Sel. Topics Signal Process.*, **2**(1):18–27, February 2008.
- [VKW17] Deepak Vasisht, Zerina Kapetanovic, Jongho Won, Xinxin Jin, Ranveer Chandra, Sudipta Sinha, Ashish Kapoor, Madhusudhan Sudarshan, and Sean Stratman. “FarmBeats: An IoT Platform for Data-Driven Agriculture.” In *14th USENIX Symposium on Networked Systems Design and Implementation (NSDI 17)*, pp. 515–529, Boston, MA, 2017. USENIX Association.
- [WA12] Steven Weber, Jeffrey G Andrews, et al. “Transmission capacity of wireless networks.” *Foundations and Trends® in Networking*, **5**(2–3):109–281, February 2012.
- [WAV16] WAVIoT. “WaveIoT NB-Fi Technology: Products and Tech Description.” Technical report, June 2016.
- [WDZ16] H. Wei, N. Deng, W. Zhou, and M. Haenggi. “Approximate SIR Analysis in General Heterogeneous Cellular Networks.” *IEEE Trans. Commun.*, **64**(3):1259–1273, March 2016.
- [Wei] Weightless. “<http://www.weightless.org/>.”

- [WG03] Zhengdao Wang and G. B. Giannakis. "A simple and general parameterization quantifying performance in fading channels." *IEEE Trans. Commun.*, **51**(8):1389–1398, August 2003.
- [WLA17] Y.-E. Wang, X. Lin, A. Adhikary, A. Grovlen, Y. Sui, Y. Blankenship, J. Bergman, and H. S. Razaghi. "A Primer on 3GPP Narrowband Internet of Things." *IEEE Commun. Mag.*, **55**(3):117–123, March 2017.
- [WND17] P. K. Wali, A. A. N, and D. Das. "Optimal Time-Spatial Randomization Techniques for Energy Efficient IoT Access in LTE-Advanced." *IEEE Trans. Veh. Technol.*, **66**(8):7346–7359, August 2017.
- [WTH17] Y. Wang, W. P. Tay, and W. Hu. "A Multitask Diffusion Strategy With Optimized Inter-cluster Cooperation." *IEEE J. Sel. Topics Signal Process.*, **11**(3):504–517, April 2017.
- [WW15] Z. Wang and V. W. S. Wong. "Optimal Access Class Barring for Stationary Machine Type Communication Devices With Timing Advance Information." *IEEE Trans. Wireless Commun.*, **14**(10):5374–5387, October 2015.
- [XYW18] J. Xu, J. Yao, L. Wang, Z. Ming, K. Wu, and L. Chen. "Narrowband Internet of Things: Evolutions, Technologies, and Open Issues." *IEEE Internet Things J.*, **5**(3):1449–1462, June 2018.
- [XZX15] Xiong Xiong, Kan Zheng, Rongtao Xu, Wei Xiang, and Periklis Chatzimisios. "Low power wide area machine-to-machine networks: key techniques and prototype." *IEEE Commun. Mag.*, **53**(9):64–71, September 2015.
- [YJC17] Z. Yuan, J. Jin, J. Chen, L. Sun, and G. Muntean. "ComProSe: Shaping Future Public Safety Communities with ProSe-Based UAVs." *IEEE Commun. Mag.*, **55**(12):165–171, December 2017.
- [YJS18] Z. Yuan, J. Jin, L. Sun, K. Chin, and G. Muntean. "Ultra-Reliable IoT Communications with UAVs: A Swarm Use Case." *IEEE Commun. Mag.*, **56**(12):90–96, December 2018.
- [YRC13] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. G. Andrews. "User Association for Load Balancing in Heterogeneous Cellular Networks." *IEEE Trans. Wireless Commun.*, **12**(6):2706–2716, June 2013.
- [ZMQ18] X. Zhang, Y. Ma, H. Qi, Y. Gao, Z. Xie, Z. Xie, M. Zhang, X. Wang, G. Wei, and Z. Li. "Distributed Compressive Sensing Augmented Wideband Spectrum Sharing for Cognitive IoT." *IEEE Internet Things J.*, **5**(4):3234–3245, August 2018.
- [ZQG17] Y. Zhong, T. Q. S. Quek, and X. Ge. "Heterogeneous Cellular Networks With Spatio-Temporal Traffic: Delay Analysis and Scheduling." *IEEE J. Sel. Areas Commun.*, **35**(6):1373–1386, June 2017.

- [ZSB16] A. Zappone, L. Sanguinetti, G. Bacci, E. Jorswieck, and M. Debbah. “Energy-Efficient Power Control: A Look at 5G Wireless Technologies.” *IEEE Trans. Signal Process.*, **64**(7):1668–1683, April 2016.
- [ZYY15] W. Zhang, Y. Yang, and C. K. Yeo. “Cluster-Based Cooperative Spectrum Sensing Assignment Strategy for Heterogeneous Cognitive Radio Network.” *IEEE Trans. Veh. Technol.*, **64**(6):2637–2647, June 2015.
- [ZZ17] Y. Zeng and R. Zhang. “Energy-Efficient UAV Communication With Trajectory Optimization.” *IEEE Trans. Wireless Commun.*, **16**(6):3747–3760, June 2017.
- [ZZL16] Y. Zeng, R. Zhang, and T. J. Lim. “Throughput Maximization for UAV-Enabled Mobile Relaying Systems.” *IEEE Trans. Commun.*, **64**(12):4983–4996, December 2016.
- [ZZZ18] C. Zhan, Y. Zeng, and R. Zhang. “Energy-Efficient Data Collection in UAV Enabled Wireless Sensor Network.” *IEEE Wireless Commun. Lett.*, **7**(3):328–331, June 2018.